# THREE ESSAYS ON CORPORATE FINANCE AND MACHINE LEARNING

by

Zifen Zeng

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Business Administration

Charlotte

2024

Approved by:

______________________________
Dr. Gene C. Lai

______________________________
Dr. I-Hsuan Ethan Chiang

______________________________
Dr. Tao-Hsien Dolly King

______________________________
Dr. Rob Roy McGregor

# ABSTRACT

ZIFEN ZENG. Three Essays on Corporate Finance and Machine Learning.
(Under the direction of DR. GENE C. LAI)

This dissertation consists of three essays on corporate finance and machine learning. The first essay investigates the relation between CEO conscientiousness and reserve management in U.S. property-liability insurers. The psychology literature claims that conscientiousness is one of the strongest predictors of work-related behavior. I find that CEO conscientiousness is negatively associated with reserve errors in the upper tail of the conditional distribution (at 75th percentile and higher), indicating insurers with more conscientious CEOs reserve less than insurers with less conscientious CEOs at a higher level of reserve errors to lower the cost of excess reserve rather than conservatism when reserve errors are extremely conservative. The evidence also shows that the negative relation is mitigated when insurers face high financial risk. Furthermore, more conscientious CEOs reserve less than less conscientious CEOs after SOX (compared with before SOX) when insurers face higher financial risk, possibly because they are more responsible for financial statements. The evidence is consistent with one feature of conscientiousness: following the rules and norms. Finally, more conscientious CEOs are better rewarded than less conscientious CEOs.

The second essay studies the relation between corporate opacity and net premium written as a proxy of policyholders' purchase behavior in U.S. property-liability insurers. I find that policyholders are willing to buy policies from less opaque insurers. The evidence also shows that policyholders are more sensitive to information about insurers' financial risk when they are less opaque. Additionally, policyholders are aware of insufficient protection by the guaranty fund. It further suggests that opacity significantly influences the purchase behavior of commercial lines,

due to the involvement of brokers and agents who possess in-depth knowledge of insurers' financial situations and product policies. Thus, insurers' opacity plays a crucial role in shaping policyholders' purchase behavior.

The third essay applies machine learning methods to detect physicians. Physician fraud takes an important portion of healthcare fraud which needs continuous assessment and revision of the control methods. Using a large dataset from a life insurer in Taiwan, I construct 32 features and use multiple methods, including the neural network and RUSBoost methods to detect fraudulent physicians. Based on the neural network model, I further analyze the importance of features in detecting fraudulent physicians. Addressing the imbalanced data issue, the AUROC score of the neural network model is 0.781 for physicians with multiple claims. I find the cost savings range from 16.3% to 36.9% assuming the fraud rate of fraudulent physicians' total claim amounts ranging from 30% to 70%. I also find the important features to identify fraudulent physicians are associated with physicians clustering in the eastern area of Taiwan, the percentage of insureds whose age are less than 18, the percentage of surgeries due to illness as opposed to accidents, and whether the physician can perform difficult surgeries. Finally, the evidence implies fraudulent physicians use the "steal a little, all the time" strategy to avoid being caught. Besides cost savings, this study can benefit the life insurer by speeding up the claim review process, narrowing down the investigation range, and excluding suspicious physicians as external reviewers.

# ACKNOWLEDGEMENTS

I cannot express how much I am grateful for the support and kindness of my committee chair, Dr. Gene C. Lai. Thank you for introducing me to the idea of research. You told me research is like a marathon; I need to persist with it and keep working. Research is a journey of ups and downs; sometimes it works, but often it doesn’t. What I need to do is my best and let it be. Your guidance and attitude towards research have been the greatest inspirations on my research journey. You are not only the best mentor in research but also in my life. You show me the true role model of an inspirational scholar and how to be a patient, kind, encouraging, and persistent person.

I would like to thank Dr. Ethan Chiang, Dr. Yongqiang Chu, Dr. Dolly King, and Dr. David Mauer for your continuous support in my job search. Your guidance equipped me well. All of you witnessed my entire challenging job search journey, from interviews to the job talk.

I am also grateful to my committee members, Dr. Ethan Chiang, Dr. Dolly King, and Dr. Rob Roy McGregor. Thank you so much for your insightful comments and invaluable time spent on my dissertation.

I also want to extend my thanks to all the faculty members at the Belk College of Business, who have been an integral part of my Ph.D. journey and have witnessed my growth. To my Ph.D. cohort, Zixin Jiang: I will never forget the days in our office at Friday 388, where you witnessed my tears and joys.

# DEDICATION

To Mom, Hui Lin, and Dad, Qingmiao Zeng: for introducing me to the idea of a sweet home and showing me how deeply I am beloved. To Grandmother, Jinmei Zhou, and Grandfather, Shanqun Zeng: for your unconditional love and for always being by my side. All of you have shown me what love is, making me a braver and more caring person. You have proven that love will always flourish and never fade.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# INTRODUCTION

This dissertation consists of three essays on corporate finance and machine learning. It explores three lines of research: the impact of CEO personality traits on reserve estimation (Chapter 1); the influence of insurers' opacity on policyholders' purchase behavior (Chapter 2); and the use of machine learning methods to detect fraudulent physicians (Chapter 3).

