HEC MONTRÉAL
École affiliée à l'Université de Montréal

Three articles on fraud detection for auto insurance companies: A cost sensitive approach mixing machine learning and economic theory

par
Raphaël Zerbato

Thèse présentée en vue de l'obtention du grade de Ph. D. en administration
(spécialisation Economie appliquée)

Novembre 2024

HEC MONTRÉAL

École affiliée à l'Université de Montréal

Cette thèse intitulée :

Three articles on fraud detection for auto insurance companies:
A cost sensitive approach mixing machine learning and economic theory

Présentée par :

Raphaël Zerbato

a été évaluée par un jury composé des personnes suivantes :

Nicolas Sahuguet
HEC Montréal
Président-rapporteur

Georges Dionne
HEC Montréal
Directeur de recherche

François Bellavance
HEC Montréal
Codirecteur de recherche

Pierre-Carl Michaud
HEC Montréal
Membre du jury

Pierre Picard
École Polytechnique, Paris
Examinateur externe

Jean-Guy Simonato
HEC Montréal
Représentant du directeur de HEC Montréal

Résumé

Nous améliorons le Modèle d'Audit Optimale (OAM) développé par Dionne et al., 2009. Semblable aux autres techniques d'apprentissage sensible aux coûts, OAM est un système expert recommandant une politique d'audite minimisant les coûts associés à la fraude. Ces coûts de fraude se divisent en trois types. Le premier est associé à l'audit de réclamations réellement frauduleuses : les vrai-positifs. Le deuxième concerne les réclamations injustement auditées : les faux-positifs. Le troisième vient des réclamations non-auditées bien que frauduleuses : les faux-négatifs. OAM diffère des autres modèles sensibles aux coûts car il lie le niveau de fraude attendu à l'agressivité de la politique d'audit choisie. L'idée est qu'une politique d'audit agressive dissuadera de futurs fraudeurs.

Cette thèse améliore OAM en trois points. Le premier article introduit un nouvel OAM, σOAM. σOAM utilise directement le signal émis par un algorithme de classification ce qui permet une meilleure optimisation de la politique d'audite par rapport à OAM. σOAM est adapté au nouveau algorithme de classification d'apprentissage machine comme le modèle de Boosting de Gradient Extrême (XGBoost). σOAM obtient de meilleurs résultats que les autres systèmes experts sensibles aux coûts dans la littérature. Le deuxième article incorpore une mesure d'aversion au risque obtenue du modèle de Cohen and Einav (2007) à l'entrainement des différents algorithmes de classification du premier article. La variable de d'aversion au risque est systématique classée parmi les plus prédictive

améliorant significativement les performances obtenues du premier article. Le but est de montrer empiriquement que l'ajout du paramètre d'aversion au risque apporte des informations supplémentaires pour prédire la fraude, ce qui confirme les prévisions des modèles théoriques liant l'aversion au risque à la propension à frauder

Le troisième article se concentre sur la modélisation de la décision de fraude du réclamant. La pénalité subie par l'assuré fraudant est présumée inconnue de l'assureur. Cette approche reflète une réalité où les assureurs n'ont que peu de pouvoir sur les sanctions imposables en cas de fraude. Cette pénalité peut être interprétée comme un coût moral que l'assuré fraudeur subie. La distribution de cette perte personnelle est inférée des données. La forme de cette distribution permet d'estimer l'élasticité du taux de fraude par rapport à la politique d'audite choisie, ce qui était un paramètre exogène dans l'OAM original. Les résultats obtenus sont comparés aux résultats de modèles similaires présents dans la littérature, notamment Picard (1996), Boyer (1999) et Boyer and Peter (2020). Les conclusions communes sont que les réclamations les plus chères doivent être auditées en priorités car elles sont les plus susceptibles d'être fausses. De nouveaux résultats indiquent que les réclamants avec premium élevé ont une proportion plus élevée fraudeurs potentiels ainsi qu'un taux de fraude d'équilibre positif.

Mots-clés

Assurance automobile, détection de fraude, système expert, classificateur, théorie économique, apprentissage machine, aversion au risque, coût moral, économétrie appliquée.

Méthodes de recherche

Données tabulaires, modèle théorique micro-économique, optimisation, apprentissage machine, économétrie appliquée.

Abstract

We improve the Optimal Audit Model (OAM) developed by Dionne et al. (2009). OAM is an expert system often used by insurance companies to determine whether it is optimal to audit a claim. Similar to other cost-sensitive learning techniques, OAM minimizes the costs associated to fraud. These fraud costs are divided into three types. The first type is associated with auditing true positives: claims that are correctly predicted as fraudulent. The second type of fraud costs is incurred for unjustly audited claims: false positives. The third type of fraud costs comes from claims that are not audited despite being fraudulent: false negatives. OAM differs from other cost-sensitive models because it links the expected level of fraud to the aggressiveness of the chosen audit policy. The idea is that an aggressive audit policy will deter future fraudsters. OAM uses the signal emitted by a claim to construct the optimal audit policy and obtain a deterrence effect.

This thesis improves OAM in three ways: The first paper introduces a new OAM, σOAM. σOAM directly uses the signal emitted by the classification algorithm, allowing a significant reduction in the fraud costs at the optimum compared to OAM. σOAM relies on the Extreme Gradient Boosting (XGBoost) model, a more recent classification algorithm than those used in OAM. σOAM achieves better results than other commonly cited cost-sensitive models in the literature. The second paper introduces a risk aversion measure obtained by applying the model developed by Cohen and Einav (2007) on a Canadian

database with new independent variables. This risk aversion measure is used in training the various classification algorithms of the first paper. The risk aversion variable is systematically ranked among the most predictive variables for detecting fraud, and the performance improvement due to the introduction of this new variable is significant. The goal is to empirically demonstrate that adding the risk aversion parameter provides additional information for predicting fraud, which confirms the predictions of theoretical models linking risk aversion to the propensity to commit fraud.

The third paper focuses on modeling the claimant's fraud decision. The penalty incurred by the fraudulent claimant is presumed unknown to the insurer. This approach reflects a reality in which insurers have limited power to impose sanctions beyond canceling fraudulent claims. The additional penalty can be interpreted as a moral cost or personal disutility that the fraudulent claimant should endure. It was considered as a constant parameter in the previous literature, including OAM. The distribution of this personal loss is inferred from the data. The shape of the distribution provides an estimate of the elasticity of the fraud rate with respect to the chosen audit policy. The results obtained by this new approach are compared with results from similar models present in the literature, including Picard (1996), Boyer (1999), and Boyer and Peter (2020). Common conclusions are that the most expensive claims should be prioritized for auditing as they are most likely to be false. New results indicate that claimants deemed the highest risk (with high insurance premiums) have a higher proportion of claimants likely to commit fraud and a positive residual fraud rate at equilibrium due to higher audit costs.

Keywords

Auto insurance, Fraud detection, Expert system and classifier, Economic theory, Machine learning, Risk aversion, Moral cost, Economic audit model.

Keywords

Auto insurance, Fraud detection, Expert system and classifier, Economic theory, Machine learning, Risk aversion, Moral cost, Economic audit model.

Research Methods

Tabular data, Microeconomic theoretical model, Optimization, Machine learning, Applied Econometrics.

Contents

List of Tables

List of Figures

List of Acronyms

AB Accident Benefit

AP All-Perils

ARA Absolute Risk Aversion

RRA Relative Risk Aversion

CAD Canadian Dollars

CARA Constant Absolute Risk Aversion

CRRA Constant Relative Risk Aversion

CSV Costly State Verification

FTP Full Term Premium

Cov Collision Coverage

OAM Optimal Audit Model

OLS Ordinary Least Square

σOAM New Optimal Audit Model

XGB Extreme Gradient Boosting

ML Machine-Learning

MLE Maximum Likelihood Estimation

TP True Positive

FP False Positive

TN True Negative

FN False Negative

TPR True Positive Rate

SA Sensitivity Analysis

$\mathscr{GDP}$ General Pareto Distribution

Acknowledgements

Je dois cette thèse à mes directeurs, les Professeurs Georges Dionne et François Bellavance. Leurs encouragements et bienveillances, leurs précieux conseils et leurs visions ont été primordiales dans l'accomplissement de ce travail de recherche. Je suis également sincerèment reconnaissant au Professeur Dionne pour le financement octroyé ainsi que la générosité avec laquelle il transmet ses grandes connaissances.

Ce travail a été possible grâce à la précieuse collaboration du Data-Lab d'Intact Assurance. À ce titre, je me dois de remercier Simon Valois et Samuel Levesque pour leurs conseils toujours à-propos.

Cette thèse représente la conclusion de six années de travail qui furent parfois éprouvantes mais toujours stimulantes. Je suis reconnaissant à mes parents, Pascale et Olivier, de m'avoir permis d'arriver jusqu'au ici et de ne jamais avoir désespéré de me voir un jour terminer. Un grand merci à mon frère Sébastien, qui finalement gradua avant moi. J'ai une pensée particulière pour ma compagne Ling, qui au jour le jour fut de toutes les batailles. Merci pour ta patience. Finalement un grand merci à toutes les personnes m'ayant accompagnées au cours de ces 6 années, Tangi, Damien, Albane, Valentin, Edouard, Jump et tous les autres...

Raphaël Zerbato

General Introduction

As noted by Tennyson and Salsas-Forn (2002), two types of approaches are prominent in the scientific literature against insurance fraud. The first one is based on economic theory and aims to produce an optimal contract that would dissuade claimants to misreport their claims. Different types of game are used to describe the interactions between the principal, here the insurer, and the agent, here the claimant. Based on sound microeconomic principles, these games provide rich insights on the optimal long term audit policy by capturing critical elements of the fraud problem such as the key incentives that lead to fraud. A key insight from such models is the dual role of the audit policy: to stop the normal processing of fraudulent claims and to deter fraud before its occurs. Although insightful, this theoretical approach is not easily applicable for insurance companies (see: Townsend (1979), Subelj et al. (2011), Bodaghiand and Teimourpour (2018) Crocker and Morgan (1998), Gollier (1987), Mookherjee and Png (1989) , Fagart and Picard (1999), Boyer (1999), Boyer and Peter (2020), Picard (1996), Picard (2013), Picard et al. (2024),Krawczyk (2009), Fudenberg et al. (1990))

The second approach focuses on expert systems to flag suspicious claims that require auditing. Empirically, insurance companies invest massively in in-house fraud detection processes. This processes are composed of various individual sub-processes such as business rules entailing automatic audit, adjusters alerting on unusual activities or classifica-

tion algorithms flagging suspicious claims to be audited. These classification algorithms, also known as expert systems, produce a ranking through a fraud score for each claim processed through them. However there are intrinsic difficulties when it comes to labeled data on insurance fraud. One difficulty lies in knowing the proportion of fraudulent claims correctly labeled. Any cost-sensitive expert system trained with this limited information can only recommend an audit policy that minimizes the expected cost of the already known fraud. As such, expert systems are useful tools but lack the capacity to account for the deterrence effect if the goal is to set-up an optimal audit policy.

This thesis aims to help bridge the gap between these two approaches. The first article introduces σOAM an extended version of the Optimal Audit Model (OAM) developed by Dionne et al. (2009) which directly uses the signal emitted by the classification algorithm, allowing a significant reduction in the fraud costs at the optimum compared to OAM. We add machine-learning (ML) classifiers as the base predictors in σOAM and compare its performances to other cost-sensitive methods and the original OAM. We carry out the tests over 3 different data-base configurations and with 4 different types of classifiers.

Our results show that σOAM outperforms its traditional cost-sensitive competitors when ML classifiers are used as base predictors. In the default configuration, the total cost of fraud is estimated at $9,844,643$\$. When the Logistic model is applied $504,509$\$ and $2,073,495$\$ are saved respectively in OAM and in the σOAM. The best ML classifier, Dart, without a cost-sensitive threshold selection saves $3,176,912$\$. σOAM with the Dart produces the best performance with $3,530,975$\$. σOAM significantly outperforms the original OAM and other tested cost-sensitive algorithms. Furthermore, σOAM displays better performances than the Elkan-Threshold framework as the model accounts for residual fraud through the deterrence effect produced by the optimal audit policy. To our knowledge, this is foreign to any existing cost-sensitive algorithm.

In our second article, we apply the first model of Cohen and Einav (2007) using Cana-

dian data on choices of insurance policy related to Collision coverages to estimate insured risk aversion. We observe that insureds deemed to have a higher risk of suffering accidents with bodily injuries have an average lower risk aversion. The risk aversion measure thus created is used as an added feature in the training of the expert system presented in the first article leading to a significant increase in classification performances and in savings over different scenarios. We also show that the newly added feature systematically ranks among the top predictors to detect fraud.

The third article aims to measure the deterrence effect of a chosen audit policy on a claimant's decision to commit fraud. In the original OAM, the deterrence effect was controlled by an exogenous parameter. However, in the third article, the deterrence effect is estimated through the hidden expected cost to the claimants, using signals from the claims to update the principal's belief about the fraud rate conditional on having a claim. The resulting audit policy is analyzed and confronted to findings from previous key studies. The results indicate that full deterrence was optimal for most groups of claimants, though it required varying audit levels depending on the shape of the expected loss distribution.

Our findings are consistent with those of Mookherjee and Png (1989), Fagart and Picard (1999) and Boyer and Peter (2020): higher compensation claims should be audited more frequently, as the most influential factors in the final audit policy are the cost of investigation and the potential reward. Inexpensive claims from 'Dishonest' types are deterred to an average of 80% by an audit policy as low as 0.00001, making further auditing unnecessary. On the other hand, 'High-risk' claimants with expensive claims have fraud present at equilibrium due to higher investigation costs and lower average risk aversion, making full deterrence impractical.

Although rewards and costs are pivotal in deciding the optimal audit policy, a higher risk aversion parameter results in a lower optimal audit. Our framework also demonstrates that the percentage of 'Dishonest' types increases as claimants transition to 'High-risk'.

Contrary to Boyer and Peter (2020), the optimal audit policy does not increase with risk type. In fact, 'Low-risk' claimants with expensive claims are audited even more than 'High-risk' claimants with similar claims.

Chapter 1

Auto-insurance fraud: a new cost-sensitive expert system.

Abstract

Ex-post moral hazard leads to less than optimal risk allocations between economic agents. In the insurance context, fraud is defined as an agent misrepresenting his true loss. Economists have demonstrated that the role of an audit strategy is twofold: to detect fraudulent claims as early as possible and to deter potential future fraud. In recent years, insurance companies have invested heavily in integrating machine-learning (ML) algorithms into their existing fraud detection processes as advanced expert systems. While powerful, such supervised expert systems cannot serve as the sole base of an audit policy as they fail to estimate the overall fraud rate faced by an insurance company. The Optimal Audit Model (OAM) developed by Dionne et al. (2009) is a cost-sensitive expert system that accounts for the elasticity of the fraud rate in response to a change in audit policy. Our contributions are fourfold: (1) we adapt the OAM model to function with any potential machine-learning

algorithms, significantly reducing the total expected cost of fraud; (2) we modify the signal treatment of OAM to create σOAM, a new cost-sensitive framework that fully leverages the enhanced predictive capabilities of modern ML classifiers; (3) we demonstrate that σOAM outperforms other widely tested cost-sensitive expert systems by being more aggressive toward groups of claimants with higher expected fraud rates; (4) we evidence the increased importance of law firms in patterns identifying fraud.

1.1 Introduction

Over the years, insurance companies have spent a considerable amount of resources establishing in-house business processes to audit suspicious claims. As presented in Figure , the current fraud detection process is comprised of sequential layers, each feeding information to the next one. The critical *triage* layer is composed of human teams who decide if a claim should undergo a full audit. Given the volume of claims supplied, a triage team has very little time to evaluate each claim. An insurance claim is a complex, high-dimensional entity, making short assessments inefficient. As automated expert systems offer an instantaneous overview of a high number of dimensions and produce consistent recommendations, they are prime candidates to assist triage experts in their workflow. However, an audit decision must first and foremost be a business decision. We propose a new expert system that integrates machine learning (ML) classifiers to produce an estimated fraud probability. Once a fraud probability is estimated, an audit decision is recommended based on three monetary factors: the cost of investigation, the expected cost of the claim, and the estimated deterrence effect on potential fraud entailed by the chosen audit policy.

Our study sits at the crossroad between economic and data-science literature. The use of expert systems in insurance fraud detection has been theorized in economic liter-

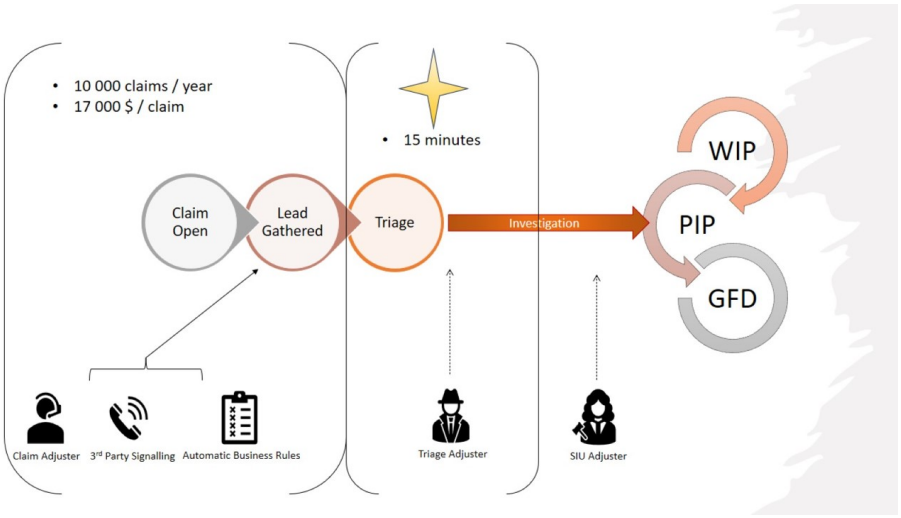


Figure 1.1: Current fraud detection process.

ature following the assessment that an optimal contract induces an audit-zone [1]. Classifying algorithms such as logistic regression, tree, Bayesian classifiers, boosted tree, etc,[2] have been tested as expert systems in this setting. A comprehensive review of the use of data-science in automobile insurance fraud is offered in Banulescu-Radu and Kougblenou (2024).

We modify and extend the Optimal Audit Model (OAM) developed by Dionne et al. (2009) to the Extreme Gradient Boosting (XGB) framework [3]. Our main contribution is to introduce σOAM, an extension of the OAM that is adapted to function with Machine-Learning (ML) models better utilizes the signal emitted by the XGB model. We compare this new cost-sensitive model to existing cost-sensitive algorithms and analyse the observed increase in cost savings. This study also contributes to the red-flags indicators literature, which focuses on selecting key features that best predict fraud [4].

[1] See Picard (2013), Mookherjee and Png (1989), Townsend (1979)
[2] Viaene et al. (2002, 2004), Majhi et al. (2019), Sun et al. (2007)
[3] See Chen and Guestrin (2016)
[4] See Weisberg and Derrig (1998), Belhadji et al. (2000), Dionne and Belhadji (1996)

The first article introduces σOAM an extended version of the Optimal Audit Model (OAM) developed by Dionne et al. (2009) which functions with machine-learning classifiers and directly uses the signal emitted by the classification algorithm, allowing for a significant reduction in the fraud costs at the optimum compared to OAM. Then, we compare σOAM's performances to other cost-sensitive methods and the original OAM. We carry out the tests over 3 different data-base configurations and with 4 different types of classifiers.

The rest of this paper develops as follow: the next section presents the literature review. Section 1.3 introduces the database upon which the model was trained. We then introduce the OAM and our modified version σOAM, its parametrization methodology and its adaptation to the XGB framework. Section 3.5.2 is composed of five different subsections. They present: the list of red-flag indicators ubiquitous in our models, the performance of the base classifiers, the performances of the four tested cost-sensitive frameworks and a deeper analysis of the previous two best performing frameworks. Finally, we present a sensitivity analysis to assess the influence of the parametrization on our global results. Section 1.6 concludes our study.

We show that the OAM model benefits more from the introduction of machine-learning models compared to the logistic regression model. We note an increase in performance measured both in dollars saved and in classification metrics. The extension σOAM increases these performances by directly relying on the output of the classifiers to create the discrete vector signal defined in Dionne et al. (2009). σOAM outperforms the two previously tested cost-sensitive frameworks by focusing the audit specifically on groups of claimants with higher potential of fraud. We also show the growing involvement of law firms in the claim process.

1.2 Literature Review and Main Contribution

1.2.1 Literature Review

According to the definition provided by the Canadian government, fraud to insurance consists in *"...any act committed with the intent to fraudulently obtain payment from an insurer. Insurance crimes range in severity, from slightly exaggerating claims to deliberately causing accidents or damage"*[5]. As this definition points out, if insurers could have perfect information about the outcome of a claim, fraud would not be such a major concern. An insurance claim is a complex object, involving many actors such as claimants, medical teams, body-shops, legal advisors, third-parties, etc. The signal produced by each claim is necessarily multi-dimensional and therefore potentially distorted at many levels. Insurance companies have invested substantial human and financial resources to reduce their information gap with the insureds.

The two main parties of each insurance contract are the insured and the insurance company. Insureds find value in insurance products because they allow for the smoothing of wealth in case of unpredicted events. Effectively, the insured, or agent, enters into a contract with an insurance company, or principal, to share an exogenous risk between the two parties. An insurance contract is characterised by a lost amount x, a reimbursement function $I(x)$, and a premium P paid by the agent to the principal. Costs associated with the processing of a claim can be modeled either as a constant C or as a function of the loss $C(x)$.

Designing an optimal audit strategy has been a major focus in the economic literature since economists started to model interactions between principals and agents through con-

[5] https://www.canada.ca/en/financial-consumer-agency/services/financial-toolkit/insurance/insurance-3/11.html

tract theory. Concerning the insurance industry, the main focus has been to characterise an optimal contract producing a Pareto-efficient allocation of risk between each party (Arrow (1964, 1974)). In particular, Raviv (1979) showed that contracts with deductibles strictly dominate coinsurance contracts in the presence of proportional transaction costs and absence of asymmetric information.

The loss of a claim is not strictly exogenous as the agent can select a level of effort that minimizes his expected loss. When effort is not observable by the insurer, this situation leads to a moral hazard problem. As the principal can only partially observe the agent's activity, the optimal contract must provide an incentive to increase the agent's effort. Mirrlees (1976) shows that a deductible is a valid incentive to encourage such a high level of effort. Holmstrom (1979) proves that moral hazard warrants the use of a monitoring set-up to further strengthen the agent's commitment.

Incorporating monitoring costs, Townsend (1979) produced the now-canonical Costly State Verification model (CSV). In this framework, only the agent is fully aware of the loss and can be tempted to dissimulate its true value to the principal. In the context of the insurance industry, the insured could be tempted to exaggerate the magnitude of the loss. The model assumes that auditing is always reliable in unveiling the true nature of the loss.

The desired outcome is an optimal contract, consistent with the participation and profit constrains, where lying about the realization of the loss is always a dominated strategy from the agent's perspective. A CSV optimal contract is characterized by an audit region which triggers an automatic investigation for any loss falling into it. In an insurance setup, the CSV model recreates a deductible type of contract. This creates a non-audit region that includes small losses that have negative actuarial value once the audit cost is included.

Many authors further refined the CSV model. Gollier (1987) includes a fixed cost structure of audit creating nuisance claims with negative actuarial values which cannot be audited and hence are not insured. Expanding on the principal-agent model developed

by Grossman and Hart (1983), Mookherjee and Png (1989) derived the conditions under which random auditing dominates deterministic auditing. The authors showed that substantive penalties to audited fraudulent claims act as a sufficient deterrent to potential fraudsters and reduce the global cost of audit.

Furthermore, when claimants have the possibility to manipulate signals[6] through third-party activities[7], Crocker and Morgan (1998) shows that the resulting optimal contract entails a level of misrepresentation for all losses but the extreme ones. The authors demonstrate that if the misrepresentation monotonically increases in the loss level, then the optimal contract is also monotonic and provides over-insurance for small losses and under-insurance for large losses. Dionne and Gagné (2001) empirically highlighted the relation between a high deductible and a propensity to inflate losses in the auto-insurance industry.

Since the volume of claims prohibits any credible promise of a full audit of any significant loss region, random auditing seems to be a dominating strategy (Mookherjee and Png (1989)). So far, the auditing decision was solely based on the amount claimed. Any methods improving the random odds to audit a fraudulent claim would simply further decrease the cost endured by insurance companies. Insurance claims are complex objects involving a large number of different parties. Levying extra information through expert systems is a cost-efficient and scalable practice widely used in the industry.

A large part of the empirical literature focuses on independent variables, or red-flag selection to detect fraud. Weisberg and Derrig (1998) proposed a consolidated list of red-flag indicators for bodily injury claims after conducting interviews with adjusters and special investigators. Belhadji et al. (2000), following the study done in Dionne and Belhadji (1996), used a Probit model to identify a list of significant indicators to predict fraud in the property damage auto-insurance claims.

[6] See Subelj et al. (2011), Bodaghiand and Teimourpour (2018)

[7] Garage, medical examiner, towing, etc.

Since its development by Chen and Guestrin (2016), the XGBoost classifier has been well tested on insurance datasets (see Pesantez-Narvaez et al. (2019)). XGBoost is an ensemble learning framework where base learners are used sequentially to navigate a selected loss function. Each base learner is used to further refine the classification of the sum of the predictions made by all the previous base learners, focusing on rectifying classification errors of the previous iterations. Base learners are here either trees or linear regression. To interpret the results obtained by XGB, Shapley-Value is often used to rank the independent variables in terms of their contribution to the prediction.

Shapley value analysis was developed by Lloyd Shapley in the game theory literature and is now widely used to interpret machine learning models. As described in Lundberg et al. (2018) and Lundberg and Lee (2017), SHapley Additive exPlanation, or SHAP, value analysis is a consistent and model-agnostic method to evaluate the importance of individual features in the prediction. The fundamental idea behind SHAP method is to sequentially add the contribution ϕ_i of each feature to the general expected value $E[f(x)] = 1$ given a vector of characteristics x to recreate the model's prediction $f(x)$. For a non-linear model $f(x)$, the order in which features are introduced matters. Hence, all combinations of the order of features need to be evaluated and averaged to obtain a global estimate of the contribution of each individual feature to the model's output. Individual feature contributions are then ranked by their absolute averaged score. We use the R package *"SHAPforxgboost"*[8] to isolate a set of size n top contributing variables based on their averaged absolute contribution.

[8]https://liuyanguu.github.io/post/2019/07/18/visualization-of-shap-for-xgboost/references

1.2.2 Main Contribution

The selection of a classifier used as the expert system becomes now a crucial aspect. A large list of classifiers has been studied, such as regressions, trees and boosted trees, Bayesian predictors, shallow and deep neural networks, SVM, KNN, etc.[9]. Supervised and unsupervised versions of these learning methods have both been applied[10]. A classifier's optimality varies regarding the various parameters of the problem, such as interpretability, data availability, noise in data, prediction robustness, and cost reduction. Cost-sensitive classifiers are a type of classifier which aims to minimize the cost associated with the predictions, rather than maximizing a classification metric[11].

Dionne et al. (2009) offer a unified approach to fraud detection by incorporating the score signal into an asymmetric information setting with monitoring costs similar to Townsend CSV model. The Optimal Audit Model (OAM) allows for ex-post signal manipulation by the claimant but also takes into account the deterrence effect of an audit policy, the idea being that if an insurance company decides to fully audit every claim, no one is likely to ever fraud again. The OAM is akin to other cost-sensitive frameworks because it selects a threshold based on a cost metric rather than prediction accuracy and, to our knowledge, is the only cost-sensitive model accounting for the entire cost of fraud through the deterrence effect.

In Dionne et al. (2009), the optimal threshold is derived from the set of binary predictors with significance levels above 99% in a logistic regression to predict fraud. OAM was tested on a sample consisting of 2,000 original claims from a European company. The estimated total cost of fraud for the company over a one-year period amounted to 622 million euros. The model contributes to a 60% reduction in the expected total cost

[9] See Viaene et al. (2002, 2004), Majhi et al. (2019)
[10] See Nian et al. (2016) and Brockett et al. (2002)
[11] See Sun et al. (2007), Höppner et al. (2022)

for the company. Thanks to the inclusion of residual fraud, the model demonstrates that investigation units should not be evaluated merely on profit, as it is optimal to audit claims with negative expected profit for deterrence effect. Our paper prolongs the work of Dionne et al. (2009) by incorporating new machine learning classifiers into the OAM framework, offering an extension to the original theoretical OAM and testing it against modern cost-sensitive frameworks.

Two cost-sensitive frameworks are tested: CSboost and Elkan threshold. CSboost is a cost-sensitive version of XGBoost developed by Höppner et al. (2022). Using the results obtained by Elkan (2001), the authors modified the XGBoost algorithm to a cost-sensitive boosting algorithm where each newly added weak learner is trained to minimize the Average Expected Cost. The algorithm is implemented via the R package *"CSboost"*, developed by Höppner et al. (2022). The algorithm does not allow for missing entries in the data, resulting in an extra loss of observations.

Elkan-Set threshold is a cost-sensitive method developed by Elkan (2001). The method consists of selecting a prediction threshold p^* that minimizes the expected cost of audit, accounting for the various costs of classification, including the cost of a false-positive classification. The optimal prediction threshold is calculated on a calibrated output. We use the isotonic regression developed by Jiang et al. (2011) as the calibrating function.

Our modified OAM, σOAM, saves 34.6% of the cost of fraud, which is evaluated as the distance between perfect prediction and the corner solution of not auditing. The second-best performing cost-sensitive model using an ML classifier manages a gain of 30.2%. The difference in performance is due to the deterrence effect of the OAM framework, which results in targeting groups of claimants where more cases of false negatives are present.

1.3 Data

The database, provided by a Canadian insurance company, is comprised of 70,813 claims with *Accident Benefit* (AB) elements opened from 2012 to 2020. An AB claim consists of anyone *"... who has been injured in a motor vehicle collision can claim for accident benefits. Pedestrians and bicyclists who have been in contact with a vehicle qualify for accident benefits."*[12]

Particular care has been taken in selecting the independent variables among the 800 features originally available. In our context, only variables available early in a claim's life were selected. This requirement severely impairs our final performance but eventually better reflects the true models' performance in a real situation of fraud detection where the model needs to be run as close as possible to the opening of the claim. A claim's life cycle spans from its opening to the day it is closed. Our limitations arise from two reasons. The first reason has to do with the difficulty of recovering any amount already disbursed to a fraudulent claim, reducing any potential gain from an audit. The second reason pertains to information availability. If a claim is suspicious to an adjuster, the said adjuster can ask to obtain various extra-specific variables. These variables will artificially inflate our classifier's performance, acting as biased predictors.

1.3.1 Dependent variable

Claims are labeled by three classes: *"Suspicious"* with 2% of the population, *"Fraud"* with 3%, and *"No_label"* with 95% of the population. *"Suspicious"* and *"Fraud"* claims are observations that have been transferred to the National Fraud Tracking System (NSTF) and not discarded after 5 days. *"Suspicious"* claims are claims where clear suspicious

[12] https://www.aboutkidshealth.ca/article?contentid=1266&language=english

activities were detected, but no definite amount of money was retrieved or disbursed, as opposed to *"Fraud"* claims.

Claims are sent to the triage phase following the appearance of one or several leads. The main sources of leads, in terms of volume, are human adjusters who handle the regular process of claims, in-house and external detection models, and the watch-list suppliers list. These various sources produce different qualities of leads, resulting in diverse shades of suspiciousness. It was, unfortunately, not possible to group observations according to their source origin with our database.

1.3.2 Literature indicators

A part of the literature on fraud detection for bodily injuries has been dedicated to the discovery of meaningful early indicators. Weisberg and Derrig (1998, 1991) asked adjusters and investigators to grade claims respectively on a suspicious/fraud scale. Each possible subset from the indicators list of up to ten variables was then fitted through a regression analysis. The best five models to predict suspicious/fraudulent claims were kept and further analyzed. The authors noted a core set of 29 indicators among the 65 tested variables.

A few remarks should be made concerning the indicators from Weisberg and Derrig (1998). The authors aimed to recreate the classification made by human teams to understand the factors that were consistently looked upon. These indicators were created through interviews where adjusters were answering *Yes* or *No* questions related to their claims. Some indicators are therefore subjective and only appear late in the claim, such as: *"Claimant was uncooperative"* or *Long disability for a minor injury*. These types of indicators cannot be used in the framework of this study, as our goal is to develop an early detection system. Furthermore, observations are here aggregated at the contract level,

making the tracking of a single injury difficult. Finally, the data extraction process itself needs to follow strict regulations concerning confidentiality and security, making specific data extractions difficult. Appendix B presents all 65 variables used in Weisberg and Derrig (1998). It indicates the 22 indicators that can be recreated using our partner database.

1.3.3 Final raw variables dimensions

Once as many of the known indicators are replicated from the insurer's database, the first step in filtering the remaining features is to discriminate the variables known to be late in the detection process, too sparsely populated, incoherent, or redundant. There is an asymmetry of knowledge between suspicious/fraudulent and normally processed claims, as extra information is often required for the former. It is hence not surprising to observe an uneven distribution of information, with post-fact variables being more and more sparsely populated. Around 65% of the feature variables have less than 50% of observations missing, resulting in around 370 sparse variables that needed to be discarded.

The final 126 extracted features can be grouped into 10 distinct sets, controlling for diverse aspects of a claim. Appendix D presents them before preprocessing and feature engineering. Some variables are not directly used by the classifiers but are necessary for the preprocessing and feature engineering steps. Appendix E and Appendix F give a detailed focus on analyzing the features composing the final database. Features are analyzed following two categories: continuous or categorical. The final database before one-hot encoding is composed of 70,412 observations with 126 feature variables. After one-hot encoding the factor variables, the number of feature variables rises to 207. One-hot encoding consists of breaking up a categorical variable comprised of many categories into a set of dummy variables, one for each previous category.

1.4 Models

1.4.1 Logistic Regression

Equation (1.1) presents the logistic model. $\hat{y}$ is the latent output of the linear model. In a traditional classification problem, the *True* region is defined by the threshold: when $\hat{y} > 0.5$. Variables are grouped into families controlling for different dimensions of the claim.

$$\begin{aligned}\hat{y} = \alpha_1 Veh\_Age + \alpha_2 PolReport + \alpha_3 CoseOfLoss + \alpha_4 HowReported \\ + \alpha_5 PointOfImpact + \alpha_6 Occupation + \alpha_7 LawVar + \alpha_8 Towing \\ + \alpha_9 AccidentCharact + \alpha_{10} VehCat + \alpha_{11} VehMake + \alpha_{12} BodyShop \\ + \alpha_{13} Educ + \alpha_{14} Reserve + \alpha_{15} Injuries + \varepsilon\end{aligned} \tag{1.1}$$

Table D.8 in Appendix D presents the details of the independent variables chosen for the regression.

1.4.2 XGB Classifiers

Developed by Chen and Guestrin (2016), ExtremGradientBoosting (XGB) is a supervised learning algorithm which can be applied to classify instances in classes. XGB is an ensemble learning framework that relies on the iterative use of weak classifiers [13] to correct its loss function.

Given a convex and differentiable loss function $l(y_i, \hat{y}_i)$, where $\hat{y}_i$ represents the model's prediction, XGBoost introduces a base learner $f_t(\mathbf{x})$ at iteration t in order to minimize the approximate general loss function $\hat{t}^t$. Here, $\mathbf{x}$ denotes the vector of independent variables.

[13] Commonly classification tree or linear regression

The goal at each iteration is to refine the model by adding $f_t(\mathbf{x})$, which further reduces the loss function, improving the model's overall predictive accuracy.

$$\hat{\imath}^t = \Sigma_{i=1}^n \left[l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{1.2}$$

Eq(1.2) is the second order Taylor expansion of the exact loss function at point $f_t(\mathbf{x_i})$ where y is the true label, $\hat{y}$ is the assigned label by the model, $\mathbf{x_i}$ is the vector of independent variables for observation i, and f_t is the t^{th} learner added to the prediction of $\hat{y}$. The $\Omega(f_t)$ is a regularization term described in (1.3). The regularization $\Omega(f_t)$ penalizes the model by the number of leaves T and the weight given to each observations at a given leaf w. The parameters γ and λ are left to the discretion of the user and respectively control the size of the base learner and the smoothing of individual predictions.

$$\Omega(f_t) = \gamma * T + \frac{1}{2} \lambda ||w||^2 \tag{1.3}$$

Eq(1.2) adds to the previous loss function $l(y_i, \hat{y}_i^{t-1})$, the added weight of the t^{th} base learner $f_t(\mathbf{x_i})$ for observation $i, i \in [1, 2, ..., n]$. Note that the term $l(y_i, \hat{y}_i^{t-1})$ in (1.2) is a constant and hence does not contribute to the optimization. Therefore (1.2) simplifies to (1.4):

$$\hat{\imath}^t = \Sigma_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{1.4}$$

where g_i and h_i are respectively the first and second order derivative of $l(y_i, \hat{y}_i^t)$ at point $\mathbf{x_i}$. The extra weight added by the t^{th} base learner for observation i is simply the weight of the leaf containing the observation i. (1.4) can be rewritten in (1.5) and (1.6). Let $I_j = \{i / q(\mathbf{x_i}) = j\}$ the leaf j containing the observation $\mathbf{x_i}$:

$$\hat{\imath}^t = \Sigma_{i=1}^{n}\left[g_i w_i + \frac{1}{2}h_i w_i^2\right] + \gamma T + \frac{1}{2}\lambda\Sigma_{j=1}^{T} w_j^2 \tag{1.5}$$

$$= \Sigma_j^T\left[\Sigma_{i\in I_j}(g_i)w_j + \frac{1}{2}(\Sigma_{i\in I_j}h_i + \gamma)w_j^2)\right] + \gamma T \tag{1.6}$$

where (1.6) is obtained from (1.5) by summing the added weight by leaf instead of by observation. For a fixed structure of tree $q(\mathbf{x})$ for each base learner, one can extract from (1.6) the optimal form for w_j^*, the weight of leaf j:

$$w_j^* = -\frac{\Sigma_{i\in I_j} g_i}{\Sigma_{i\in I_j} h_i + \lambda} \tag{1.7}$$

Replacing the optimal weight w_j^* in (1.6) results in the scoring function $\hat{\imath}^t$. $\hat{\imath}^t$ is used to select which base learner $f^t(\mathbf{x})$ is optimal to add at iteration t as described in (1.8).

$$\hat{\imath}^t(q) = -\frac{1}{2}\Sigma_{j=1}^{T}\frac{\Sigma_{i\in I_j} g_i}{\Sigma_{i\in I_j} h_i + \lambda} + \gamma T \tag{1.8}$$

Appendix G further details the splitting procedure to select the variables in each base classification tree.The hyper-parameters are selected after a grid search cross-validation procedure using a statistical approach of the F1 score performance measure developed by Goutte and Gaussier (2005) and detailed in Appendix H with results in Appendix I [14]

1.4.3 Optimal Audit Model (OAM)

This section presents the OAM of Dionne et al. (2009) and its extension σOAM. An audit strategy is defined as a function mapping each claim to a recommendation of audit. This function uses as input a signal emitted by a base classifier such as a logistic regression

[14] $F_1 = \frac{2\text{tp}}{2\text{tp} + \text{fp} + \text{fn}}$, where tp = true-positive, fp = false-positive and fn = false-negative

or any other machine learning classifier. An audit function is said *"Optimal"* when it minimizes the total expected cost of fraud.

OAM differs from typical cost-sensitive models as it encompasses an economic approach at its roots. The fraudulent behavior is modeled at the customer level by a bivariable utility function $\mathscr{U}(W,\omega)$ where W refers to the wealth of the customer and ω refers to a fixed moral cost associated with the action of fraud. We assume $\mathscr{U}^{'}(W) \geq 0$ and $\mathscr{U}^{''}(W) \leq 0$ and $\mathscr{U}^{'}(\omega) \leq 0$. ω follows a distribution on $\mathbb{R}+$ over the entire population of claimants.