The first chapter investigates the relationship between the CEOs' conscientiousness trait and reserve management within U.S. property-liability insurance insurers. Insurance companies, having collected premiums at beginning of a policy, must set up reserves for future claim payments. Some claims, such as auto liability claims, can span several years before settlement. Essentially, the loss reserve is an estimate of these unpaid claims. Moreover, it represents the largest liability on the balance sheet, underscoring the importance of reserve estimation for insurers. How are these reserves determined? Initially, actuaries provide a range of estimates, but the estimation is finalized by CEO's discretion. One feature of the loss reserve is that after the initial disclosure of reserve estimation, insurers are required to report the revised reserve estimation in subsequent years. This difference between initial and revised estimates is the definition of reserve errors.

There is a tradeoff in reserve estimation. On the one hand, if insurers reserve too little, they will face a higher insolvency risk due to potential insufficient funds to cover future claims. On the other hand, over-reserving can be costly because insurers have fewer free cash flows available for investment in positive NPV projects.

On average, property-liability insurance companies tend to reserve more to avoid insolvency, as it adversely affects not only stockholders but also policyholders. Thus, reserving more is important, but reserving too much is not optimal.

Previous literature has extensively examined reserve management from various perspectives due to its importance as a corporate policy. I want to explore this topic from a new angle, focusing on the conscientiousness trait, one of the “Big Five” personality traits. The big five traits are a system description of personality. Conscientious CEOs are more responsible. They also tend to follow rules. I believe responsibility and following rules are associated with reserve estimates. The other four traits are agreeableness, which means people are trusting and want to avoid conflicts. Extraversion means people are talkative and outgoing. Emotional stability means people can deal with stress. Openness to experience means people are thoughtful and creative. By their definitions, these four traits don’t directly relate to reserve estimation. Therefore, my theoretical focus will be specifically on the conscientiousness trait.

I measure conscientiousness based on the CEOs’ spoken language. And it’s from the conference calls Q&A portion, identified by CEO names. Because the Q&A portion is less scripted. The conscientiousness trait is calculated by using machine learning algorithms based on the literature by Mairesse et al., 2007.  The advantage of this method is that it does not rely on keywords. Instead, it is based on the sentence structure.

I apply quantile regression to examine the association between reserve errors and conscientiousness because the incentives of reserve estimation may differ across different levels of reserve errors. Interestingly, I find that when insurers reserve too much (reserve errors at the 75th quantiles and higher (80th, 85th, 90th, and 95th)), conscientious CEOs will lower the reserve estimation to to be more responsible for stockholders’ wealth instead of conserveness at higher levels of reserve errors so that insurers do not over-reserve too much. For economic significance, I find one standard deviation increase in conscientiousness will decrease the reserve by 1.317% of total assets, which is about 2 million dollars on average. Moreover, I find that this negative relationship will be mitigated when insurers face high financial risks. This is when the responsibility feature comes into play because conscientious CEOs are responsible to stakeholders. Furthermore, since conscientious people are more likely to follow rules, I find that after SOX, they tend to pursue more accurate estimations to meet the requirements. Finally, I find that insurers pay higher compensation to more conscientious CEOs.

The contribution of chapter 1 is that I examine reserve errors from a new angle: the conscientiousness of CEOs, based on text analysis technology. I find that CEOs' conscientiousness trait influences the important corporate policy -- reserve estimation.

The second chapter studies whether the opacity of insurers shapes a typical policyholder's purchase behavior. The information quality (opacity) is valuable for current and prospective policyholders (debtholders) because policyholders would not be paid in full if insurers become insolvent. Yet, prospective policyholders may not have the incentives and expertise to understand the quality of information about insurers' financial health due to the complexity of the liability structure of insurers. Thus, whether the opacity of insurers has an impact on insurance purchase behavior is an empirical question. The lack of transparency may lower policyholders' utility regarding the information risk of whether an insurer's information is reliable. Transparency can enhance policyholders' belief that their future claims would be paid in full when losses are incurred.

Policyholders are concerned about insurers' insolvency, especially when the information is accurate. Thus, I examine the interaction effect between financial risk and opacity on insurance purchase behaviors. I argue that insurers' opacity will mitigate the negative effect between financial risk and insurance purchase behavior.

Policyholders suffer losses when insurers become insolvent even though there is a state-level guaranty association. Because the guaranty fund provides limited coverage and not all lines of business are protected. Thus, policyholders have more incentive to identify safer and more reliable insurers to avoid future losses due to the financial failure of insurers with the presence of the guaranty fund. In other words, policyholders are sensitive to opacity even though their policies are protected by the guaranty fund.

For lines of business not protected by the guaranty fund, these are riskier for policyholders. This is because these lines of business have, in general, high risk and are covered by limited insurers. Since policyholders are aware of the risk of not being protected by the guaranty fund, they would be more sensitive to the opacity when insurers face high financial risk. Therefore, policyholders of policies that are not protected by the guaranty fund will be more sensitive to financial risk when opacity changes.

Commercial lines insurance products protect against business-specific operational and liability risks, playing an important role in the risk management strategies of business entities. The complexity of business entities' risk situations arises from the diverse risks and requirements of different business types, necessitating the involvement of brokers and agents. Brokers and agents, who can assess all possible unique insurance situations and possess useful and in-depth knowledge about insurers' financial situations and their products, provide the most accurate and relevant advice to meet clients' unique needs. Consequently, brokers and agents not only ensure that policies fit clients' needs but also verify whether insurers can fulfill their obligations to pay claims to policyholders. Thus, brokers and agents are more likely to recommend products from less opaque insurers, as they can access more information.

I find that policyholders are willing to buy policies from less opaque insurers. The evidence also shows that policyholders are more sensitive to information about insurers' financial risk when they are less opaque. Additionally, policyholders are aware of insufficient protection by the guaranty fund. It further suggests that opacity significantly influences the purchase behavior of commercial lines, due to the involvement of brokers and agents who possess in-depth knowledge of insurers' financial situations and product policies.

The third chapter studies how to detect fraudulent physicians using supervised machine learning algorithms and analyzes the importance of features. Healthcare costs have become a major expenditure in the U.S. since 1980 (Li et al., 2008). They can grow more in the future because of an aging population and advancing health technology. Besides the direct financial losses, frauds also severely hinder the health care system from providing quality services because frauds reduce the funds available to the health care system. Therefore, effective fraud detection is vital in reducing cost and improving the quality of healthcare services. Physicians play the most critical role among service providers because physicians determine the type of treatments (surgery vs. non-surgery) and length of hospital stays. Thus, focusing on physician fraud detection can generate a higher saving potential.

Constructing features and detecting fraudulent physicians from legitimate ones have important implications for insurers: saving costs, speeding up the claim review process, narrowing down the fraud investigation range, and excluding suspicious physicians as external reviewers. Because the goal is to detect fraudulent physicians, the data is reconstructed from the claim level to the physician level. There are 32 features divided into six groups: claim basics, physician characteristics, fraud strategies, early signals, insured characteristics, and agent characteristics.

Imbalance data refers to a dataset within which the number of minority class observations is far less than the majority class. Imbalanced data issue is common in the rare events detection problems such as fraud detection. To address the imbalanced data issue, I choose two data sampling methods: the random under sampling method for RUSBoot model and the class weights method for the neural network model. Random under sampling is used to rebalance the data by randomly removing observations from the majority class. The class weights method assigns weights to classes inversely proportional to their frequencies. A higher class weight means that the model emphasizes this class and penalizes mistakes in it more heavily.

Both error-based and cost-based methods are used to measure the model's performance. AUROC is the error-based method which is used as the indicator for the discriminating power of the classifier. It is the area under the receiver operating characteristic curve which plots the true positive rate against the false positive rate at different decision-making thresholds. AUROC equals 1, indicating the classifier performs perfectly, while 0.5 indicates randomly. The more AUROC is closer to 1, the better the classifier is. The cost-based method is to estimate the cost savings achieved by using machine learning to predict fraudulent claims.

Based on the neural network model, I further analyze the importance of features in detecting fraudulent physicians. Addressing the imbalanced data issue, the AUROC score of the neural network model is 0.781 for physicians with multiple claims. I find the cost savings range from 16.3% to 36.9% assuming the fraud rate of fraudulent physicians' total claim amounts ranging from 30% to 70%. I also find the important features to identify fraudulent physicians are associated with physicians clustering in the eastern area of Taiwan, the percentage of insureds whose age are less than 18, the percentage of surgeries due to illness as opposed to accidents, and whether the physician can perform difficult surgeries. Finally, the evidence implies fraudulent physicians use the "steal a little, all the time" strategy to avoid being caught. Besides cost savings, our study can benefit the life insurer through speeding up the claim review process, narrowing down the investigation range, and excluding suspicious physicians as external reviewers.

# CHAPTER 1: CEO CONSCIENTIOUSNESS AND RESERVE MANAGEMENT: EVIDENCE FROM U.S. PROPERTY-LIABILITY INSURERS

## 1.1. Introduction

This paper examines the relation between CEO consciousness and reserve management for property-liability insurers. A growing literature suggests that managers' attitudes and beliefs, such as confidence, optimism, risk aversion, and ability, significantly impact corporate policies and performance (e.g., Abdel-Khalik, 2007; Peterson et al., 2009; Galasso and Simcoe, 2011). The literature also shows that managers' traits such as MBA degrees, birth cohort, and execution-related abilities (Fast, Aggressive, Persistence, Proactive, Work Ethic, High Standards), and Big Five traits (agreeableness, conscientiousness, extraversion, emotional stability, and openness to experience) have a significant impact on decision-making or job (firm) performance (Bertrand and Schoar, 2003; Almlund et al., 2011; Kaplan et al., 2012; Green et al., 2019).[1,2] For example, Gow et al. (2016) examine the association between Big Five traits and corporate policies (e.g., financing and investment decisions and book-to-market ratio). They find CEO conscientiousness is positively associated with book-to-market and somewhat associated with net leverage. Their evidence also shows that CEO openness influences R&D intensity, investment, book-to-market, and leverage.[3]