Figure 1.2 displays the decision facing a claimant. The claimant will defraud if and only if his expected utility is higher than the status quo, that is, (2.17) holds. Let A be the event that a claim is audited. If a claim is audited, its true nature is revealed. The expected utility is a function of: $P(A/F)$, the probability of audit conditional on the claim being fraudulent; B, a financial penalty imposed on a claimant caught being fraudulent; t, the average compensation the claimant can expect from a successful fraudulent claim; and W_0, the claimant's wealth after the insurance premium is paid.

$$(1-P(A/F))\mathscr{U}(W_0+t,\omega))+P(A/F)\mathscr{U}(W_0-B,\omega)) \geq \mathscr{U}(W_0,0) \tag{1.9}$$

Each claimant is segmented into θ groups. In the original study, θ segmentation uses claimant's characteristics known by the insurer independent of the distribution of moral

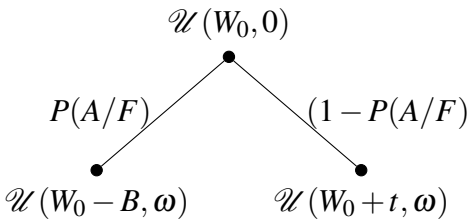


Figure 1.2: Claimant's choice

cost ω. In this study, θ are formed following the amount of AB Premium charged to the claimants, reflecting the risk segmenting done by the insurance company on its pool of claimants.

When (2.17) becomes an equality, the claimant is indifferent between the certain and uncertain state. The function $\phi(P(A|F)) : [0,1] \rightarrow \mathbb{R}+$, links the probability $P(A|F)$, the probability of audit conditionnal on the fact that a claim is fraudulent, to the moral cost ω corresponding to such indifference. Appendix J gives additional details on the characteristics of the $\phi(P)$ function.

An audit strategy is determined by two key factors: the θ group to which the claimant belongs and the signal emitted by the claim itself. Each claim's signal is represented by a vector σ of binary variables with dimension k. Since every $\sigma_i \in \Sigma$ is a binary variable, the total number of possible signal configurations is finite, with a cardinality of $2^k = l$. These signals, derived from the base classifier, reflect its top predictors. Let P_i^F and P_i^N represent the probabilities that the signal σ takes configuration i, conditional on the claim being fraudulent (F) or non-fraudulent (N), respectively.

$$P_i^F = P(\sigma = \sigma_i|F) \tag{1.10}$$

$$P_i^N = P(\sigma = \sigma_i|N) \tag{1.11}$$

Signals emitted are then ordered by their historical ratio $\frac{P_i^F}{P_i^N}$ as displayed in (2.14).

$$\frac{P_1^F}{P_1^F} < \frac{P_2^F}{P_2^N} < ... < \frac{P_i^F}{P_i^N} < ... < \frac{P_l^F}{P_l^N} \tag{1.12}$$

Dionne et al. (2009) demonstrate that the optimization problem of selecting the optimal audit policy for each θ involves determining a rank i in (2.14). Claims with a signal ranked higher than i should be audited, while those with a signal ranked lower than i

should be approved without further investigation. The chosen rank i reflects the level of aggressiveness of the audit policy for a given θ group.

Let $q : \Theta \times \Sigma \to [0,1]$ denote the audit strategy implemented by the insurance company. The audit strategy $q(\theta, \sigma)$ specifies whether a claim with signal σ should be audited based on the claimant's group θ_i. Define $Q^F(\theta)$ and $Q^N(\theta)$ as the respective probabilities of auditing a fraudulent or non-fraudulent claim for a claimant of type θ. These probabilities, $Q^F(\theta)$ and $Q^N(\theta)$, are determined by the chosen audit strategy $q(\theta, \sigma)$, as shown in (1.13) and (1.14).

$$Q^F(\theta) = \Sigma_{i=1}^{l}(P_i^F q(\theta, \sigma_i)) \tag{1.13}$$

$$Q^N(\theta) = \Sigma_{i=1}^{l}(P_i^N q(\theta, \sigma_i)) \tag{1.14}$$

The recommended audit policy aims to minimize the total expected cost of fraud, which consists of two main components: the expected cost of investigation (2.18) and the expected cost of residual fraud (2.19). The model excludes opportunistic fraud and assumes that a claimant either fully fabricates a fraudulent injury claim or that the injury is genuine. Therefore, π represents the probability of a claim being a legitimate Accident Benefit (AB) claim. Let $H(\omega/\theta)$ denote the conditional cumulative distribution of the moral cost for claimants in group θ, and $h(\omega/\theta)$ its corresponding density function.

$$c\pi E_\theta[Q^N(\theta)] + c(1-\pi)E_\theta[Q^F(\theta)H(\phi(Q^F(\theta)/\theta))] \tag{1.15}$$

$$t(1-\pi)E_\theta[(1-Q^F(\theta))H(\phi(Q^F(\theta))/\theta)) \tag{1.16}$$

Adding (2.18) and (2.19) and minimizing the sum over $q(\theta, \sigma)$ gives the global opti-

mization problem in (2.20).

$$Min_{q(\theta,\sigma)}E_\theta\Bigg[c\pi[Q^N(\theta)]+c(1-\pi)Q^F(\theta)H(\phi(Q^F(\theta)/\theta)) \\ +t(1-\pi)\Big((1-Q^F(\theta))H(\phi(Q^F(\theta))/\theta))\Big)\Bigg] \tag{1.17}$$

Proposition 1 of Dionne et al. (2009) shows that, under limited conditions, the problem of solving (2.20) consists in finding an optimal severity index i in (2.14) for each θ. Then the optimization problem becomes (2.21):

$$Min_{i_\theta}E_\theta\Bigg[c\pi\mu(i^*_\theta)+(1-\pi)H(\phi(\lambda(i^*_\theta))/\theta) \\ \Big(c\lambda(i^*_\theta)+t(1-\lambda(i^*_\theta))\Big)\Bigg] \tag{1.18}$$

where $\gamma(i)=\Sigma_{i=1}^{l}(P_i^F)$ and $\mu(i)=\Sigma_{i=1}^{l}(P_i^F)$. Due to the discretization of the top five predictors to form the σ signal vectors, it is possible to have a claim in the test-set whose signal σ is unknown in the training set. For such observations no predictions is possible as the optimal threshold *"i"* in (2.14) cannot be trained.

1.4.4 Parametrization of OAM

Dionne et al. (2009) describes an implementation methodology that is followed here. The authors start by approximating the average fraud rate Γ by extracting $\hat{\pi}$, the observed probability of filling a claim, fraudulent or not. Assuming a global fraud rate of $P(F/\theta)$, the authors deduce the probability of observing a fraudulent claim $\hat{\Gamma}(0,\theta)=\hat{\pi}P(F/\theta)$ and the probability of a genuine accident $\pi=\hat{\pi}(1-P(F/\theta))$.

$$\Gamma(Q,\theta)=(1-\pi)H(\phi(Q^F_\theta)/\theta) \tag{1.19}$$

$$\eta(Q,\theta)=\frac{-Q^F_\theta\phi'(Q^F_\theta)h(\phi(Q^F_\theta)/\theta)}{H(\phi(Q^F_\theta)/\theta)} \tag{1.20}$$

$\hat{\Gamma}(0,\theta)$ is the probability of observing a fraudulent claim observed with the actual audit policy. To estimate the effect of a deviation of the current audit policy we need to estimate the elasticity of the fraud rate with respect to the investigation strategy: $\eta(Q,\theta)$ in (1.20). Dionne et al. (2009) highlights the difficulty of estimating such function as it depends on the estimation of the moral cost distribution and on a parametrization of the function ϕ: $H(\phi(q)/\theta)$. The authors opted for the following approximation in (1.21):

$$\Gamma(Q,\theta) = \hat{\Gamma}(\theta)(1-Q)^{\gamma(\theta)} \tag{1.21}$$

(1.21) is used to approximate (2.21) into (1.22). The parameter $\gamma(\theta)$ controls the deterrence effect per θ or the sensitivity of a group's fraud rate to a change in the audit policy.

$$\begin{aligned} Min_{i_\theta} E_\theta \Bigg[& c\pi\mu(i_\theta^*) + \hat{\Gamma}(\theta)(1-\lambda(i_\theta^*))^{\gamma(\theta)} \\ & \Big(c\lambda(i_\theta^*) + t(1-\lambda(i_\theta^*))\Big)\Bigg] \end{aligned} \tag{1.22}$$

In our study, θ groups are formed by increasing deciles of Accident Benefit[15] (AB) premiums charged to the insured. Each observation is given an index from 1 to 10, 10 being the group with highest AB premium paid by the insured. Table 1.1 gives the population of claims per θ with our data as well as the thresholds forming each group.

The probability of filling an AB claim, $\hat{\pi}(\theta)$, is parameterized using all the active policies in 2019. Active policies are classified and summed per θ groups. All AB claims with a date of loss in 2019 are extracted from the main train database, grouped and count per θ. The ratio of the two extractions gives an estimate of the probability that an active policy would fill an AB claim per θ. The global probability of filling an AB claim is about 1% for the year 2019. This is lower than the average probability of being involved in an

[15] AB claims are claim with bodily injuries.

Table 1.1: Population and parameters per θ

θ	Pop	$\hat{\pi}(\theta)$	Thres$	P(F/θ)	$\pi(\theta)$	$\Gamma(0,\theta)$	$t(\theta)$	$c(\theta)$
1	15153	0.68%	0	11.6%	0.60%	0.07%	16997	2219
2	5303	0.66%	221	11.0%	0.59%	0.07%	15538	1639
3	5036	0.70%	283	11.7%	0.62%	0.08%	15032	1698
4	4970	0.78%	341	12.4%	0.68%	0.09%	14604	1662
5	5011	0.77%	402	13.4%	0.66%	0.10%	14813	1793
6	5349	0.95%	472	14.1%	0.81%	0.13%	14821	1868
7	5813	1.0%	558	15.4%	0.85%	0.15%	14653	1941
8	6478	1.2%	670	16.8%	1.0%	0.21%	15227	2112
9	7836	1.4%	834	18.6%	1.1%	0.26%	14947	2144
10	9463	1.5%	1137	22.3%	1.2%	0.35%	15061	2365

Note: Column *Pop* is the population per θ in the Train-set. $\hat{\pi}(\theta)$ is the probability of filling an AB claim. *Thres$* is the AB premium lower limit that form the θ group. *P(F/θ)* is the probability of fraud. $\pi(\theta)$ is the probability of accident. $\Gamma(0,\theta)$ is the probability of observing a fraudulent claim. $t(\theta)$ is the average compensation per claim and $c(\theta)$ is the average cost of investigation both in Canadian dollars.

accident of about 1.9%, given by the local government in 2018 [16]. Table 1.1 shows an expected increase in filling rate $\hat{\pi}$ in relation with the premium group of the claimant. A higher AB premium does reflect a higher risk of injury overall as is depicted in Table 1.1.

In Dionne et al. (2009), $P(F)$, the probability of fraud, was given by the industrial partner for a general representative customer at 8%. In consultation with our industrial partner, we assume a global fraud rate $P(F)$ of 15% and distribute it at θ levels following the methodology presented in Appendix K. We later conduct a sensitivity analysis for this base assumption. All results presented below in this article are calculated with this 15% hypothesis. We also consider that all investigated claims were truly fraudulent although no amount of money could be recovered. This hypothesis is also relaxed later when the target population is limited to claims from which an amount of dollars was recovered.

Table 1.1 shows a steady increase in the probability of fraud by index of θ group.

[16] https://files.ontario.ca/mto-2/mto-orsar-2018-en-2021-11-18.pdf

Claims with higher AB premium have a higher tendency to be fraudulent than claims belonging to the first decile of AB premium. The observed fraud rate in the database is about 5%, before the distribution of the residual fraud of 15% presented in Appendix K. 37% of the observed 5% is concentrated in the last AB premium decile. The residual fraud rate is assumed to be uniformly distributed at around 9.5% per θ.

The average compensation per claim, $t(\theta)$, is directly observed at the claim level. The database is comprised of 153 sub-files disbursement variables which track any amount disbursed by the insurer on a claim at the sub-policy level. The total amount is sum by θ and then averaged. Due to outliers, we impose a cap of 50000\$ in total compensation per claim[17]. This is not a strong assumption as any claim reaching a higher amount is automatically audited. We observe higher compensations for claims belonging to smaller AB premium. This is explained by a density of compensation with heavier tale for smaller AB premiums.

Estimated with the help of our industrial partner, the average cost of investigation, $c(\theta)$, is comprised of three different sub-costs: fixed cost of salaries of an adjuster plus online investigation $(700\$ + 300\$)$ per claim, the extra medical examination asked by the partner for each claim directly observed in the data and the extra surveillance cost in the interval $[4000, 7000]$.

We are left with estimating the cost and the expected cost per claim of extra-surveillance which corresponds to expenditures such as hiring private detectives. Appendix L presents a detailed breakdown of the methodology upon which the expected extra-surveillance costs are estimated. The major drive of the global investigation cost is the extra-medical expenditure which is directly observed in the data and increases for θ groups with higher AB premiums.

[17]The 3rd quantile is at 20000\$

1.4.5 Accident signal σ

After having parameterized the OAM, we need to input the signal emitted by the classifier. The original model from Dionne et al. (2009) allows for σ accident signals to only be a set of binary variables. The optimization process is hence simplified to selecting the optimal threshold *"i"* in (2.14)[18]. Dionne et al. (2009) used a logistic regression and selected the top significant variables at the 1% level to form the signal vector σ. Since variables were only binaries, they formed a finite set of possible signals. In our case, main predictors can be continuous or multilevel. In such cases, the variables are either discretized by quartiles or one-hot encoded to create a finite set of possible signal vectors σ. We use Shapley value analysis to select the top predictors among the ML models tested. Unlike in Dionne et al. (2009), such features in our data are most likely not binary. One-hot encoding is carried-out during the preprocessing for all factorial features extracted. Continuous variables are segmented into quantiles and then one-hot encoded themselves to form the signal vector.

1.4.6 σOAM

We propose an extension to OAM called σOAM. In σOAM the signal σ is created directly from the output of the classifier. An ML classifiers produces a continuous output that is used to rank observations. Traditionally, a cut-off function is optimized to select a threshold that maximizes a chosen classification metrics. We use the optimization function (1.22) as the cutoff selector. To obtain a finite set of potential signals, we discretize the output vector of the ML classifiers into centiles which form the signal sub-groups. The optimization problem is similar to (2.21) but *"i"* corresponds to the index of the signal sub-group in (2.14) that minimizes the expected cost of fraud. This extension provides several advantages.

[18] Proposition 1 in Dionne et al. (2009)

The first advantage relate to a steep improvement in performances. An upgraded ML classifiers will directly impact the accuracy of the predictions of the OAM layer and increase the performances in terms of dollars saved. The original OAM's F1 score relies solely on the information contained in the top variables contributing the most to the ML classifier's prediction. However, two ML classifiers may have a similar list of ranked predictors but different classification performances. Table B.6 in Appendix B shows the top predictors of each model in every configuration. In the Default configuration, which corresponds to the raw database presented in Section 1.3, the XGB framework with the Linear and Tree booster have almost an identical list of top predictors. Much of the $\{Model, Configuration\}$ pair have a similar list of top predictors. We later show in Table 1.4 of Section 3.5.2 that the decrease from columns F1 to $F1_{OAM}$ after applying the optimization function (1.22) is much more pronounced for OAM than for σOAM.

Table 1.2 details the different predictions made by the linear booster in the default configuration using σOAM or OAM. There are a total of respectively 120 and 100 i subgroups for each model. Column Rank shows the global aggressiveness of both OAM and σOAM. A higher rank means that a signal with a lower ratio $\frac{P_i^F}{P_i^N}$ is audited. We see that σOAM audits from ranks 14 ($i = 86$) to 16 ($i = 82$) onward, sending to be audited about 67.6% of the fraudulent claims and 12 % of the non-fraudulent claims. OAM model is more aggressive, auditing claims with signal ranking from 34, ($i = 22$) to rank 38 ($i = 100$). This audit policy leads to a similar audit rate of fraudulent claims of 66.8% while a higher 19.4% of non-fraudulent claims are audited. This difference of 7.4% in the percentage of non-fraudulent audited claims makes σOAM outperforms OAM in terms of amount saved.

There are additional advantages to using σOAM to OAM. Results shows that σOAM is more robust than OAM to different configurations but is also less difficult to parameterize as different choice of discretization of selected continuous variables would lead to different sets of results. Another nice feature is that predictions are always possible on a test-set as

all signals are always comprised within the training-set. In OAM if a continuous variable is selected among the top predictors, it is tempting to increase the precision of the model by creating smaller subgroups of index *i* instead of the 4 quartiles used in this study. However, this will increase the probability of receiving unknown groups of signals from the test-set making predictions impossible. σOAM is not hindered by such dilemma and will directly benefit from any upgrade made to the classifying technology.

1.5 Results

Two types of results are obtained in the optimal audit model. The first consists in the F1 score generated by the base classifiers [19] as a measure on the efficiency of the classification of claims on the test-set. The second is the cost-saving estimate resulting from the use of the OAM. The test-set is comprised of 14080 observations in the default configuration and is created through a stratified-sampling procedure implemented by the R-package *"SamplingStrata"* developed by Barcaroli (2014)[20]. Appendix M gives the list of target variables used to obtain an accurate representative test-set.

Table 1.3 presents the 3 configurations tested here. The *"Default"* configuration includes all the 208 features variables after preprocessing. The *"Fraud-Only"* configuration removes all observations that are labeled *"Suspicious"*[21]. Finally, the *"No-Law"* configuration removes all variables related to the presence of a law firm in the claim process. Law variables[22] represent a potential bias as a claimant might solicitate a lawyer because he was signified by the insurer of a potential investigation. Because of this bias, the default

[19] XGBoost tree,dart and linear and the Logistic regression

[20] For a comprehensive introduction to the package please refer to: https://www.istat.it/it/files/2014/06/SamplingStrata-An-R-Package-for-the-Optimization-of-Stratified-Sampling.pdf

[21] All observations where no amount of $ where recovered, around 50% of the target population

[22] See Table D.8 in Appendix D

Table 1.2: σOAM and *OAM* Linear booster in default configuration

	σOAM					OAM				
θ	Rank	*i* groups	Thres	% F Audit	% NF Audit	Rank	*i* groups	Thres	% F Audit	% NF Audit
1	11	90	86	0.58	0.08	31	19	22	0.60	0.16
2	12	89	86	0.61	0.09	32	101	104	0.60	0.17
3	13	88	86	0.64	0.10	33	130	104	0.60	0.17
4	14	86	86	0.66	0.11	34	22	31	0.64	0.18
5	15	87	86	0.68	0.12	35	104	31	0.66	0.19
6	16	82	86	0.69	0.13	36	31	31	0.66	0.19
7	17	85	86	0.71	0.14	37	26	31	0.69	0.19
8	18	80	86	0.72	0.15	38	100	26	0.69	0.20
9	19	84	87	0.74	0.16	39	21	26	0.70	0.21
10	20	83	82	0.75	0.17	40	18	100	0.70	0.22

Note: Rank is the increasing indexed σ subgroups created by each claim. As the rank increases, the True Positive Rate decreases. Thres is the optimal signal threshold below which no claim should be audited. % F Audit and % NF Audit are respectively the percentage of fraudulent and non-fraudulent claims audited.

Table 1.3: Base classifier Default Configuration

Config Name	Classifiers	# features	AB_label	Law
Default	Tree/Dart/Linear & logis	190	Investigated	Present
No-Law (NL)	Tree/Dart/Linear & logis	186	Investigated	Absent
Fraud-Only (FO)	Tree/Dart/Linear & logis	190	Fraud	Present

Note: Column *"Config Name"* displays the name of the configuration. *"Classifiers"* is the type of classifiers used. *"# features"* is the number of variables used in the train-set. In column AB_label, "Investigated" indicates that all suspicious claims are considered fraudulent. In column AB_label "Fraud" indicates that only observations with an amount of dollars saved are used as the target population. *"Law"* is the presence or absence of law variables in the training-set.

configuration already censors the law variables of observations that had lawyers arriving 36 days or after the date of the loss. Despite this censorship, law variables are constantly among the top predictors of fraud and a *"No_law"* configuration is tested to gauge the robustness of the models.

1.5.1 Red-flags indicators

Out of 126 variables[23] in the dataset, there is a set of 25 core variables which are ubiquitous in predicting fraudulent claims. Out of these 25 variables, 8 of them are present in the list of top 8 main predictors between 5 and up to 11 of the 12 tested models. Despite the censorship of 36 days mentionned above, we notice that 4 of these 8 features control for different aspects of lawyers on a claim. Selected law variables control for law-firms with: a high specialty in insurance claims, their presence on a claim, the number of lawyers present and the number of days before their first arrival on a claim[24]. In the *"Default"* configuration, these variables are set to 0 if the first lawyer arrived on the case after 36 days of the loss as a way to control for lawyers hired because a claim was already declared suspicious.

The other 3 main predicting variables are: the number of years the claimant has had his driving licence, the number of days since the claimant has been insured with the industrial

[23] See Appendix D

[24] Law_rep_agrr, Law_rep_none, DTd_los_law, Nbr_lawyers

partner, [25] and abnormal channels of reporting the accident to the insurer [26]. The *"No-Law"* configuration adds red flags related to the types of injuries such as the presence of a back injury, a high psychological trauma and the delay between the date of loss and the first physical injury declared [27].

Between the 4 different classifiers tested in both *"Default"* and *"Fraud-Only"* configurations, there are 40 different potential predictors to be extracted. Appendix A shows that a succinct list of 8 variables already fills 36 slots of the 40 possible slots. Only one red flag is added to the list [28]. σ signals in OAM between the two configurations are hence very similar as they are formed by the top 5 predictors of each model.

Weisberg and Derrig (1998) listed 29 core predictors to detect fraud in bodily-injury auto-insurance claims. Due to difference in granularity between the data in our study and the indicators controlled by Weisberg and Derrig (1998), a one-to-one mapping is difficult. However, out of the 25 main predictors of this study, 6 belong to dimensions cited by adjuster as indicative of fraud in Weisberg and Derrig (1998). Another 3 core indicators were tested but not selected in Weisberg and Derrig's core indicators. All but 2 of the 8 variables that form the most important subset of core predictors belong to the red-flags list selected by Weisberg and Derrig (1998). Notably Law variables were tested in Weisberg and Derrig (1998) but were not cited by adjuster at the time as key indicators. This underlines a clear evolution in the industry as the importance of law firms in claims settlement has increased significantly, sometime leading to fraudulent practices as evidenced in this study. Appendix A lists all the 25 key predictors.

[25] The two variables are correlated because of regular client turnover
[26] Drv_licence_years, DTd_dol_Oicd, HR_unknown
[27] Inj_Back, MostSeverePsyMax, DTd_FAB_Dol
[28] Law_rep_moderate

Table 1.4: Optimal Audit Model results

			OAM		σOAM		Elkan		CSboost	
Config	Classifier	F1	$F1_{OAM}$	Sav$	$F1_{OAM}$	Sav$	$F1_{OAM}$	Sav$	$F1_{OAM}$	Sav$
Default	Logistic	0.336	0.287	540,509	0.31	2,073,495	0.350	2,659,740	×	×
	Dart	0.401	0.302	2,220,976	0.385	3,530,975	0.402	2,973,675	0.297	1,482,911
	Linear	0.398	0.251	849,518	0.361	3,270,789	0.389	3,034,209	0.297	1,397,911
	Tree	0.384	0.27	1,521,527	0.358	2,289,313	0.339	2,580,260	0.299	1,402,410
Fraud-Only	Logistic	0.323	0.183	-1,250,041	0.229	-122,484	0.31	1,154,900	×	×
	Dart	0.383	0.212	-326,884	0.306	1,257,100	0.349	1,566,847	0.28	751,628
	Linear	0.356	0.172	-1,735,863	0.272	780,840	0.358	1,504,290	0.28	762,392
	Tree	0.318	0.197	-563,346	0.318	1,143,741	0.293	1,077,348	0.28	751,628
No-Law	Logistic	0.153	0.195	-1,490,538	0.222	461,022	0.257	1,058,786	×	×
	Dart	0.301	0.185	-1,048,756	0.287	1,738,962	0.308	1,991,014	0.095	313,487
	Linear	0.281	0.167	-2,668,173	0.263	1,106,568	0.265	1,397,390	0.091	298,356
	Tree	0.254	0.181	-1,890,283	0.221	1,073,631	0.242	1,076,887	0.091	12,653

Note: Columns $\{Config, Classifiers\}$ introduce the configuration and the classifier. σOAM and OAM are both estimated with a γ parameter of 2. *Default* configuration has all variables related to the presence of lawyers and type of law firms on the claims in the training set and also includes all suspicious claims as fraudulent. *Fraud-Only* limits the target population to claims with amount of money recovered. *No-Law* removes all variables related to law from the training set. The first row introduces the Cost-sensitive framework OAM, σOAM, Elkan and CSBoost,.... *F1* is the F1-score of the algorithm used on the test-set. Column $F1_{OAM}$ is the F1 score after the minimization of cost is applied and *Sav*$ is the amount saved to the corner solution of not-auditing introduced in Table 1.5. A negative amount reflects that the corner solution of *No-audit* is more profitable.

1.5.2 Classifiers Results

The stand-alone $F1$ column in Table 1.4 presents the raw classification scores that reflects how well each model performs at the task of distinguishing between a suspicious and a regular claim. Table 1.4 shows an overall better performance of the XGB family classifiers (Dart, Linear and Tree) compared with the Logistic regression with higher classification performances for each classifiers in every configuration except the pair *{"Fraud-Only",Tree}*. Out of the 3 base classifiers, Dart shows the overall best performance with an average F1-score of 36.1. In comparison, the average F1-score obtained by the Logistic regression is 27.0.

The average best result is obtained in the *Default* configuration with an average F1-score of 0.37 over the four classifiers. We can notice an important decrease in all F1-scores from the *Default* to the *Fraud-Only* or *No-Law* configurations. ML classifiers outperforms the Logistic regression in pure classification results in 11 out of the 12 tested pairs of $\{Config, Classifiers\}$. Overall the Dart booster is more stable and has a higher classification performance than the 3 other tested based classifiers.

Table 1.4 shows that σOAM outperforms OAM. The average $F1_{OAM}$ for the two models are respectively 29.4 and 21.6. We clearly see the stability of using σOAM compared to OAM. Comparing the Linear and Tree classifier in the Default configuration in Table 1.4, we see a slight advantage in classification performances for the Linear booster with a F1 score of 39.8 compared to 38.4. After the cost sensitive OAM is applied, the two classifiers have inverted $F1_{OAM}$ performances, respectively 25.1 vs 27. This results in the Linear booster having worst results in term of amount saved in OAM with 849,518$ to 1,521,527$ to the corner solution of not-audit. As σOAM selects its optimal threshold directly on the output of the classifiers, the higher prediction power of the Linear booster is less distorted with an $F1_{OAM}$ of respectively 36.1 and 35.8. Similarly higher amounts are saved as the classification becomes less degraded.

1.5.3 Performances of the cost-sensitive frameworks

Following Bahnsen et al. (2016), we assess the performance of the 4 cost-sensitive frameworks (OAM, σOAM, Elkan and CSBoost) through the cost entailed by their usage compared with the default situation of not-auditing. This allows for a fair comparison of performances across configurations as test-sets may differ. Table 1.5 presents the expected costs associated with the 3 corner solutions[29] of perfectly predicting fraud, auditing every claim, and not auditing any claim for each configurations/cost-sensitive framework. (1.23) through (1.25) show how each cost in Table 1.5 is calculated. Although the same test-set is used across all models, some corner costs differ. By construction, *"Fraud-Only"* configuration has a different test-set reference cost. Moreover, Logistic models do not accept missing values resulting in further discarding some observations and variables. Columns Sav$ in Table 1.4 present the result of the difference between the minimized cost of fraud obtained by the corresponding classifiers in (1.22), and the corner solution of *No-Audit*.

$$All - Audit = (TP + FN) \cdot \bar{c} + (FP + TN) \cdot (\bar{c} + \bar{t}) \tag{1.23}$$

$$No - Audit = (FP + FN + TP + TN) \cdot \bar{t}^{'}) \tag{1.24}$$

$$Perfect - Prediction = (TP + FN) \cdot \bar{t} + (FP + TN) \cdot \bar{c} \tag{1.25}$$

where $\bar{t}$ and $\bar{c}$ are respectively the average compensation and cost of investigation per θ. FP, TP, FN and TN are *False-Positive*, *True-Positive*, *False-Negative* and *True-Negative* instances.

Analysing column Sav$ in Table 1.4, the best performing cost-sensitive framework is the σOAM model paired with the Dart classifier saving $3,530,975$$, around 35% of the excess cost due to fraud and auditing valid claims. Regardless of the configuration and the

[29] Not accounting for residual fraud

Table 1.5: Test-DataBase key costs of corner solutions in $

Model	Population	Perfect-Prediction	All-Audit	No-Audit
Def/NL	14,073	207,545,209	234,465,168	217,389,852
FO	13,741	209,807,439	237,073,839	215,634,936
FOLogi	11,872	179,397,549	202,898,204	184,496,828
NLLogi	12,411	180,991,409	204,654,570	190,037,764
DefLogi	12,335	180,102,741	203,655,137	188,608,623
CSboost Def/NL	11,087	160,086,400	180,790,992	168,436,333
CSboost FO	10,531	157,856,590	178,308,085	162,433,178

Note: Abbreviations Def, NL and FO refer respectively to the configuration, *Default*, *No-Law* and *Fraud-Only* presented in Table 1.4. Abbreviations Logi and CSboost refer to the classifiers used: Logistic regression or Cost-Sensitive boosting. *Population* shows the number of observations for each test-set. *Perfect-Prediction* is the cost of perfectly predicting all known fraudulent claims. *All-Audit* corresponds to auditing every claims and *no-Audit* is the cost of not having a fraud detection process.

classifier, σOAM and Elkan cost-sensitive frameworks always produce the first and second top amount saved. As mentioned in Table 1.4, *Default* configuration has all variables related to the presence of lawyers and type of law firms on the claims in the training set and also includes all suspicious claims as fraudulent. *Fraud-Only* limits the target population to claims with amount of money recovered. *No-Law* removes all variables related to law from the training set.

Table 1.4 shows that over the 12 pairs *{Config,Classifiers}*, σOAM model averages savings of $1,550,329$\$ of the extra cost, and $1,799,102$\$ excluding Logistic model, respectively 17.4% and 21.4% of the extra cost. Similarly, Elkan model averages savings of $1,839,612$\$ and $1,911,324$\$ excluding Logistic model, respectively 18.1% and 22.6% of the extra cost. Although Elkan model is more stable than σOAM when the configura-

tion changes from *Default* to *Fraud-Only* or *No-Law*, σOAM produces the top two results with pairs *{Default,Dart}* and *{Default, Linear}* and outperforms Elkan in the *"Default"* configuration. The following section expands on the comparisons between these two cost-sensitive frameworks.

The vector of the top 5 predictors selected in the *"Default"* and *"Fraud-Only"* configurations are almost the same for each classifier(see Table A.5 in Appendix A). Consequently in the *OAM* framework, the main difference between the two configurations is their target population. The average F1-scores obtained by the 4 base classifiers [30] in Table 1.4 decreases only slightly from 0.379 in the *"Default"* configuration to 0.345 in the *"Fraud-Only"* configuration.

However, the decrease in the average F1-score of the 4 base classifiers between the *"Default"* and *"Fraud-Only"* configurations in the OAM framework is steeper from 0.27 to 0.19 as observed in Table 1.4. Given all the similarities between these two configurations, this steeper decrease once the OAM framework is applied indicates that the top selected predictors are significantly less powerful to classify observations labeled as *"Fraud"* than they are to classify *"Suspicious"* observations.

In Table 1.4, except for 5 instances out of 27, *Save*\$ results obtained by machine-learning classifiers are higher than their Logistic counterparts. Out of all classifiers, Dart has the highest averaged saved amount of $1,370,995$\$, 16.4% of the extra cost, and the top two highest performances across all 3 configurations and 4 cost-sensitive frameworks. The Logistic regression has the lowest averaged saved amount of $565,043$\$, 5% of the extra-cost of fraud, across its 3 cost-sensitive frameworks. Overall machine-learning classifiers produce higher and more stable savings than the Logistic regression.

In σOAM, the signal is created directly from the output of the algorithm instead of derived from the top 5 predictors. The resulting predictions produce better performances

[30] Logis, Dart, Tree, Linear

both in terms of F1-score and Sav$ in Table 1.4 with an average increase of respectively 37% and 199% across all configurations. The predictions are also much more robust to changes in configurations with an average F1-score increase from OAM of 37% and 47% for configurations *Fraud-Only* and *No-Law*. The increase in performances occurs for all classifiers but further highlights the gap in performances between the logistic regression and ML predictors. Overall, the increase in performances obtained from OAM to σOAM far outweighs the increase in performances obtained by an ML classifiers to the logistic regression, which emphasizes the role of an optimal cost-sensitive model.

1.5.4 Elkan vs σOAM

Until now, comparison between cost-sensitive frameworks were carried out as if there were no residual fraud. Since the global probability of fraud given by the partner is 15%, the aggressiveness of the σOAM is higher than optimal if the observed fraud rate of 5% is assumed to be the true rate. Classical cost-sensitive models' performances make such assumption. To fully compare σOAM and Elkan model, we need to assume that the observed fraud rate of 5% labelled in the data is true.

At a 5% global fraud rate, we observe in Table 1.6 that σOAM saves $3,363,068$\$ compared to not-auditing, representing about 34.6% of the extra cost of fraud. Elkan model saves $2,973,675$\$, 30.2% of the extra cost of fraud, a reduction of $389,393$\$ when compared to σOAM. Table 1.6 details the difference between the audit policies of the two frameworks.

The F1-scores of Elkan σOAM are similar at around 0.40. σOAM audit policy targets more aggressively the θ groups that have a higher fraud rate. On the other hand, Elkan model sets its threshold solely based on the cost-ratio. Given that investigation costs are at an order of magnitude less than compensation, σOAM focuses on diminishing the an-

Table 1.6: σOAM vs Elkan with an assumed 5% fraud rate

			Elkan					σOAM				
			Sav$:2,973,675$, F1Score: 0.40					Sav$:3,363,068$; F1Score: 0.40				
θ	$\frac{t}{c}$	F%	TP	FP	TN	FN	Thrs	TP	FP	TN	FN	Thrs
1	7.66	1.76	15	28	3171	47	0.130	15	24	3175	47	0.190
2	9.43	1.15	3	17	1013	6	0.106	1	5	1025	8	0.374
3	8.84	1.90	8	21	943	7	0.113	5	9	955	10	0.190
4	8.78	2.62	13	16	916	17	0.114	12	12	920	18	0.130
5	8.29	3.78	15	28	898	27	0.121	15	28	898	27	0.141
6	7.95	4.54	19	31	934	27	0.126	21	35	930	25	0.141
7	7.56	5.90	20	45	1014	42	0.132	27	81	978	35	0.101
8	7.19	7.50	42	54	1146	51	0.139	53	123	1077	40	0.109
9	6.96	9.56	50	75	1376	81	0.144	78	200	1251	53	0.109
10	6.35	13.6	112	105	1485	155	0.158	183	363	1227	84	0.092

Note: columns $\frac{t}{c}$ is the ratio of the compensation to the investigation cost per θ. $F\%$ is the fraud-rate and TP, FP, TN and FN are respectively the true-positive, false-positive, true-negative and false-negative number of claims. $Thrs$ is the calibrated probability threshold above which a claim is audited. A higher threshold means a less aggressive audit policy.

ticipated number of false negative instances thanks to its capacity to anticipate subsequent fraud rates.

Tables P.15 and P.16 in Appendix P present similar results for *"Fraud-Only"* and *"No-Law"*. Assuming the observed fraud rate to be correct also increases the performance in these other 2 configurations. In the *"Fraud-Only"* configuration, the fraud rate is set to 3% as explained in Section 1.3. Under these conditions, σOAM performs similarly to Elkan model with a saving of $1,536,980$ compared with $1,566,847$. Under the *"No-Law"* configuration, σOAM framework achieves a $2,023,254$ saving compared to the $1,991,014$

performance of the Elkan framework. Tables P.15 and P.16 in Appendix P show σOAM displays an increased aggressiveness vis-à-vis θ with higher fraud rate, limiting the number of false-negative and behaving similarly than in the *"Default"* configuration.

1.5.5 Sensitivity analysis

Table 1.7 presents the sensitivity of the results obtained in Table 1.4 to assumptions made on the global fraud rate and γ, the sensitivity of the fraud rate to the audit policy. The sensitivity analysis is carried out with Dart base learner and σOAM as the cost-sensitive framework. We observe an inverse relationship between γ and the aggressiveness of the OAM as shown by column $\#Adt$: number of audited claims. Because of the higher deterrence effect, fewer observations need to be audited to obtain an optimal cost. Inversely, a higher global fraud rate leads to a more aggressive audit policy as the cost of residual fraud increases. Cost performances are globally stable across the different parameters.

1.6 Conclusion

We have extended OAM to recent machine-learning classifiers and compared its performances to other cost-sensitive methods. ML classifiers produce better and more stable performances than the Logistic regression when incorporated into OAM. Our second contribution was to extend the OAM model to a more general fraud detection model σOAM that produces better classification and saving performances across the three different configurations and four classifiers tested.

In the default configuration, the total cost of fraud is estimated at $9,844,643$\$. When the Logistic model is applied $504,509$\$ and $2,073,495$\$ are saved respectively in *OAM* and in the σOAM . The best ML classifier, Dart, without a cost-sensitive threshold selection

Table 1.7: σOAM with Dart base learner, Sensitivity Analysis at 15% fraud rate

P(F):		5% Fraud		10% Fraud		15% Fraud		20% Fraud	
Conf	γ	Sav$	#Adt	Sav$	#Adt	Sav$	#Adt	Sav$	#Adt
FO	1	1,246,400	1931	1,110,314	2547	717,753	3085	271,410	3529
FO	2	1,481,432	1549	1,307,614	1984	1,257,100	2327	1,143,011	2550
FO	3	1,633,840	1293	1,453,016	1972	1,423,141	1820	1,312,336	1972
NL	1	1,902,179	1163	1,529,541	1394	1,083,588	1741	708,874	2075
NL	2	2,023,254	746	2,042,559	1116	1,738,962	1215	1,611,014	1362
NL	3	2,008,734	683	2,104,657	832	2,085,044	984	2,028,554	1089
Def	1	3,523,823	1623	3,523,823	2023	3,180,470	2386	2,987,844	2682
Def	2	3,363,068	1290	3,473,061	1513	3,530,975	1726	3,442,445	1915
Def	3	3,160,843	997	3,431,421	1245	3,495,284	1386	3,506,196	1527

Note: column *Conf* shows the configuration's name. γ is the elasticity of the fraud rate to the optimal audit policy. Sav$ is the amount of dollars saved to the corner solution of non-auditing for a global fraud rate $P(F)$ of $X\%$, #Adt is the number of claims audited

saves $3,176,912$\$. σOAM with the Dart produces the best performance with $3,530,975$\$. σOAM significantly outperforms the original OAM and other tested cost-sensitive algorithms. Furthermore, σOAM shows better performances than the Elkan-Threshold framework when for residual fraud and a deterrence effect that is, to our knowledge, foreign to any existing cost-sensitive algorithm.

We also bring forward evidence of the new importance taken by law firms in claim settlements. A more accurate parametrization would lead to further improvements. OAM offers a framework to calculate an optimal audit policy from any type of continuous output. However, detecting fraudulent claims could be enhanced by combining different class of data and algorithms. A key enabling factor to fraud is the influence of claim suppliers such

as medical examiners, lawyers or mechanics. Combining complex network predictions on graph data with XGB predictions and tabular data could enhance the prediction power of the output and further refine OAM. The sensitivity of the fraud rate to the audit policy is approximated by the parameter γ that would benefit from empirical insights to its behavior. Further refining the claimant's choice to fraud by introducing a measure of risk aversion for the insured and an endogenous deterrence effect in the setting of the optimal economic model could also be interesting developments.