The literature, however, has not examined the association between the manager's personality traits and the corporate policies of insurers, with one exception. Berry-Stölzle et al. (2018) find that CEO overconfidence is negatively associated with loss reserves.[4] To fill this gap,

---

[1] Big Five traits are the most widely used as personality proxies in the psychology literature.
[2] Green et al. (2019) find a positive relation between executive extraversion and firm outcomes, indicating that the more extraverted CEOs have better career development and firm outcomes.
[3] Their other results (Table X in their paper) show that agreeableness and extraversion are not associated with firm policies and book-to-market ratio in the models with all control variables and fixed effects.
[4] It should be noted that overconfidence is not one of the big five traits according to the literature.

we examine the relation between CEOs' conscientiousness, one of the Big Five traits, and loss reserve management. Unlike Gow et al. (2016), we mainly focus on conscientiousness (one of the Big Five traits) because we choose (a) specific trait(s) that is (are) associated with reserve estimates.[5] Among the Big Five traits, we choose conscientiousness because its characteristics, such as being painstaking, cautious, and responsible, play an essential role in work-related behaviors (Specht et al., 2011). In addition, Roberts et al. (2009) define conscientiousness as "individual differences in the propensity to follow socially prescribed norms for impulse control, to be goal-directed, planful, able to delay gratification, and to follow norms and rules." This characteristic is important when we examine the impact of the Sarbanes-Oxley Act. Roberts et al. (2009) also indicate that conscientiousness is associated with better economic and workplace outcomes related to firm performance. Finally, unlike managers' attitudes and beliefs, conscientiousness is stable in a person's life span (Specht et al., 2011).[6]

Among corporate policies for insurers, we examine the loss reserve estimates for the following reasons. The loss reserve is the largest liability on the balance sheet for property-liability insurers. Therefore, the reserve estimate is an important corporate policy. Insurance company actuaries recommend an acceptable range for loss reserves, and managers make the final decision on the loss-reserve estimate (e.g., Hsu et al., 2019). This discretion is work-related behavior that requires responsibility. We argue that the loss reserve estimate is likely influenced by personality traits such as managerial conscientiousness. We also suggest that a conscientious CEO not only has discretion about the reserve estimate, but her conscientious management style also influences the actuaries that estimate the reserve estimates. As a result, the reserve estimate of an insurer with a conscientious CEO is expected to be responsible.

[5] Recall that Gow et al. (2016) investigate the relation between all Big Five traits and various corporate policies.
[6] Conscientiousness is also associated with great career success (Judge et al., 1999; Kern and Friedman, 2008).

Additionally, the loss reserve estimation is disclosed and revised every year for the ten years after the initial report. We can examine ex-post whether the original reserve estimate is overstated or understated. Specifically, the difference between the original report and the revised estimation is called reserve error, which reflects the manager's discretion during the initial report period.

High financial risk is a major concern for all types of firms, especially for firms in the financial industry, such as the insurance industry. Since the conscientiousness trait also exhibits characteristics of being more responsible and cautious, it is interesting to explore how conscientious CEOs choose their reserve management policy when an insurer faces high financial risk. We suggest that when an insurer faces higher financial risk, the insurer with a more conscientious CEO is likely to reserve more.

Finally, following the literature (e.g., Ho et al., 2013; Dah et al., 2014; Banerjee et al., 2015), we use the Sarbanes-Oxley Act (SOX) of 2002, an exogenous shock, as an identification strategy for our study. Specifically, we investigate the impact of the passage of SOX on the relationship between CEO conscientiousness and reserve management when insurers have high financial risk. SOX requires the CEO to issue a statement certifying that her/his company's financial statements and disclosures are fairly present in all material respects. Since the psychology literature suggests that conscientiousness indicates a propensity to follow social rules and norms (Roberts et al., 2009), we argue that financial statements certified by more conscientious CEOs are likely to be more responsible after SOX.

We use publicly traded insurers as our sample because the conscientiousness measure can be calculated only for publicly traded insurers. The final sample size is 244 insurer-years (29 insurers) from 2002 to 2015. We employ the quantile regression method for our analysis due to the positively skewed distribution of reserve errors (Grace and Leverty, 2010). In addition, our analysis shows that the reserve estimate decision varies with different levels of reserve estimates.[7] Our results show that the negative relation between CEO conscientiousness and reserve error is significant in the upper tail of the conditional distribution ($75^{th}$, $80^{th}$, $85^{th}$, $90^{th}$, and $95^{th}$) of reserve errors, implying conscientious CEOs tend to reserve less at the upper tail of reserve errors. A possible reason is that conscientious CEOs are more responsible for the stockholders' wealth. To maximize stockholders' wealth, at a high level of reserve estimates, insurers with more conscientious CEOs are likely to reserve less because reserving too much is costly.[8] Reserves are typically composed in liquid assets (such as cash), which have a lower rate of return. To mitigate the endogeneity issue due to the possibility of non-random hiring of conscientious CEOs in firms, we use the propensity score matching (PSM) method to match low conscientious CEOs with high conscientious CEOs and ensure that firms in the treatment (with high conscientious CEOs) and matched (with low conscientious CEOs) groups are similar in observable insurers characteristics. Our baseline results remain robust using the PSM approach. In our additional analysis, we also find that CEOs' conscientiousness prevails over CFOs' conscientiousness in reducing reserve errors.

We also find that the negative relation between conscientiousness and reserve errors in the upper tail is mitigated when insurers face high financial risk, measured by Expected shortfall (ES) and Value at risk (VaR). This evidence shows that conscientious CEOs reserve more when insurers face higher financial risk, which is consistent with consciousness's characteristics, i.e., being responsible and cautious. Furthermore, the evidence suggests the conscientious CEO lowers the reserve error in the upper tail to lower the cost of holding liquid assets after SOX when the financial

---

[7] Please see Section 2.3 for details.
[8] Our summary statistics show higher positive reserve errors at the upper tail.

risk is high. The overall results are consistent with the conscientiousness trait features such as being responsible and rules abiding. Finally, we find that insurers pay higher compensation to more conscientious CEOs.

Our study contributes to the growing literature that explores the relation between CEOs' traits and corporate policies. Specifically, our study is the first to use multivariate analysis to examine the relation between CEO Big five traits (specifically, conscientiousness) and corporate policy (reserve management). Berry-Stölzle et al. (2018) find that CEO overconfidence is negatively associated with loss reserves. But overconfidence is not one of the big five traits. In addition, we use machine learning algorithms to calculate the conscientiousness score based on CEOs' spoken language instead of questionnaires. More importantly, the algorithms that we use do not solely rely on keyword counts to determine whether a CEO is conscientious. Rather, we compute a conscientious score based on the linguistic spoken style rooted in personality (e.g., using filler words, such as like and well) and not related to conversation content. Thus, the CEO cannot use specific keywords related to conscientiousness to fake that s/he is conscientious. We also provide evidence on the interaction effect between conscientiousness and our variables of interest, including financial risk and SOX. Finally, our evidence has an implication for choosing future CEOs. If the board of directors cares about the CEO over reserve too much, then the board should choose a conscientious CEO.

This paper is organized as follows. Section 2 reviews the related literature linking managerial personality to reserve management, discusses conscientiousness and reserve error measures, and proposes hypotheses. Section 3 describes how we measure conscientiousness and reserve errors, the empirical methodology framework, and data sources. Section 4 presents the summary statistics of the sample and empirical results. Section 5 concludes.

## 1.2. Conscientiousness, Reserve Errors, and Hypothesis

### 1.2.1. Big Five Traits and Conscientiousness

The Big Five traits (agreeableness, conscientiousness, extraversion, emotional stability, and openness to experience) framework represents a system description of personalities, which are continuously stable across the life span and can predict the significance of behavioral differences (Barrick and Mount, 1991; Roberts et al. 2009; Specht et al., 2011).

We focus on conscientiousness (one of the Big Five traits), which is considered as the most relevant predictor of job performance in the psychology literature (Barrick and Mount, 1991; Mount et al., 1998; Furnham et al., 1999; Barrick et al. 2001; Specht et al., 2011; Bleidorn et al., 2018). Conscientious people tend to have an orientation to detail and are responsible, and are good at analysis, carefulness, and precision (e.g., Costa and McCrae, 1992; John and Srivastava, 1999). We suggest conscientiousness is related to reserve estimations because CEOs are required to be responsible in finalizing the reserve estimation.

### 1.2.2. Reserve Estimates and Errors

Insurers underwrite the risk in return for the premiums received at the beginning of the policy period, but they do not pay out the losses at the beginning of the policy period. In other words, insurers do not earn the whole premiums when received. Instead, insurers, on average, pay out losses throughout the policy period. There are time gaps between the premiums received and the claims arising, between the claims arising and the loss's payments, and between the loss's payments and the balance sheet date. Insurers set up a reserve to pay for future losses.

Under statutory accounting principles (SAP), insurers estimate the liabilities for the unpaid claim occurring before the balance sheet date. This estimated liability is called loss reserve, which represents the largest liability on a property-liability insurer's balance sheet. Estimating the reserve is challenging because predicting future losses and claims is difficult. While past claims' information can be helpful, past claims cannot precisely predict future claims because the estimate is fraught with uncertainty. After actuaries provide a range of loss reserve estimates, managers make final decisions about the reserve estimates. In other words, the loss reserve is subject to the manager's discretion.

One unique feature of loss reserve is that after the initial estimation, insurers need to revise their loss reserve estimations when new information about the claim arrives. The difference between the original loss reserve and the revised loss reserve is called reserve error which provides an ideal measure of whether the original loss reserve is over-stated or under-stated and reflects the information of the manager's discretion.

### 1.2.3. Hypothesis Development

Among the Big Five personality traits, the conscientiousness trait of a CEO likely offers features that can influence reserve estimates because more conscientious CEOs are likely to be more responsible, cautious, and painstaking than less conscientious CEOs.

#### 1.2.3.1. Reserve Estimate Decisions: A Risk-Taking versus Conservatism Hypothesis

In this section, we discuss two channels that affect CEO reserve estimate decisions: the risk-taking and the conservatism hypothesis.