Bibliography

Arrow, K. J. (1964). “The Role of Securities in the Optimal Allocation of Risk-bearing”. In: pp. 91–96. ISSN: 00346527, 1467937X.

— (1974). “Optimal insurance and generalized deductibles”. In: *Scandinavian Actuarial Journal* 1, pp. 1–42. DOI: 10.1080/03461238.1974.10408659.

Bahnsen, A. Correa, D. Aouada, A. Stojanovic, et al. (2016). “Feature engineering strategies for credit card fraud detection”. In: *Expert Systems with Applications* 51, pp. 134–142. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2015.12.030.

Banulescu-Radu, D. and Y. Kougblenou (2024). “Data science for insurance fraud detection: a review”. In: *Handbook of Insurance*. Springer. Chap. Forthcoming.

Barcaroli, G. (2014). “"SamplingStrata: An R Package for the Optimization of Stratified Sampling"”. In: vol. "61". "4", "1–24". DOI: "10.18637/jss.v061.i04".

Belhadji, E., G. Dionne, and F. Tarkhani (2000). “A Model for the Detection of Insurance Fraud”. In: *The Geneva Papers on Risk and Insurance - Issues and Practice* 25, pp. 517–538.

Bodaghiand, A. and B. Teimourpour (2018). “Automobile Insurance Fraud Detection Using Social Network Analysis: Case Studies in Social Networks and Beyond”. In: *Applications of Data Management and Analysis*, pp. 11–16. DOI: 10.1007/978-3-319-95810-1_2.

Brockett, P., R. Derrig, L. Golden, et al. (2002). “Fraud Classification Using Principal Component Analysis of RIDITs”. In: *Journal of Risk and Insurance* 69, pp. 341–371. DOI: 10.1111/1539-6975.00027.

Chen, T. and C. Guestrin (2016). “XGBoost: A Scalable Tree Boosting System”. In: *CoRR* abs/1603.02754. DOI: 10.1145/2939672.2939785. arXiv: 1603.02754.

Crocker, K. J. and J. Morgan (1998). “Is Honesty the Best Policy? Curtailing Insurance Fraud through Optimal Incentive Contracts”. In: *Journal of Political Economy* 106.2, pp. 355–375. ISSN: 00223808, 1537534X.

Dionne, G. and E. Belhadji (1996). “Évaluation de la fraude à l’assurance automobile au Québec”. In: *Assurance* 64, pp. 365–394.

Dionne, G. and R. Gagné (2001). “Deductible Contracts against Fraudulent Claims: Evidence from Automobile Insurance”. In: *The Review of Economics and Statistics* 83.2, pp. 290–301. ISSN: 00346535, 15309142.

Dionne, G., F. Giuliano, and P. Picard (2009). “Optimal Auditing With Scoring: Theory and Application to Insurance Fraud”. In: *Management Science* 55, pp. 58–70. DOI: 10.1287/mnsc.1080.0905.

Elkan, C. (2001). “The Foundations of Cost-Sensitive Learning”. In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence: 4-10 August 2001; Seattle* 1.

Gollier, C. (1987). “Pareto-optimal risk sharing with fixed costs per claim”. In: *Scandinavian Actuarial Journal* 1987.1-2, pp. 62–73. DOI: 10.1080/03461238.1987.10413818.

Goutte, C. and E. Gaussier (2005). “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation”. In: *Lecture Notes in Computer Science* 3408, pp. 345–359. DOI: 10.1007/978-3-540-31865-1_25.

Grossman, S. J. and O. D. Hart (1983). “An Analysis of the Principal-Agent Problem”. In: *Econometrica* 51.1, pp. 7–45. ISSN: 00129682, 14680262.

Holmstrom, B. (1979). “Moral Hazard and Observability”. In: *The Bell Journal of Economics* 10, pp. 74–91.

Höppner, S., B. Baesens, W. Verbeke, et al. (2022). “Instance-dependent cost-sensitive learning for detecting transfer fraud”. In: *European Journal of Operational Research* 297.1, pp. 291–300. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2021.05.028.

Jiang, X., M. Osl, J. Kim, et al. (2011). “Smooth Isotonic Regression: A New Method to Calibrate Predictive Models”. In: *AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science* 2011, pp. 16–20.

Lundberg, S., G.G. Erion, and S.I Lee (2018). “Consistent Individualized Feature Attribution for Tree Ensembles”. In: *CoRR* abs/1802.03888. arXiv: 1802.03888. URL: http://arxiv.org/abs/1802.03888.

Lundberg, S. and S.I. Lee (2017). “A unified approach to interpreting model predictions”. In: *CoRR* abs/1705.07874. arXiv: 1705.07874. URL: http://arxiv.org/abs/1705.07874.

Majhi, S., S. Bhatachharya, R. Pradhan, et al. (2019). “Fuzzy clustering using salp swarm algorithm for automobile insurance fraud detection”. In: *Journal of Intelligent & Fuzzy Systems* 36, pp. 2333–2344. DOI: 10.3233/JIFS-169944.

Mirrlees, J. (1976). “The Optimal Structure of Authority and Incentives Within an Organization”. In: *Bell Journal of Economics* 7, pp. 105–131. DOI: 10.2307/3003192.

Mookherjee, D. and I. Png (1989). “Optimal Auditing, Insurance, and Redistribution”. In: *The Quarterly Journal of Economics* 104.2, pp. 399–415.

Nian, K., H. Zhang, A. Tayal, et al. (2016). "Auto insurance fraud detection using unsupervised spectral ranking for anomaly". In: *The Journal of Finance and Data Science* 2.1, pp. 58–75. DOI: 10.1016/j.jfds.2016.03.001.

Pesantez-Narvaez, J., M. Guillen, and M. Alcañiz (2019). "Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression". In: *Risks* 7.2. ISSN: 2227-9091. DOI: 10.3390/risks7020070.

Picard, P. (2013). "Economic analysis of insurance fraud". In: *Handbook of Insurance: Second Edition*, pp. 349–395. DOI: 10.1007/978-1-4614-0155-1_13.

Raviv, A. (1979). "The Design of an Optimal Insurance Policy". In: *The American Economic Review* 69.1, pp. 84–96. ISSN: 00028282.

Subelj, L., S. Furlan, and M. Bajec (2011). "An expert system for detecting automobile insurance fraud using social network analysis". In: *Expert Systems with Applications* 38.1, pp. 1039–1052. DOI: 10.1016/j.eswa.2010.07.143.

Sun, Y., M. S. Kamel, Andrew K.C. Wong, et al. (2007). "Cost-sensitive boosting for classification of imbalanced data". In: *Pattern Recognition* 40.12, pp. 3358–3378. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2007.04.009.

Townsend, R. M. (1979). "Optimal contracts and competitive markets with costly state verification". In: *Journal of Economic Theory* 21.2, pp. 265–293. ISSN: 0022-0531. DOI: 10.1016/0022-0531(79)90031-0.

Viaene, S., R. Derrig, B. Baesens, et al. (2002). "A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection". In: *Journal of Risk and Insurance* 69, pp. 373–421. DOI: 10.1111/1539-6975.00023.

Viaene, S., R. Derrig, and G. Dedene (2004). "A Case Study of Applying Boosting Naive Bayes to Claim Fraud Diagnosis". In: *Knowledge and Data Engineering, IEEE Transactions on* 16(5), pp. 612–620. DOI: 10.1109/TKDE.2004.1277822.

Weisberg, H. and R. Derrig (1998). “Quantitative methods for detecting fraudulent automobile bodily injury claims”. In: *Risques* 35, pp. 75–101.

— (1991). “Fraud and Automobile Insurance: A Report on Bodily Injury Liability Claims in Massachusetts”. In: *Journal of Insurance Regulation* 9.4, p. 497.

Chapter 2

Risk Aversion in Insurance Fraud Detection.

Abstract

Risk aversion has been well studied in the economic literature, as a person's characteristic to avoid uncertainty. Committing fraud is like participating to a lottery. The decision to commit fraud is in essence the choice by a claimant of proceeding with a normal claim with certainty or trying to fraud for the uncertainty of higher gains but greater risk. We revisit upon the model introduced by Cohen and Einav (2007) to derive a measure of risk aversion based on the insured's selection of collision coverage. When we add the measure of risk aversion to the Optimal Audit Model, OAM, we show that the insurer's classification performance and savings increase. The newly added feature ranks systematically among the top predictors for detecting fraud.

2.1 Introduction

In Canada, auto insurance premiums are estimated to be between 5 and 15% higher due to the cost of fraud. Using considerable resources, insurance companies continuously improve their expert systems to efficiently flag fraudulent claims in an effort to reduce the expected cost of fraud. Improving the performance of these expert systems has become a critical challenge as a way to reduce the audit cost and, in turn, create a higher deterrence effect.

The inclination of a claimant towards committing fraud is inherently similar to participate to a gamble, a calculated decision influenced by risk that must be modeled with precision. The role of risk aversion in insurance demand [1] has been rigorously analyzed. Nevertheless, its impact on fraudulent behavior has predominantly been studied in the realm of tax evasion, and occasionally within the purview of insurance fraud [2]. A better modeling of the key variables in the fraud decision process is important as it should improve the prediction performance of the expert systems used to detect fraudulent claims. We use a model developed by Cohen and Einav (2007) to infer risk aversion from insurance data. We apply their model to a Canadian database made available by a partner insurance company.

Zerbato et al. (2024) extended Optimal Audit Model (OAM), a cost-sensitive expert system designed to model the relationship between the audit policy and the expected resulting level of fraud to include new machine learning classifiers. Dionne et al. (2009) showed that the prediction power of OAM was able to further reduce the cost incurred by fraud. In this study, we use the newly developed OAM, σOAM, which integrates an

[1] Outreville (2014), Halek and Eisenhauer (2001), Loubergé and Dionne (2024), Sydnor (2010), Schlesinger (2013)

[2] Eisenhauer (2008), Ghosh and Crain (1995), Boyer and Peter (2020)

Extreme Gradient Boosting (XGB) classifier, to evaluate whether explicit consideration of risk aversion enhances the classification performance and reduces the total cost of fraud. The empirical findings indicate that the integration of the risk aversion variable into the training set results in a 1.5% increase in classification performance within the XGB framework. Additionally, σOAM realizes a 3.5% enhancement in cost reduction, thereby further decreasing the total cost of fraud by 10,000 CAD, an additional 1.2% of the cost induced by fraud, compared to σOAM without the risk aversion variable.

The rest of the document is segmented as follows: Section 2.2 introduces the literature review on risk aversion. Section 2.3 focuses on the different collision coverage choices made by the insureds and the relation between Bodily-Injury (or Accident-Benefit (AB)) and collision premiums. Section 2.4 introduces the methodology developed by Cohen and Einav (2007) to infer η, the risk aversion parameter from insurance data. Finally, Section 2.5 presents the results and is divided into three subsections: Subsection 2.5.1 details the regression results to estimate the yearly risk of collision for each insured; Subsection 2.5.2 describes the probit model used to estimate risk aversion; and Subsection 2.5.3 documents the increase in performance obtained from adding risk aversion to σOAM. Section 2.6 concludes this study.

2.2 Literature Review and Main Contribution

2.2.1 Literature Review

Risk aversion was first introduced by Pratt (1964) and **Arrow1964**; Arrow (1973) as a measure of utility loss when a risk-averse person faces uncertainty. Higher risk aversion leads to a higher reluctance to participate in fair uncertain games. Risk aversion in the context of insurance has been widely studied in the literature as a key determinant of

an individual's decision to buy insurance. Schlesinger (2013) reviews the effect of risk aversion on insurance demand. Outreville (2014) compiles a comprehensive overview of the key behavioral factors explaining risk aversion and of current estimation methods of risk aversion in the insurance industry.

However, to our knowledge, few studies have focused on the influence of risk aversion in an insurance fraud setting and fewer have considered risk aversion has an endogenous component of the fraud rate. Picard (1996) introduced a risk aversion parameter in a theoretical model tackling the audit game with and without commitment. The model is further enriched in Picard et al. (2024) where two types of risk averse individuals are introduced: *"High Risk Aversion"* and *"Low Risk Aversion"*. *"High Risk Aversion"* insureds choose higher coverage than their *"Low Risk Aversion"* counterparts. The authors' model implies that higher deductible contracts attract less risk averse insureds and makes fraud more attractive. Empirically, several key characteristics such as the magnitude of risk aversion, its relationship to the level of wealth, the condition upon which it can be derived, and its demographic characteristics, are heavily debated in the literature. Until now, risk aversion has been estimated from responses submitted to questionnaires or from observed decisions of the household, and from insurance company data.

A long list of authors focused on characterising the marginal evolution of risk aversion to wealth. Holt and Laury (2002) offered a coherent framework for estimating risk aversion in the small and in the large. Through a series of controlled interviews where subjects were placed in both fictive and real lotteries with increasingly risky rewarding and certain situations, the authors evaluated the level of risk aversion through the number of *safe* choices made by each subject. By increasing the scale of each lottery, the authors noticed that subjects were less and less keen to take risk once the certain amount became high enough. This observation indicates increasing Relative Risk Aversion (RRA) that is best modeled by a *power-expo* utility function $U(x) = \frac{1-exp(-\alpha x^{1-r})}{\alpha}$ where $r = 0.269$ is the

parameter of constant relative risk aversion and $\alpha = 0.029$ is the absolute risk aversion coefficient. When both parameters are positive, the function has the satisfying properties of increasing RRA and decreasing Absolute Risk Aversion (ARA).

Another early study by Friend and Blume (1975) used cross-sectional data from a household survey on risky assets ownership and other demographic variables including wealth. The authors proposed a structural model of investment where demand for risky assets is generated by the household's problem to maximize its wealth across its life horizon. When fitted to the survey data, the magnitude of the RRA parameter is estimated to be greater than 1, and households display signs of decreasing RRA as households' portfolio become more risky when wealth increases.

Chetty (2006) used labor data to estimate the magnitude of the RRA parameter. The authors noticed a link between the ratio of the agent's labour supply to the labour supply elasticity of income and the risk aversion parameter. The numerator is intrinsically related to the marginal utility of wealth whilst the denominator is linked to the derivative of the marginal utility of wealth. Using documented techniques to estimate these two parameters from labour survey data, the authors were able to identify the measure of the curvature of the utility function by looking at individual labour-leisure choices as decision under uncertainty. The results show a tight bound on relative risk aversion of at most 1.25.

Guiso and Paiella (2008) measured the coefficient of RRA through an Italian households survey data containing a specific question on an hypothetical lottery as well as data on wealth, investment portfolio and job status. The authors tested the characteristics of the relationship between risk aversion with deterministic and random starting wealth level. Evidences from the survey suggest that both constant ARA and RRA poorly describe the curvature of the utility curve. ARA is indeed a decreasing convex function of wealth but it decreases at a much slower pace than constant RRA would suggest.

Chiappori and Paiella (2011) study the composition of households' portfolio in risky

assets. Using panel data, the authors are able to test the effect of changes in wealth, in asset prices, volatility, and returns on a measure of risk aversion at individual and population level. Results show evidence of a Constant Relative Risk Aversion (CRRA) as exogenous shocks on wealth do not affect the composition of portfolio. When equity and housing are included as risky assets, the elasticity of wealth to the share of risky assets becomes positive. A possible explanation is the iliquid nature of real estate assets that prevent homeowners from deleveraging fast when wealth suffers a negative shock. Finally, the authors infer the joint distribution of wealth and risk aversion which allows testing the CRRA assumption. Risk aversion varies greatly among households and is weakly negatively correlated with wealth. The assumption of independence between the distribution of wealth and risk aversion is therefore a valid approximation.

Risk aversion parameter can be inferred following the base model of Cohen and Einav (2007). The authors used Israeli auto-insurance data on the choice of deductibles and on the risk level to infer the risk aversion parameter of the insureds. Two methods are proposed to deduce risk aversion. The first model assumes the risk level to be perfectly predicted and uses that estimate in a Probit regression where the risk aversion parameter is the latent output. The presence of adverse selection in the Israeli industry makes the first model misspecified. The authors offer an alternative approach using a Gibbs sampler to generate synthetic draws of risk level and risk aversion under adverse selection. They rely on the local insurance industry's specificity to create exogenous chocs in the premium/deductible menus to nonparametrically identify the distribution of risk aversion. Since we do not consider adverse selection, we use the first model with collision premium and deductible choice variables that are not present in training the classifiers to get an estimate of the risk aversion parameters. We further detail the model in a following section.

So far, the focus has been to estimate the risk aversion parameter, its magnitude, determinants and its relation to wealth. Few studies have researched on the implication of

risk aversion on the propensity to fraud. Muller et al. (2011) aim to derive a range of realistic audit policies and fraud strategies, guaranteeing market participation at the moment of signing for both the insurer and insured sides. Their modeling of the claimant's problem includes a risk aversion parameter whose influence is notable in determining the possible set of outcomes. The resulting optimal audit policy and fraud strategy resemble a threshold function with a large(low) audit rates above(below) a certain probability of fraud. Upper and lower limits of the audit strategy are set by the participation constraint of the claimant. Muller et al. (2011) show that claimants with higher risk aversion accept more intense scrutiny from the insurer while still participating in the market. However the effect of risk aversion on the fraud rate is not estimated.

Moreno et al. (2006) suggests moving away from combating insurance fraud with a pure audit framework. Instead, the authors develop a model based on a self-selection principle utilizing a bonus/malus mechanism which acts as the sole deterrence against fraud. The solution is a set of feasible contracts defined by an increase in premium following a period where a claim was submitted. The set of contracts needs to satisfy constraints such as participation on both sides of the contract and incentives for loss declaration. The authors showed that, at the converged solution, the post-claim premium approximates the risk aversion coefficient. More risk-adverse claimants will not necessitate a higher premium increase to deter them from frauding.

2.2.2 Main Contribution

Lesch and Brinkmann (2011), Ribeiro et al. (2020), and Tennyson (2008) list four main and mutually non-exclusive families of explanations on the reasons prompting insurance claimants to fraud. The first, second, and third families of explanations relate, respectively, to the *"Economic and Contractual"* potential gains, the *"Macrosocial"* environment, and

the *"Moral Sociological Views"* surrounding the claimants. The first two families depend on the local or global economic situation and the potential gain obtainable from fraud. The third family of explanations is tied to the degree of acceptability of fraud in the claimant's entourage and the understanding of the social role of insurance companies in society. The fourth main family, *"Moral Psychological Views"*, is related to personality aspects such as moral development, self-control, and risk aversion. Lesch and Brinkmann (2011) and Tennyson (2008) propose qualitative analysis of the fraud factors. Ribeiro et al. (2020) uses survey data from Portugal to test the prevalence of some factors in the decision but does not test for the importance of risk aversion itself in the decision to fraud.

Using Cohen and Einav (2007)'s first model, we infer a measure of risk aversion from our partner insurance database. We revisit the findings from Cohen and Einav (2007) by applying an upgraded model to a Canadian dataset with new independent variables and examining the resulting insights. After estimating the risk aversion measure, we test its effect across various expert systems designed for fraud detection. The analysis of these results aims to establish risk aversion as a significant factor influencing the decision to commit fraud, while also highlighting its role in enhancing the performance of the detection models.

2.3 Data review

Our data set includes claims associated with bodily injury fraud. To independently estimate risk aversion from data already employed to identify fraud, we utilize data related to insurance contract choice concerning property damage coverage. In the Canadian context, there are multiple forms of property damage coverage. Our analysis is confined to Collision and All-Perils coverages, as these involve scenarios where the claimant is entirely liable for the damages incurred. After excluding claims with missing coverage and/or

premium information, we removed 7822 claims, resulting in a final sample size of 62589 observations. Unlike in Israel, both collision and all-perils coverages are optional resulting in certain observations having no Collision premiums or Collision deductibles when insureds found it unnecessary to cover these specific risks.

Table 2.1 and Figure 2.1, respectively, display the population per deductible chosen and the premium charged. Table 2.1 shows a clear imbalance between deductible levels with a clear preference for $\{0, 500, 1000\}$ levels in both Collision and All-Perils coverages. In Table 2.1, row *"Ded"* shows how claims are aggregated: if deductible $\in [1000, +\infty[= 1000$ if deductible $\in]0, 500[= 500$. When a value is present for each coverage, then the higher value is kept to be aggregated [3].

Table 2.1: Deductible Choice per Coverage

$	0	100	250	300	500	1000	1500	2000	2500	3000	5000	10000
Coll	25817	1	20	683	26808	9275	12	78	34	7	1	0
A-P	47124	1	6	268	9553	5569	29	66	96	2	28	2
Ded	10324				37146	15119						

Note: Each column corresponds to a deductible choice available. Rows *Coll* and *A-P* are the deductibles chosen by the insured for Collision or All-Perils coverage, respectively. The final row, Ded, is the aggregated deductible option.

Figure 2.1 illustrates the distribution of premiums paid for the three deductibles. Each distribution exhibits long-tail characteristics that have been truncated for display purposes. The variables FTP_COLL and FTP_AP represent the premium densities for Collision and All-Perils coverage, respectively. The aggregated premium density, FTP_Cov, serves as the focal point of this study. The empirical density of FTP_Cov is bimodal, with peaks at 0 and approximately 240. Notably, 10,521 observations report a null FTP_Cov. As indicated in Table 2.1, 10,324 observations have zero deductible. Of these, 10,279 obser-

[3] Only for 174 observations

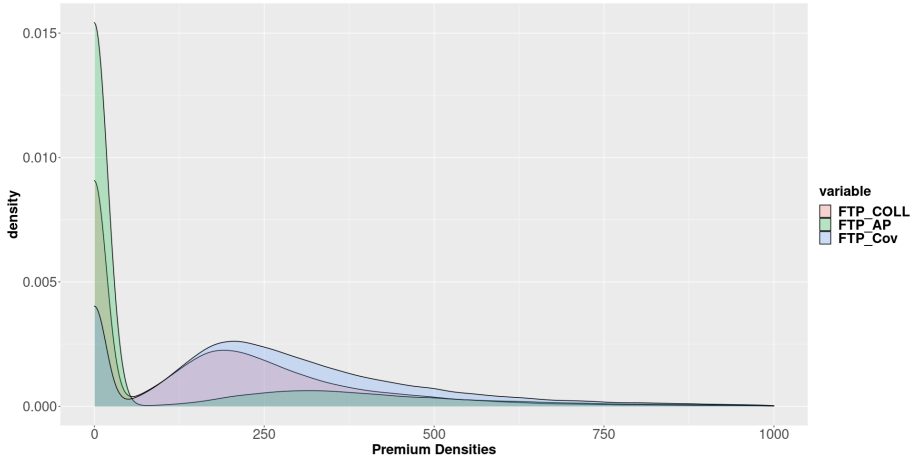


Figure 2.1: Empirical Density of Premiums

vations form the intersection of both groups, enabling the isolation of insured individuals who did not opt for any form of Collision coverage. This subset constitutes 16.4% of the entire dataset. Additionally, 241 observations have a genuine null deductible with a positive premium. Furthermore, 45 observations are excluded due to the inconsistency of having a zero premium accompanied by a positive deductible. Table 2.2 presents the population distribution in the final three groups of claimants.

Table 2.2: Choice of Collision Coverage

	Deductible			
FTP_Cov	0	500	1000	NA
$FTP\_Cov > 0$	45	36981	15043	0
$FTP\_Cov = 0$	10279	165	76	1

Note: FTP_Cov indicates the premium collision coverage. If $FTP\_Cov = 0$, no collision coverage is chosen. Deductible gives the chosen level of coverage. Each cell presents the population of each subgroups.

2.3.1 Relation between FTP_Cov and FTP_AB

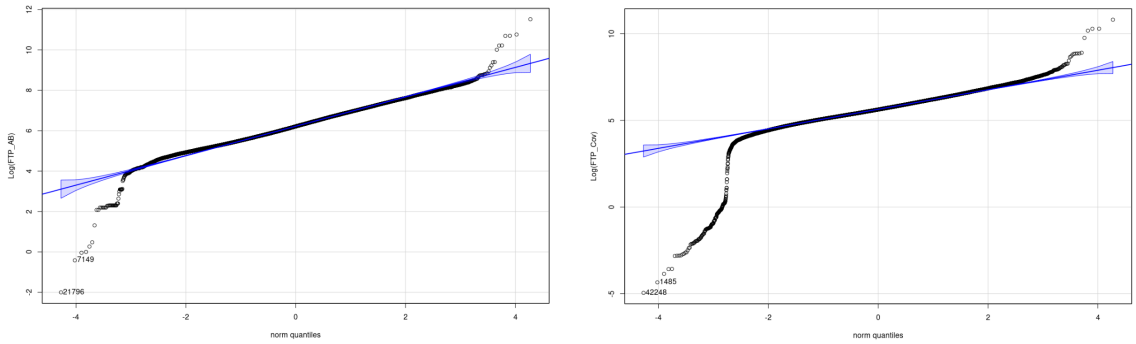
We investigate the relationship between the premium associated with bodily injury coverage, denoted as FTP_AB, and the aggregated collision premium, FTP_Cov. The θ groups identified in Zerbato et al. (2024) consist of claimants classified based on their Accident Benefit (AB) premium levels. According to Zerbato et al. (2024), there is a clear increase in the observed fraud rate with rising AB premiums. Because we estimate the risk aversion parameters using the collision premiums charged, we need to explore the correlation between the AB and collision premiums to determine whether any increase in classification performances comes from this relationship.

In Section 2.5, we incorporate our risk aversion measure as a new input in the expert system developed by Zerbato et al. (2024). This expert system recommends an audit policy for each risk group θ, which is formed from quantiles of FTP_AB. When conditioned on being positive, both FTP_{Cov} and FTP_{AB}, which are respectively the collision and bodily injury premium, approximately follow a LogNormal distribution, as illustrated in Figure 2.2. The extreme lower and upper quantiles, particularly for $Log(FTP_Cov)$, do not conform closely to a normal distribution; however, a significant portion of the central quantiles lies within the confidence interval. Using the maximum likelihood estimate, we find that $log(FTP_Cov) \sim \mathscr{LN}(5.62, 0.68)$, where 5.62 and 0.68 represent the location and scale parameters, respectively. A similar estimation for FTP_AB yields parameters $log(FTP_AB) \sim \mathscr{LN}(6.23, 0.706)$.

Table 2.3 presents three measures of correlation tested between FTP_AB and FTP_Cov. The three measures suggest a modest positive correlation between the two variables. Figure 2.3 shows the OLS results and the non-parametric regressions of FTP_Cov_{scaled} over FTP_AB_{scaled}. The red dotted line is the Loess nonparametric curve surrounded by the 90% variability interval of the position of the regression function. The full green line is

Table 2.3: Correlation between FTP_AB and FTP_Cov

Test	Correlation	P.Value
Spearman	$\rho = 0.473$	$< 2.2e-16$
Pearson	$r = 0.411$	$< 2.2e-16$
Kendall	$\tau = 0.327$	$< 2.2e-16$



(a) Normal QQPlot $Log(FTP\_AB)$ (b) Normal QQPlot $Log(FTP\_Cov)$

Figure 2.2: Normal QQ Plots

the OLS regression given by (2.1):

$$FTP\_Cov_{scaled} = 0.448 * FTP\_AB_{scaled} - 0.003418 \tag{2.1}$$

The *SSR* obtained by the Loess regression sits at 1052.819 for an estimated number of parameters of 5.67, a span of 0.75. The *SSR* obtained by the OLS is at 27076.85. The nonparametric Loess curve shows that a positive linear relation is a good estimation of the relationship between FTP_Cov_{scaled} and FTP_AB_{scaled}. The loess curve is less steep than the OLS regression which estimates a coefficient of 0.448. This relationship is consistent with the medium degree of correlation measured in Table 2.3, but indicates a decreasing degree of correlation as the premiums increase. The moderate correlation suggests that the

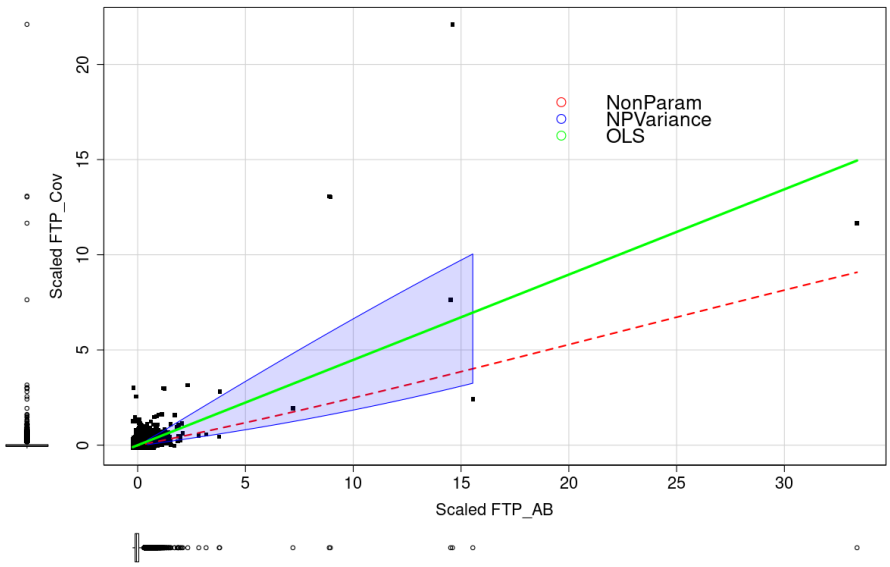


Figure 2.3: OLS and non parametric regression of FTP_Cov to FTP_AB

improved classification performance observed in Section 2.5 is not solely attributable to the correlation between the premium variables.

2.4 Inferring risk aversion parameter η

Pratt (1964) developed the local absolute risk aversion measure, Eq.(2.2), derived from the premium equivalent of removing an uncertain fair lottery.

$$\eta = -\frac{u''(w)}{u'(w)} \tag{2.2}$$

where w is the initial wealth and u is the insured's utility function.

Cohen and Einav (2007) inferred the $r(w)$ of auto-insurance insureds from their contract choices in an Israeli insurance company dataset of collision coverage. An insurance contract is defined by a premium p and a deductible d. A low deductible means a higher

coverage. For an exposure t and given two insurance contracts, one low deductible $\{p^l, d^l\}$ and one high deductible $\{p^h, d^h\}$, the insured chooses the low deductible contract if and only if his expected utility is higher given the risk level λ: (2.4) holds true where the expected utility $v(p,d)$ is defined by (2.3). A higher λ would encourage a higher choice of coverage hence the choice of the low deductible $\{p^l, d^l\}$. It follows that the $p^l > p^h$.

From (2.4), Cohen and Einav (2007) obtains that the choice of low deductible contract takes place only if Eq.(2.6) is verified.

$$v(p,d) = (1-\lambda t)u(w-pt) + (\lambda t)u(w-pt-d) \tag{2.3}$$

$$v(p^l, d^l) \geq v(p^h, d^h) \tag{2.4}$$

$$r(x) = \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\bar{d}} \tag{2.5}$$

$$\eta \geq r(x), \tag{2.6}$$

where $\bar{d} = .5(d^h + d^l)$, $\Delta d = d^h - d^l$, $\Delta p = p^l - p^h$ and λ is the risk level. Eq.(2.5) is the threshold of risk aversion that dictates the choice of the low deductible contract. Given a risk level λ, if the estimated risk aversion parameter η of an insured is greater than the insured's threshold $r(x)$, he will choose the low deductible or high coverage contract.

The Israeli insurance system structurally differs from the Canadian system because the former's deductible choices depend on the premium charged by the insurer following (2.7), as opposed to a fix choice of deductible for the Canadian database. Once the *regular* premium is estimated, the insured can opt for a low deductible amounting to 60% of the *regular* deductible for an increased premium of 6%. The minimum *cap* on the deductible is also regularly adjusted by the Israeli company, creating exogenous changes in Cohen and Einav (2007)'s database.

$$d_{it} = min\{\frac{1}{2}p_{it}, cap_t\} \tag{2.7}$$

Such changes are used to nonparametrically identify the distribution of risk aversion from different contract choices under adverse selection with similar observables, which are absent in the Canadian context. We rely on Cohen and Einav (2007)'s first model, which does not allow for non-observable heterogeneity in the risk level estimate. The chosen model relies on an econometric model described by five equations. Eq.(2.8), Eq.(2.9), and Eq.(2.10) assume a bi-variate log-normal distribution for the pair: $\{\lambda_i, \eta_i\}$, respectively, the risk level per year and the risk aversion parameter. λ and η are latent variables that can be inferred from a claim-emitting process (2.11) and the choice of deductible (2.12).

$$ln(\lambda_i) = \boldsymbol{x}_i^{'}\beta + \varepsilon_i \tag{2.8}$$

$$ln(\eta_i) = \boldsymbol{x}_i^{'}\gamma + \upsilon_i \tag{2.9}$$

$$\begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix} \sim^{iid} \mathscr{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2, \rho\sigma_\lambda\sigma_\eta \\ \rho\sigma_\lambda\sigma_\eta, \sigma_\eta^2 \end{bmatrix} \right) \tag{2.10}$$

$$claims_i \sim Poisson(\lambda_i, t_i) \tag{2.11}$$

$$\begin{aligned} Pr(choice_i = low) &= Pr\left(\eta_i \geq \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\bar{d}} \right) \\ &= Pr\left(exp(\boldsymbol{x}_i^{'}\beta + \varepsilon_i) \geq \frac{\frac{\Delta p}{exp(\boldsymbol{x}_i^{'}\gamma + \upsilon_i)\Delta d} - 1}{\bar{d}} \right) \end{aligned} \tag{2.12}$$

Supposing λ is perfectly identified by the insurer then $\varepsilon = 0$ in (2.8). We are left with a sequential estimation of the risk level λ using the Poisson family regression model and the risk aversion parameter η using a probit model, conditional on the observables $\boldsymbol{x}_i^{'}$. Once λ is estimated through (2.8) and (2.11), the variable $r(x) = \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\bar{d}}$ is computed and included in $\boldsymbol{x}_i^{'}$ to estimate η with (2.9).

2.5 Results

2.5.1 Estimation of risk-level λ

λ is the collision probability per year. Our data includes the number of collision claims over the last 10 years for each insured. To obtain a yearly risk rate, we offset each observation by 10 or by their years of driving experience if it is less than 10, following (2.13).

$$\begin{aligned} log(\hat{\lambda}) =& \beta_1 \Big(\varepsilon - ln(NbrYrsDrive) \Big) \\ &+ \beta_2 \text{Pre\_Injuries} + \beta_3 \text{Age} + \beta_4 \text{Income} \\ &+ \beta_5 \text{Nbr\_Risk} + \beta_6 \text{Nbr\_Potential\_Drivers} \\ &+ \beta_7 \text{Annual\_Km} + \beta_8 \text{Vehicle\_Yr} \\ &+ \beta_9 \text{Vehicle\_Cat} + \beta_{10} \text{Vehicle\_Make} \\ &+ \beta_{11} \text{Veh\_Class\_Cd} + \beta_{12} \text{Employment} \\ &+ \beta_{13} \text{Education} + \beta_{14} \text{Veh\_Rating} \end{aligned} \tag{2.13}$$

Eq.(2.13) details some individual variables and some *families* of variables. Pre_Injuries is composed of two variables controlling for the presence of known physical or psychological conditions suffered by the insureds. Age, Income are stand-alone variables controlling for the average age and income of all the insureds on the policy. Variables Nbr_Risk_On_Prem_Qty and Nbr_Potential_Drivers control for the number of identified risk factors and the number of drivers used to calculate the charged premium which are not fully disclosed by the insurance company. Annual_Km is the average yearly distance the car is expected to travel and Vehicle_Yr is the year the car was manufactured.

Vehicle_Cat has 15 dummy categories controlling for the type of vehicle in the policy that range from *"Mini Vans"* to *"Sport Luxury"*. Vehicle_Make comprises 31 categories

that list the vehicle's manufacturer. Veh_Class_Cd has 18 dummy categories that are coded in 14 classes and 4 subclasses of vehicle. The interpretation of these classes is voluntarily left unclear by the insurance company. Employment comprises 24 categories that reflect the occupation of each insured on the claim and their type of labor (manual, sedentary, etc.). Education has six dummy categories controlling for the level of education from high school to university. Veh_Rating is comprised of three categories giving a rating to the vehicle measuring its score on three types of coverage: collision, comprehensive, and direct compensation.

Our goal is to obtain the yearly risk rate of collision λ. The estimated dependent variable $\hat{\lambda}$ is given by the number of collisions that the claimant suffered in the last ten years. We test the Negative-Binomial, Poisson and Quasi-Poisson regressions to estimate the yearly risk-level λ. Table 2.4 shows the set of variables significant at 1% in the regressions for Poisson regressions with and without offset modifier described in (2.13). The offset is given by the number of years of driving experience, capped at 10, to obtain a yearly collision rate. There are 21,662 reported collision accident in the database over 10 years for a total population of 62,756 observations. Over 74% of the AB claimants never had a collision accident over the last 10 years. Of the 26% of AB claimants with collision reported, more than 50% had only a single collision while some had up to 7 reported accident.

Cohen and Einav (2007) demonstrate that variables related to the intensity of car usage are negatively associated with the probability of collision. In our analysis, this variable is represented by the average annual distance traveled (Annual_km), which shows a positive, though small, effect. Cohen and Einav (2007) also find that family policies are less prone to collisions. Since we cannot directly control for marital status, the number of potential drivers serves as a reliable proxy and reveals a similar relationship. Additionally, Cohen and Einav (2007) report a negative correlation between age, gender, education, and the

likelihood of accidents. In our findings, age is significantly linked to the probability of collision, whereas gender and education do not rank among the top predictors. Interestingly, both retired individuals and students exhibit a lower risk level. Regarding vehicle characteristics, Cohen and Einav (2007) state that "bigger, more expensive, older, and non-commercial cars are more likely to be involved in an accident." However, in our study, the age of the vehicle is not a significant factor, and owning an SUV (either large or small) reduces the likelihood of collisions overall. On the other hand, cars manufactured in Asia are significantly associated with a higher collision rate.

Table 2.4: Regression on the Number of Collisions

	Poisson No Offset (1)	*Poisson Offset* (2)
Pre_Phys	0.088^{***}	0.079^{***}
	-0.014	(0.014)
Pre_Psy	0.171^{***}	0.166^{***}
	-0.019	(0.019)
Age	0.004^{***}	0.002^{***}
	-0.0005	(0.0005)
Nbr_Potential_Drivers	-0.124^{***}	-0.140^{***}
	-0.011	(0.012)
Nbr_Risk_On_Prem_Qty	-0.035^{***}	-0.041^{***}
	-0.013	(0.013)
Annual_km	1.10^{-5***}	1.10^{-6***}
	(1.10^{-5})	(1.10^{-6})
Veh_Comp_Rat	0.005^{***}	0.005^{***}
	-0.001	(0.001)
Cat_Eco-Domestic	-0.391^{***}	-0.377^{***}
	-0.143	(0.143)
Cat_Large SUV	-0.384^{**}	-0.385^{**}
	-0.163	(0.163)
Cat_Small SUV	-0.330^{**}	-0.327^{**}
	-0.139	(0.139)
Veh_ASIAN	0.538^{**}	0.517^{**}
	-0.231	(0.231)
Retired	-0.171^{***}	-0.208^{***}
	-0.058	(0.058)
Student	-0.162^{**}	-0.193^{***}
	-0.064	(0.064)
Offset	No	Yes
Observations	57,434	57,429
Log Likelihood	-45,332	-44,973
Akaike Inf. Crit.	90,870	90,154

Note: Model (1) has no offset and (2) is offset by the driving experience. The selected model to estimate λ is (1). $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. The underlined model "Poisson Offset".