**Risk-Taking Hypothesis**

Black and Scholes (1973) suggest that a firm's equity can be considered as a European call option. Galai and Masulis (1976) show that the value of the stock is an increasing function of the variance of stock returns (the other factors are the firm's value, the riskless interest rate, and the time to liquidation). Higher firm risk can maximize shareholders' wealth due to their limited liability. As the goal of a CEO is to maximize shareholders' wealth, CEOs may increase firm risk by reserving less to increase the value of equity call options. Based on this discussion, the risk-taking perspective suggests that CEOs tend to reserve less to maximize stockholders' wealth.

**Conservatism Hypothesis**

In general, CEOs of insurers are conservative in reserve estimates, as evidenced by the empirical results of all of the papers examining reserve estimates (Berry-Stölzle et al., 2018; Hsu et al., 2019). The literature shows that the mean of reserve errors is positive. Conservative estimates can lower insolvency risk because more reserve provides more buffer to pay future claims. Therefore, more reserves can protect stakeholders from insolvency. In addition, from the agency cost perspective, lowering insolvency risk can protect CEOs' jobs. If an insurer becomes insolvent, the reputation of the insurer's CEO will be affected. For this reason, CEOs have an incentive to reserve more from the agency cost perspective. Another reason for conservative estimates is that insurers are under state regulations. Regulators' main concern of insurers' operation is the insolvency risk. To address regulators' concerns, insurers typically have conservative reserves, which are generally higher than the minimum required by regulation. In summary, CEOs tend to reserve more than expected future payments. We refer to this type of reserving behavior as the conservatism hypothesis.

**Conflicts between the Risk-Taking Hypothesis and the Conservatism Hypothesis**

There is a conflict regarding reserve management between the conservatism hypothesis and the risk-taking hypothesis. While the risk-taking hypothesis predicts reserving less, the conservatism hypothesis predicts reserving more.

#### 1.2.3.2 Optimal Reserve Estimate Decision

We next discuss the optimal reserve estimate decision. Since reserves are set up for future payments, insurers need to invest these anticipated future payments in liquid assets (e.g., cash), resulting in lower rates of return and/or fewer positive net present value projects. Consequently, the cost associated with additional reserves increases as the reserves increase. On the other hand, the marginal benefits of additional reserves are associated with lower expected insolvency costs (both direct and indirect costs). With additional reserves, the buffer of paying future losses is higher, resulting in a lower insolvency probability and, in turn, lower expected insolvency costs.

When CEOs make decisions on the reserve estimates, they consider both the marginal costs associated with holding liquid assets and the marginal benefits of reducing insolvency costs due to holding additional reserves. Therefore, there is an optimal size of reserve estimates as shown in Figure 1.1.

Figure 1.1 shows a trade-off between the cost of holding liquid assets and the cost of insolvency.

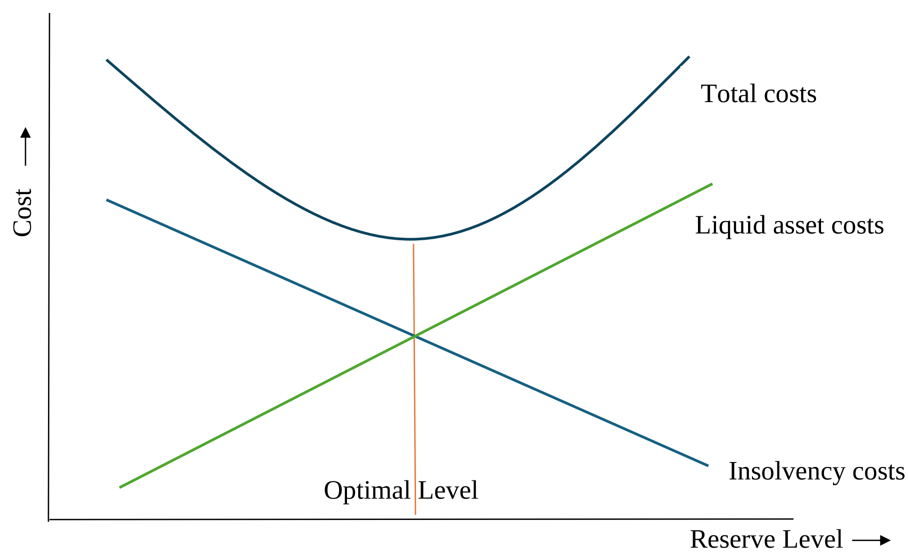



Figure 1.1: Trade-off between the cost of insolvency and the cost of holding liquid assets

On the one hand, when the reserve estimate is lower than the optimal level, the cost of holding liquid assets is less than the expected cost of insolvency. On the other hand, when the reserve estimate is higher than the optimal level, the cost of holding liquid assets is higher than the cost of expected cost of insolvency. At the point of optimal reserve estimate, the cost of holding liquid assets is equal to the cost of insolvency or the total cost of holding liquid assets and insolvency is minimal.

#### 1.2.3.3 CEO Conscientiousness and Reserve Estimates

Based on Figure 1.1, we discuss the relationship between conscientiousness and reserve estimates by two scenarios: when reserve estimates are less than the optimal level and when reserve estimates are higher than the optimal level.