Table 2.5 shows the output of the latent variable from the predictions of the two models presented in Table 2.4. The average collision rate is estimated to be around 6% by the partner insurer. We observe that, as expected, the *"No-Offset"* regression underestimates the average risk level at 3.6% while the *"Offset"* regression has an average risk level of 5.7%, in line with what is expected from our partner estimate. Given a risk level λ, we estimate the covariate for each insured using (2.5). A description of $r(x)$ and $log(r(x))$ is given in Table 2.5.

Table 2.5: Risk level yearly rate and Threshold

		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
λ No_Offset	Poisson	0	.028	.037	.036	.045	.321
λ Offset	Poisson	0	.044	.056	.057	.067	.315
$r(x)$ Threshold	Raw	-0.001	0	.001	.003	.002	45.482
	Log	-16	-7.5	-6.7	-6.8	-6.0	3.817

Note: This table shows the key statistics on the location of the density of the predicted λ, the yearly risk of collision. No_offset is the regression where the number of claims in the last 10 years is not simply divided by 10 to obtain a yearly rate. Offset weight the number of collision in 10 years by the driving experience of the insureds. $r(x)$ Threshold line shows the dispersion of the threshold r(x) obtained after applying (2.5) to the λ obtained from Poisson with Offset.

2.5.2 Estimation of risk aversion parameter η

We estimate η using the Probit model described in (2.13) with the added covariate (2.5). $\delta p = p^l - p^h$, where p^l is the premium charged if the insured selected the low deductible 500$ and p^h is the premium charged if the insured selected the high deductible of 1000$. As mentioned in Section 2.3, the regular premium is calculated for a deductible of 500$. If the insured chooses a higher/lower deductible, a mark-down/mark-up is applied to the regular premium. Each observed premium is hence reweighted once to its "regular" amount with the multipliers practiced when the policy was last renegotiated.

In the Israeli floating deductible system, $\frac{\Delta p}{\Delta d}$ is fixed for any insured as the mark-up and mark-down are applied to both deductible and premium. Cohen and Einav (2007) deduced that a risk-neutral claimant, $\eta = 0$, will choose a low deductible contract if his claim rate is higher than 30% regardless of its cost. A similar estimation in the Canadian context using the average collision premium of 260$ and a fixed deductible difference of 500$, gives that a risk-neutral claimant will select the low deductible contract if his yearly collision rate is greater than 7.8%. According to different estimates of λ in Table 2.5, between 12% and 25% of the population has a higher collision rate per year.

Due to the fixed cost structure of the deductibles, there are 11,546 observations with negative thresholds and 5,271 with an absent threshold due to missing information in the regression λ. Insureds with a negative threshold will always take the highest possible coverage regardless of their yearly risk rate. The covariate $ln(r(x))$ removes these 16,818 observations. We hence train two different models: Model (1) includes the population who selected not to have a collision coverage as having chosen the lowest deductible coverage between 500 and 1000. Model (2) excludes that population entirely.

Tables Q.17 and Q.18 in Appendix Q introduce the set of variables significant at the 1% level for the two models presented. We compare the results obtained with the insights obtained by Cohen and Einav (2007)'s Table 3 column 2. Contrary to Cohen and Einav (2007), our model does not find gender nor age as significant factors in determining risk aversion. We observe that the family variable [4] is positively correlated to risk aversion in both our model and Cohen and Einav (2007). Luxury vehicles tend to increase the risk aversion parameter, while the influence of income remains inconclusive in our models, as in Cohen and Einav (2007). Owning a European vehicle strongly indicates lower risk aversion, while controlling for imported, economy, or mid-size cars increases risk aversion. Experienced drivers, older vehicles, and higher annual mileage each contribute to

[4] Nbr_Potential_Drivers is our proxy

higher levels of risk aversion among the insured.

Similarly to Cohen and Einav (2007), $log(r(x))$ has a significant negative effect on the estimate of risk aversion. Given a level of risk λ, $r(x)$, given by (2.5), is the "value" of risk aversion that makes an insured indifferent between a low and a high coverage. If an individual has a higher risk aversion than $r(x)$, he then chooses the high coverage. Hence, when $r(x)$ increases, the probability of choosing a low deductible decreases as shown in Table Q.18.

Table 2.6 presents the distribution of negative and absent thresholds based on coverage choices. Of the negative threshold observations, 8,456 (73%) are from insured individuals who consistently opted for a low deductible of 500. An additional 2,162 (18%) negative threshold observations come from those who chose no coverage, making the collision premium used to calculate the threshold a prediction, as noted in Section 2.3. While these negative threshold observations cannot be used to infer η, they generally align with the coverage choices of the insured. Furthermore, Table 2.6 indicates that most of the cases with no risk level predictions correspond to claimants lacking collision coverage.

Table 2.6: Negative Threshold Analysis

	Deductible Choice			
$Threshold < 0$	0	500	1000	No Coverage
False	0	27,784	13,822	4,166
True	23	8,456	905	2,162
No λ	22	906	392	3,951

Note: This table shows the distribution of observations due to negative threshold per group of coverage.

Table 2.7 displays some statistics on the estimation of η. Model (1)/(2) is the Probit regression where the No-coverage population is excluded/included. We expect to see insureds who chose no-coverage to be on average less risk averse than insureds who have

chosen a 1000$ deductible and in turn to insureds who picked a 500$ deductibles. We clearly see such a relationship: For Model (1) and (2), we see a translation to the right of the distribution of η as the chosen coverage increases. The general density also displays such coherence: we notice a lower density of our risk-aversion estimate for Model (1) than for Model (2).

Table 2.7: η Predictions

		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Model (1) I	Log(η)	-7.07	-0.38	0.88	0.81	2.14	3.91
	η	0.00085	0.680	2.419	6.958	8.527	49.972
	η for No Coverage	0.00085	0.17	0.60	3.16	2.34	49.88
	η for Ded_1000	0.0025	0.45	1.35	4.51	4.49	49.93
	η for Ded_500	0.0058	1.24	4.14	8.98	12.32	49.97
Model (2) E	Log(η)	-5.97	-0.21	1.01	0.95	2.23	3.91
	η	0.00255	0.81	2.73	7.36	9.24	49.97
	η for Ded_1000	0.0025	0.46	1.36	4.51	4.49	49.93
	η for Ded_500	0.0059	1.24	4.15	8.99	12.33	49.97

Note: Log(η) and η are respectively the risk aversion and the log of risk aversion. η for No Coverage, Ded_1000 and Ded_500 show the distribution of the inferred risk aversion for each group: individuals with no coverage, individuals with high deductible 1000$ and individual with low deductible 500$.

Given an estimate of η, we test if our predictions correctly predict the choice of coverage. We should observe a choice of low deductible only if $\eta \geq r(x)$. Table 2.8 shows the confusion matrix of the two models. We see that the predictions overshoot the threshold for almost all observations. Given our estimate of η and λ, all insureds with low coverages should switch to high coverage as (2.6) is verified. In Table 2.8, these observations are 17,915 for Model (1) and 13,732 for Model (2). Our estimates of η are likely too high for these insureds.

We further inspect the distance $\eta - r(x)$. Table 2.9 shows the distributions of the distances for true-positive observations, for insureds who chose the highest coverage with $\eta > r(x)$, and false-positive observations, for insureds who chose the lowest coverage with

Table 2.8: Confusion Matrix

		Model (1)		Model (2)	
Prediction		0	1	0	1
Choice	0	14	17915	6	13732
	1	3	27709	0	27655

Note: Row Prediction corresponds to observations where $\eta \geq r(x)$ meaning that the model predicts the insured should opt for the lower deductible. Row Choice is the actual choice observed in the data. 1/0 means high/low coverage contract.

$\eta \leq r(x)$. Several observations can be drawn. First: true-positive distances are much more spread with higher variance and higher mean than false-positive distances. Second: the distribution of distances is higher for true-positive than for false-positive. This indicates that the overestimated η for insureds who chose the lowest coverage is still likely lower than the estimated η for true-positive, allowing to still effectively split the two populations. Similar conclusions can be drawn for both Model (1) and (2).

Table 2.9: Misclassification Error

		Mean	Std	1st Qu.	Median	3rd Qu.
Model (1) I	FP	16	279	0.39	1.27	4.90
	TP	321	36996	1.61	6.34	25.78
Model (2) E	FP	12	156	0.68	1.73	5.07
	TP	93	3432	2.35	6.87	21.91

Note: Column Mean is the average distance $\eta - r(x)$ and Std is its standard deviation. Column 1st Qu. is the location of the 25% quantile of the distance $\eta - r(x)$. Median is the median and Column 3rd Qu. is the location of the 75% quantile.

For an equivalent collision risk level λ, a higher risk aversion entails a higher coverage

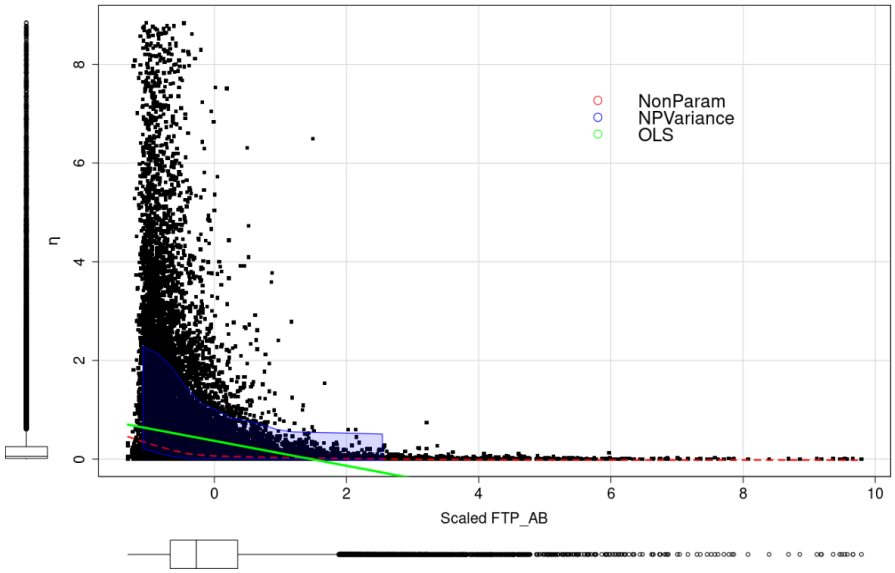
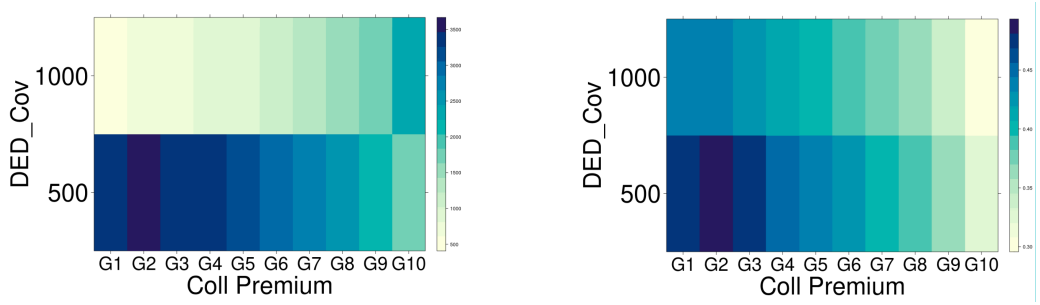


Figure 2.4: Correlation between Scaled FTP_AB and η

leading to a higher collision premium. We check the correlation between FTP_AB and the estimated η. We find a small negative correlation as expected from the medium positive relation between FTP_AB and FTP_Cov. The Pearson correlation stands at -0.26. Figure 2.4 shows the scatter-plot of the scaled AB premium and risk aversion η. The red curve shows the OLS regression result while the green curve is the non-parametric loess regression with a blue confidence interval. Figure 2.4 confirms the negative correlation between the two variables. This negative correlation rapidly decreases as the premium increases, indicating that insureds deemed to have a higher risk of bodily injury have a lower risk aversion.

As mentioned in Section 2.2, Picard et al. (2024)'s model includes two types of risk averse agents. In their results, higher coverage contracts attract more risk averse insureds and are less prone to fraudulent behaviors. Figures 2.5a and 2.5b show respectively the population and the averaged η per type of coverage contract. The X-axis is the deciles



(a) Population per contract type

(b) η per contract type

Figure 2.5: Analyse per contract type

of "regular" collision premium, the actuarial premium charged before the choice of deductible, and the Y-axis is the chosen collision deductibles. The low negative correlation between FTP_AB and η shows that the risk aversion parameter brings extra information to predict fraud than its potential link to the AB premium.

Figure 2.5a shows that insureds with lower actuarial risks choose higher coverage. This result is surprising as one would expect higher risk insureds to opt for a higher coverage. One possible explanation may lie with Figure 2.5b which displays the averaged η per contract type. We see that η is constantly decreasing in both premium and deductible choices. All else being equal, and in line with Picard et al. (2024)'s predictions, more risk-averse insureds would choose higher coverage. This effect of risk aversion appears sufficient to counterbalance a lower actuarial risk of collision.

Figure 2.6 shows the observed fraud rate per type of contract. Once again Picard et al. (2024)'s conclusions hold true. We see a higher observed fraud rate associated with higher deductibles. This fraud rate is also constantly increasing with the actuarial premium representing the Collision risk estimated by the insurance company.

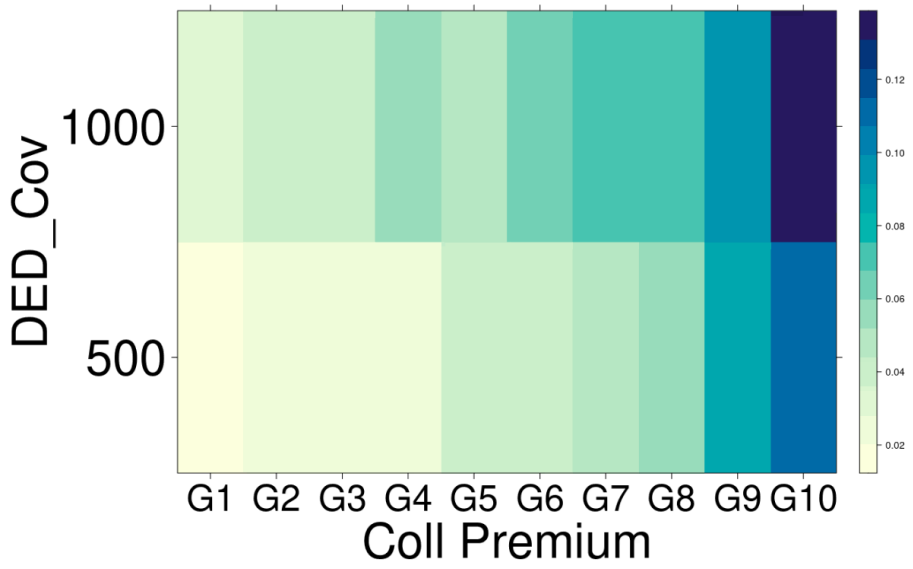


Figure 2.6: Fraud rate per contract type

2.5.3 σOAM with η

The Optimal Audit Model (OAM) is a cost-sensitive model that was developed by Dionne et al. (2009). σOAM is an extension of OAM introduced by Zerbato et al. (2024). Both OAM and σOAM produce an audit strategy that provides an audit recommendation based on a set of independent variables emitted by each claim. The goal of the audit strategy is to minimize the total expected cost of fraud by using the signal emitted by a classifier, such as logistic regression or other machine learning classifiers.

The resulting audit strategy is based on two dimensions: the θ group to which the claimant belongs and the signal emitted by the claim itself. In OAM, a claim signal consists of a vector of binary variables σ of dimension k that is formed from the discretized top predictors of the classifier. Since all $\sigma \in \Sigma$ are made up of binary variables, there is a finite universe of cardinality signals $2^k = l$ that can be ranked according to (2.14). Zerbato et al. (2024) extended OAM to σOAM by directly discretizing the calibrated output of the

base classifier to form the σ signal of a claim which can then also be ranked following (2.14).

$$\frac{P_1^F}{P_1^N} < \frac{P_2^F}{P_2^N} < ... < \frac{P_i^F}{P_i^N} < ... < \frac{P_l^F}{P_l^N} \tag{2.14}$$

where P_i^F and P_i^N be the respective probabilities of the signal σ to take configuration i conditional on the claim being fraudulent (F) or non-fraudulent (N).

$$P_i^F = P(\sigma = \sigma_i / F) \tag{2.15}$$

$$P_i^N = P(\sigma = \sigma_i / N) \tag{2.16}$$

Fraudulent behavior is modeled through a utility function $\mathscr{U}(W, \omega)$, where W is wealth, and ω represents the moral cost of fraud. We assume that $\mathscr{U}'(W) \geq 0$, $\mathscr{U}''(W) \leq 0$ and $\mathscr{U}'(\omega) \leq 0$. Claimants will fraud if their expected utility, given by (2.17), exceeds the status quo.

$$(1 - P(A/F))\mathscr{U}(W_0 + t, \omega)) + P(A/F)\mathscr{U}(W_0 - B, \omega)) \geq \mathscr{U}(W_0, 0) \tag{2.17}$$

where t is the compensation asked, W_0 is the initial claimant's wealth and B is the claimant's penalty linked to fraud.

The recommended audit policy aims to minimize the total expected cost of fraud, which can be broken down into two components: the expected cost of investigation, shown in (2.18), and the expected cost of residual fraud, represented in (2.19). The model does not account for opportunistic fraud, where an individual might exaggerate an actual injury. Instead, it assumes that a claim involving an injury is either completely fabricated or entirely legitimate. As such, π represents the probability that an AB (Accident Benefit)

claim is genuine. Let $H(\omega/\theta)$ be the conditional cumulative distribution function of the moral cost of fraud for each group θ, and $h(\omega/\theta)$ its corresponding density.

$$c\pi E_\theta[Q^N(\theta)] + c(1-\pi)E_\theta[Q^F(\theta)H(\phi(Q^F(\theta)/\theta))] \tag{2.18}$$

$$t(1-\pi)E_\theta[(1-Q^F(\theta))H(\phi(Q^F(\theta))/\theta)) \tag{2.19}$$

Adding (2.18) and (2.19) and minimizing the sum over $q(\theta,\sigma)$ gives the global optimization problem in (2.20).

$$\begin{aligned} Min_{q(\theta,\sigma)}E_\theta \Bigg[& c\pi[Q^N(\theta)] + c(1-\pi)Q^F(\theta)H(\phi(Q^f(\theta)/\theta)) \\ & + t(1-\pi)\Big((1-Q^F(\theta))H(\phi(Q^F(\theta))/\theta))\Big)\Bigg] \end{aligned} \tag{2.20}$$

Proposition 1 of Dionne et al. (2009) shows that, under limited conditions, the problem of solving (2.20) consists of finding an optimal severity index i in (2.14) for each θ. Then the optimization problem becomes (2.21):

$$\begin{aligned} Min_{i_\theta}E_\theta \Bigg[& c\pi\mu(i^*_\theta) + (1-\pi)H(\phi(\lambda(i^*_\theta))/\theta) \\ & \Big(c\lambda(i^*_\theta) + t(1-\lambda(i^*_\theta))\Big)\Bigg] \end{aligned} \tag{2.21}$$

where $\gamma(i) = \Sigma_{i=1}^{l}(P_i^F)$ and $\mu(i) = \Sigma_{i=1}^{l}(P_i^F)$. Chapter 1 further describes the parametrization methodology in detail.

Zerbato et al. (2024) used the Extreme Gradient Boosting (XGB) framework as a base classifier in OAM and σOAM. Here, we use η as an additional feature-engineered variable in the same training set used in Zerbato et al. (2024). We compare the results obtained between 5 different scenarios. Scenario 0 is the base scenario in which the XGB model training is carried out without the η but on a training database without the observations

for which no estimate of η could be made. Scenario 1 trains the model with the η variable that is introduced in (2.13) and detailed in Table Q.17 and Q.18 in Appendix Q. Scenario 2 trains the model with η but removes any variable used in inferring η in the general training. Scenario 3 does not include η but the variable FTP_AB as a test to compare with the classification performances obtained with η. Scenario 4 combines the premium variable FTP_AB and η.

Section 2.3 identified a moderate correlation between FTP_Cov and FTP_AB. According to Zerbato et al. (2024), the detected fraud rate increases with higher AB premium groups θ. When adding a scenario involving FTP_AB, we find that η provides additional information beyond what FTP_AB alone offers. In line with Zerbato et al. (2024), we compare three key metrics: the F1 classification performance of the XGB model on the test set, the F1 score after applying σOAM, and the amount of savings compared to the baseline scenario of no-audit.

Table 2.10: σOAM results when η is added

Scenario	Training-Set	F1	$F1_{\sigma OAM}$	Savings $	% of Cost-Saved
(0)	Base Scenario	46.2	45.6	2,128,000	37.8%
(1)	η with η vars	45.3	45.0	2,137,942	38.0%
(2)	η with No η vars	46.9	46.1	2,195,562	39.0%
(3)	Base + FTP_AB	45.5	45.5	2,118,984	37.1%
(4)	Base + FTP_AB + η	46.2	45.7	2,123,792	37.7%

Note: Column Scenario and Training-set show the different configuration tested. The training set removes observations for which η cannot be estimated. η vars are the variables introduced in (2.13) and some of them are presented in Table Q.17 and Q.18 in Appendix Q. The Base Scenario is the XGB framework with a Dart classifier with hyper-parameters similar to Zerbato et al. (2024). Column F1 and $F1_{\sigma OAM}$ are respectively the F1 scores of the XGB on the test set before and after σOAM is applied. Savings is the amount saved to the corner solution of no audit. % of Cost-Saved is the percentage of cost saved of the total cost induced by fraud obtained by subtracting the corner solution of no-audit to the corner solution of perfect predictions. Row Base + FTP_AB is the base scenario with the added premium charged for AB coverage as an additional training variable.

Table 2.10 displays the results obtained for each training scenario. We see an increase in performance when the η or FTP_AB features are added to the model. This increase is felt through all 3 key performance measures. We observe a change in classification performances of -2% from an F1 score of 46.2 to 45.3, $+1.5\%$ from 46.2 to 46.9, -1.6% from 46.2 to 45.5 and no change in classification for Scenario (1), (2), (3) and (4) respectively. The increase is the strongest in Scenario (2) when the risk aversion feature is added to the training and variables used to infer it are removed. We see a similar pattern when it comes to classification performances after the σOAM is applied, as displayed in Column σOAM in Table 2.10. Column "% of cost saved" shows that Scenario (1) and (2) do better than Base Scenario by respectively 0.4% and 3.1% in absolute savings, a progress of 0.2% and 1.2% in the total cost of fraud as seen in Column "% of Cost-Saved". Scenario (3) and (4) test the influence of the premium variable FTP_AB in the prediction of fraud. Both scenarios do not perform better than Base Scenario which indicates that the increase of performance resulting from including η as a training variable does not come from the correlation between η and FTP_AB.

Table 2.11 presents the top 10 predictors of fraud according to the Shapley value analysis of the trained XGB model. Shapley value analysis was developed by Lloyd Shapley in the game theory literature and is now widely used to interpret machine learning models. As described in Lundberg et al. (2018) and Lundberg and Lee (2017), SHapley Additive exPlanation, or SHAP, value analysis is a consistent and model agnostic method to evaluate the importance of individual features in the prediction. The fundamental idea behind SHAP method is to sequentially add the contribution of each feature to the general expected value, given a vector of characteristics X. The goal is to recreate the model's prediction and test different combinations of features to obtain a global estimate of the contribution of each individual feature to the model's output. Individual feature contributions are then ranked by their absolute averaged score. We use the R package *"SHAP-*

Table 2.11: Top 10 predictors of fraud per scenario

Scenario	(0)	(1)	(2)	(3)	(4)
No_Lawyer	1	1	1	1	1
Aggressive Lawyer	2	2	2	2	2
# days between loss and client inception	3	3	3	4	4
# Lawyers	4	5	5	5	5
# days between loss and first Lawyer	5	6	6	6	7
Reserve when AB sides opened	6	8	7	8	9
# days between AB sides and Loss	7	7	8	7	8
# claims on policy	8	9	9	9	10
Reserve at 10 days after first AB	9	10	10	10	> 10
# days before serious injuries	10	> 10	> 10	> 10	> 10
η	NA	4	4	NA	6
FTP_AB	NA	NA	NA	3	3

Note: Table 2.11 presents the ranking of the top 10 predictors of each scenario presented in Table 2.10. NA means that the variable was not tested. Each variable is given a rank from 1 to 10 representing its rank by contribution in the fraud prediction. Each column (0) to (4) represents each Scenario presented in Table 2.10.

forxgboost"[5] to isolate a set of size n top contributing variables based on their averaged absolute contribution.

Both η and FTP_AB appear in respectively fourth and third place, just after variables controlling if no lawyer was involved in the claim, if an aggressive law firm was present or the number of days between the date of loss and the first date the claimant became insured with the company. This is consistent with the results obtained in Zerbato et al. (2024) which highlighted the importance of law variables in predicting fraud. These results are true for all three scenarios presented in Table 2.10. Figure 2.7 displays η and FTP_AB relation to the probability of observing fraud. As expected, FTP_AB and η have an inverse influence on the predicted fraud rate. We find that a higher tolerance for risk has a strong positive influence on the classifier to predict fraud. This influence decreases rapidly as risk

[5]https://liuyanguu.github.io/post/2019/07/18/visualization-of-shap-for-xgboost/references

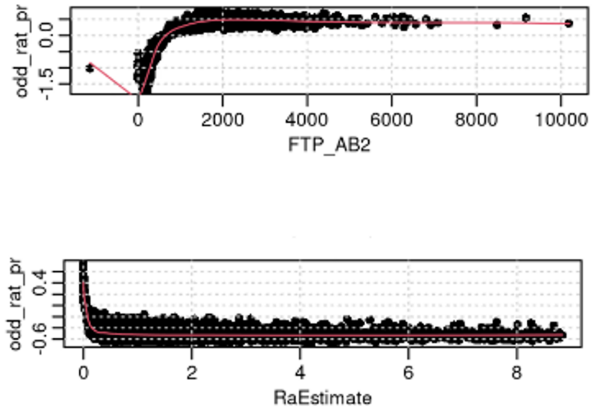


Figure 2.7: Contribution of η and FTP_AB to prediction using Shape Analysis

aversion increases. We find a mirror relation with the AB premium.

2.6 Conclusion

We have estimated the risk aversion parameter by applying the first model of Cohen and Einav (2007) and using Canadian data on choices of insurance policy related to Collision coverages. We show that a medium correlation exists between the AB premium and the Collision premium. Such correlation is inverted with the risk aversion characteristic. We hence observe that insureds deemed to have a higher risk of suffering an accident with bodily injuries, have an average lower risk aversion. The resulting η variable is used as an added feature in the training of the expert system presented in Zerbato et al. (2024). We observe an increase in classification performances and in savings when the risk aversion feature is added to the base scenario. We also show that the newly added feature

systematically ranks among the top predictors to detect fraud.

Bibliography

Arrow, K. J. (1964). “The Role of Securities in the Optimal Allocation of Risk-bearing”. In: pp. 91–96. ISSN: 00346527, 1467937X.

— (1973). “Optimal insurance and generalized deductibles”. In: *Scandinavian Actuarial Journal* 1, pp. 1–42.

Boyer, M. and R. Peter (2020). “Insurance Fraud in a Rothschild–Stiglitz World”. In: *Journal of Risk and Insurance* 87.1, pp. 117–142.

Chetty, R. (2006). “A New Method of Estimating Risk Aversion”. In: *American Economic Review* 96.5, pp. 1821–1834.

Chiappori, P. A. and M. Paiella (2011). “Relative Risk Aversion Is Constant: Evidence From Panel Data”. In: *Journal of the European Economic Association* 9.6, pp. 1021–1052.

Cohen, A. and L. Einav (2007). “Estimating Risk Preferences from Deductible Choice”. In: *American Economic Review* 97.3, pp. 745–788. DOI: 10.1257/aer.97.3.745.

Dionne, G., F. Giuliano, and P. Picard (2009). “Optimal Auditing With Scoring: Theory and Application to Insurance Fraud”. In: *Management Science* 55, pp. 58–70. DOI: 10.1287/mnsc.1080.0905.

Eisenhauer, J. G. (2008). “Ethical Preferences, Risk Aversion, and Taxpayer Behavior”. In: *Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics)* 37.1, pp. 45–63.

Friend, I. and M. E Blume (1975). “The Demand for Risky Assets”. In: *American Economic Review* 65.5, pp. 900–922.

Ghosh, D. and T. L. Crain (1995). “Ethical Standards, Attitudes toward Risk, and Intentional Noncompliance: An Experimental Investigation”. In: *Journal of Business Ethics* 14.5, pp. 353–365.

Guiso, L. and M. Paiella (2008). “Risk Aversion, Wealth, and Background Risk”. In: *Journal of the European Economic Association* 6.6, pp. 1109–1150. (Visited on 03/25/2024).

Halek, M. and J. G. Eisenhauer (2001). “Demography of Risk Aversion”. In: *The Journal of Risk and Insurance* 68.1, pp. 1–24.

Holt, C. A. and S. K. Laury (2002). “Risk Aversion and Incentive Effects”. In: *American Economic Review* 92.5, pp. 1644–1655.

Jiang, X., M. Osl, J. Kim, et al. (2011). “Smooth Isotonic Regression: A New Method to Calibrate Predictive Models”. In: *AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science* 2011, pp. 16–20.

Lesch, W. and J. Brinkmann (2011). “Consumer Insurance Fraud/Abuse as Co-creation and Co-responsibility: A New Paradigm”. In: *Journal of Business Ethics* 103, pp. 17–32.

Loubergé, H. and G. Dionne (2024). *Developments in risk and insurance economics: The past 50 years*. Tech. rep. HEC Montreal, Canada Research Chair in Risk Management.

Lundberg, S., G.G. Erion, and S.I Lee (2018). “Consistent Individualized Feature Attribution for Tree Ensembles”. In: *CoRR* abs/1802.03888. arXiv: 1802.03888. URL: http://arxiv.org/abs/1802.03888.

Lundberg, S. and S.I. Lee (2017). "A unified approach to interpreting model predictions". In: *CoRR* abs/1705.07874. arXiv: 1705.07874. URL: http://arxiv.org/abs/1705.07874.

Moreno, I., F. Vázquez, and R. Watt (2006). "Can Bonus-Malus Alleviate Insurance Fraud". In: *Journal of Risk and Insurance* 73, pp. 123–151.

Muller, K., H. Schmeiser, and J. Wagner (2011). "Insurance Claims Fraud: Optimal Auditing Strategies in Insurance Companies". In.

Outreville, J. F. (2014). "Risk Aversion, Risk Behavior, and Demand for Insurance: A Survey". In: *Journal of Insurance Issues* 37.2, pp. 158–186.

Picard, P. (1996). "Auditing Claims in the Insurance Market with Fraud: The Credibility Issue". In: *Journal of Public Economics* 63.1, pp. 27–56.

Picard, P., J. Wang, and K. C. Wang (2024). *The Collusion Between Policyholders and Car Dealers and the Manipulation of Insurance Claims*. Tech. rep. Ecole Polytechnique, France.

Pratt, J. W. (1964). "Risk Aversion in the Small and in the Large". In: *Econometrica* 32.1/2, pp. 122–136.

Ribeiro, R., B. Silva, C. Pimenta, et al. (2020). "Why do consumers perpetrate fraudulent behaviors in insurance?" In: *Crime, Law and Social Change* 73.

Schlesinger, H. (2013). "The theory of insurance demand". In: *Handbook of insurance*, pp. 167–184.

Sydnor, J. (2010). "(Over)insuring Modest Risks". In: *American Economic Journal: Applied Economics* 2.4, pp. 177–199.

Tennyson, S. (2008). "Moral, Social, and Economic Dimensions of Insurance Claims Fraud". In: *Social Research* 75.4, pp. 1181–1204.

Tennyson, S. and P. Salsas-Forn (2002). “Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives”. In: *Journal of Risk and Insurance* 69.3, pp. 289–308.

Zerbato, R., G. Dionne, and F. Bellavance (2024). *Auto-insurance Fraud: a new cost-sensitive expert system*. Tech. rep. Canada Research Chair in Risk Management, HEC Montreal.

Chapter 3

Auto Insurance Fraud: An endogenous measure of the deterrence effect.

Abstract

We modify and extend the Optimal Audit Model (OAM) designed by Dionne et al. (2009) to obtain a measure of the deterrence effect induced by an audit policy. Our extension is three-fold: We first show that OAM corresponds to the perfect Bayesian equilibrium of a sequential game between the insurer and the insured where the personal fraud penalty, or the moral cost of fraud, is private knowledge to the insured and is measured in utility terms. Under an explicit utility assumption and using a measure of risk aversion estimated from the data, we are able to infer the reaction function of the insureds to a given audit policy. We then determine the overall fraud rate based on the current audit policy and estimate the deterrence effect from the optimal Stackelberg equilibrium audit strategy for each group of claimants. We compare this optimal audit policy with the theoretical results from key contributions in the literature and formulate recommendations on how to upgrade

the current audit strategy of an insurer with a better estimate of its deterrence effect.

3.1 Introduction

The deterrence effect is defined as the inhibition of criminal behavior through the threat or application of punishment.[1]. In insurance, the role of an audit policy serves two purposes: to prevent the payout of fraudulent claims and to create a dissuasive effect, discouraging potential fraudsters. Insurers invest heavily in fraud detection processes, which together constitute their audit policy. However, optimizing this policy can be challenging. To improve efficiency and outcomes, insurers often turn to theoretical insights and statistical models to guide their auditing strategies.

The Optimal Audit Model (OAM), developed by Dionne et al. (2009), integrates a theoretical economic framework with statistical modeling to generate an optimal audit policy. In this study, we modify and extend the OAM to incorporate a measure of the deterrence effect caused by an audit policy. Leveraging data from a partner insurance company, we infer the distribution of the moral cost of fraud, ω, representing the insured's personal penalty for committing fraud. With an explicit utility function, we are able to infer the insureds' reaction to fraud penalties. Using predictions of conditional fraud rates derived from a machine learning algorithm, we estimate the overall fraud rate under the current audit policy and determine the optimal audit policy based on a Stackelberg equilibrium approach, incorporating the audit policy's deterrence effect. We then compare this optimal policy with key theoretical contributions in the literature and offer recommendations to enhance the partner's current audit strategy.

The structure of this paper is as follows: The next section provides a review of relevant literature. Section 3.3 introduces the sequential game that corresponds to the OAM

[1] https://www.merriam-webster.com/dictionary/deterrence

framework. In Section 3.4, we highlight key findings from the literature and replicate these results using our dataset. Section 3.5 is divided into four subsections, which cover: 1) the parametrization process used to estimate the density of expected loss; 2) an evaluation of three different scenarios under the current audit policy and their impact on the proportion of *Dishonest* types across risk levels; 3) the derivation of the optimal audit policy and its associated deterrence effect for each claimant group; and 4) a sensitivity analysis to examine how the parametrization affects our overall findings. Finally, Section 3.6 concludes the study.

We confirm two established findings from the literature within our model and introduce three novel results. The validated findings are: 1) Expensive claims should be audited more frequently than inexpensive ones, but *High-risk* claimants with expensive claims should not face full audit-induced deterrence. 2) Fraud persists at equilibrium because not all *Dishonest* insureds are fully deterred. Our new contributions are: 3) The proportion of *Dishonest* claimants is higher among *High-risk* individuals. 4) Risk aversion plays a significant role in the decision to commit fraud. 5) Rewards and penalties are the most decisive factors in determining the optimal audit level, suggesting that distributing claim payments in delayed installments could significantly enhance the deterrence effect.

3.2 Literature review and Main contribution

3.2.1 Literature review

Tennyson and Salsas-Forn (2002) observes that two strands of literature exist on the topic of insurance fraud. One, grounded in optimal contracting and game theory, seeks to minimize the total expected cost of fraud, which encompasses the cost of fabricated claims. The other focuses on statistical models that help build expert systems to flag suspicious

claims for auditing.

Labeling data on insurance fraud presents inherent challenges. One major issue is determining the correct proportion of fraudulent claims in the dataset. A cost-sensitive expert system trained on this incomplete information can only recommend an audit policy aimed at minimizing the expected cost of the known fraud cases. Therefore, while expert systems are valuable, they remain incomplete if the objective is to establish an optimal audit policy.

Tennyson and Salsas-Forn (2002) points out that a key insight of economic theory is the emphasis on deterrence as a central aspect of combating fraud: *"... the primary role of auditing in an optimally designed system is the deterrence..."* of fraud *"... rather than its detection."*[2]. However, theoretical models are often highly stylized and offer general guidance rather than specific and actionable solutions. Tennyson and Salsas-Forn (2002) examine whether these general theoretical observations are reflected in actual fraud detection practices. The authors conclude that certain theoretical insights are indeed borne out in empirical findings: more expensive claims have a higher probability of being audited, and *build-up* claims[3] also face increased audit scrutiny. Furthermore, the occurrence of false positives is interpreted as evidence that insurance companies are already utilizing audits primarily as a deterrence tool.

To deter fraud, we must identify its most influential parameters. Lesch and Brinkmann (2011), Ribeiro et al. (2020), and Tennyson (2008) explore the four main and mutually non-exclusive types of rationalization for fraudulent behavior among insurance customers. The first and second types relate to the *Economic and Contractual* and to *Macrosocial* approaches. Both emphasize fraud as a rational assessment of potential economic gain. The main factors influencing the decision to commit fraud are the loss/benefit ratio, the

[2]Page 4 of Tennyson and Salsas-Forn (2002)

[3]For example, claims involving injuries that are more difficult to assess (e.g., sprains)

probability of audit, and the overall economic situation.

The third type concerns the *Moral Sociological Views* of claimants. Important factors include social norms and relationships, such as the *prevalence and acceptance of insurance fraud* and the *poor public image or lack of awareness of the societal role of insurance companies*. One might argue that the audit rate demonstrates that insurance companies are actively working to resolve inefficiencies caused by fraud, rather than simply passing the costs on to customers. The fourth type is the only approach that does not rely on current audit policies as a key factor: *Moral Psychological Views*. This reasoning involves individual characteristics such as risk aversion, with the main driver being the personal reward gained from committing fraud, sometimes even disregarding the factual context. Of the four types, only *Moral Sociological Views* does not treat audit policy as a key factor.

The deterrence effect of audit policy is central to the theoretical body of literature discussed by Tennyson and Salsas-Forn (2002). Picard (2013) clearly illustrates this in his comprehensive review. An audit policy becomes optimal when it balances three conflicting forces: the cost incurred (or saved) from wrongly (or correctly) auditing claims, and the cost of undetected fraudulent claims, the rate of which is directly influenced by the chosen audit policy.

Townsend (1979) developed the Costly State Verification (CSV) model, which, like all its extensions[4], in the context of deterministic audits, relies on a credible deterrent audit policy to induce truth-telling from the insured. In the case of random auditing, Mookherjee and Png (1989) and Fagart and Picard (1999) demonstrate that the penalty imposed on an audited fraudulent insured plays a pivotal role in formulating the optimal contract. First, if no upper bound is placed on the penalty, arbitrarily large fines would effectively deter any potential fraud while requiring only a minimal audit policy. Additionally, the penalty

[4]Subelj et al. (2011), Bodaghiand and Teimourpour (2018), Crocker and Morgan (1998), Gollier (1987)

cannot be linked to the reported loss, as this could create compensation zones where truth-telling is not optimal.

Mookherjee and Png (1989) derive a general optimal contract where the penalty incurred equals the size of the actual loss suffered by the fraudulent claimant. If the claimant is found to be lying, the claim is automatically dismissed. Their optimal contract has several distinctive properties, one of which includes a reward indemnity for insureds who truthfully report their losses when audited. However, this particular feature is not empirically observed. The resulting optimal audit strategy is random for all reported losses above the lowest possible indemnity, with a higher probability of audit as the claimed indemnity increases. As noted by Picard (2013), Mookherjee and Png (1989)'s optimal contract is difficult to characterize beyond some general results, making it impractical for application in the current insurance industry.