**Reserve Estimates Are Less Than the Optimal Level**

When insurers have a low level of reserve estimates that are less than the optimal level, less conscientious CEOs are likely to have conservative reserves due to concerns about insolvency risk, job security, and regulation, in line with the conservatism hypothesis. More conscientious CEOs are also likely to have more conservative reserves for similar reasons. On the other hand, more conscientious CEO might have the incentive to reserve less as suggested by the risk-taking hypothesis. Based on the above analysis, we infer that, at a low level of reserve estimates, the reserve estimates of insurers with more conscientious CEOs do not differ from those of insurers with less conscientious CEOs.

**Reserve Estimates Are Higher Than the Optimal Level**

We next discuss the relationship between conscientiousness and reserve estimates when reserve estimates are higher than the optimal level (the costs of holding liquid assets are higher than the costs of insolvency). More conscientious CEOs are more responsible for stockholders' wealth than less conscientious CEOs because one important feature of conscientiousness is responsibility. To maximize stockholders' wealth, at a high level of reserve estimates, insurers with more conscientious CEOs are likely to take higher risks by reducing reserve estimates than insurers with less conscientious CEOs. As mentioned above, taking higher risks increases the shareholders' wealth due to the limited liability. It should also be noted that with the high level of reserve estimates, the probability of insolvency is very low, thereby reducing the expected insolvency cost.

Based on the above argument, we suggest that at higher levels of reserve estimates, insurers with more conscientious CEOs would reserve less than insurers with less conscientious CEOs. To develop a testable hypothesis, we use reserve errors as a proxy for reserve estimates because it is difficult to estimate the optimal level of reserve. Reserve errors are defined as the difference between the aggregate loss reserve at time *t* and the reestimated aggregate loss reserve at time *t*+5, then scaled by total admitted assets. Reserve estimation can be verified ex-post because, after the initial estimation, insurers need to revise their loss reserve estimations when new claim information arrives. Thus, reserve error provides an ideal measure of whether the original loss reserve is overestimated or underestimated, reflecting the manager's discretion. Our hypothesis about the relationship between conscientiousness and reserve errors is stated below.

*Hypothesis*: At high levels of reserve errors, conscientiousness is negatively related to reserve errors.

## 1.3. Data and Methodology

### 1.3.1. Data

Our initial sample consists of all publicly traded property-liability insurers of which the CEOs' spoken language is available from the question and answer (Q&A) portion of conference call transcripts from 2001 to 2018. We obtain conference call transcripts from LexisNexis. The reserve error and other financial data are from the National Association of Insurance Commissioners (NAIC). CEO characteristics are from Execucomp. Firm risk variables such as Value at risk, Expected shortfall, and Distance to default are calculated from CRSP. Based on the data availability, our final sample contains 244 insurer-year observations (29 unique insurers) from 2002 to 2015.[9]

#### 1.3.1.1. Big Five Traits and Conscientiousness Measure

The section provides the measurement of conscientiousness of CEOs. The Big Five traits (e.g., conscientiousness) can be measured by the language used because people differ in their talking styles (Allport and Odbert, 1936; Pennebaker and King, 1999; Mehl et al., 2006; Mairesse et al., 2007; Gow et al., 2016). Pennebaker and King (1999) also find that the Big Five traits are highly correlated with linguistic features. Specifically, linguistic features such as using sentimental words, verb tense, causal words, words per sentence, and speech rate reflect personality traits (Pennebaker and King, 1999; Pennebaker et al., 2001; Mehl et al., 2001; Mehl et al., 2006; Mairesse et al., 2007; Gow et al., 2016).

Mairesse et al. (2007) develop four well-performed algorithms for scoring Big Five traits using continuous scales. In each algorithm, they use 88 Linguistic Inquiry and Word Count (LIWC) features (Pennebaker et al., 2001) and 14 Machine Readable Cataloguing (MRC) features (Coltheart, 1981) to train the model to get the traits scores.[10] Furthermore, they confirm that conscientiousness can be well measured from spoken language.[11] Specifically, conscientiousness is negatively related to discrepancies words (e.g., should, would, and could), exclusive words (e.g.,

---

[9] We ended our sample period in 2015 because we need to have 5-year window to calculate reserve errors. For the initial sample, there are 42 unique insurers with conscientiousness and reserve errors data. There are 33 unique insurers after combining a set of control variables. Less firms are missing control variables; there are 29 unique insurers in the final sample to run the regression.

[10] The list of features can be found in Table 6 of Mairesse et al. (2007). Features mean independent variables used in the training algorithm and the dependent variable is individual personality trait score.