Fagart and Picard (1999) fully characterize the optimal contract under random auditing, assuming a fixed penalty and a CARA (Constant Absolute Risk Aversion) utility function. The authors confirm Mookherjee and Png (1989)'s result that an optimal contract should reward truthful reporting when an audit occurs. However, they also find that the reward should decrease as the amount of compensation claimed increases. The optimal contract takes the form of a fixed deductible, with an audit probability that makes the claimant indifferent between committing fraud and remaining in a certain state of not claiming, regardless of whether an actual accident occurred.

Several studies also examine the crucial role of audit commitment. Two types of games can emerge: First, insurers can credibly commit to an audit policy by taking a first-move advantage, gaining a Stackelberg advantage[5]. Alternatively, a no-commitment audit policy arises in a simultaneous game or when the insurer can only react to the insured's action. In

[5]This advantage occurs in practice when the insurer minimizes its cost function while knowing the response function of the insured.

such cases, the equilibrium reached is less efficient compared to that of full commitment.

In Picard (1996)'s auditing game, the insurer has a first-move advantage and can credibly commit to an audit policy. There are two types of insureds: opportunist and honest. The insured's type is private information, known only to the insured. Honest insureds face a high personal moral cost for committing fraud, while opportunists will choose to commit fraud if the audit policy lacks sufficient deterrence. A fixed and exogenous penalty is imposed when cheating is detected. The true loss, the cost of the audit, and the compensation are fixed. The optimal audit policy can fully deter fraud if the audit cost is sufficiently low and the proportion of opportunistic claimants is either locally high or null, depending on the contract type.

On the contracting game side, Picard (1996) assumes Wilson (1977)'s definition of market equilibrium as the smallest possible set of profitable contracts. The author shows that under this definition, only pooling contracts are viable at equilibrium and that such contracts need to be equivalent for both honest and opportunist claimants. Because of the pooling result, the proportion of opportunist claimants subscribing to a given contract will determine if no audit or full deterrence is optimal and if there is fraud at equilibrium for a given contract. Here the general proportion of opportunist claimants is an exogenous parameter of the population. Boyer (1999) obtains a similar conclusion in a context where the insurer's Stackelberg advantage is removed.

Under the assumption of constant Relative Risk Aversion (RRA), Boyer and Peter (2020) introduce high/low-risk agents into a no-commitment game. The authors show that, ceteris-paribus, high-risk claimants have a higher incentive to fraud. Under public information on risk type and constant or increasing relative risk aversion, the average compensation for high-risk claimants is increasing in their propensity to declare a loss and results in a higher incentive to audit for the insurer. This leads to two interesting and conflicting forces. On the one hand, high-risks have more incentive to fraud due to

high potential gains. On the other hand, they are also more likely to be in a real accident. The insurance contract at equilibrium is a separated one where high-risks pay a higher premium for a similar indemnity, whether risk types are known or not. When risk types are private, rationing of low-risks contracts can occur (up to no-coverage) from the insurer to avoid adverse selection situations. This low indemnity reduces the value for insurers to audit low-risk individuals, making it even more likely for low-risk types to commit fraud. Consequently, when risk types are private information, the overall level of fraud in the economy is expected to increase.

In the claiming game, the no-commitment situation results in a lower global expected utility for honest claimants as fraud remains present at equilibrium and some of its cost is passed to them through higher premiums. Krawczyk (2009), using Fudenberg et al. (1990)'s folk theorem, compares the no-commitment and the commitment games in a repetitive set-up where the insurer faces a different short-lived claimant for a long number of periods. The key question is the possibility for the principal to recreate a *commitment effect* through actions repetition, hence diminishing the inefficiency arising from no-commitment.

As stated in Krawczyk (2009), Fudenberg et al. (1990)'s folk theorem insights make each short-lived agent always follows his immediate best response since no punishment can be enforced. A key assumption in this context is the observability of the principal's optimal mixed strategy. If observable, the short-lived agents will in turn deter the principal from deviating from it by immediately switching their best action responses. If the principal decides not to audit any longer, she will be punished by agents frauding immediately. If the strategy is hidden, the agents will become aware of a change in strategy as the deviation from the principal's anticipated cumulative payoff will widened. Agents will then adjust their own best response accordingly. Fudenberg et al. (1990) show that if the strategy is hidden, the highest payoff obtainable by the principal is hence the lowest payoff

possibly attainable in her strategy support.

Krawczyk (2009) shows that under full observability of the mixed strategy, there exists an equilibrium path that erases the inefficiency caused by the lack of commitment. Following Fudenberg et al. (1990), when agents play lawfully and when only the principal's actions are observable, the payoff associated with *Audit* is the maximum obtainable by the insurer. Therefore the insurer might as well audit all claims and all reports will be truthful.

Krawczyk (2009) remarks that in this highly stylised game, if there is only one short-lived agent facing the insurer at each period, the observability of a mixed strategy is disputable as the claim is either audited or not. More realistically, an insurer faces a high number of claimants simultaneously in each period. Krawczyk (2009) shows that as this number increases, the principal recreates the mixed strategy effect in the aggregate and tends toward a commitment equilibrium. Krawczyk (2009) further refines his model by allowing the short-lived agent to only observe the actions taken by the principal across the laps of the m last periods. The author shows that the best possible equilibrium's payoff for the principal is then situated between the long-term no-commitment and full commitment payoff, somewhat mitigating the inefficiency of the absence of full commitment.

3.2.2 Main Contributions

Following Tennyson and Salsas-Forn (2002), we find in Section 3.4 that some predictions in the literature are not supported empirically. Firstly, the range of available contracts is more limited than the literature suggests, as there is only a finite set of deductibles to choose from. Additionally, premiums are calculated according to strict actuarial rules, considering factors such as the type of car, the policyholder's age, the number of drivers, etc. In practice, the insurance contracting process is determined independently of fraud-related concerns. Therefore, we focus our analysis on the claims process, assuming the

contracting process is independent of fraud issues. However, the reverse is not true, and we show that the contracting process significantly influences the claims process.

Secondly, the immediate penalty for a claimant caught lying rarely exceeds the cancellation of any remaining payments, as noted in Mookherjee and Png (1989). Further consequences may arise when the claimant renegotiates their contract, typically in the form of higher premiums. However, the true impact of these consequences largely remains private knowledge for the insured.

As seen in Section 3.2.1, penalties play a crucial role in random audit models. If the penalty is merely equal to the actual loss, then submitting build-up claims is no longer a rational method of committing fraud. The fraudulent claimant would be better off entirely fabricating the claim, as this would minimize the potential penalty. The obvious counterargument is that build-up claims might be harder to detect than completely fabricated ones. For the purposes of this study, we assume that the fraudulent portion of a claim is entirely fabricated. Since we are dealing with bodily injury fraud, we assume the fraudulent claimants were never injured in the accident.

Thirdly, insurance companies collect a substantial amount of data on both claimants and claims. The availability of this data, combined with advancements in machine learning, has improved the performance of expert systems. In practice, insurance companies use these algorithms to detect fraudulent patterns in large datasets and flag suspicious claims for special investigation. As Tennyson and Salsas-Forn (2002) notes, while this empirical approach does not account for the deterrence aspect of fighting fraud, it can reliably predict a fraud score for each claim.

Dionne et al. (2009) developed the Optimal Audit Model (OAM) that links the deterrence effect of optimal auditing to the benefits of the scoring expert system. Our extensions to the model are three folds: 1) we show that OAM corresponds to the perfect Bayesian equilibrium of a sequential game of fraud between the insurer and the insured where the

penalty of the moral cost is private knowledge to the insured and is measured in utility terms. 2) We estimate the elasticity of the fraud rate to the audit policy by inferring the distribution of the expected loss and the average risk aversion for each group of claimants; and 3) From the inferred distributions, we are able to estimate the current total fraud rate given the current audit policy through the deterrence effect that such audit policy entails.

3.3 Extension of OAM

We provide an extension of the Optimal Audit Model (OAM) designed by Dionne et al. (2009). The cost function that is minimized in the original OAM has the following form:

$$Min_{Q^f}\Big[c\pi E_\theta[Q^n(\theta)] \tag{3.1a}$$

$$+c(1-\pi)E_\theta[Q^f(\theta)H(\phi(Q^f(\theta)/\theta))] \tag{3.1b}$$

$$+t(1-\pi)E_\theta[(1-Q^f(\theta))H(\phi(Q^f(\theta))/\theta))]\Big], \tag{3.1c}$$

where Q^f is the audit probability conditional on the observation being fraudulent, Q^n is the probability of audit conditional on the observation not being fraudulent. c is the cost of auditing a claim, t is the average payment for a claim not audited, π is the probability of accident, θ is a sub-population of claimants, ϕ is the mapping function between the optimal audit strategy and the threshold moral cost, H is the moral cost's CDF for a given θ, $H(\phi(Q^f(\theta)/\theta))$ represents the probability of individuals in the θ sub-population who will commit fraud under a given audit strategy. Line (3.1a) refers to the cost of false-positive claims, which are claims that are audited but not fraudulent. Line (3.1b) refers to the cost of true-positive claims, which are claims that are both audited and fraudulent. Line (3.1c) refers to the cost of false negative claims, which are fraudulent claims that are not verified.

Eq (3.2) is the probability of observing a fraudulent claim with the insurer's actual audit policy. Dionne et al. (2009) estimated (3.2) with (3.4). $\hat{\Gamma}(\theta)$ was parameterised by assessing the probability of filling a claim and required an assumed global fraud rate. The elasticity of the fraud rate to the audit policy $\eta(Q,\theta)$ was estimated by $(1-Q)^{\gamma(\theta)}$, leading to (3.4) as the final estimate of $\Gamma(\hat{Q},\theta)$.

$$\Gamma(Q,\theta) = (1-\pi)H(\phi(Q_\theta^f)/\theta) \tag{3.2}$$

$$\eta(Q,\theta) = \frac{-Q_\theta^f \phi'(Q_\theta^f) h(\phi(Q_\theta^f)/\theta)}{H(\phi(Q_\theta^f)/\theta)} \tag{3.3}$$

$$\Gamma(\hat{Q},\theta) = \hat{\Gamma}(\theta)(1-Q)^{\gamma(\theta)} \tag{3.4}$$

Figure 3.1 presents a sequential game reminiscent of the ones described in Picard (1996), Boyer (1999) and Boyer and Peter (2020). In the following sections, we start by solving this sequential game with the hidden expected cost ω. We then formally show that (3.1) corresponds to the sequential game in Figure 3.1.

3.3.1 Sequential Equilibrium

Figure 3.1 illustrates the sequential claiming game in the Optimal Audit Model (OAM). We introduce new parameters: ψ, the probability of fraud committed by claimants; P_2, the audit probability when no claim is made. We show that P_2 is a dominated strategy; P_1, the audit probability when a claim is submitted; and S, the insurer's belief that fraud has occurred given the existence of a claim.

This setup represents a sequential game where Nature, denoted as (N), first determines whether an accident occurs, with probability π. The next player is the Claimant, (C), who knows whether an accident has happened and decides either to commit fraud with probability ψ or to make a legitimate claim with probability $(1-\lambda)$. Finally, the third player,

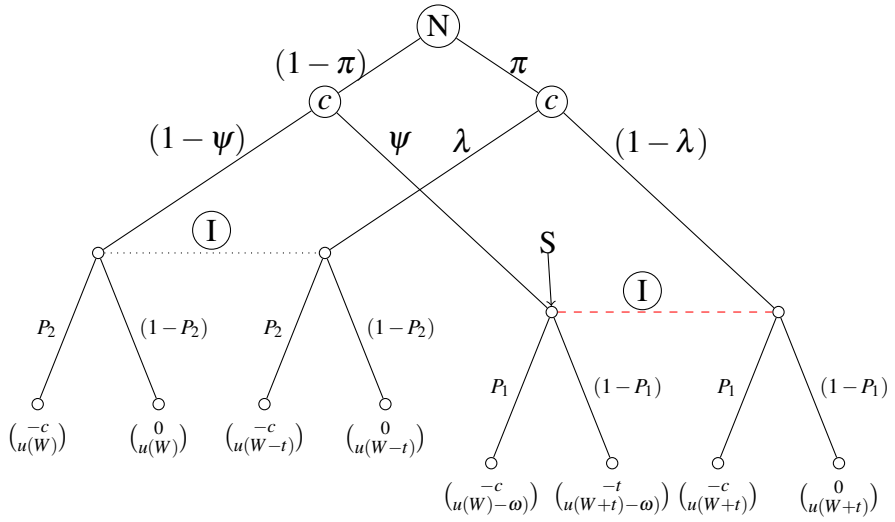


Figure 3.1: Sequential Tree

Note: (N) is Nature's move to decide if an accident occurs with probability π. (c) is the claimants decision to fill a claim with probability ψ, the unconditional fraud rate, and $(1-\lambda)$ if an accident occurs or not. (I) is the insurer deciding to audit conditional on a claim being filled P_2 or not P_1. Information set are symbolised by a dashed line. S is the behavioral fraud rate.

the Insurer, (I), can only observe whether a claim has been submitted. The Insurer's information sets, represented by the dotted lines in Figure 3.1, dictate the available actions: either Audit or No-Audit. The payoff outcomes at each leaf of the game tree represent the final payoffs, with the first payoff corresponding to the Insurer and the second to the Claimant.

Our game differs slightly from each of the previous games mentioned. In Picard (1996) and Boyer (1999), there are *Honest* or *Dishonest* claimants whose proportions are fixed and can be known or unknown. In our model, there are an infinite number of types of claimants defined by their personal fraud penalty or moral cost, ω, which is private knowledge to the insured. The proportion of *Honest* claimants now depends on the set of extreme parameters [6], and the inferred distribution of expected loss for each subgroup of claimant.

[6] See Section 3.5.1 on $P0$. The parameters are: maximum compensation, lowest risk aversion, current

Following Dionne et al. (2009), we model the expected penalty for fraud to appear as soon as the claimant decides to commit fraud regardless of the outcome. This can be interpreted potentially as a "moral-cost" if the claimant is not audited. Section 3.5.1 details how H, the cumulative distribution of ω per θ, is inferred.

The main difference between Picard (1996) and Boyer (1999) is that in the former, the insurer has the possibility to commit to an audit policy by minimizing her expected cost knowing the reaction function of the claimants. While in the second, the insurer cannot anticipate the reaction of the insureds and therefore loses the Stackelberg advantage. Our framework also gives this Stackelberg advantage to the insurer. The insurer updates her belief S through a technology that gives a probability of fraud conditional on the signal emitted by each claim. In this study, the technology is a calibrated Machine Learning (ML) classifier. For a given group of claimants, once the *local* probability of fraud is known, the insurer discovers the effect of his current audit policy which corresponds to her position on H. Knowing the position on H allows the insurer to anticipate the reaction of claimants to a change in his audit policy, hence the Stackelberg advantage. Similarly to Boyer and Peter (2020)'s model, we divide our claimants into 10 θ groups corresponding to risk types. The higher the θ index, the greater is the estimated risk of claims where claimants suffered bodily injuries. These claims are known as Accident Benefit claims or AB claims.

Figure 3.1 shows two dominated strategies similar to the one obtained in Boyer and Peter (2020): It is sub-optimal for a customer not to declare a loss when an accident occurs leading to $\lambda = 0$ that is $u(W+t) \geq u(W-t)$. It is sub-optimal to investigate a file when no claim has been declared and hence $P_2 = 0$ because $-c \leq 0$.

To characterize the equilibrium, we start by solving the Insurer's problem described in

audit policy sets to 0, and the minimum wealth possible

(3.5):

$$\begin{aligned} &Min_{P_1}\left[(1-\pi)\psi\left(P_1 c+(1-P_1)t\right)+\pi P_1 c\right] \\ &\text{or } Min_{P_1}\left[P_1\left((1-\pi)\psi c+\pi c\right)+(1-P_1)\left((1-\pi)\psi t\right)\right] \end{aligned} \tag{3.5}$$

Three cases can occur:

$$\text{If } (1-\pi)\psi c+\pi c > (1-\pi)\psi t \text{ then } P_1 = 0 \tag{3.6}$$

$$\text{If } (1-\pi)\psi c+\pi c < (1-\pi)\psi t \text{ then } P_1 = 1 \tag{3.7}$$

$$\text{If } (1-\pi)\psi c+\pi c = (1-\pi)\psi t \text{ then } P_1 \in [0,1] \tag{3.8}$$

Similar to Picard (1996), Boyer (1999), and Boyer and Peter (2020), if the personal fraud penalty ω is known to the insurer, the perfect Bayes-Nash mixed equilibrium has the form:

$$\begin{aligned} &(1-\pi)\psi c+\pi c = (1-\pi)\psi t \\ &\psi = \frac{\pi c}{(1-\pi)(t-c)} \end{aligned} \tag{3.9}$$

The representative claimant's problem is given by:

$$Max_{\psi}\left[(1-\psi)u(W)+\psi\left(P_1(u(W)-k)+(1-P_1)(u(W+t)-k)\right)\right] \tag{3.10}$$

Three cases can occur:

$$\text{If } u(W) > P_1(u(W)-\omega)+(1-P_1)(u(W+t)-\omega) \text{ then } \psi = 0 \tag{3.11}$$

$$\text{If } u(W) < P_1(u(W)-\omega)+(1-P_1)(u(W+t)-\omega) \text{ then } \psi = 1 \tag{3.12}$$

$$\text{If } u(W) = P_1(u(W)-\omega)+(1-P_1)(u(W+t)-\omega) \text{ then } \psi \in [0,1] \tag{3.13}$$

Since the same audit strategy is applied to a heterogeneous group of claimants, equations (3.11), (3.12), and (3.13) will arise with respective probabilities $(1-H(\omega^*))$ and $H(\omega^*)$, where $\omega^* = (1-P_1)(u(W+t)-u(W))$, and H represents the cumulative distribution function of ω over the group of claimants. This implies that the insurer's belief about the current audit policy's position on H will dictate the direction and adjustment of future audit strategies.

3.3.2 New OAM cost function

In this section, we show that the OAM expected cost of fraud in (3.1) corresponds to the same minimization problem as in (3.5), the minimization problem of the sequential tree. Eq.(3.1) has three components: true-positive (TP), false-positive (FP) and false-negative (FN), described in (3.14).

$$\begin{aligned}
ExpCost[TP/\theta] &= c * P(Audit \cap ClaimFraud/\theta) \\
&= c * P(ClaimFraud/\theta) * P(Audit/ClaimFraud,\theta) \\
ExpCost[FP/\theta] &= c * P(Audit \cap NoFraud/\theta, N_i) \\
&= c * P(ClaimNoFraud/\theta) * P(Audit/ClaimNoFraud,\theta) \\
ExpCost[FN/\theta] &= t * P(NoAudit \cap ClaimFraud/\theta) \\
&= t * P(ClaimFraud/\theta) * P(NoAudit/ClaimFraud,\theta)
\end{aligned} \tag{3.14}$$

We show that both equation (3.1) and (3.5) provide identical measures of the expected cost of fraud. Let us reintroduce the conditional audit probabilities:

$$Q^n(\theta) = P(Audit/ClaimNoFraud,\theta) \tag{3.15}$$

$$Q^f(\theta) = P(Audit/ClaimFraud,\theta) \tag{3.16}$$

The audit policy consists in selecting the optimal percentage of claims to inspect at random in each subgroup of claimants. We use the hyper-geometric law to describe the number of successes in n random draws without replacement from a population of cardinality N with K as the number of successes. Let $X \sim$ Hyper-geometric(K,N,n), then $P(X=k) = \frac{\binom{N}{k} * \binom{N-K}{n-k}}{\binom{N}{n}}$ and $E(X) = n\frac{K}{N}$ gives the average number of successes for n random draws or the average number of True-Positive, $\overline{TP}$, for a given n.

$$\overline{TP} = n\frac{K}{N} \tag{3.17}$$

From (3.17), we get for a given n the average numbers of False-Positive and False-Negative respectively:

$$\overline{FP} = n - n\frac{K}{N} \tag{3.18a}$$

$$\overline{FN} = K - n\frac{K}{N} \tag{3.18b}$$

From (3.17), (3.18a) and (3.18b), we can deduce :

$$P(Audit/ClaimFraud,\theta) = \frac{\overline{TP}}{K} = \frac{n}{K} * \frac{K}{N} = \frac{n}{N} \tag{3.19a}$$

$$P(Audit/ClaimNoFraud,\theta)) = \frac{\overline{FP}}{N-K} = \frac{n*(1-\frac{K}{N})}{N-K} = \frac{n*(1-\frac{K}{N})*\frac{1}{N}}{(N-K)*\frac{1}{N}} = \frac{n}{N} \tag{3.19b}$$

$$P(NoAudit/ClaimFraud,\theta) = \frac{\overline{FN}}{K} = (K - n\frac{K}{N}) * \frac{1}{K} = (1 - \frac{n}{N}) \tag{3.19c}$$

Recognizing that $\frac{n}{N}$ is simply the fraction of claims to be audited belonging to a subgroup, we have shown that for random draws, $Q^n(\theta) = Q^f(\theta) = P(\theta)$.

$P(ClaimNoFraud/\theta)$ and $P(ClaimFraud/\theta)$ in (3.14) are $H(\phi(Q^f(\theta)/\theta))$ and π in (3.1), where $\phi(P)$ is (3.20) as defined in Dionne et al. (2009). From the sequential tree, we know that not filling a claim when an accident occurs is a dominated strategy leading to $\lambda = 0$ and $P(ClaimNoFraud/\theta) = \pi$. From the claimant's side of the equilibrium, the choice to fraud ψ is either 1, 0, or mixed depending on which of (3.11), (3.12) (3.13) is verified. The first case happens with probability: $(1 - H(\omega^*))$, while the second and third case happen with probability $H(\omega^*)$.

$$\phi(P) = (1-P)\big(u(w+t) - u(w)\big) \tag{3.20}$$

Replacing (3.19a) through (3.19c) and probability $(1-H(\omega^*))$ and $H(\omega^*)$ into (3.14), we get (3.21a) to (3.21c) which are equivalent to the insurer problem of minimizing his expected cost in Section 3.3.1:

$$Min_{Pc}\pi P \tag{3.21a}$$

$$+c(1-\pi)PH(\phi(P/\theta)) \tag{3.21b}$$

$$+t(1-\pi)(1-P)H(\phi(P/\theta)) \tag{3.21c}$$

$$\text{s.t.} P \in [0,1] \tag{3.21d}$$

This complete the proof of equivalence between the OAM cost function (3.1) from Dionne et al. (2009) and the insurer's problem (3.5) from Boyer (1999).

3.4 Data Review

Out of consultations with the insurance company, we can estimate that the current global audit policy has a True-Positive-Rate (TPR) between 33 and 50 %. As our database does not allow to precisely identify which claims are false-positive under the current audit policy, we have to make an assumption on how auditing divides itself across the different groups. We test some of the insights provided in Section 3.2 on our data-base. We use the results to base our different scenarios on the current audit policy presented in Section 3.5.2. We observe a known fraud rate of 5% in the insurer's portfolio of claim with bodily injury elements.

We specifically focus on the optimal random auditing contract (Mookherjee and Png (1989) and Fagart and Picard (1999)), with games with honest and dishonest claimants (Picard (1996) and Boyer (1999)) and with games with risky and non-risky claimants

(Boyer and Peter (2020)). These games offer a set-up closely related to the one developed here.

The four main tested effects are listed along with the studies where they can be found:

Effect 1. A higher asked claim leads to a higher equilibrium audit by the insurer (Mookherjee and Png (1989), Fagart and Picard (1999), Boyer and Peter (2020)) .

Effect 2. When there are two types *Honest* and *Dishonest*, a pooling equilibrium is reached as the optimal contract cannot separate them whether their proportions are known or not (Picard (1996), Boyer (1999)).

Effect 3. When accident risk-types are known, holding the cost of audit c and the compensation for a claim t constant, *High-risks* claimants have a higher fraud rate than *Low-risks* claimants. However at equilibrium, the coverage t[7] is higher for *High-risks* than for *Low-risks* which has a reducing effect on the equilibrium fraud rate, making the *High-risks* equilibrium fraud rate uncertain as it creates a higher audit incentive for the principal (Boyer and Peter (2020)).

Effect 4. When there is adverse selection, the *High-risks* contracts remain the same but *Low-risks* obtain lower compensations leading to a lesser incentive to be audited on the insurer's part which in turn results in an increased fraud rate. (Boyer and Peter (2020))

Table 3.1 verifies how these four effects are found in our database. Column $\hat{\pi}$ is the probability of filling a claim and is increasing in the premium AB, or θ indexes. Claims are gathered by their AB premiums in θ groups. A higher index indicates a higher paid insurance premium, hence a higher risk of accident with injury. An insurance contract is defined by a premium and a coverage amount. Empirically, the "regular" premium

[7] In Boyer and Peter (2020) coverage and compensation are interchangeable

amount is determined through actuarial models. The coverage amount is full insurance on marginal damages with a fixed deductible. The "regular" premium amount is then increased or decreased as a function of the choice of deductible made by the claimant. A higher(lower) choice of deducible results in less(more) coverage. In Canada, AB coverage is mandatory and without any possible choice of deductible. We use Collision deductibles as a measure of the insured's choice of level of coverage.

The compensation t relates to the specific amount disbursed on a claim. In Table 3.1, Columns t, *t.diff* and *t.p.val* respectively display the average compensation per group, the difference between the average compensation disbursed for a regular claim and a fraudulent claim, and the significance level of that difference. The reserve AB is the amount expected to be paid by the insurer at the beginning of the claim. Columns *Reserve*, *Res.diff* and *Res.p.val* display similar information for the reserve.

Table 3.1: Test of the four main effects

θ	Pop	Fraud	$\hat{\pi}$	t	t.diff	t.p.val	Reserve	Res.diff	Res.p.val	No Cov	% Ded 500	% Ded 1000
1	6596	35	0.56%	12182	1761	0.49	12803	-305	0.92	9.42	79.43	11.16
2	5229	58	0.66%	11443	-1471	0.43	10970	-1259	0.45	8.35	77.30	14.35
3	4964	92	0.70%	11266	2227	0.02	11057	2112	0	9.30	73.99	16.71
4	4886	130	0.78%	10950	98	0.94	10565	-573	0.68	10.46	70.41	19.13
5	4933	191	0.77%	11251	1970	0.02	10520	636	0.36	12.39	66.18	21.43
6	5249	237	0.95%	11708	2116	$<$0.001	10248	-431	0.66	14.60	61.19	24.21
7	5723	334	1.00%	11706	2217	$<$0.001	10432	-344	0.54	15.03	58.49	26.48
8	6379	474	1.20%	12362	2678	$<$0.001	10306	-482	0.4	17.77	52.77	29.46
9	7714	719	1.40%	12046	2780	$<$0.001	9918	103	0.75	21.53	46.52	31.95
10	9356	1282	1.50%	12572	4246	$<$0.001	9817	122	0.62	30.59	33.04	36.37

Note: Column $\hat{\pi}$ is the probability of filling a claim calculated by taking the average ratio of filled claims to the number of active policies for a given year. Column *Fraud* shows the number of observed fraudulent claims. Columns *t*, *t.diff* and *t.p.val* respectively display the average compensation per group, the difference between the average compensation disbursed for a regular claim and a fraudulent claim, and the t. p.value significance. Columns *Reserve*, *Res.diff* and *Res.p.val* are similar but for Reserve amounts. Columns No-Cov, %Ded 500 and %Ded 1000 shows the respective shares of population for each collision deductible choice of coverage.

From Effect 1. we expect to see more expensive claims audited more often. To estimate the true cost of a claim, we sum all the disbursed payments t made until the date of extractions. To correctly estimate Effect 1., we need to account for any potential bias due to incomplete disbursements because of detected fraud. Therefore, we also analyse the amount put in reserve when the AB element of the claim is opened as it is more closely related to the expected cost of the claim when the audit decision is made. Table 3.1 shows that both the reserve and the disbursed amount are stable across risk types. Columns *t.p.val* and *Res.p.val* test if the average compensation/reserve of fraudulent claims is similar to the average compensation/reserve of the regular claims. Column *t.p.val* shows that for higher risk groups θ_6 to θ_{10} the difference is significantly positive, making the regular claims more expensive compared to the fraudulent ones. Column *Res.diff* even shows that high-risk claimants (θ_9 and θ_{10}) have an average reserved amount of less than 10,000$, lower than that of other groups. This average amount is however undervalued compared to the final disbursed amount t which is on par with other θ groups while the average Reserve for θ_9 and θ_{10} are significantly lower.

Within premium θ groups, we further divide claims into four different levels of compensation [8]. We use a *t.test*, to analyse if within each combination of $\{\theta, t\}$, the average amount put in reserve for regular claims is similar to the one put in reserve for fraudulent claims. The results are presented in Table 3.6 in Appendix R. We observe little disparity of reserve or of paid compensation between fraudulent and non-fraudulent claims holding risk level constant [9].

Figure 3.2 displays the observed fraud rate per θ and compensation groups. Each curve corresponds to a premium group which index indicates its AB risk level. A higher index corresponds to a higher risk. We clearly see that θ_{10}, the top red dashed curve, has a

[8](11.9,2.92e+03],(2.92e+03,6.04e+03],(6.04e+03,1.52e+04],(1.52e+04,1.97e+06]

[9]Only for high risk claimants for $\theta_7, \theta_8, \theta_9, \theta_{10}$ and within the highest compensation group

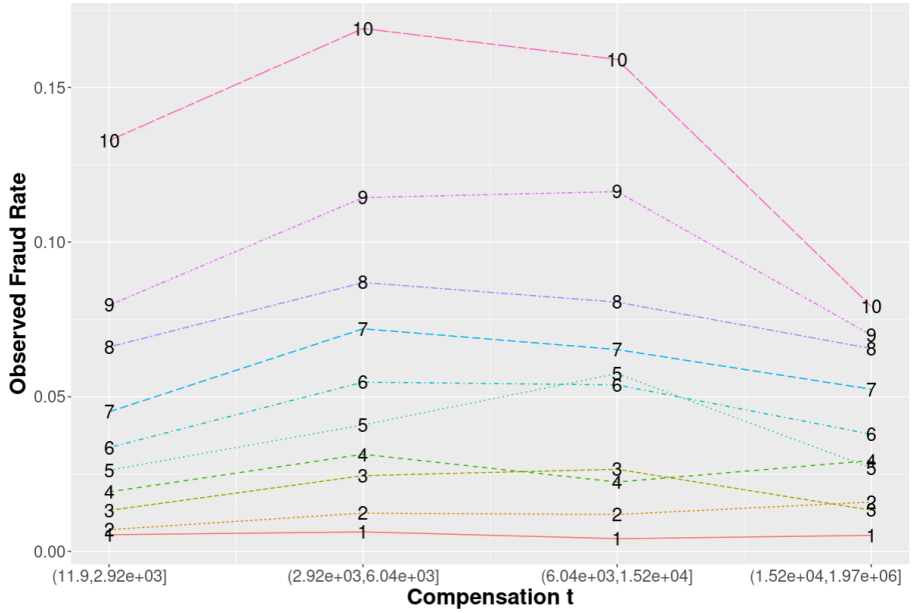
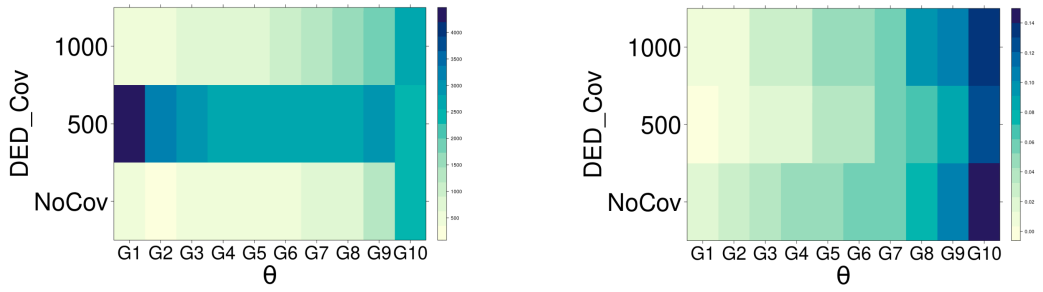


Figure 3.2: Audit Rate per θ and t from 2016

$X-axis$ is for the 4 groups of compensation t. $Y-axis$ is the observed fraud rate per group. Each curve is labeled by its θ index, a higher index indicates a higher charged AB premium.

higher observed fraud rate than θ_9, the light purple dotted curve, across all compensation groups. We see a decreasing observed fraud rate as θ index decreases with almost no overlapping. There is hence little indications in our database that audited claims are more costly than regular claims. Assuming a constant True Positive Rate across groups of $\{\theta, t\}$, the primary drive of the audit policy seems to be the risk level rather than the cost of the claim. For Effect 1. to be true, it would require false-positive claims to be substantially more expensive than regular claims in order to bring the average price of audited claims higher than the one of regular claims. However, because we cannot observe the false-positive claims, we cannot completely rule out Effect 1..

Effect 2. predicts that there are no contract type with only *Honest* type users. Figure 3.3b shows that fraudulent claims are detected across all combinations of premium and deductible, confirming Effect 2..



(a) Population per contract type

(b) Observed Fraud Rate per contract type

Figure 3.3: Analysis per contract type
X-axis is the θ AB premium groups in increasing order. Y-axis is the collision coverage choice. The left hand side figure 3.3a shows the number of claimants. The right hand side figure 3.3b shows the observed fraud rate.

From Effect 3., we expect that a claim belonging to a higher θ has a higher chance of being audited. As displayed in column *Fraud* in Table 3.1, the percentage of detected fraud strictly increases with the group's risk level from 0.68% to 13.7%. Assuming a constant or increasing True-Positive-Rate (TPR) across θ, this increasing detected fraud rate indicates at least a similar or increasing audit policy toward *High-risk* claimants. This finding is consistent with the conclusion of Effect 3.: *High-risk* claimants seem to be more often audited than *Low-risk* claimants.

Columns *No Cov*, *% Ded 500* and *% Ded 1000* in Table 3.1 shows, as Figure 3.3a, as the risk group θ rises, the percentage of insureds with a 500 deductible coverage constantly decreases from 79.43% to 33.04%. As risk types increase, insureds mostly shift to a lower coverage with either no collision coverage or 1000 deductible. Overall, our findings contradict Effect 3.'s conclusion that coverage is higher for *High-risk* claimants. A similar conclusion is drawn when the same analysis is conducted with groups formed by decile of Coverage premium in place of θ groups as shown in Figure T.10 in Appendix T. We conclude that we cannot find strong evidence in the Canadian context of Effect 3.: *High-risk* claimants do not have a higher coverage than *Low-risk* claimants and *High-risk* claims

are not on average more expensive.

Effect 4. predicts that when risk-types are unknown, there is a rationing of *Low-risks* through a lower coverage. The ambiguity between compensation and coverage in Boyer and Peter (2020) is again an issue as *Low-risks* do choose higher coverage so offering lower coverages might scare them but compensations are stable across risk types. However, we do observe a lower detected fraud rate for *Low-risk* claimants that could arise from a lower audit policy. Effect 4. would be true if we could verify that there is a low audit policy for *Low-risk* claimants which would results in a high hidden fraud rate. A possible explanation could be the absence of adverse selection in the portfolio of our partner.

3.5 Application

We apply the modified OAM to our dataset and derive the current equilibrium fraud rate and the optimal audit policy. In this application, we use a *CARA* exponential utility function (3.22) where η is the relative risk aversion coefficient, ω is the insured expected moral cost or penalty of fraud and W is the insured's wealth level.

$$U(W,\omega) = -\frac{exp(-\eta W)}{\eta} - \omega \tag{3.22}$$

The exponential utility leads to (3.23). If (3.23) is verified it will result in the claimant frauding decision:

$$(1-p)\Big[\frac{-exp(-\eta(W+t))}{\eta} - \omega\Big] + p\Big[\frac{-exp(-\eta W)}{\eta} - \omega\Big] \geq -\frac{exp(-\eta W)}{\eta} \tag{3.23}$$

where p is the audit probability, and t is the compensation asked.

Rearranging (3.23) in terms of the expected personal penalty of fraud ω, we get that a claimant frauds if and only if the inequality (3.24) holds true. (3.24) is the equivalent of

(3.20) with an exponential utility function. Hypothesis 1 in Section 3.5.1 makes sure that the skill level is never a biding constraint in committing a fraud.

$$\omega \leq \frac{-(1-p)exp(-\eta W)}{\eta}\Big[exp(-\eta t)-1\Big] \tag{3.24}$$

3.5.1 Parametrizations

In this section, we infer the cumulative density of the personal fraud penalty ω. We assume that there is one density per risk level θ. We use (3.25) to extract the limit expected penalty tolerable ω^* by the claimant to fraud. When applied to discovered fraudulent claims, each ω^* is the upper limit of an interval $A = [0, \omega^*]$ which contains the claimant's private expected penalty, $\hat{\omega}$.

$$\omega^* = \frac{-(1-p)exp(-\eta W)}{\eta}\Big[exp(-\eta t)-1\Big] \tag{3.25}$$

In Boyer (1999), Boyer and Peter (2020), and Picard (1996), the fraudulent part of a claim is fully fabricated by the insured but the compensation asked is fixed. We slightly relax this assumption later on. The optimal compensation asked by a claimant is given by (3.26).

$$\hat{t}^* = argMax_{\hat{t}}(1-P(\hat{t})) * u(w+\hat{t}) + P(\hat{t}) * u(w) - \hat{\omega} \tag{3.26}$$

Our optimal audit policy is fixed per subgroup of claimants making $P(\hat{t})$ constant over the four different intervals of compensation I for each each θ displayed in (3.27) through (3.30). We also assume that the current audit policy behaves in a similar fashion.

$$I1 = [0, 2.92e^{+03}] \tag{3.27}$$

$$I2 = [2.92e^{+03}, 6.04e^{+03}] \tag{3.28}$$

$$I3 = [6.04e^{+03}, 1.52e^{+04}] \tag{3.29}$$

$$I4 = [1.52e^{+04}, 1.97e^{+06}] \tag{3.30}$$

Hypothesis 1 *For simplicity, we suppose that each claimant has access to a set of skills that allows him to claim an amount $\hat{t}$ such that $\hat{t}^* \in i$, i being an interval of I_k, $i \subset I_k, k \in \{1,2,3,4\}$, and $max(i) \approx> \phi(P_{current})$, making the skills distribution very close to the $\hat{\omega}$ and $\hat{\omega} \approx \omega^*$.*

Because $i \subset I_k, k \in \{1,2,3,4\}$, Hypothesis 1 eliminates claimants strategies consisting in choosing a lower interval that presents a lower audit probability. Since the probability of audit is constant over i, the claimant's choice has to be $max(i)$, the maximum possible amount the claimant can asked for his given set of skills. We assume $max(i)$ to be closed to his $\hat{\omega}$, his personal expected loss. Hypothesis 1 ensures that the unknown level of skills to craft a claim of value $\hat{t}$ is never an obstacle to fraud. Using Hypothesis 1, we can approximate the expected loss by $\phi(\hat{t})$.

As seen above, if a true accident occurs then it is not optimal for the claimant to build-up the claim. The detected fraud rate of 5% is conditional on a claim being filled. However the unconditional fraud rate is of the order of 0.01%, making fraud a tail phenomena. We apply Extreme Value Theory (EVT), specifically the Pickands–Balkema–De Haan (PBH) theorem (Pickands (1975), Balkema and Haan (1974)), to infer the behavior of the tail of the distribution of ω.

The PBH theorem describes the distribution of $X - \mu$, the excess of a random variable X above a given threshold μ as described in (3.31). The authors show that for a sufficiently

high threshold, the excess follows a distribution $F_u(y)$ that is efficiently approximated by the Generalised Pareto Distribution (GPD) if the original distribution F belongs to a large list of densities [10]. Once the threshold μ is fixed, the GPD parameters are then estimated via maximum likelihood.

$$F_u(y) = P(X - \mu \leq y | X > \mu) = \frac{F(\mu + y) - F(\mu)}{1 - F(\mu)} \tag{3.31}$$

Our random variable X is $log(\omega)$ given by (3.32) . Hypotheses 1 ensure that $\phi(\hat{t}) \approx > \omega$ for all fraudulent claims detected in the database. The maximum likelihood is obtained from $\phi(\eta, \hat{t}, p, W)$ to determine $\{\eta_\omega, \sigma_\omega\}$ in (3.33). The threshold μ is $P0$ and represents the start of the tail of the complete distribution of expected loss. We defined $P0$ as the threshold formed by the most favorable case possible of fraud: No-Audit and the maximum amount asked $\hat{t}$ as seen in (3.34). Insureds with a fraud personal penalty superior to $P0$ will never fraud. Similar to Picard (1996) and Boyer (1999), de-facto we also have two types of insureds *Honest* and *Dishonest*. However unlike in Picard (1996) and Boyer (1999), their proportions vary by θ and depend on the observed parameters of the group as shown in Section 3.5.2.

$$\phi(\eta, \hat{t}, p, W) = log(\omega) = log(1-p) - log(\eta) + log\Big[1 - exp(-\eta\hat{t})\Big] - \eta W \tag{3.32}$$

$$log(\omega) - P0) \sim \mathscr{GPD}(0, \eta_\omega, \sigma_\omega) \tag{3.33}$$

$$P0 = \phi(mean(\eta), max(\hat{t}), 0, mean(W)) \tag{3.34}$$

Figure 3.4 displays $log(\omega)$ for each combination of risk level and asked compensations $\hat{t}$. The series of black dots are the limit threshold $P0$ for each θ groups. To calculate

[10] The list includes the normal, lognormal, beta, exponential, gamma, Student t, uniform.

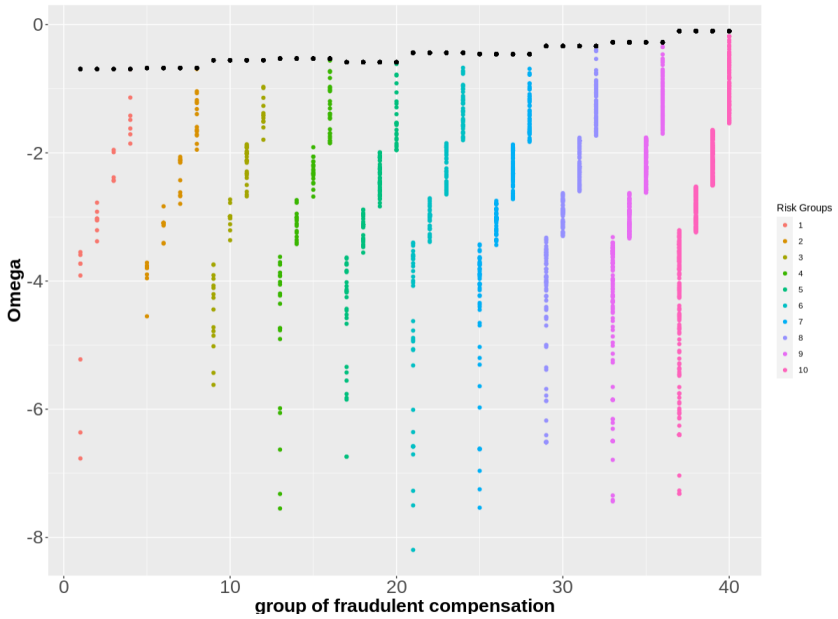


Figure 3.4: Threshold ω for fraudulent claims per for each pair $\{\theta, \hat{t}\}$
X-axis displays each subgroup formed by the combination of $\{\theta, t\}$. Y-axis is the $log(\omega)$ calculated for each fraudulent claimant by applying (3.32).

the MLE of the GDP distributions, we rely on the functions from the package *ReIns*[11] developed by Albrecher et al. (2017) in R.

We acknowledge that Hypothesis 1 is strong but it is necessary for the fraud decision to solely depend on the audit policy and the expected penalty and not on the distribution of skills. To relax Hypothesis 1, data on the current audit policy would be needed to create a lower bound on interval A. Inference techniques with non-parametric maximum likelihood for interval data could then be used to estimate a cumulative density H.

Once the cumulative density H is inferred, given a current audit policy and an average $\bar{\hat{t}}$, we can retrieve a position on H, the cumulative $\mathscr{GPD}$ of the distribution of excess $log(\omega) - P0$ for each subgroups using (3.35).

[11] https://CRAN.R-project.org/package=ReIns

$$P(log(\omega) - PO \leq P_{Current}/\eta_k, \bar{\hat{t}}_k) = H(P_{Current}/\eta_k, \bar{\hat{t}}_k) \tag{3.35}$$

where $k \in [1, 10]$ is the θ index.

However H only describes the behavior of the tail of the true distribution of ω. As such, (3.35) gives the probability of fraud conditional on already being in the tail. To obtain an unconditional estimate of ψ, the subgroup fraud rate in Figure 3.1, we need to scale $\hat{\psi} = H(P_{Current}/\eta_k, \bar{\hat{t}}_k)$ with weight ξ so that it becomes coherent with the observed probability of filling a claim, $\hat{\pi}$ presented in Table 3.2. We hence get $\psi = \hat{\psi} * \xi$ where ψ is the unconditional fraud rate for the group.

From the sequential tree in Figure 3.1, we have that (3.36) gives $\hat{\pi}$, the probability of filling a claim, which we can rearrange into (3.37), expressing π in terms of $\hat{\pi}$ and ψ. From Figure 3.1, (3.38) gives S, the fraud rate given a claim is filled.

$$\hat{\pi} = \psi - \pi * \psi + \pi \tag{3.36}$$

$$\pi = \frac{\hat{\pi} - \psi}{1 - \psi} \tag{3.37}$$

$$S = \frac{(1 - \pi) * \psi}{(1 - \pi) * \psi + \pi} \tag{3.38}$$

From (3.37) we get $(1 - \pi) = \frac{1 - \hat{\pi}}{1 - \psi}$. Calling $\hat{\psi} = H(P_{Current}/\eta_k, \bar{\hat{t}}_k)$ and substituting (3.37) in (3.38) and ψ by $\hat{\psi} * \xi$ we obtain (3.39):

$$S = \frac{\hat{\psi} * \xi (1 - \hat{\pi})}{\hat{\pi} * (1 - \hat{\psi} * \xi)} \tag{3.39}$$

An estimate of S for each subgroup is given by the calibrated output of the ML classifiers. We chose the weights ξ so that (3.39) holds true. The weights ξ represent the percentage of the current total population that commit fraud given our estimate of S and $\hat{\pi}$. Isolating ξ from (3.39), we get:

$$\xi = \frac{S\hat{\pi}}{\hat{\psi}(1-\hat{\pi}) + S\hat{\pi}\hat{\psi}} \tag{3.40}$$

3.5.2 Density of ω and proportion of *Dishonest/Honest* claimants per risk type

Table 3.2 shows the parametrization of the model for an assumed audit policy given by (3.41) where $k_i \in \{1, 1.75, 2.5, 3.25\}, i \in \{1,2,3,4\}$, i being the compensation t group described in (3.27) through (3.30). We hence assume a strictly increasing current audit policy in t, irrespective of the risk group θ. Figure 3.5 displays the appearance of each of the excess density for the personal fraud penalty ω. Only columns $\hat{\pi}$ in %, π in %, and ψ in % in Table 3.2 are affected by the different assumptions on P, the current audit policy. Section 3.5.3 show the impact of changing the current audit policy assumption on the results.

$$P_{Current} = \frac{\#Fraud}{\#Claims} * k_i, k_i \in \{1, 1.75, 2.5, 3.25\}, i \in \{1,2,3,4\} \tag{3.41}$$

Several observations can be drawn from Table 3.2. Column $W\$$ shows a strict decrease in wealth with risk-level θ. We see that the average scaled income for θ_1 claimants stands at 0.92 and decrease to 0.76 for θ_{10}. The income proxy in the database consists in the Statistic Canada average revenue in the claimant's full postal code area. Column P_0[12] is the threshold parameter defining the start of the left tail of the full distribution of ω. We see that as the risk level θ increases, $P0$ increases. Assuming the global distribution of ω does not drastically differ from one θ to another, a higher threshold indicates a higher percentage of *Dishonest* type. In other words, groups of *High-risk* may have a higher percentage of *Dishonest* types among them. Eq.(3.42) through (3.46) display the influence

[12] $P0 \equiv \mu_\omega$ the location parameter of a $\mathscr{GPD}$

Table 3.2: Population and parameters per θ for Scenario 1

θ	η_ω	σ_ω	P_0	$W\$$	$t\$$	$c\$$	$max(t)$	η	$\hat{\pi}$ in %	π in %	ψ in %
1	-0.496	3.34	-0.69	0.92	0.25	0.03	1.27	0.62	0.57	0.56	0.009
2	-0.818	3.19	-0.68	0.93	0.21	0.03	1.11	0.54	0.66	0.65	0.015
3	-0.683	3.51	-0.56	0.92	0.23	0.03	1.18	0.49	0.71	0.69	0.020
4	-0.409	3.12	-0.53	0.91	0.21	0.03	1.11	0.44	0.78	0.76	0.023
5	-0.521	3.32	-0.59	0.91	0.18	0.03	0.97	0.40	0.77	0.74	0.030
6	-0.393	3.21	-0.44	0.89	0.21	0.03	1.12	0.39	0.95	0.91	0.042
7	-0.431	3.15	-0.46	0.87	0.21	0.04	1.09	0.39	1.01	0.96	0.056
8	-0.543	3.55	-0.34	0.86	0.23	0.04	1.19	0.36	1.28	1.20	0.082
9	-0.463	3.38	-0.28	0.83	0.22	0.04	1.18	0.31	1.44	1.33	0.111
10	-0.506	3.70	-0.10	0.76	0.25	0.05	1.29	0.26	1.59	1.43	0.160

Note: Column θ is the premium group risk. η_ω, σ_ω, P_0 are respectively the shape, scale and location parameters of the $\mathscr{GPD}$ distribution. Columns $W\$$, $t\$$, $max(t)$, and $c\$$ are respectively the scaled wealth in CAD, the compensation, the max compensation and the cost of audit of a representative claimant. $t\$$, $c\$$, $max(t)$ variables are scaled in unit of wealth W. η is the scaled average risk aversion parameter obtained with a similar model to Zerbato et al. (2024). $\hat{\pi}$ is the filling rate per θ observed from the data. π and ψ are respectively the estimated probability of accident and fraud rate, averaged at θ, inferred by the model.

of each input on the proportion of *Honest*/*Dishonest* type in each θ groups. Eq.(3.43) shows that the increase in $P0$ directly results from the wealth decrease of *High-risk* types.

Column η represents the scaled average relative risk aversion (RRA) parameter estimated by the same model as in Zerbato et al. (2024). The estimation of η employs a method from Cohen and Einav (2007) applied to our Canadian database, utilizing collision choices regarding deductibles and the premiums charged. Consequently, η is not directly inferred from information on Accident Benefits (AB) coverage.

Table 3.2 indicates that the RRA parameter η decreases strictly as risk levels increase. Eq.(3.44) illustrates that the impact of η on $P0$ is ambiguous. As long as the condition in (3.45) holds, the proportion of *Honest* claimants will increase with higher η, indicating that a more risk-averse group will contain fewer *Dishonest* individuals. However, as wealth W decreases, the deterrence effect associated with risk aversion diminishes. Notably, Table

3.2 shows that *High-risk* θ groups exhibit a higher tolerance for risk, resulting in an overall ambiguous effect regarding (3.44).

Moreover, equation (3.46) reveals that higher compensation levels correspond to an increased percentage of *Dishonest* claimants within each group. Nevertheless, column $max(t)$ demonstrates that the maximum compensation requested remains relatively constant across the θ groups, indicating that it does not significantly influence our estimations.

$$\phi(\eta, max(t), W) = P0 = -log(\eta) + log\Big[1 - exp(-\eta * max(t))\Big] - \eta W \tag{3.42}$$

$$\frac{\partial \phi(\eta, max(t), W)}{\partial W} = -\eta < 0 \tag{3.43}$$

$$\frac{\partial \phi(\eta, max(t), W)}{\partial \eta} = -\frac{1}{\eta} + \frac{max(t)}{exp(\eta * max(t)) - 1} - W \tag{3.44}$$

$$\frac{max(t)}{exp(\eta * max(t)) - 1} - \frac{1}{\eta} < W \tag{3.45}$$

$$\frac{\partial \phi(\eta, max(t), W)}{\partial t} = \frac{max(t)}{exp(\eta * max(t)) - 1} > 0 \tag{3.46}$$

Figure 3.5 display the $\mathscr{GDP}$ distribution inferred for each θ groups. Columns η_ω and σ_ω in Table 3.2 control respectively the shape and scale of the densities. These estimates are relatively stable across θ. θ_2 and θ_3 have a more pronounced concavity due to a lower η_ω. θ_7 has the lowest σ_ω and hence the highest mode.

Column $\hat{\pi}$ (in %) represents the observed filling rate, which strictly increases with the risk level θ, as indicated in Table 3.2. Column ψ (in %) denotes the current unconditional fraud rate for each risk level θ, given the assumed audit policy illustrated in (3.41). Among the different θ levels, θ_{10} exhibits the highest ψ, while θ_1 has the lowest. The value of ψ also strictly increases with θ, aligning with a higher percentage of *Dishonest* claimants and Effect 3. Column π (in %) signifies the accident rate, which also strictly increases with θ and indicates an average probability of a true AB accident at 0.67%.

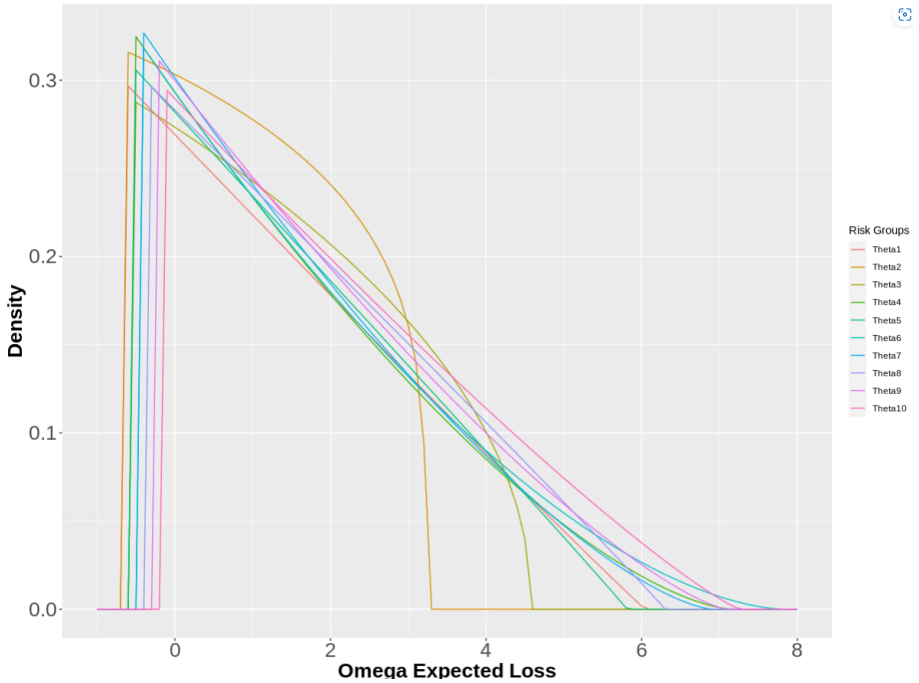


Figure 3.5: $\mathscr{GPD}$ of $log(\omega) - P0$ for each θ

Figure 3.5 presents the 10 estimated tails of $log(\omega) - P0$ density for each of the AB premium group θ.

3.5.3 Deterrence effect and audit policy for each $\{\theta, t\}$

Table 3.3 presents the optimal audit policy obtained from minimizing (3.47a) under constrain (3.47b), for a representative claimant. The solutions are presented for each combination of $\{\theta, t\}$. We tested different scenarios on the current audit policy. These scenarios are based upon the 4 effects discussed in Section 3.4.

$$Min_P\Big(c\pi P + (1-\pi)H(\phi(P/\theta))(cP + t(1-P))\Big) \tag{3.47a}$$

$$\text{s.t.} P \in [0, max(P)] \tag{3.47b}$$

$max(P)$ is determined by $H(\phi(P)) = 1$ which is the audit policy deterring the claimant with the lowest expected cost of the group. Eq.(3.47a) is solved for each pair of $\{\theta, t\}$ using the R package *Nolptr* [13].

As illustrated in Figure 3.2, the observed fraud rate on the y-axis strictly increases with the risk level θ and remains relatively constant across various compensation amounts t. The different scenarios tested maintain the distribution of the discovered fraud rate across the combinations of θ, t, but weight them according to an audit scenario as described in (3.41). Consultations with the insurance company indicate that the current True Positive Rate (TPR) falls between 33% and 50%. Based on the observed fraud rate, we can estimate a realistic global audit rate ranging from $[11.6\%, 17.5\%]$ across the entire database. The different weights k_i are also selected to conform to this overall audit range.

Following on Effect 1, Scenario 1 (S1) presented in (3.48) posits that the current audit policy P prioritizes expensive claims. In this scenario, the weights k are selected to increase by 0.75 within each θ group as the overall claim cost rises. Scenario 2 (S2) outlined in (3.49) features a P with a constant True Positive Rate, defined as $k_i = 2$ for all $i \in 1,2,3,4$ and all θ. Scenario 3 (S3) discussed in (3.50) models Effect 3 by adjusting P to concentrate on *High-Risk* claimants. In this case, weights k_i, where $i \in [1, 10]$, increase by 0.15 as the θ index rises. The weights k_i are applied across all four compensation groups within each θ.

$$\text{Scenario 1: } K \in \{1, 1.75, 2.5, 3.25\} \tag{3.48}$$

$$\text{Scenario 2: } K \in \{2, 2, 2, 2\} \tag{3.49}$$

$$\text{Scenario 3: } K \in \{1 + 0.15 * i; i \in [1:10]\} \tag{3.50}$$

[13] https://CRAN.R-project.org/package=Nloptr

Table 3.3 presents noteworthy insights. Column Opt^* displays the optimal audit policy recommended by the model, which corresponds to the ideal proportion of claims that should be randomly audited for each pair of $\{\theta,\hat{t}\}$. Column $Deter\%_{Opt}$ indicates the deterrence effect achieved by Opt^*, representing the percentage of "Dishonest" claimants who will refrain from committing fraud. Column P denotes the assumed current audit policy applied to the subgroup $\{\theta_i,t_j\}$, $i \in [1,10]$ and $j \in [1,4]$. Column $Deter\%_P$ reflects the deterrence effect produced by the current audit policy P. It is evident that the assumptions regarding P do not significantly affect the optimal audit policy obtained from the model. Comparing the optimal audit policies Opt^* from S1 to S2 reveals an average decrease of 0.003 with a standard deviation of 0.011. Similar findings are observed across all comparisons between scenarios.

A notable disparity exists between the assumed current audit policy P and the recommended optimal audit policy Opt^*. The policy P is derived from the observed fraud rates for each pair of $\theta,\hat{t}$ while adhering to the overall audit capacity of the insurance company. Across all scenarios—S1, S2, and S3—P primarily targets *High-Risk* claimants, with particular emphasis on the groups $\theta_{8,9,10}$. Audit rates for these *High-Risk* groups consistently exceed 15%. Figure 3.7 below provides a detailed analysis of Opt^* for S1.

According to (3.47), P influences the audit recommendation through estimates of the true accident rate π and the parameters of the $\mathscr{GPD}$. A stricter current audit policy P results in a lower fraud rate conditioned on having a claim S for the same filing rate $\hat{\pi}$, which in turn leads to a higher probability of a true accident π. However, since our estimates of S and $\hat{\pi}$ are fixed, this effect is offset by ξ in (3.40), resulting in constant values for π and ψ across all scenarios, as shown in Table 3.3.

The only remaining influence of a change in P passes through the ML estimation of the $\mathscr{GPD}$ parameters. The highest differences in Opt^* arise for groups $\{\theta_4,t_1\}$ and $\{\theta_5,t_1\}$. Focusing on $\{\theta_5,t_1\}$ in Table 3.3, Figure 3.6 shows that S3 produces a more concave

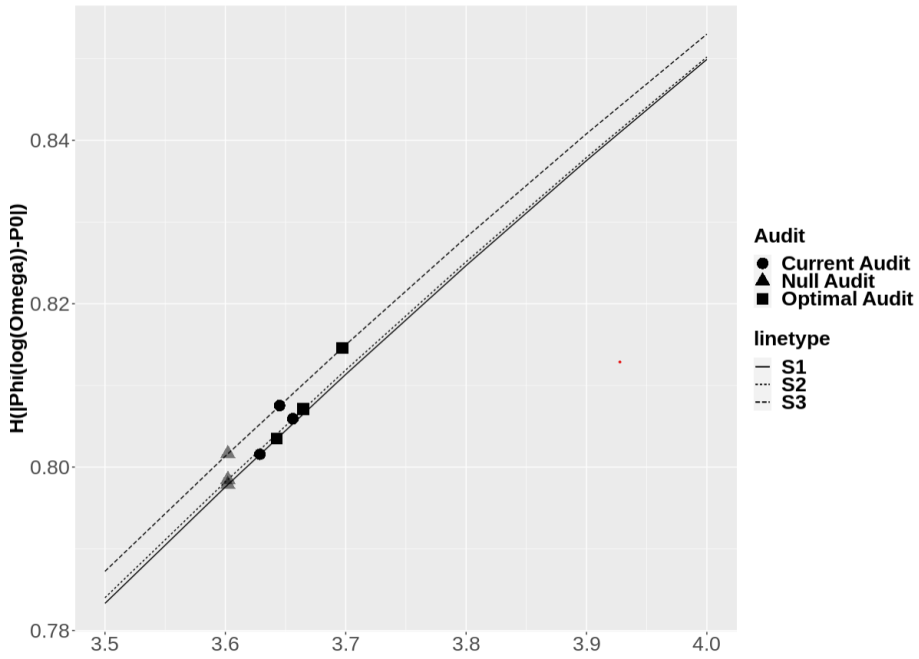


Figure 3.6: Comparison of Scenario 1, 2, 3 for $\{\theta_5, t_1\}$
X-axis are the values of $\omega - P0$ obtained by (3.32). Y-axis is the cumulative distribution function. Each curve represents a audit scenario indicated in the legend for the group $\{\theta_5, t_1\}$. The ■, • and ▲ are the locations of respectively Opt^*, P and No_Audit. Figure 3.6 shows the effect of the curvature of the tail of the density of ω on the deterrence effect of the audit policy. S3 is the highest curve with the most pronounced curvature which entails the strongest deterrence effect and lead to the most aggressive audit policy.

shaped distribution of expected loss with values $S3\eta_\omega = -0.511$, $S2\eta_\omega = -0.515$ and $S1\eta_\omega = -0.521$. Because of the steepest slope of the estimated $\mathscr{CDF}$, the S3 audit policy deters faster than in the other two scenarios, leading to a higher Opt^* of 9%.

Figure 3.7 displays the average cost ratio $\frac{t}{c}$ per group $\{\theta, t\}$ for S1. Each dot is labeled below with its Opt^* in black and above by the achieved deterrence effect in red. Note that similar to Table 3.3, Opt^* is never exactly 0. While the cost ratio strictly increases with the compensation groups, this increase is much lower for *High-risk* claimants due to an increase in the average cost of investigation. The downward slope is observed also for t_3

[14] but is not as pronounced.

A mixed strategy is applied by the principal to the group of agents belonging to a given $\{\theta,t\}$. Each agent only plays a pure strategy depending on his personal expected loss. The group of agents as a whole is considered to play a mixed-strategy if some agents fraud at equilibrium. Figure 3.7 distinguishes two types of equilibrium as 7 subgroups $\{\theta,t\}$ are highlighted for having fraud present at equilibrium. These groups belong either to the lowest t groups with a relatively high cost ratio or to the highest t groups with a relatively low cost ratio. A sufficiently high/low cost ratio modifies the equilibrium of the subgroup from no/full deterrence to a mixed equilibrium by incentivizing a higher/lower audit policy. In Boyer and Peter (2020) *High-risk* claimants have naturally a higher fraud rate and a higher coverage which increases the audit incentive for the insurer. In our study, we also conclude that *High-risk* claimants have a positive fraud rate at equilibrium for expensive claims but due to a relatively higher cost of audit. Focusing on θ_2, Figure 3.7 shows the effect of the difference in parameters of $\mathscr{GDP}$ inferred in Figure 3.5 in Section 3.5.2. A lower η_ω leads to a more concave shape density which increases the deterrence potential of the audit policy. As a result, a full deterrence is achieved with lower audit for groups $\{\theta_2,t_{2,3,4}\}$ and it becomes optimal to fully deter with $Opt^* = 28.6\%$ even for the lowest compensation group $\{\theta_2,t_1\}$.

As mentioned above, claims are gathered by the compensation asked into 4 different groups $\{t_1 = 2,915, t_2 = 6,038, t_3 = 14,476, t_4 = 63,136\}$ introduced by Column t_G in Table 3.3. Column P in Table 3.3 displays the tested current audit policy per group $\{\theta,t\}$. In S1, P is assumed to focus on expensive claims as depicted in (3.48). Under S1, the assumed current audit policy P on inexpensive claims t_1 averages at 4.28%, deterring 78.9% of *Dishonest* agents from frauding. t_4 high expensive claims are assumed to be more audited under S1 with an average P at 12.85%. However, the achieved deterrence

[14] Black Square observations

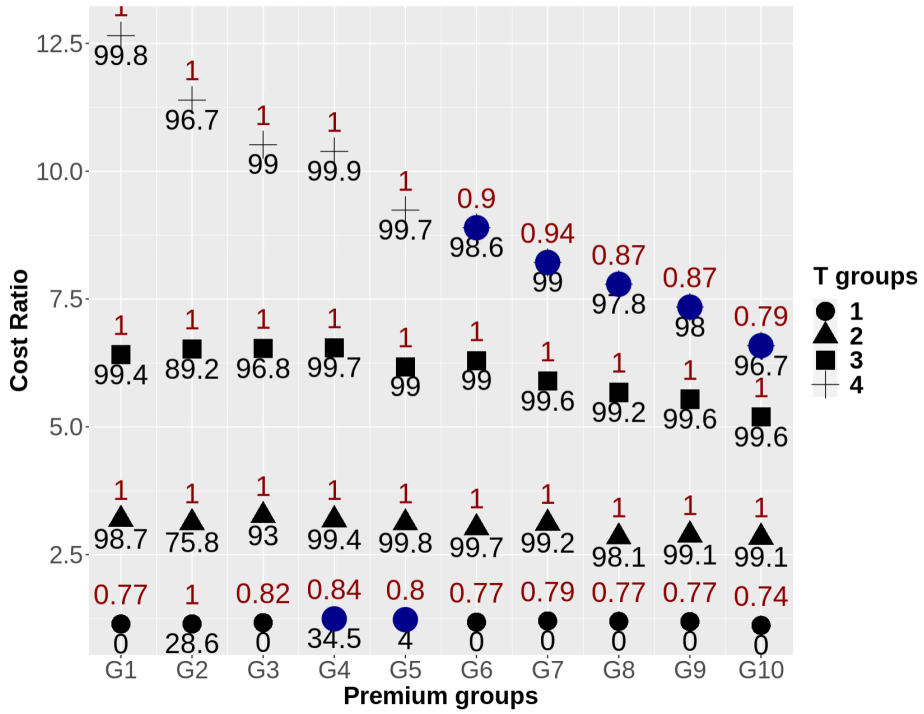


Figure 3.7: Deterrence effect and Opt^* per group of $\{\theta, t\}$ for S1

Figure 3.7 introduces the recommended optimal policy under Scenario 1. Each dot is a group formed by the AB premium θ in the x-axis and the compensation group t described in the legend. The y-axis is the cost-ratio $\frac{t}{c}$ formed by the reward asked and the cost of audit. Each dot is labeled below with its Opt^* in black and above by the achieved deterrence effect in red.

effect only amounts to 18.8% of the *Dishonest* type for an audit 3 times higher compared to t_1.

If we remove the special case of θ_2, σOAM preconizes for inexpensive claims $t = 1$ a similar average deterrence of 78.4% with a similar average level of audit Opt^* of 4.3%. However σOAM focuses the audit where the cost ratio is slightly higher for $\{\theta_{4,5}, t = 1\}$ which optimize the cost of audit. If we include θ_2, Opt^* raises to 6.7% for a 80.7% deterrence effect due to the shape of the ω density of θ_2 described above. For t_4 expensive claims, the average recommended deterrence effect is 93.2% obtained with an average

audit level of 98.5%. For t_2 and t_3 respectively, the average levels of audit are 95.7% and 98.2%, deterring all *Dishonest* agents. These results show the important effect of the average compensation t on deterring which is at least on par with the audit policy. Section 3.5.4 further inspects the sensitivity of audit to a change in t.

We also find that *Low-risk* types are more easily deterred than *High-risk* types which to our knowledge is a new a result in the literature and is coherent with Zerbato et al. (2024)'s findings that *High-risks* are less risk averse than *Low-risks*. S3 assumes that *High-risk* claimants are more audited than *Low-risk* claimants. Under S3, θ_1 claimants' average audit level stands at 0.5% to achieve a deterrence effect of 48%. θ_{10} claimants' are audited to 31.7% to achieve a similar deterrence effect of 52.7%, a 4% augmentation from the deterrence effect obtained for θ_1 but with an audit policy a 100 times higher. The optimal audit policies recommended by σOAM for respectively θ_1 and θ_{10} are similar and amount to 74.4% and 74.3% to achieve a deterrence effect of respectively 94% and 87% of the *Dishonest* type. These results hold true, although in less elevated proportions under S1 and an alternative S1 scenario where the audit weights are revised with value $\{3.25, 2.5, 1.75, 1\}$. The main difference between *Low-risk* and *High-risk* lies in the estimate of η, the risk aversion parameter.

Table V.22 in Appendix U displays Opt^* if all subgroups share a unique density of moral costs. The identical tail is obtained by averaging all parameters η_ω, σ_ω and setting the obtained tail as the unique tail for each subgroup's expected loss. As the shape of the tails are identical, it is equivalent to setting a constant deterrence effect for each subgroups. Table V.22 compares the resulting Opt^* to the one obtained under S1. The main differences appear for the pair $\{\theta_{4,5}, t = 1\}$ who have a higher relative cost ratio as displayed in Figure 3.7. Both optimal audit policy strongly decrease from respectively 38.6% to 11.3% and 11.2% to 1.4% to achieve the same deterrence effect. Consequently averaging the shape of the tail distribution marginally contributed to ease the deterrence effect in theses two

pairs.

Table 3.3: Optimal Audit per group of $\{\theta, t\}$

θ	t_G	$\frac{t}{c}$	$\hat{S}_{\%}$	W	$\pi_{\%}$	$\psi_{\%}$	S1 in %				S2 in %				S3 in %			
							P	Det_P	Opt^*	Det_{Opt}	P	Det_P	Opt^*	Det_{Opt}	P	Det_P	Opt^*	Det_{Opt}
1	2915	1.14	1.08	0.91	0.559	0.006	0.5	77.2	≈ 0	77.2	1.1	77.4	≈ 0	77.2	0.5	77.4	≈ 0	77.3
	6038	3.19	1.27	0.93	0.557	0.007	1.1	59.2	98.7	100	1.3	59.3	98.7	100	0.6	59.3	98.7	100
	14479	6.41	1.55	0.92	0.556	0.009	1	42.5	99.4	100	0.8	42.5	99.4	100	0.4	42.5	99.4	100
	63136	12.7	2.48	0.92	0.551	0.014	1.7	13	99.8	100	1	12.8	99.8	98	0.5	12.8	99.8	100
2	2915	1.14	1.52	0.92	0.654	0.01	0.7	95.1	28.6	100	1.4	95.3	29.2	100	0.8	95.3	28.8	100
	6038	3.12	1.76	0.95	0.653	0.012	2.2	71.4	75.8	100	2.5	71.6	76	100	1.4	71.6	75.9	100
	14479	6.52	1.78	0.95	0.653	0.012	3	50.5	89.2	100	2.4	50.5	89.3	100	1.4	50.5	89.2	100
	63136	11.4	3.8	0.92	0.639	0.025	5.2	17.1	96.6	100	3.2	16.3	96.7	100	1.8	16.3	96.7	100
3	2915	1.17	1.84	0.9	0.693	0.013	1.3	82.4	≈ 0	82.1	2.7	82.7	≈ 0	82.1	1.7	82.7	≈ 0	82.4
	6038	3.26	2.1	0.93	0.691	0.015	4.3	62.8	93	100	4.9	62.9	93.1	100	3.2	62.9	93.1	100
	14479	6.53	2.86	0.93	0.686	0.02	6.6	45.7	96.8	100	5.3	45.3	96.9	100	3.5	45.3	96.9	100
	63136	10.5	4.43	0.91	0.675	0.031	4.3	14.9	99	100	2.7	14.4	99.1	100	1.7	14.4	99	100
4	2915	1.24	1.96	0.91	0.767	0.015	1.9	78.7	34.5	83.5	3.8	79.1	35.4	83.7	2.8	79.1	37.3	84.3
	6038	3.18	2.47	0.91	0.763	0.019	5.5	63.5	99.4	100	6.3	63.7	99.4	100	4.6	63.7	99.4	100
	14479	6.55	2.55	0.92	0.763	0.02	5.6	47.9	99.7	100	4.5	47.8	99.3	100	3.3	47.8	99.7	100
	63136	10.4	4.78	0.9	0.745	0.038	9.5	19.5	99.9	100	5.9	18.1	99.9	100	4.3	18.1	98.2	89.6
5	2915	1.23	2.45	0.91	0.748	0.019	2.6	80.2	4	80.3	5.3	80.8	6	80.7	4.2	80.8	9.1	81.5
	6038	3.12	2.95	0.94	0.745	0.023	7.2	64.2	97.8	100	8.2	64.5	97.9	100	6.5	64.5	97.9	100
	14479	6.17	4.25	0.91	0.735	0.033	14.4	49.2	99	100	11.5	48.4	99.1	98.6	9.2	48.4	99.1	100
	63136	9.24	5.83	0.88	0.723	0.045	8.8	18.4	99.7	100	5.4	17.3	99.7	99.3	4.3	17.3	99.7	100
6	2915	1.18	2.9	0.89	0.923	0.028	3.4	77.6	≈ 0	77.1	6.7	78.1	≈ 0	77.2	5.9	78.1	≈ 0	77.4

to be continued on the next page

Table 3.3: (continued from previous page)

θ	t_G	$\frac{t}{c}$	$\hat{S}_{\%}$	W	$\pi_{\%}$	$\psi_{\%}$	S1 in %				S2 in %				S3 in %			
							P	Det_P	Opt^*	Det_{Opt}	P	Det_P	Opt^*	Det_{Opt}	P	Det_P	Opt^*	Det_{Opt}
	6038	3.02	4.23	0.89	0.91	0.041	9.6	62.9	99.6	100	10.9	63.3	99.4	100	9.6	63.3	99.7	100
	14479	6.3	4.26	0.91	0.91	0.041	13.5	48.2	99.8	99.9	10.8	47.5	99.9	99.9	9.4	47.5	99.9	100
	63136	8.9	6.15	0.89	0.892	0.059	12.3	19.7	99	91.9	7.6	18.2	99	91.6	6.6	18.2	99.2	93.2
7	2915	1.2	3.96	0.87	0.973	0.041	4.5	79.6	≈ 0	79	9	80.3	≈ 0	79	8.6	80.3	≈ 0	79.2
	6038	3.11	4.72	0.88	0.965	0.048	12.6	65.3	99.1	100	14.4	65.8	99.2	100	13.7	65.8	99.2	100
	14479	5.9	5.56	0.9	0.956	0.057	16.3	50.3	99.6	100	13.1	49.6	99.7	99.6	12.4	49.6	99.6	100
	63136	8.22	7.6	0.85	0.936	0.078	17	21.4	98.9	93.4	10.5	19.3	99.9	100	10	19.3	99.9	100
8	2915	1.19	4.25	0.86	1.23	0.055	6.6	77.7	≈ 0	77.0	13.2	78.7	≈ 0	76.7	13.6	78.7	≈ 0	76.7
	6038	2.84	5.32	0.9	1.21	0.069	15.2	62.9	98.1	100	17.4	63.5	98.3	100	17.8	63.5	98.3	100
	14479	5.67	6.41	0.85	1.2	0.083	20.1	47.6	99.2	100	16.1	46.7	99.3	100	16.5	46.7	98.3	97.8
	63136	7.79	9.24	0.85	1.16	0.12	21.3	20.8	97.7	86.4	13.1	18.5	97.2	90.9	13.5	18.5	96.7	81.7
9	2915	1.19	5.64	0.83	1.36	0.082	8	77.8	≈ 0	76.7	15.9	78.8	≈ 0	76.7	17.5	78.8	≈ 0	76.3
	6038	2.88	7.26	0.85	1.33	0.106	20	64.8	99.1	100	22.9	65.5	99.2	100	25.2	65.5	99.2	100
	14479	5.54	7.92	0.83	1.32	0.115	29.1	51.4	99.6	100	23.3	50.1	99.7	100	25.6	50.1	99.7	100
	63136	7.34	9.72	0.81	1.3	0.142	22.7	22.9	98	87	14	20.4	98	88	15.4	20.4	98	86.7
10	2915	1.12	7.82	0.76	1.46	0.126	13.3	76.3	≈ 0	74.3	26.6	77.6	≈ 0	74	31.2	77.6	≈ 0	72.7
	6038	2.84	9.54	0.76	1.44	0.153	29.6	63.5	99.1	100	33.8	64.5	99.3	90.6	39.7	64.5	99.3	100
	14479	5.19	10.7	0.78	1.42	0.173	39.8	51.5	99.6	100	31.8	49.5	99.7	100	37.4	49.5	99.7	100
	63136	6.59	11.7	0.74	1.4	0.188	25.7	21.9	96.4	77.7	15.8	19.1	96.5	77.7	18.6	19.1	96.2	75.4

The first row shows the assumption made on the current audit policy. Columns θ and t_G show the sub-group. $\frac{t}{c}$ is the cost ratio, $\hat{S}$ is the conditional fraud rate given by the ML classifiers and W is the wealth. ψ is the unconditional fraud rate, π is the accident rate and P is the assumed current audit policy. Opt^* is the optimal audit policy calculated by the model. Deter% is the deterrence effect as the percentage of *Dishonest* claimants that will not fraud because of Opt^*.

3.5.4 Sensitivity Analysis of the audit strategy and the deterrence effect

Our framework has a total of 6 different parameters directly observed from the data and an extra 4 parameters inferred subsequently. We conduct a sensitivity analysis (SA) by alternatively multiplying each parameter by 1.5 and observing the resulting optimal audit policy and deterrence effect compared to the default situation of Scenario 2 which assumes that the current observed fraud rate multiplied by 2 is the current audit rate P. The global audit policy under S2 sits at 74.9% for a deterrence effect of 93.36% of the *Dishonest* type.

The tested parameters are displayed in column *Param* of Table 3.4. Columns *Audit* and *Deterrence* show the sign and magnitude of the effect of a 50% increase in the raw input on the global optimal audit strategy and the global optimal deterrence effect, respectively. Table 3.4 is divided into two categories: the top five rows represent direct parameters observed from the data, which are used to infer the subsequent four rows of indirect parameters. Any change in the indirect parameters cannot occur while holding all other factors constant. Nonetheless, we test them with this caveat in mind.

The direct parameter that is the most influential in the final audit policy is the average compensation t. It has the highest positive impact on the audit policy with a 20% global increase from S2. However a higher t induces a higher fraud rate, limiting the deterrence effect of the 20% increase in the audit policy to a 5.61% increase in the deterrence effect. The average audit cost c is the second most influential direct parameters. When it increases by 50%, the optimal audit policy decreases by 18%. Contrary to t, c has no influence outside the audit policy on the deterrence effect. A decrease of 18% in the audit policy hence entails a direct 14% increase in the number of *Dishonest* types that will fraud.

Parameter $\hat{S}$ is the calibrated output of the ML classifier used to update the insurer's belief about S. An increase of 50% indicates that the signal provided by the claims in this

subgroup suggests a higher probability of fraud based on numerous observable characteristics of each claim. Column $\hat{S}$ in % in Table 3.3 displays the actual predictions made by the ML classifiers for each group θ, t. Contrary to t and c, $\hat{S}$ has a limited effect on Opt^*. $\hat{S}$ determines the accident rate π through (3.36) to (3.38). However, we observe that π has an ambiguous influence on the optimization problem (3.47). Table 3.4 further shows that π does not significantly influence the determination of the optimal audit strategy. Consequently, the observed conditional fraud rate has a limited impact on the formulation of the audit policy.

An increase in the risk-aversion parameter η reduces the global Opt^* by 2.6%, which in turn leads to a decrease in deterrence of 1.2%. Surprisingly, the net global effect is an increase in the number of fraud cases due to a reduced incentive to audit. It is noteworthy that the local effects of an increased η vary across the θ risk-type groups. As mentioned in Table 3.2, *High-risk* types are less risk-averse than *Low-risk* types, and the decrease is uniform. θ_{10} has the lowest average η at 0.26 and is the only θ affected by the change in η. This change resulted in a 30% drop in audits, leading to a reduced deterrence effect of 23.1%, as shown in Tables S.19 and S.20 in Appendix S.

P_0 is the indirect parameter that has the most influence on both the audit policy and the deterrence effect. P_0 represents the threshold between *Honest* and *Dishonest* types for a given θ. P_0 is negative. When P_0 decreases, the proportion of *Dishonest* claimants diminishes, leading to a significant drop in global Opt^* of 22%. The deterrence effect also decreases due to this reduced audit by 18%. Parameter η_ω controls the concavity of the density of expected loss and is negative in this context. We have already observed its effect for θ_2 in Section 3.5.3. A lower η_ω eases the deterrence effect, which rises by 2.9%, even though Opt^* decreases by 4.3%.

Table 3.4: Sensitivity Analysis under S2

Param	Δ Audit		Δ Deterrence	
	Inc/Dec	Magnitude	Inc/Dec	Magnitude
t * 1.5	↑	20.47%	↑	5.61%
c * 1.5	↓	−17.98%	↓	−13.77%
$\hat{S}$*1.5	≈	0%	≈	0%
η * 1.5	↓	−2.6%	↓	−1.2%
W*1.5	≈	0%	≈	0%
η_ω*1.5	↓	−4.3%	↑	2.9%
σ_ω*1.5	↓	−5.4%	↓	−23%
$P0$*1.5	↓	−21.8%	↓	−18.5%
π*1.5	≈	0%	≈	0%

Table 3.4 shows the effect of a change in the different input parameters governing the audit recommendation and its deterrence effect. Each parameter is in turn increased by 1.5 everything else kept equal. Δ Audit and Δ Deterrence columns display respectively the sign and the magnitude of this arbitrary change on the results.

3.6 Conclusion

The purpose of this paper is to extend the OAM model developed by Dionne et al. (2009) with an endogenous measure of the deterrence effect. Dionne et al. (2009) links the two distinct bodies of literature on fraud detection by incorporating claims' signal into a theoretical audit model. In the original OAM model, the deterrence effect was assumed cosntant. The deterrence effect is now estimated through the hidden expected cost of the claimants, using signals emitted by the claims to update the belief of the principal on the fraud rate conditional on having a claim. The resulting audit policy has been analysed and confronted to findings from previous key studies.

We find that full deterrence is optimal for most groups of claimants. This full deterrence is however reached with varying levels of audit depending on the shape of the ex-

pected loss distribution. We demonstrate similar finding than Mookherjee and Png (1989), Fagart and Picard (1999) and Boyer and Peter (2020): higher compensation claims should be more audited as the most influential inputs on the final audit policy are the cost of investigation and the potential reward. *Dishonest* type with inexpensive claims are deterred to an average of 80% from an audit policy of the order of 1.10^{-5}, rendering any additional auditing superfluous.

High-risk claimants with expensive claims have fraud present at equilibrium in part due to a higher cost of investigation and a lower average risk-aversion making full deterrence prohibitive. Even though rewards and costs are more influential in deciding the optimal audit policy, a higher risk aversion parameter leads to a lower optimal audit. Our framework also shows that the percentage of *Dishonest* types increases as claimants become more *High-risk*. Contrary to Boyer and Peter (2020), the optimal audit policy does not increase with risk type. *Low-risk* claimants with expensive claims are even more audited than *High-risk* claimants with expensive claims. This may be explained by the absence of adverse selection in the portfolio of our partner.

Bibliography

Albrecher, H., J. Beirlant, and J.L. Teugels (2017). *Reinsurance: Actuarial and Statistical Aspects*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9780470772683. URL: https://books.google.ca/books?id=HZjObwAACAAJ.

Arrow, K. J. (1964). "The Role of Securities in the Optimal Allocation of Risk-bearing". In: pp. 91–96. ISSN: 00346527, 1467937X.

— (1974). "Optimal insurance and generalized deductibles". In: *Scandinavian Actuarial Journal* 1, pp. 1–42. DOI: 10.1080/03461238.1974.10408659.

Bahnsen, A. Correa, D. Aouada, A. Stojanovic, et al. (2016). "Feature engineering strategies for credit card fraud detection". In: *Expert Systems with Applications* 51, pp. 134–142. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2015.12.030.

Balkema, A. A. and L. de. Haan (1974). "Residual Life Time at Great Age". In: 2.5, pp. 792–804.

Banulescu-Radu, D. and Y. Kougblenou (2024). "Data science for insurance fraud detection: a review". In: *Handbook of Insurance*. Springer. Chap. Forthcoming.

Bodaghiand, A. and B. Teimourpour (2018). "Automobile Insurance Fraud Detection Using Social Network Analysis: Case Studies in Social Networks and Beyond". In: *Applications of Data Management and Analysis*, pp. 11–16. DOI: 10.1007/978-3-319-95810-1_2.

Boyer, M. (1999). “When is The Proportion of Criminal Elements Irrelevant? A Study of Insurance Fraud When Insurers Cannot Commit”. In: *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*. Ed. by Georges Dionne and Claire Laberge-Nadeau, pp. 151–173.

Boyer, M. and R. Peter (2020). “Insurance Fraud in a Rothschild–Stiglitz World”. In: *Journal of Risk and Insurance* 87.1, pp. 117–142.

Cohen, A. and L. Einav (2007). “Estimating Risk Preferences from Deductible Choice”. In: *American Economic Review* 97.3, pp. 745–788. DOI: 10.1257/aer.97.3.745.

Crocker, K. J. and J. Morgan (1998). “Is Honesty the Best Policy? Curtailing Insurance Fraud through Optimal Incentive Contracts”. In: *Journal of Political Economy* 106.2, pp. 355–375. ISSN: 00223808, 1537534X.

Dionne, G., F. Giuliano, and P. Picard (2009). “Optimal Auditing With Scoring: Theory and Application to Insurance Fraud”. In: *Management Science* 55, pp. 58–70. DOI: 10.1287/mnsc.1080.0905.

Elkan, C. (2001). “The Foundations of Cost-Sensitive Learning”. In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence: 4-10 August 2001; Seattle* 1.

Fagart, M.C. and P. Picard (1999). “Optimal Insurance Under Random Auditing”. In: *The Geneva Papers on Risk and Insurance Theory* 24.1, pp. 29–54.

Fudenberg, D., D. M. Kreps, and E. S. Maskin (1990). “Repeated Games with Long-run and Short-run Players”. In: *The Review of Economic Studies* 57.4, pp. 555–573.

Gollier, C. (1987). “Pareto-optimal risk sharing with fixed costs per claim”. In: *Scandinavian Actuarial Journal* 1987.1-2, pp. 62–73. DOI: 10.1080/03461238.1987.10413818.

Grossman, S. J. and O. D. Hart (1983). “An Analysis of the Principal-Agent Problem”. In: *Econometrica* 51.1, pp. 7–45. ISSN: 00129682, 14680262.

Holmstrom, B. (1979). “Moral Hazard and Observability”. In: *The Bell Journal of Economics* 10, pp. 74–91.

Höppner, S., B. Baesens, W. Verbeke, et al. (2022). “Instance-dependent cost-sensitive learning for detecting transfer fraud”. In: *European Journal of Operational Research* 297.1, pp. 291–300. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2021.05.028.

Jiang, X., M. Osl, J. Kim, et al. (2011). “Smooth Isotonic Regression: A New Method to Calibrate Predictive Models”. In: *AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science* 2011, pp. 16–20.

Krawczyk, M. (2009). “The Role of Repetition and Observability in Deterring Insurance Fraud”. In: *The Geneva Risk and Insurance Review* 34.1, pp. 74–87. (Visited on 07/10/2024).

Lesch, W. and J. Brinkmann (2011). “Consumer Insurance Fraud/Abuse as Co-creation and Co-responsibility: A New Paradigm”. In: *Journal of Business Ethics* 103, pp. 17–32.

Mirrlees, J. (1976). “The Optimal Structure of Authority and Incentives Within an Organization”. In: *Bell Journal of Economics* 7, pp. 105–131. DOI: 10.2307/3003192.

Mookherjee, D. and I. Png (1989). “Optimal Auditing, Insurance, and Redistribution”. In: *The Quarterly Journal of Economics* 104.2, pp. 399–415.

Picard, P. (1996). “Auditing Claims in the Insurance Market with Fraud: The Credibility Issue”. In: *Journal of Public Economics* 63.1, pp. 27–56.

— (2013). “Economic analysis of insurance fraud”. In: *Handbook of Insurance: Second Edition*, pp. 349–395. DOI: 10.1007/978-1-4614-0155-1_13.

Pickands, J. (1975). “Statistical Inference Using Extreme Order Statistics”. In: *The Annals of Statistics* 3.1, pp. 119–131. (Visited on 07/18/2024).

Raviv, A. (1979). “The Design of an Optimal Insurance Policy”. In: *The American Economic Review* 69.1, pp. 84–96. ISSN: 00028282.

Ribeiro, R., B. Silva, C. Pimenta, et al. (2020). “Why do consumers perpetrate fraudulent behaviors in insurance?” In: *Crime, Law and Social Change* 73.

Subelj, L., S. Furlan, and M. Bajec (2011). “An expert system for detecting automobile insurance fraud using social network analysis”. In: *Expert Systems with Applications* 38.1, pp. 1039–1052. DOI: 10.1016/j.eswa.2010.07.143.

Tennyson, S. (2008). “Moral, Social, and Economic Dimensions of Insurance Claims Fraud”. In: *Social Research* 75.4, pp. 1181–1204.

Tennyson, S. and P. Salsas-Forn (2002). “Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives”. In: *Journal of Risk and Insurance* 69.3, pp. 289–308.

Townsend, R. M. (1979). “Optimal contracts and competitive markets with costly state verification”. In: *Journal of Economic Theory* 21.2, pp. 265–293. ISSN: 0022-0531. DOI: 10.1016/0022-0531(79)90031-0.

Wilson, C. (1977). “A model of insurance markets with incomplete information”. In: *Journal of Economic Theory* 16.2, pp. 167–207.

General Conclusion

In the first article, we introduced σOAM, an extension to the cost-sensitive expert system Optimal Audit Model (OAM) developed by Dionne et al. (2009). Results have shown that our cost-sensitive approach σOAM outperforms more traditional cost-sensitive models as it produces a more aggressive audit strategy for groups of claimant more prone to fraud. The addition of machine-learning classifiers improved the performances both in term of classifications and in minimizing the cost of fraud.

The second article focused on obtaining a measure of risk aversion. We replicated Cohen and Einav (2007)'s model in a Canadian context to obtain a risk aversion parameter for each insured. Results in Article 2 showed that risk aversion added predictive power to our classifiers used in Article 1 and was systematically among the most predictive variables in detecting fraud. σOAM's performances were thus improved showing the importance of risk aversion in the fraud process.

In the third article, we refined the modeling of the claimant's decision to fraud by introducing a personal penalty suffered from fabricating a claim. After inferring the distribution of such penalty, we obtained a measure of the deterrence effect of the optimal audit policy. This new model's recommendations were tested against four theoretical insights from the literature. Our recommendations are consistent with 2 of them. Several new findings were also evidenced such as the primordial role of the reward and the risk aversion

on the deterrence effect entailed by the audit policy.

Appendix A

Table A.5 presents the main predictors extracted from the base classifiers. The first column introduces the name of the feature. Features marked with $\textcircled{\checkmark}$ are among the top five predictors for the corresponding base classifiers. Features marked with ✓are among the top eight predictors while features marked with $\times$ are not in the top eight predicting variables. The column *Derrig* indicates that the main predictor belongs to the traditional red-flags for body injury claims listed by Weisberg and Derrig (1998). A boxed reference ($\boxed{ACC8}$) indicates the red-flags belong to the core list indicators selected by Weisberg and Derrig (1998). A regular reference (CLT5) indicates that the corresponding indicator was tested in Weisberg and Derrig (1998)[15].

[15] Appendix A1 3.6, table B.6

Table A.5: Optimal Audit Model results

VarName	Tree	Dart	Line	Logis	FOLogis	FODart	FOTree	FOLine	NLLogis	NLDart	NLLine	NLTree	Derrig
Law_rep_none	✓	✓	✓	✓	✓	✓	✓	✓	NA	NA	NA	NA	CLT5
DTd_dol_Oicd	✓	✓	✓	×	×	✓	✓	✓	×	✓	✓	✓	INS7
DRV_licence_yrs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	NA
ToL	✓	×	×	×	×	×	✓	×	×	×	×	✓	*ACC*8
HR_Unknown	✓	×	✓	×	×	×	✓	✓	×	✓	✓	✓	*INJ*1
Nbr_Lawyers	✓	✓	✓	✓	✓	✓	✓	✓	NA	NA	NA	NA	NA
Law_rep_agrr	✓	✓	×	✓	✓	✓	✓	×	NA	NA	NA	NA	CLT5
DTd_los_law	✓	✓	×	✓	✓	✓	✓	×	NA	NA	NA	NA	CLT1
Shop_Cat_Others	×	✓	×	×	×	×	×	×	×	✓	×	×	*CLT*2
Law_Rep_Mod	×	×	×	×	✓	✓	×	×	NA	NA	NA	NA	CLT5
Nbr_Claims_Pol	×	×	×	×	×	✓	×	×	✓	×	×	×	*ACC*8
Unk_ToL	×	×	✓	×	×	×	×	✓	×	×	✓	×	NA
Min_injsum_rat	×	×	✓	×	×	×	×	×	×	×	✓	×	*INJ*1
Sum_injmax	×	×	✓	×	×	×	×	✓	×	×	✓	×	*INJ*1
Min_injmin	×	×	×	×	×	×	×	✓	×	×	✓	×	*INJ*1
TotPartCad_rat	×	×	×	×	×	×	×	×	×	×	×	✓	*ACC*10
Inj_Back	×	✓	×	✓	×	×	×	×	✓	✓	✓	✓	*INJ*1
HR_Other	×	×	×	×	×	×	×	×	✓	×	×	×	NA
HR_247	×	×	×	×	×	×	×	×	✓	×	×	×	NA
HR_fax	×	×	×	×	×	×	×	×	✓	×	×	×	NA
HR_phone	×	×	×	✓	×	×	×	×	✓	×	×	×	NA
MostSeverePsyMax	×	×	×	×	×	×	×	×	×	✓	×	×	*INJ*2
DTd_FAB_DoL	×	×	×	×	×	×	×	×	×	✓	×	×	*INJ*6
Nbr_Days_Law_Miss	×	×	×	✓	✓	×	×	×	×	×	×	×	CLT1
male_clmnt	×	×	×	×	×	×	×	×	×	✓	×	×	NA

Appendix B

Table B.6: List of indicators used by Weisberg and Derrig (1998). The column *Suspicious* and *Fraud* indicates the indicator stood-out respectively for the adjusters or the investigators. The *In study* column indicates the name of the corresponding variable created to match the indicator in this study.

Code	Indicators	Suspicious	Fraud	In study
ACC1	No report by police officer at scene	Yes	Yes	Yes
ACC2	No witnesses to accident	Yes	No	No
ACC3	Rear-end collision	No	No	Yes
ACC4	Single vehicle accident	No	No	Yes
ACC5	Controlled intersection collision	No	No	Yes
ACC6	Claimant was in parked vehicle	No	No	Yes
ACC7	Two drivers were related or friends	No	No	No
ACC8	Late-night accident	No	Yes	Yes
ACC9	No plausible explanation for accident	Yes	Yes	Yes[23]
ACC10	Claimant in and old, low-value vehicle	Yes	No	Yes

Continued on next page

Table B.6: List of indicators used by Weisberg and Derrig (1998). The column *Suspicious* and *Fraud* indicates the indicator stood-out respectively for the adjusters or the investigators. The *In study* column indicates the name of the corresponding variable created to match the indicator in this study. (Continued)

Code	Indicators	Suspicious	Fraud	In study
ACC11	A rental vehicle involved in accident	No	No	No [25]
ACC12	No tow from scene despite severely damaged car	No	No	Yes
ACC13	Site investigation raised questions	No	No	No [27]
ACC14	Property damage was inconsistent with accident	No	No	Yes
ACC15	Very minor impact collision	Yes	No	Yes
ACC16	Claimant vehicle stopped short	No	Yes	Yes
ACC17	Claimant vehicle made unexpected maneuver	No	No	Yes
ACC18	Insured/claimant versions differ	No	No	No
ACC19	Insured felt set up, denied fault	Yes	No	No
CLT1	Retained an attorney very quickly	No	No	Yes
CLT2	Had a history of previous claims	No	No	Yes
CLT3	Gave address as hotel or PO Box	No	No	No
CLT4	Was an out-of-state resident	No	No	No
CLT5	Retained a “high-volume” attorney	No	No	Yes
CLT6	Was difficult to contact/uncooperative	No	No	No
CLT7	Was one of three or more claimants in vehicle	No	Yes	Yes

Continued on next page

Table B.6: List of indicators used by Weisberg and Derrig (1998). The column *Suspicious* and *Fraud* indicates the indicator stood-out respectively for the adjusters or the investigators. The *In study* column indicates the name of the corresponding variable created to match the indicator in this study. (Continued)

Code	Indicators	Suspicious	Fraud	In study
CLT8	Was resident of high claim town	No	No	No
CLT9	Avoided use of telephone or mail	No	No	No
CLT10	Was unemployed	No	No	Yes
CLT11	Appeared to be “claims-wise”	Yes	Yes	No
INS1	Had a history of previous claims	No	No	Yes
INS2	Gave address as hotel or PO Box	No	No	Yes
INS3	Readily accepted fault for accident	No	No	No
INS4	Was acquainted with other-vehicle occupants	No	No	No
INS5	Was not willing to provide a sworn statement	No	No	No
INS6	Was difficult to contact/uncooperative	Yes	No	P-F
INS7	Accident occurred soon after policy effective date	No	No	Yes
INS8	Appeared to be “claims-wise	No	No	No
INJ1	Injury consisted of strain/sprain only	No	Yes	No
INJ2	No objective evidence of injury	Yes	Yes	No
INJ3	Police report showed no injury or pain	No	Yes	No
INJ4	Claimant refused to appear for IME	No	Yes	P-F

Continued on next page

Table B.6: List of indicators used by Weisberg and Derrig (1998). The column *Suspicious* and *Fraud* indicates the indicator stood-out respectively for the adjusters or the investigators. The *In study* column indicates the name of the corresponding variable created to match the indicator in this study. (Continued)

Code	Indicators	Suspicious	Fraud	In study
INJ5	No emergency treatment was given for the injury	No	No	No
INJ6	Non-emergency treatment was delayed	No	Yes	No
INJ7	First non-emergency treatment was by a DC	No	No	No
INJ8	Activity check cast doubt on injury	No	No	P-F
INJ9	Injuries were inconsistent with police report	Yes	No	No
INJ10	IME suggests injury was unrelated to accident	No	No	No
INJ11	Unusual injury for this auto accident	No	No	No
INJ12	Evidence of an alternative cause of injury	No	Yes	No
TRT1	Large number of visits to a chiropractor	Yes	No	P-F
TRT2	DC provided 3 or more modalities on most visits	Yes	No	P-F
TRT3	Large number of visits to a physical therapist	No	No	P-F
TRT4	MRI or CT scan but no inpatient hospital charges	No	No	P-F
TRT5	Use of “high volume” medical provider	No	No	Yes
TRT6	Significant gaps in course of treatment	No	No	P-F
TRT8	Treatment was unusually prolonged	No	No	P-F
TRT7	IME questioned extent of treatment	No	No	P-F

Continued on next page

Table B.6: List of indicators used by Weisberg and Derrig (1998). The column *Suspicious* and *Fraud* indicates the indicator stood-out respectively for the adjusters or the investigators. The *In study* column indicates the name of the corresponding variable created to match the indicator in this study. (Continued)

Code	Indicators	Suspicious	Fraud	In study
TRT9	Medical audit raised questions about charges	No	No	P-F
LW1	Claimant worked for self or family member	No	Yes	Yes
LW2	Employer wage differs from claimed wage loss	No	No	P-F
LW3	Claimant recently started employment	No	No	No
LW4	Employer unknown/hard to reach	No	No	No
LW5	Lost wages statement looked unofficial	No	No	P-F
LW6	Long disability for a minor injury	Yes	No	P-F

Appendix C

Table C.7 displays the number of variable per family, the number of such variables discarded and the reason as to why these variables are discarded.

Table C.7: Reasons presented refer to the discarded variables. *Low information*: most observations missing. *Post-Fact*: known too late in the life cycle of the claim.

Family of var	# var discarded	Reason to discard	# included vars
Disbursed amount per pol	51	Post-Fact	0
Occupation	0	All used	29
Disability status	11	Post-Fact	0
Specific policy[35] *info*	104	Post-Fact[37]	16
Litige-status	4	Post-Fact	0
Invest-status	3	Post-Fact	0
Work-Function & -Status	15	Low information/PF	0
Lawyer	20	Low information	4
Third_Party	25	Low information	1
Garage	35	Low information	2
Doctor	20	Low information	0
Towing	22	Low information	1
Custodian	20	Low information	0
Driver	20	Low information	2
Claimant	22	Low information	7
Insured	32	Low information	0
Repairer	10	Low information	0
IP	15	Low information	0

Continued on next page

Table C.7: Reasons presented refer to the discarded variables. *Low information*: most observations missing. *Post-Fact*: known too late in the life cycle of the claim. (Continued)

Family of var	# var discarded	Reason to discard	# included vars
Damages	50	Redundant	4
Loss	7	Low information	4
Klm Driven	2	Redundant	1
Sev of injury	12	Low information/Redu	35
Inception date	1	Redundant	2
Vehicle	23	Low information	4
Premium	10	used for theta	0
Deductible	9	used for theta	0
NBR[39]	39	one is used	
Drivable	1	Redundant	1
Reserve	9	Post-Fact	3
Drivable	1	Redundant	0
Dart	1	Redundant	2
EDU	0	all used	5
Changes in coverage	22	Low information	2
Paid res & exp	9	Post-Fact/Redu	3
Police-Report	0	all used	1

Continued on next page

Table C.7: Reasons presented refer to the discarded variables. *Low information*: most observations missing. *Post-Fact*: known too late in the life cycle of the claim. (Continued)

Family of var	# var discarded	Reason to discard	# included vars
Highway	0	all used	1
$No_collision$	0	all used	1
DTd_loss_law	0	all used	1
How_reported	0	all used	1
Cause_of_loss	0	all used	1
Point of Impact	0	all used	1

Appendix D

The 10 groups are: *Date* comprised of 7 variables that control the difference in days between the date of loss and the inception of the policy, the first AB exposure, the first Property-Damage (PD) exposure and the date of the first lawyer to appear on the claims.

The second type of variables relates to some characteristics of the policy or the claimants. It is comprised of 8 variables, that control for the number of male and female claimants, the number of claims on the policy, the oldest claimant in the claim, the number of drivers allowed on the policy, the number of years of driving experience the main beneficiary of the contract has and the number of recent changes in the AB coverage choice.

The third and fourth types are related respectively to the level of education of the main beneficiary and his type of employment. The fifth type is comprised of variables pertaining to the type of law firm involved on the claim. Such variables track the level of *"aggressiveness"* with which law firms specialized in insurance claims disputes settle the case.

Table D.8: List of variables extracted before preprocessing and feature engineering.

Variable Names
Date Variables

Continued on next page

Table D.8: List of variables extracted before preprocessing and feature engineering. (Continued)

Variable Names			
Date_Loss	First_AB	First_PD	First_Law
Last_Move	PD_Acc_Date	Orig_Incept	
Claimant and Policy Characteristics			
Fem_Clmnt	Male_Clmnt	Nbr_Claims_Pol	Age_Clmnt
Nbr_Drv_Prem	Drv_Lic_Yrs	Endt_AB_Chg	Chg_AB_End
Claimants Education			
Edu_LHS	Edu_Colg	Edu_HS	Edu_Miss
Edu_Uni			
Claimants Employment			
Occ_Retired	Occ_Mid_Mgmt	Occ_Clerk	Occ_Sales
Occ_Sr_Mgmt	Occ_Trk_Drv	Occ_Mach_Op	Occ_Mec
Occ_Const_Trades	Occ_Lab	Occ_Dentist	Occ_Phys
Occ_Teacher	Occ_Eng	Occ_Farmer	Occ_FireF
Occ_Home	Occ_Ins_Adj	Occ_Lawyer	Occ_Nurse
Occ_Sec	Occ_Psy	Occ_Police	Occ_Ins_Agt
Occ_Retail	Occ_SEmp	Occ_Stud	Occ_Oth
Occ_Unk	Occ_Prof	Occ_Child	
Law Variables			

Continued on next page

Table D.8: List of variables extracted before preprocessing and feature engineering. (Continued)

Variable Names			
Nbr_Lawyer	Law_Rep_Agr	Law_Rep_Mod	Law_Rep_None
Accident Characteristics			
Police_Rpt	Cause_Loss	How_Rpt	POI[43]
Ann_Km	Nbr_Clmnt	Nbr_Clmnt_AB	Nbr_Tow
Highway	Time_Loss	Loss_Prov	LossType
Insd_Fault%	Shop_Cat	Rep_Prov	Nbr_3rdPty
Nbr_Drv			
Vehicle and Damages			
Tot_Repl_Lab_CAD	Tot_Lab_CAD	Tot_Repl_Lab	Tot_Part_CAD
Veh_Age	Veh_Cat	Veh_Type	Driv_Audit
Total_Loss	Veh_Price		
Injury Variables			
Nbr_Inj_Max	Sum_Inj	Psy	Phys
Drv_Inj	Insd_Inj	Pre_Phys	Pre_Psy
Days_Pre_Psy	Days_Pre_Phys	MostSev_Phys_Min	MostSev_Psy_Min
MostSev_Psy_Max	MostSev_Phys_Max	Min_Psy	Min_Phys
Max_Phys	Max_Psy	Max_Inj_Max	Min_Inj_Max
Sum_Inj_Max	Nbr_Inj_Max	Max_Inj_Min	Min_Inj_Min

Continued on next page

Table D.8: List of variables extracted before preprocessing and feature engineering. (Continued)

Variable Names			
Sum_Inj_Min	Nbr_Inj_Min	Max_Inj_Sum	Min_Inj_Sum
Nbr_Inj_Sum	Nbr_Inj	Inj_Back	Inj_Chest
Inj_Neck	Inj_Shldr	Inj_Head	Inj_Oth
Reserve and Disbursement Variables			
Dart_Sum	Dart_Max	Res_AB_Open	Res_10d_AB

The sixth family of variables controls for some specific accident characteristics such as the type of accident ("rear-end","improper-left-turn", etc...), the manner with which the accident was reported to the insurance company, the point of impact on the insured car and other characteristics such as the body shop chosen by the claimant. The seventh type of variables records the severity of the damages and some indications on the vehicle involved in the crash. It has 10 distinct variables. The first four accounts for the estimated cost of labor and the cost of replacing the various part of the vehicle. The remaining 6 variables relate to the type, age, price and drivability of the vehicle. The two next families of variables control respectively for the severity of injury sustained in the accident and the estimated financial cost placed in reserve by the insurance company.

The variable *Cause_of_Loss*, from the **Accident characteristics** family, controls the reason behind the accident: *"rear-end accident"*, 36% of observations, *"improper-left-turn"*, 9.5%, *"Bicyclist/pedestrian"*, etc. From the same family, *How_reported* controls for the means by which our Industrial Partner was made aware of the incident. *POIPrimary* describes the primary point of impact on the claimant's vehicle while *Shop_Category* identifies the level of previous collaboration with the garage associated with the accident.

The family **Injury variables** depicts the severity of the injuries. There are two types of injuries: psychological and physical. 62000 claims have at a least one listed physical injuries and only 14000 claims have at least one psychological injuries. Injuries are listed by person, body location and trauma severity as reported after a medical examination. As claims are aggregated at the contract level, there might be multiple injured per claim. Therefore, variables such as *min_injmax* or *max_injsum* refer to the least sever injury respectively (*min_*) sustained by the maximum injured, or the most injured claimant (inj*max*), and to the sum of the all most sever injuries sustained on the claim. The same logic applies to all similar variables. *Pre_Psycho* and *Pre_Physical* indicates if there are any physical or psychological injuries listed for the policy holder of the contract while *psy* and *phys* control for the number of psychological and physical injuries on the claim. The family **Reserve and disbursement variables** control for the dollars put in reserve following our industrial partner actuarial predicted cost of treatment: *Res_at_AB_open* and such estimate after ten days. Disbursed amounts refer to amount spent on two specified policies: *amp* and *air*. Policy *amp* refers to medical payments and policy *air* refers to income replacement. As many people may be injured on the same claim, the prefix *max/min* refers again to the average amount paid for the maximum/minimum injured person on the claim. The prefix *freq* refers to the frequency of payments so far and the suffix number corresponds to the number of days at which the snapshot has been taken. We only limit ourselves up to fifteen days after the date of loss to avoid capturing post-fact information. A number of variables can be categorized as global. Variables, such as *sum_injsum*, *Res_at_AB_open*, *phys* and *psy* or the variables capturing the severity of the accident such as *TotalReplaceLaborCAD* give an overview of the claim by controlling for the general severity and number of injures, the estimated global cost of the AB side and the severity

of the damages sustained by the vehicle.

Appendix E

Appendix E focuses on analysing the features composing the final database. Features are analysed following two categories: continuous or categorical. The final database before one-hot encoding is composed of 70412 observations with 126 feature variables. After one-hot encoding the factor variables, the number of feature variables rises to 207.

This section analyses the categorical variables at value level present in the database. There are in total a 62 categorical variables, and after one-hot encoding, a total of 426 level-variables of which 345 have investigated/fraud observations. Each level-variable forms a subset of the general population. In this section, each such subset's investigated/fraud rate is compared to the remaining population's investigated/fraud rate through a two-proportion z-test The tested null hypotheses is that the investigated/fraud rate is equal in the sub-sample formed by observations bearing such level and in the remaining population. The z-test formula is as follow:

$$z = \frac{i_{sample} - i_{pop}}{\left(\frac{i_{tot} * n_{tot}}{n_{sample}} + \frac{i_{tot} * n_{tot}}{n_{pop}}\right)^{\frac{1}{2}}} \tag{E.51}$$

Table E.9 below presents the results of the top twenty five levels ranked by decreasing *P.value*[44]. The first column present the variable and the tested level separated by a

[44]There are respectively 114 and 141 levels with investigation or fraud rate higher than the remaining population at the 1% level

"_". Columns *"InvesSamp"* and *"InvesPop"* are respectively the percentage of investigated claims belonging to the tested level and the remaining population respectively. The column *conf.low* is the 95% lowest bound of the investigation rate for the level population. The *SampPop* is the number of claims in the level. The P-value column is not presented because all P-values are below one per thousands. The same analysis is conducted with the different preprocessing of *AB_fraud_label* where all *"Investigated"* claims are considered as *"No_label"*. The columns *FraudSamp*, *FraudPop*, *FraudConf.low* are similar to the one described before but for the subset of *"Fraud"* claims.

As an example the first line of Table E.9 shows: *"law_rep_agrr_1"* which refers to the variable *"aggressive law firm is on the claim"* with "_1", indicating that one such law firm is on the claim. There are 2546 claims with such level and this subset has an investigation rate of 33%. The remaining population of 67866 has an investigated rate of above 4%. With a 95% confidence interval, the lowest estimate of the investigated rate for the subset of claims with strictly one aggressive law firm is 27%.

Table E.9 shows the similarities between the sets of significant levels in predicting *"Investigated"* or *"Fraud"*. Out of the set of 25 variables, 19 are common between the two sets. Table E.9 also illustrates the importance taken by law variables with eleven out of the main 25 levels for both preprocessing. Among these law variables, one or more aggressive law firm being present on a claim seems to indicate a higher investigation or fraud. Observations with multiple back or neck injuries are also associated with higher investigation and/or fraud rate. Similarly observations with multiple male claimants and no female claimants have a significant higher investigated and fraud rate.

The variable *ShopCat* controls for the type of body shop used in the claim. **ShopCat_other** refers to body shops that do not belong to the rely network of our industrial partner. This level is ranked forty-one with a 1% significance when it concerns fraud and is therefore more significant for predicting investigation. The variable *OCC_HomePerson*

controls how many claimants are not working. Observations with such characteristics have a significantly higher investigating rate than the remaining population. The fraud rate difference is also significant at the 1% level but is ranked forty-nine and is out of this table scope. The variable *OCC_SelfEmployed* has an opposite pattern: both difference in fraud and investigated rate are significant at the 1% level but the ranking P.value is lower for fraud than for investigated claims. Claims with multiple accident benefit claimants, presence of male claimants and absence of female claimants all have significant difference in term of investigated rate compare to the general population.

Table E.9: Analysis of categorical variables

Index	VarName	InvesSamp	InvesPop	Conflow	FraudSamp	FraudPop	ConfLow	SampPop
1	lawRep_agrr_1	33.39	4.42	27	22.55	2.38	19	2546
2	lawRep_agrr_2	60.22	5.25	50	40.88	2.96	33	274
3	lawRep_Nopass	12.00	1.99	10	7.02	1.02	6	24441
4	#Day_MissLaw	11.71	2.18	9	6.67	1.23	5	24278
5	#Lawyer_1	13.68	3.50	1	7.03	2.17	4	13565
6	lawRep_mod_1	13.06	4.45	8	7.32	2.54	4	8323
7	INJ_BACK_2	14.03	4.99	8	8.43	2.81	5	3678
8	lawRep_mod_2	22.30	5.27	15	14.55	2.97	10	825
9	ShopCat_Other	8.37	4.34	4	-	-	-	19628
10	INJ_BACK_3	26.26	5.32	18	16.36	3.01	10	495
11	lawRep_agrr_3	68.00	5.42	51	44.00	3.08	28	50
12	INJ_OTHER_2	12.84	5.12	7	7.80	2.89	4	3115
13	male_clmnt_2	12.50	5.18	6	6.86	2.95	3	2768
14	lawRep_agrr_NA	23.56	5.35	0.15	-	-	-	433
15	INJ_NECK_2	10.86	5.12	0.05	6.30	2.90	0.03	4209
16	lawRep_none_pass_NA	28.34	5.38	0.18	17.00	3.06	0.10	247

Continued on next page

Table E.9: Analysis of categorical variables (Continued)

Index	VarName	InvesSamp	InvesPop	Conflow	FraudSamp	FraudPop	ConfLow	SampPop
17	Homepers_1	10.32	5.09	0.04	-	-	-	5030
18	InjOther_1	7.13	4.43	0.02	4.28	2.38	0.02	26893
19	#CLAIMANTS_AB_2	12.62	5.23	0.06	8.00	2.94	0.04	2250
20	#LAWYER_2	9.32	5.05	0.04	6.56	2.73	0.03	6886
21	lawRep_mod_NA	24.22	5.38	0.15	-	-	-	322
22	female_clmnt_0	7.06	4.51	0.02	-	-	-	26404
23	PREXS_Yes	6.84	4.36	0.02	-	-	-	31356
24	INJ_OTHER_3	18.92	5.37	0.11	12.95	3.03	0.07	502
25	PRE_PSYCHO_1	8.69	5.08	0.03	5.08	2.87	0.02	7461
26	OCC_RETIRED_0	5.85	2.39	0.00	0.03	62537		
27	male_clmnt_3	19.07	5.38	0.00	0.10	451		
28	INJ_BACK_1	6.60	4.46	0.00	0.02	33012		
29	male_clmnt_1	6.58	4.48	0.00	0.02	32927		
30	EDU_MISSING_3	16.81	5.37	0.00	0.09	595		
31	INJ_NECK_3	16.58	5.37	0.00	0.09	597		
32	NBR_CLAIMANTS_AB_NA	15.51	5.36	0.00	0.08	722		
33	INJ_BACK_4	31.25	5.42	0.00	0.18	112		
34	INJ_SHOULDER_2	14.50	5.35	0.00	0.07	862		
35	HR_Call_Centre	7.26	4.91	0.00	0.02	16609		
36	lawRep_agrr_4	70.59	5.45	0.00	0.44	17		
37	EDU_HIGH_SCHOOL_2	11.21	5.31	0.00	0.05	1901		
38	OCC_RETAIL_1	9.59	5.27	0.00	0.03	3161		
39	OCC_SELF_EMPLOYED_1	9.41	5.27	0.00	0.03	3367		
40	OCC_HOMEPERSON_2	16.26	5.40	0.00	0.08	449		
41	INJ_SHOULDER_1	7.13	5.03	0.00	0.02	14682		
42	lawRep_mod_3	26.05	5.43	0.00	0.14	119		

Continued on next page

Table E.9: Analysis of categorical variables (Continued)

Index	VarName	InvesSamp	InvesPop	Conflow	FraudSamp	FraudPop	ConfLow	SampPop
43	PRE_PHYSICAL_2	10.81	5.35	0.00	0.04	1536		
44	INJ_CHEST_0	5.78	3.49	0.00	0.02	60657		
45	LESS_HIGH_SCHOOL_1	9.03	5.30	0.00	0.03	3199		
46	EDU_HIGH_SCHOOL_1	6.32	5.03	0.00	0.01	17249		
47	NBR_THIRD_PARTY_0	6.00	4.41	0.00	0.01	46597		
48	VehMake_ACURA	11.66	5.37	0.00	0.05	1012		
49	VehCat_Mid__Foreign	7.38	5.18	0.00	0.02	9178		
50	NBR_CLAIMANTS_4	9.64	5.34	0.00	0.03	2030		
51	INJ_SHOULDER_2	16.51	5.44	0.00	0.12	83		
52	NBR_CLAIMANTS_4B_4	12.39	5.40	0.00	0.05	694		
53	HR_Fax	9.60	5.35	0.00	0.03	1886		
54	INJ_HEAD_0	5.74	3.78	0.00	0.02	60554		
55	VehMake_HONDA	7.72	5.26	0.00	0.02	5801		
56	NBR_CLAIMANTS_1B_2	6.95	5.16	0.00	0.01	11885		
57	VehCat_Unknown	6.52	5.06	0.00	0.01	19645		
58	OCC_OTHER_2	11.54	5.40	0.00	0.04	780		
59	PRE_PSYCHO_2	14.44	5.42	0.00	0.06	360		
60	PhysInjuryMissEv_1_Evaluation	1.55	0.96	0.00	0.04	69053		
61	NBR_CLAIMANTS_3	8.36	5.26	0.00	0.02	6806		
62	VehMake_LowPoP	6.59	5.12	0.00	0.01	16427		
63	EDU_MISSING_2	7.84	5.32	0.00	0.02	4185		
64	OCC_MACHINE_OP_1	10.10	5.39	0.00	0.03	1139		
65	male_clmnt_4	24.00	5.45	0.00	0.10	75		
66	RepStateProv_O	9.81	5.39	0.00	0.03	1244		
67	EDU_UNIVERSITY_0	3.60	3.22	0.00	0.02	66310		
68	HR_After_Hours_Service	7.41	5.31	0.00	0.01	5315		

Appendix F

Appendix F analyses the 50 continuous variables present in the database. The first part analyses the correlation structure within the feature variables. The second part tests the difference in mean of each continuous variables between subsets of Investigated/Fraud observations and No-label observations.

Figure F.8 presents all correlation above a 60% threshold. We can clearly see clusters

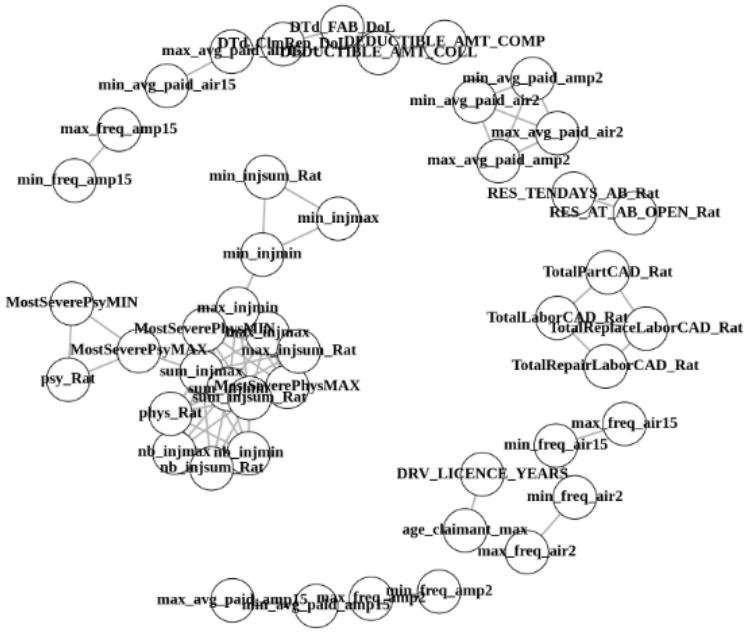


Figure F.8: Correlation Network.

of variables correlated together. The biggest cluster is comprised of variables treating the severity of injuries present on the claim. This cluster cam be further disaggregated into three smaller cluster. The most central one is linked to physical injuries with variables such as *phys_Rat*[45] and *MostSeverePhysMAX*[46] which are strongly linked with global injury indicators such as *max_injsum_Rat*[47] or *sum_injsum_Rat*[48]. This indicates that the global severity of the injuries is driven more by physical injuries rather than psychological ones. To the north of the central cluster, the first smaller cluster of three variables *min_injmin*, *min_injsum_Rat* and *min_injmax* controls for the injuries with smallest gravity. This sub-cluster is linked to the central cluster by the edge between *min_injmin* and *max_injmin* which are respectively the less and most severe injury sustained by the less injured claimant. The third sub-cluster is composed of psychological injuries and is linked to the central cluster by an edge with *sum_injmax*. The variable *sum_injmax* is the sum of all the injuries sustained by the most critically injured claimant. Psychological variables being most related to *sum_injmax* can be interpreted as that the most severely hurt claimant is most often the one suffering from psychological injuries.

The second main cluster is composed of four variables all related to damages on the car: *TotalPartCAD_Rat*, *TotalLaborCAD_Rat*, *TotalRepairLaborCAD_Rat*, *TotalPartCAD_Rat*, normalised by the price of the vehicle. The final cluster of interest shows the relation between the variables *Dtd_FAB_DoL* and *Dtd_ClmRep* which control for the difference in days between the first AB elements and the date of loss and for the difference in days between the first report of the claim and the date of loss. The high correlation between the two suggests that few claims have property damage elements opened long before the accident benefits elements. Table F.10 shows the results of student-test between

[45]Number of physical injuries sustained in the claim

[46]The most severe physical injury sustained on the claim

[47]The sum of the maximum injuries sustained by all claimants normalised by the number of AB claimants

[48]The sum of the severity of all the injuries of all the claimants normalised by the number of AB claimants

the mean of investigated and no label claims. The null hypothesis being that the sample mean, the mean of investigated claims, is equal to the population mean, the mean of the no label claims. The results are presented in Table F.10. Variables are ordered by increasing P-values. There are 45 significant variables at the 1% levels and 11 variables not significant at the 5% levels.

Table F.10: Analysis of continuous variables

Index	VarName	Pop_Mean	Pop_Sd	Sample_Mean	Sample_Sd	p.value
1	Veh_Age	7.59	5.14	8.24	5.03	0.00
4	TimeToExp	189.56	105.54	195.83	105.74	0.00
5	nb_injmin	2.05	1.28	2.53	1.42	0.00
6	SevPhysMAX	0.119	0.0638	0.13	0.0553	0.00
7	max_injmax	0.125	0.0681	0.147	0.067	0.00
8	min_injmax	0.0852	0.0439	0.0906	0.0356	0.00
9	sum_injmax	0.25	0.19	0.35	0.21	0.00
10	nb_injmax	2.23	1.29	2.94	1.38	0.00
13	NbrClaimsPol	2.12	2.29	2.31	2.88	0.00
14	ClaimantAgeMax	45.76	19.13	42.07	17.65	0.00
15	DaysBfOpenPhysMAX	93.64	144.48	101.39	178.52	0.00
19	ToL	505.97	468.64	603.13	480.66	0.00
22	phys_Rat	1.86	1.06	2.12	1.05	0.00
23	psy_Rat	0.23	0.55	0.58	0.77	0.00
24	ResAbOpenRat	9823.52	15596.60	8238.07	9724.99	0.00
25	Res10DaysAbRat	6935.84	15386.09	6013.56	8603.30	0.00
31	TotalLabCADRat	0.09	0.10	0.07	0.09	0.00
34	DTD_{LosLaw}	1.76	6.34	6.85	10.84	0.00
2	DeductAmtColl	372.99	360.60	325.20	409.46	0.00
3	FTP_AB	690.70	2857.56	1084.92	864.15	0.00

Continued on next page

Table F.10: Analysis of continuous variables (Continued)

Index	VarName	Pop_Mean	Pop_Sd	Sample_Mean	Sample_Sd	p.value
11	max_injmin	0.12	0.07	0.13	0.07	0.00
12	sum_injmin	0.23	0.19	0.29	0.21	0.00
16	SevPhysMIN	0.11	0.06	0.12	0.05	0.00
17	min_avg_paid_amp15	51.72	1102.04	8.16	176.57	0.00
18	max_avg_paid_amp15	52.37	1109.45	8.32	177.25	0.00
20	DrvLicYears	23.40	15.73	16.19	12.61	0.00
21	DaysBfPD	16.36	152.25	5.34	30.77	0.00
26	nb_injsum_Rat	2.09	1.24	2.69	1.33	0.00
27	sum_injsum_Rat	0.23	0.18	0.32	0.20	0.00
28	min_injsum_Rat	0.08	0.04	0.08	0.04	0.00
29	max_injsum_Rat	0.12	0.07	0.14	0.07	0.00
30	TotalPartCAD_Rat	0.13	0.18	0.10	0.15	0.00
32	ReplLabCADRat	0.07	0.08	0.05	0.07	0.00
33	RepLabCADRat	0.02	0.02	0.02	0.02	0.00
37	DTD$_{DoLOicd}$	2744.76	2682.88	1407.94	1844.30	0.00
38	DTd_FAB_DoL	40.63	234.52	25.23	75.41	0.00
39	SevPsyMIN	0.03	0.06	0.07	0.09	0.00
40	SevPsyMAX	0.03	0.07	0.07	0.09	0.00
41	min_avg_paid_amp2	9.34	446.36	2.84	103.41	0.01
42	min_injmin	0.08	0.04	0.08	0.04	0.01
43	max_avg_paid_amp2	9.35	446.37	2.99	104.56	0.01
44	DeductAmtComp	317.28	317.66	300.15	383.10	0.01
45	maxavgpaidair15	20.31	1497.55	5.36	97.86	0.01
46	minavgpaidair15	20.17	1497.04	5.36	97.86	0.01

Appendix G

Chen and Guestrin (2016) also use equation 1.8 to evaluate a potential split for a tree by calculating the difference in $\hat{\imath}^t(q)$ between the original tree and the original tree after the split. The authors obtain equation G.52 where I_L and I_R are instances in respectively the left and right nodes after a split.

$$\imath_{split} = \frac{1}{2}\left[\Sigma_{j=1}^{T}\frac{\Sigma_{i\in I_L} g_i}{\Sigma_{i\in I_L} h_i + \lambda} + \Sigma_{j=1}^{T}\frac{\Sigma_{i\in I_R} g_i}{\Sigma_{i\in I_R} h_i + \lambda} - \Sigma_{j=1}^{T}\frac{\Sigma_{i\in I} g_i}{\Sigma_{i\in I} h_i + \lambda}\right] + \gamma T \tag{G.52}$$

As it is impossible to list all tree structures $q(x)$, XGBoost starts from a single leaf and construct a tree base-learner by selecting a variable k from a random sub-sample V of all the feature variables available using (G.52). The authors offer two different methods to select the best value for a potential split among all variables k in the sub-sample V. The first method is called the *Exact Greedy Algorithm for Split Finding* which is impractical for many real life applications. The second method, used here, is called the *Approximate Algorithm for Split Finding*.

Approximate Algorithm for Split Finding consists in calculating weighted quantiles for each variables of the sub-sample. Using the notation from Chen and Guestrin (2016), we define $D_k = \{(x_{1k}, h_1), (x_{2k}, h_2), ..., (x_{nk}, h_n)\}$ the ordered tuple over feature variable k of each observation value paired with the corresponding second derivative. (G.53) is the rank function that weights each quantile by the value of the second derivative h.

$$r_k(z) = \frac{1}{\Sigma_{(x,h)\in D_k}(h)} \Sigma_{(x,h)\in D_k, x<z}(h) \tag{G.53}$$

Potential split points $\{s_{k1}, s_{k2}, ..., s_{kl}\}$, for each feature k are selected following a criteria dependent on ε, an hyper parameter regulating the XGBoost algorithm as described in (G.54).

$$r_k(s_{k,j}) r_k(s_{k,j+1}) \leq \varepsilon,\ s_{k1} = min_i(x_{ik}), s_{kl} = max_i(x_{ik}) \tag{G.54}$$

Once a potential split point $s_k 1$ for each feature v is selected, (G.52) is used to evaluate the $\{k, s_{ki}\}$ combination that most decreases Equation (1.2), where k refers to the feature variable chosen and s_{ki} the value upon which the split is done. Chen and Guestrin (2016) finally address issues related to multiple potential optimal split and implementation as well as sparsity robust trees that we leave to the curious reader.

In this study, we implement the XGBoost algorithm using the R-package: *"xgboost: Extreme Gradient Boosting"*[49]. Hyper-parameters η, γ, *max_depth*, λ, subsample, α, *scale_pos_weight* are fixed through a cross-validation procedure described in the following section. η, α and *scale_pos_weight* parameters are additional hyper-parameters not mentioned during the introduction of the boosting model. η is the equivalent of a learning step for a gradient descent optimization algorithm. Each new booster is equivalent to direction selection in the optimization space, while η limits the step taken in the selected direction. α is similar to the γ parameter and puts a weight on newly added leaves. The *scale_pos_weight* parameter controls for the penalty weight of mis-classifying a positive sample when the data is imbalanced. The sampling_method used is fixed to the *"uniform"* method where each columns has a similar chance of being sub-sampled.

[49] https://xgboost.readthedocs.io/en/latest/parameter.htmlglobal-configuration

Appendix H

In this section, we describe the methodology of Goutte and Gaussier (2005) followed here to select the most stable set of hyper-parameters. The authors offer a statistical interpretation to the *Precision* (H.56) and *Recall* (H.57). We follow their methodology to select the combination of hyper-parameters that would produce the most stable results on new data.

$$F1score = 2 * ((precision * recall)/(precision + recall)) \tag{H.55}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{H.56}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{H.57}$$

Table H.11: Confusion Matrix

Assignment	z	+	-
Label l		TP	FP
	-	FN	TN

(H.56) and (H.57) can be reformulated in statistical terms as $p = P(l = +/z = +)$ and $r = P(z = +/l = +)$. The authors underline the important shift in interpretation arising from the reformulation of (H.56) and (H.57): *"...in the original formulation, p and r are*

just formulas calculated from the observed data, ... , p and r (...) are now parameters of a (primitive) generative model. Thus, the usual expressions arise only as estimates of these unknown parameters."[50]

Table H.11 introduces the confusion matrix. The vertical dimension represents the prediction made by the model while the horizontal dimension represent the true label of the observation. The top right corner cell is the portion of observations that have been correctly classified as positive, hence True-Positive. The other cells are from right to left and top to bottom: False-Positive, False-Negative and True-Negative classification. Assuming that TP conditional on FP, TP and FN follow Binomial Distribution and a Beta prior for Precision and Recall, the authors obtained (H.58) through (H.62).

$$p = \frac{X}{X+Y} \tag{H.58}$$

$$r = \frac{X}{X+Z} \tag{H.59}$$

$$\text{with } X \sim \Gamma(TP+\lambda, h) \tag{H.60}$$

$$Y \sim \Gamma(FP+\lambda, h) \tag{H.61}$$

$$Z \sim \Gamma(FN+\lambda, h) \tag{H.62}$$

$$\tag{H.63}$$

From (H.58) the authors can deduce that F1 score is a combination variable as described in (H.64).

[50] page 4, paragraph 3 of Goutte and Gaussier (2005)

$$F_1 = \frac{U}{V+U} \tag{H.64}$$

$$\text{with } U \sim \Gamma(TP+\lambda, 2h) \tag{H.65}$$

$$V \sim \Gamma(FP+FN+2\lambda, h) \tag{H.66}$$

Thanks to (H.64), F1 has now a density with a set of parameters $\{TP, FN, FP, \lambda, h\}$ from which we can simulate an empirical density of the distribution of each $F1score$ obtained during the cross-validation procedure. An $F1$ score is preferred to another one following the criteria described in (H.67).

$$\hat{P}(F_1^1 > F_1^2) = \frac{1}{L}\Sigma_{i=1}^{L} I(f_i^1 > f_i^2) \tag{H.67}$$

We now present the steps that are implemented in this study to select the set of parameters that produces the best distribution of the $F1$ score.

- Retrieve TP, FP and FN for the entire Training database through the CV process
- Simulate a sample of the Gamma laws (L=10000) based on the TP, FP and FN from each set of parameters. Γ and h are arbitrarily chosen to be 2.
- Simulated Distribution of F1 score are then compared by empirically calculating the probability in (H.67).
- All pairwise score are calculated and then the dominating set of parameters is selected.
- The model is then retrained on the train database using the set of selected parameters.
- Predictions are made on the testing set.

Appendix I

The cross-validation procedure applied to XGB model searches the most performative combination of parameters. The search grid is comprised of 1458 combinations and controls for hyper-parameters described in table I.12. The first line introduces the grid itself while subsequent rows give the best parametrization selected after the cross-validation for each configuration in table 1.3. Specific parameters for *Linear* booster are: *Updater* = *coord_descent* and *Sample_Type* = *uniform*. For *Dart* booster, specific parameters are: *FeatSelecter* = *random* and *Normalize_type* = *tree*.

Table I.12: Cross-validation grid

Model	λ	η	γ	MaxDepth	Subsample	Weight	DropRate
Grid	$\{0,1,2\}$	$\{.3,.4,.5\}$	$\{.3,.4,.5\}$	$\{4,5,6\}$	$\{.7,.8\}$	$\{5,10,20\}$	$\{.08,.1,.12\}$
Tree	2	.4	.3	4	.8	5	NA
Dart	0	.3	.5	4	.8	5	.1
Linear	0	.5	.4	5	.7	5	NA
FOTree	1	.3	.4	4	.8	5	NA
FODart	1	.3	.5	4	.7	5	.1

Continued on next page

Table I.12: Cross-validation grid (Continued)

Model	λ	η	γ	MaxDepth	Subsample	Weight	DropRate
FOLinear	0	.5	.5	5	.8	5	NA
NLTree	0	.3	.5	4	.8	10	NA
NLLinear	0	.5	.4	5	.7	10	NA
NLDart	2	.4	.4	4	.8	5	.08

Appendix J

The function $\phi(P)$ is truncated at P_0 such that $\phi(P) = 0$ if $P_0 < P < 1$ where P_0 is defined by (J.68). P_0 is the audit probability at which someone with no moral cost will be deterred to fraud. Any more aggressive audit strategy that P_0 is inefficient as at any audit level such that $P > P_0$, there is simply no fraud being committed.

If the penalty to defraud B is zero, then $P_0 = 1$. The observation signifies that when there are no penalties B, individuals with no moral cost are deterred from defrauding only if the audit strategy consists in investigating all claims: $P_0 = 1$. If the penalty to defraud $B > 0$ then $0 < P_0 < 1$.

$$(1 - P_0)U(W_0 + t, 0) + P_0 U(W_0 - B, 0) = U(W_0, 0) \tag{J.68}$$

The function ϕ has characteristics as described in (J.69) through (J.71):

$$\phi(0) > 0 \tag{J.69}$$

$$\phi(P_0) = 0 \tag{J.70}$$

$$\phi^{'}(P) < 0 \text{ if } 0 < P < P_0 \tag{J.71}$$

(J.69) shows that if the audit rate is non-existent, claimants with a positive moral cost ω will be indifferent between the status-quo and fraud and (J.68) will be true. (J.70) reflects the absence of fraud when the audit level is at its maximum P_0, ω belonging to $\mathbb{R}+$. (J.71) is the rate at which an aggressive audit will deter fraud.

Appendix K

The following methodology describes the spreading and smoothing of the residual undetected fraud over all θ:

- A 15% fraud rate gives 10561 fraudulent claims out of 70142 observations.
- $10561 - 3848 = 6713$ undetected frauds.
- Undetected_fraud_rate $P(\hat{F}/\theta)$ is $\frac{6713}{(70142-3848)} = 10.00\%$ and is distributed equally on all θ
- let N be the non-fraud population of a given θ
- let I be the fraud population of a given θ
- $P(F/\theta) = \frac{P(\hat{F}/\theta)*N+I}{I+N}$

Appendix L

We present here the methodology with which the expected cost of special investigation is estimated. After discussing with our industrial partner, we base our estimate on the hypothesis that around 10% of investigated claims necessitate a deeper special investigation. Over a period of eight years[51], 1200 claims have been specially investigated resulting in a global probability of investigation of 1.7%. This global probability is then disaggregated over all θ groups following the procedure below:

- From the data, we can observe the probability $P(\theta/AEX > 0)$ where $AEX > 0$ indicates that the observation has special examination costs.
- Hypothesis: $P(\theta/AEX > 0) = P(\theta/Surveillance)$.
- By Bayes law, we get $P(Surveillance/\theta) = \frac{P(\theta/Surveillance)*P(\theta)}{P(S)}$.
- If some θ do not have observations with $AEX > 0$ because of low population then a minimum 0.1% is enforced.

The cost of special investigation is modeled after the total cost of compensation $t(\theta)$. The assumption is that a cost of special investigation would on average not exceed a certain fraction of the total compensation that the claim will cost. Moreover our industrial partner indicated that most special investigations have a cost in $[4000\$, 7000\$]$.

[51] The database lasts from 2012-2020

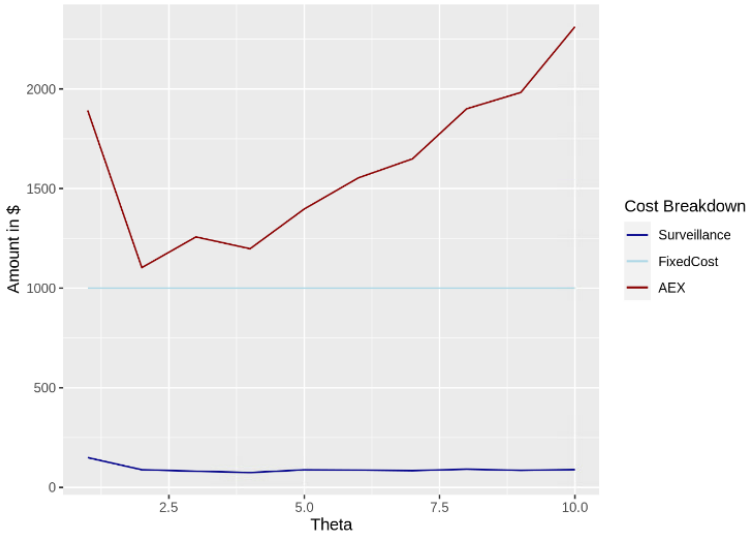


Figure L.9: Break down of global $c(\theta)$

Based on these observations, the range of compensation at claim level is calculated and segmented in fifty parts. Similarly the range of special investigation cost is segmented in fifty distinct parts. A mapping between the two range allows to obtain the final special investigation costs. The cost is obtained as follow: an observation belonging to the 1^{st} segment in t range will be clustered at midpoint of the 1^{st} segment in special surveillance and so on for each clusters. We now have a cost and a probability at the claim level. This allows to calculate the expected cost of special investigation and aggregate it at the θ level. Figure L.9 shows the break down of investigation cost $c(\theta)$ per component at θ level.

Appendix M

The algorithm stratifies the data into categories according to a list of discretized variables. Stratas are created according to dimensions: *"How_reported"*, *Cause_of_Loss* and *Occupation*. For each created strata, the algorithm samples the minimum number of observations necessary so that the first and second moments of some target variables are within a 95% interval from the full population distributions. These target dimensions are: 15 dimensions related to disbursement of policies and 6 general purposes variables: Age of claimant, the reserve when AB side is opened, the fraud characteristics, the vehicle age, the remaining time on the claimants' contract and the history of claims on the policy.

Appendix N

Table N.13: Classifier results

Model	Config	tn	fn	tp	fp	youden	specif	sensit	precis	F1score
Logi	Default	11548	502	159	126	0.230	0.989	0.241	0.558	0.336
	FraudOnly (FO)	11405	297	90	80	0.226	0.993	0.233	0.529	0.323
	NoLaw (NL)	11605	635	67	104	0.087	0.991	0.095	0.392	0.153
XGB	GBTree	12914	482	275	402	0.333	0.970	0.363	0.406	0.384
	Dart	12730	420	337	586	0.401	0.956	0.445	0.365	0.401
	Linear	12525	372	385	791	0.449	0.941	0.509	0.327	0.398
	FOTree	13159	328	110	144	0.240	0.989	0.251	0.433	0.318
	FODart	13103	287	151	200	0.330	0.985	0.345	0.430	0.383
	FOLinear	12911	258	180	392	0.381	0.971	0.411	0.315	0.356
	NLTree	12598	542	215	718	0.230	0.946	0.284	0.230	0.254
	NLDart	11835	361	396	1481	0.412	0.889	0.523	0.211	0.301
	NLLinear	11516	339	418	1800	0.417	0.865	0.552	0.188	0.281

Appendix O

This table presents columns *NR-AC*$ and *NA-AC*$ that are respectively the cost associated with the model if all non-predictable observations are all audited and if all non-predictable observations are processed normally.

Table O.14: Optimal Audit Model Results

Config	Class	$F1$	$F1_{OAM}$	%NoPred	NR-AC$	AA-AC$	NA-AC$	$ saved
Def	tree	0.384	0.270	0.07	215,868,325	216,054,768	216,038,512	1,521,527
Def	dart	0.401	0.302	0.05	215,168,876	215,302,863	215,294,116	2,220,976
Def	line	0.398	0.251	0.02	216,540,334	216,590,968	216,585,021	849,518
Def	logi	0.336	0.287	89.6	18,571,000	209,430,903	188,068,114	540,509
FO	logi	0.323	0.183	0	185,746,869	185,746,869	185,746,869	-1,250,041
FO	dart	0.383	0.212	0.05	215,961,820	216,087,086	216,072,933	-326,884
FO	tree	0.318	0.197	0.10	216,198,282	216,461,076	216,434,269	-563,346
FO	line	0.356	0.172	0.04	217,370,799	217,474,018	217,462,195	-1,735,863
NL	logi	0.153	0.195	0.016	191,528,302	191,568,561	191,563,736	-1,490,538
NL	dart	0.301	0.185	0.27	218,438,608	219,113,157	219,043,042	-1,048,756
NL	line	0.281	0.167	0	220,058,025	220,058,025	220,058,025	-2,668,173

Appendix P

Table P.15: σOAM vs Elkan in Fraud-Only

			Elkan					σOAM				
θ	$\frac{t}{c}$	F%	TP	FP	TN	FN	Thrs	TP	FP	TN	FN	Thrs
1	7.65	3.34	11	60	3256	33	0.130	7	33	3283	37	0.230
2	9.52	2.42	1	17	1039	3	0.105	1	5	1051	3	0.294
3	8.99	2.96	3	21	975	5	0.111	2	11	985	6	0.230
4	8.77	3.41	11	24	883	3	0.114	10	16	891	4	0.193
5	8.27	3.97	11	17	926	8	0.121	10	11	932	9	0.193
6	8.02	4.34	8	30	920	16	0.125	8	19	931	16	0.193
7	7.61	5.28	17	7	51	1020	0.131	15	43	1028	21	0.155
8	7.24	6.27	30	73	1049	32	0.138	29	69	1053	33	0.155
9	6.91	7.13	37	95	1277	47	0.145	39	114	1258	45	0.115
10	6.27	9.35	75	137	1433	68	0.159	87	217	1353	56	0.079

Columns $\frac{t}{c}$ is the ratio of the compensation to the investigation cost per θ. $F\%$ is the fraud-rate and TP, FP, TN and FN are respectively the true-positive, false-positive, true-negative and false-negative number of claims. $Thrs$ is the calibrated probability threshold above which a claim is audited. A higher threshold means a less aggressive audit policy.

Table P.16: σOAM vs Elkan in No-Law

			Elkan					σOAM				
θ	$\frac{t}{c}$	F%	TP	FP	TN	FN	Thrs	FP	FN	TP	TN	Thrs
1	7.66	1.76	5	29	3170	57	0.130	7	60	2	3192	0.211
2	9.43	1.15	3	22	1008	6	0.106	6	7	2	1024	0.211
3	8.84	1.90	5	26	938	10	0.113	14	12	3	950	0.203
4	8.78	2.62	13	40	892	17	0.114	25	20	10	907	0.206
5	8.29	3.78	8	46	880	34	0.121	39	34	8	887	0.200
6	7.95	4.54	22	61	904	24	0.126	58	26	20	907	0.200
7	7.56	5.90	23	83	976	39	0.132	88	40	22	971	0.113
8	7.19	7.50	40	130	1070	53	0.139	179	48	45	1021	0.089
9	6.96	9.56	55	196	1255	76	0.144	275	65	66	1176	0.075
10	6.35	13.63	139	324	1266	128	0.158	498	85	182	1092	0.064

Columns $\frac{t}{c}$ is the ratio of the compensation to the investigation cost per θ. $F\%$ is the fraud-rate and TP, FP, TN and FN are respectively the true-positive, false-positive, true-negative and false-negative number of claims. $Thrs$ is the calibrated probability threshold above which a claim is audited. A higher Threshold means a less agressive audit policy.

Appendix Q

Table Q.17: Probit Regression η - Part 1

	Dependent variable: Choice of Coverage	
	No Coverage included	*No Coverage excluded*
	(1)	(2)
Pre_Phys	-0.040^{***}	-0.059^{***}
	(0.013)	(0.014)
Pre_Psycho	-0.099^{***}	-0.131^{***}
	(0.020)	(0.022)
Nbr_Potential_Drivers	0.097^{***}	0.097^{***}
	(0.010)	(0.011)
Drv_Licence_Years	0.231^{***}	0.166^{***}
	(0.010)	(0.011)
Annual_Km	0.035^{***}	0.025^{***}
	(0.007)	(0.008)
Vehicle_Year	0.032^{***}	-0.073^{***}
	(0.011)	(0.011)
Veh_Coll_Rat	0.021^{***}	0.011^{***}
	(0.003)	(0.003)
Cat_Eco_Domestic	0.543^{***}	0.786^{***}
	(0.207)	(0.219)
Cat_Eco_Import	0.581^{***}	0.814^{***}
	(0.206)	(0.218)
Cat_Full_Pickups	0.731^{***}	0.952^{***}
	(0.208)	(0.220)
Cat_Large_SUV	0.676^{***}	0.849^{***}
	(0.217)	(0.230)
Cat_Luxury	0.673^{***}	1.017^{***}
	(0.225)	(0.239)
Cat_Mid_Domestic	0.611^{***}	0.802^{***}
	(0.207)	(0.219)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table Q.18: Probit Regression η - Part 2

	Dependent variable: Choice of Coverage	
	No Coverage included	*No Coverage excluded*
	(1)	(2)
Cat_Mid_Foreign	0.594***	0.784***
	(0.205)	(0.217)
Cat_Mini_Vans	0.682***	0.825***
	(0.206)	(0.219)
Cat_Small_Pickups	0.587***	0.735***
	(0.218)	(0.232)
Cat_Small_SUV	0.715***	0.914***
	(0.205)	(0.217)
VehOrigin_European	-0.599***	-0.868***
	(0.215)	(0.228)
Occ_Child	0.238***	0.277***
	(0.075)	(0.085)
Occ_Retired	0.208***	0.256***
	(0.065)	(0.073)
Occ_Middle_Mgmt	0.198***	0.282***
	(0.077)	(0.085)
Occ_Teacher	0.329***	0.323***
	(0.080)	(0.088)
Occ_Nurse	0.212***	0.241***
	(0.079)	(0.087)
Occ_FT_Student	0.202***	0.259***
	(0.066)	(0.074)
Occ_Other	0.171***	0.212***
	(0.063)	(0.072)
Edu_University	0.126***	0.111***
	(0.035)	(0.038)
Observations	45,641	41,393
Log Likelihood	−27,573.150	−23,294.750
Akaike Inf. Crit.	55,358.300	46,801.500
Note:		*p<0.1; **p<0.05; ***p<0.01

Appendix R

Group	Estimate	Estimate 1	Estimate 2	Statistic	p-value	Parameter	Conf. Low	Conf. High
1	4352	9558	5207	3	0	8	1234	7470
2	516	8969	8453	0	0.74	7	-3011	4042
3	4276	10531	6256	2	0.13	4	-1968	10520
4	-7594	16975	24569	-1	0.49	6	-32878	17691
5	1342	8507	7165	1	0.31	7	-1528	4211
6	110	8504	8393	0	0.89	14	-1526	1747
7	382	9423	9041	0	0.79	11	-2676	3440
8	-2061	13627	15688	-0	0.68	16	-12455	8332
9	2320	8075	5754	4	0	17	1037	3604
10	613	8351	7738	1	0.56	24	-1555	2781
11	1135	9665	8530	1	0.27	30	-926	3196
12	6452	14253	7800	4	0	20	3223	9682
13	2686	8145	5459	3	0.01	21	907	4466
14	1412	8148	6736	3	0.01	41	328	2495
15	-55	8600	8656	-0	0.95	26	-1708	1598
16	-3251	12505	15757	-1	0.51	26	-13181	6678
17	-386	8194	8580	-0	0.82	29	-3883	3112
18	263	8120	7857	0	0.85	40	-2612	3138
19	1673	8928	7255	3	0.01	73	434	2913
20	2097	12802	10705	1	0.42	28	-3163	7357
21	525	7760	7235	1	0.30	40	-490	1540
22	1009	8146	7136	2	0.03	78	85	1933
23	274	8646	8371	0	0.77	61	-1587	2135
24	-3324	12362	15686	-1	0.46	40	-12337	5690

Continued on next page

Table R.18 Continued from previous page

Group	Estimate	Estimate 1	Estimate 2	Statistic	p-value	Parameter	Conf. Low	Conf. High
25	179	8310	8130	0	0.86	53	-1784	2142
26	533	8034	7501	1	0.20	109	-287	1352
27	141	8716	8575	0	0.85	87	-1357	1639
28	261	12012	11751	0	0.89	65	-3429	3951
29	314	8001	7687	1	0.62	90	-927	1555
30	-2	7881	7883	0	1	118	-1919	1917
31	437	8037	7600	1	0.53	99	-927	1800
32	-279	12093	12372	-0	0.89	102	-4218	3661
33	589	8125	7536	1	0.41	119	-822	2000
34	1006	7990	6983	2	0.02	655	172	1840
35	1058	8430	7371	2	0.01	285	214	1902
36	1202	11133	9931	1	0.27	162	-957	3361
37	1050	7826	6776	4	0	311	458	1642
38	678	7700	7022	3	0	790	229	1128
39	624	8244	7620	2	0.10	426	-130	1378
40	-1317	10198	11516	-1	0.49	163	-5092	2458

Appendix S

Table S.19: Sensitivity Analysis

			Default		t*1.5		$\hat{S}$*1.5		c*1.5		η*1.5		W*1.5	
Θ	P	P Det	Opt	Opt Det	Opt	Opt Det	Opt	Opt Det	Opt	Opt Det	Opt	Opt Det	Opt	Opt Det
1	0.0105	0.48	0.745	0.943	0.987	1	0.743	0.933	0.742	0.919	0.745	0.943	0.745	0.943
2	0.0237	0.585	0.728	1	0.728	1	0.728	1	0.655	0.987	0.728	1	0.728	1
3	0.0388	0.514	0.723	0.955	0.923	1	0.723	0.955	0.475	0.74	0.722	0.955	0.723	0.955
4	0.0512	0.522	0.836	0.959	0.971	0.926	0.836	0.959	0.744	0.92	0.837	0.96	0.835	0.959
5	0.0758	0.528	0.756	0.948	0.977	1	0.757	0.952	0.493	0.74	0.75	0.942	0.727	0.908
6	0.09	0.518	0.746	0.921	0.996	0.997	0.746	0.924	0.721	0.875	0.747	0.928	0.746	0.922
7	0.117	0.538	0.742	0.929	0.992	1	0.747	0.948	0.497	0.738	0.745	0.938	0.747	0.948
8	0.15	0.518	0.739	0.909	0.983	1	0.737	0.901	0.457	0.695	0.744	0.942	0.739	0.91
9	0.19	0.536	0.738	0.9	0.992	1	0.741	0.906	0.468	0.704	0.714	0.886	0.738	0.899
10	0.27	0.525	0.737	0.872	0.988	0.974	0.674	0.823	0.44	0.641	0.497	0.72	0.738	0.877

Table S.20: Sensitivity Analysis 2

			η_ω*1.5		μ_ω*1.5		$P0_\omega$*1.5		π*1.5	
Θ	P	P Deter	Opt	Opt Deter	Opt	Opt Deter	Opt	Opt Deter	Opt	Opt Deter
1	0.0105	0.525	0.701	0.967	0.723	0.705	0.497	0.706	0.745	0.943
2	0.0237	0.671	0.4	1	0.732	0.869	0.495	0.766	0.728	1
3	0.0388	0.594	0.597	1	0.718	0.718	0.476	0.727	0.723	0.955
4	0.0512	0.56	0.918	1	0.728	0.738	0.857	0.966	0.825	0.952
5	0.0758	0.584	0.801	1	0.719	0.711	0.512	0.728	0.742	0.95
6	0.09	0.554	0.73	0.96	0.722	0.699	0.499	0.723	0.746	0.924
7	0.117	0.58	0.898	1	0.722	0.712	0.497	0.727	0.747	0.948
8	0.15	0.574	0.834	1	0.709	0.657	0.495	0.722	0.74	0.917
9	0.19	0.583	0.713	0.963	0.713	0.669	0.488	0.706	0.742	0.91
10	0.27	0.576	0.468	0.744	0.467	0.5	0.495	0.71	0.739	0.879

Appendix T

Appendix U

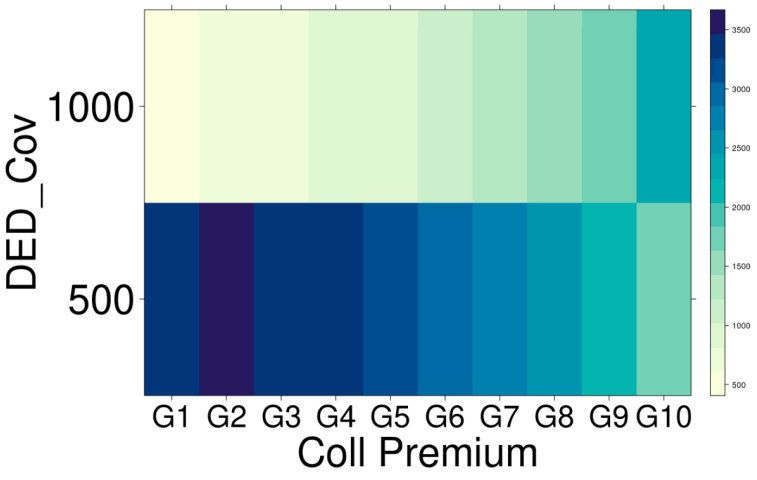


Figure T.10: Population per type of collision insurance contracts

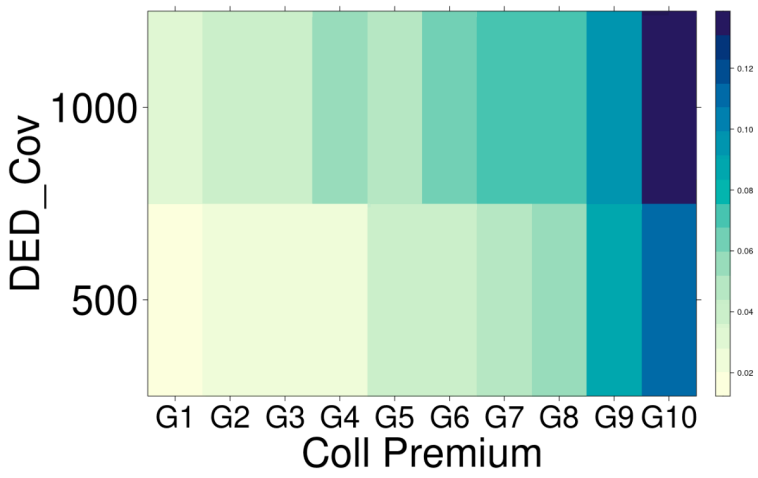


Figure T.11: Fraud Rate per type of collision contract

Table U.21: Constant Deterrence effect

θ	Opt^* for S1	Opt^* CdE	θ	Opt^* for S1	Opt^* CdE
1	0.000	0	2	1.000	0
	0.999	0.981		1.000	0.979
	1.000	0.991		0.998	0.991
	0.991	0.997		1.000	0.996
3	0.000	0	4	0.386	0.113
	1.000	0.979		1.000	0.979
	1.000	0.991		0.999	0.990
	0.999	0.922		0.998	0.997
5	0.112	0.014	6	0.001	0
	0.997	0.978		1.000	0.978
	1.000	0.99		1.000	0.990
	0.986	0.997		0.991	0.997

to be continued on the next page

Table U.21: (continued from previous page)

θ	Opt^* for S1	Opt^* CdE	θ	Opt^* for S1	Opt^* CdE
7	0.000	0	8	0.000	0
	0.998	0.978		0.998	0.978
	1.000	0.990		1.000	0.991
	1.000	0.997		0.978	0.997
9	0.000	0	10	0.000	0
	1.000	0.977		1.000	0.977
	1.000	0.990		0.999	0.990
	0.980	0.982		0.965	0.973

Appendix V

Table V.22: Constant Deterrence effect

θ	Opt^* for S1	Opt^* CdE	θ	Opt^* for S1	Opt^* CdE
1	0.000	0	2	1.000	0
	0.999	0.981		1.000	0.979
	1.000	0.991		0.998	0.991

to be continued on the next page