[11] They claim that the main feature of conscientiousness is avoidance of using negative emotion words (e.g., fear, anger, depression, sadness). The other features of conscientiousness are described below. Conscientious people talk more about job and occupation, which are defined as content related to personal concerns in LIWC. They prefer to use longer words (e.g., words longer than six letters, number of syllables in the word), words related to communication (e.g., talk, listen, share), insight words capturing the sense of understanding or learning (e.g., think, know, consider), words acquired late by children, prompts (e.g., yeah, OK, huh), positive emotion words (e.g., happy, love, nice). They use fewer swear words and fewer pronouns (e.g., I, them, itself).

exclusive, but, and without), negations (e.g., no, not, and can't), causation words (e.g., because, reason, and why), and positively related to positive emotion words (e.g., happy and nice) (Pennebaker and King, 1999).[12]

In linguistics, different word categories (e.g., filler words, longer words, insight words, discrepancies words, exclusive words, and causation words) are used in verbal communication. For example, spoken language uses some common filler words (e.g., er, ah, you know, like, and well). It should be noted that these word categories are not related to the content of conversations, rather, word categories are associated with personality traits.[13] Importantly, the language style is hard to conceal because it is naturally revealed in the conversation, and it is difficult to change the deeply rooted language style.

We download quarterly conference call transcripts for our sample insurers from Lexis Nexis from 2001–2018. We then automate an algorithm in R language to identify the CEO's spoken responses in the conference calls. We only keep CEOs' responses to the question and answer (Q&A) section of the calls because managerial responses during Q&A are likely to be less scripted (Hollander et al., 2010). To measure CEOs' conscientiousness level, we feed the CEOs' responses from the Q&A section to the well-trained linguistic algorithms developed by Mairesse et al. (2007).[14] For each CEO and a conference call, the conscientiousness trait scores are generated using the four linguistic algorithms provided by *Personality Recognizer* application, and then these scores are winsorized at 1st and 99th percentiles.[15] We average the scores from the four algorithms

---

[12] Conscientiousness is also negatively related to using swear words, negative emotion words and positively related to using longer words, insight words (e.g., realize, understand), and filler words (e.g., like, well) (Mehl et al., 2006).
[13] For example, filler words are used to calculate personality traits scores but not related to conversation content such as bright and wonderful.
[14] We use a Java command-line application *Personality Recognizer* that reads text information and estimates Big Five personality scores which are based on models analyzed in Mairesse et al. (2007). Appendix A describes the method of Mairesse et al. (2007) in detail.
[15] *Personality Recognizer* application estimates Big Five personality scores based on four different models: Linear Regression, M5' Model Tree, M5' Regression Tree, and Support Vector Machine with Linear Kernel.

to get the call-level score. At this point, we have call-level CEO conscientiousness scores; for example, if a CEO is in the firm for the past 20 quarters in our sample period attending conference calls, then we have 20 conscientiousness scores for this CEO. Some fundamental characteristics of firms around the call date and the seasonality may impact CEOs' responses during a conference call; therefore, we follow Green et al. (2019) to develop our CEO conscientiousness score.[16]

We estimate the following OLS regression to extract the variation in CEO conscientiousness affected by firm fundamentals and obtain the residual call-level conscientiousness score.

$$Call\ Conscientiousness = \beta_1 Ret_{t-63,t-2} + \beta_2 Ret_{t-1,t+1} + \beta_3 Ret_{t+2,t+63} + \beta_4 Earnings\ Call$$
$$+\beta_5 Loss + Qtr\ fixed\ effect + \varepsilon.$$

In this specification, $Ret_{t-63,t-2}$ is the stock returns in the previous quarter, $Ret_{t-1,t+1}$ is the 2-day returns around the call date, $Ret_{t+2,t+63}$ is the returns over the following quarter, *Earnings Call* is an indicator variable which is set equal to one if the conference call date occurred around the four-day window [–1,2] of the earnings announcement date (day 0), and *Loss* is an indicator variable which is set equal to one if the latest quarter before the conference call reports negative earnings. To generate a time-invariant conscientiousness measure for the CEO, we take a weighted average of all the residual call-level scores by the number of words spoken by the CEO in the Q&A section of each call. Finally, we assign this weighted average CEO-level conscientiousness measure to all the data points related to the CEO (and for the firm) to treat it as a time-invariant CEO fixed effect.

The following are sentence examples of unconscientious and conscientious people from our conference calls data sample.

[16] Green et al. (2019) estimate executive extraversion score. We do not include Meet-or-Beat and Surprise variables from IBES in our regression due to data limitations for our insurance firms.

| Unconscientious | Conscientious |
| --- | --- |
| - The first were underwriting margins were unacceptable in commercial.<br>- We don't see a solution for that right now.<br>- I'm not going to say what we are going to do.<br>- That does not say, though, when I look at my core businesses, we can't get another point out of our core businesses. We obviously can't. | - I think that as we've indicated, the $25 billion goal is achievable with all of the actions that we've laid out.<br>- We are watching very carefully the appeal process.<br>- We are taking specific steps to improve that.<br>- Additionally, as you'd expect, we conducted a variety of detailed analyses to see if there were any other unique causes to the pattern, we saw spike and we found none. |

#### 1.3.1.2. Reserve Error Measure

Property-liability insurers are required to disclose the initial reserve estimates and revised reserve estimates every year for ten years after the initial disclosure in Schedule P of the National Association of Insurance Commissions' (NAIC) annual statutory filing. This regulation allows us to compare the revised reserve estimate and the original reserve estimate to determine whether the original loss reserve is overstated or understated. The difference between the original estimated reserve and the revised reserve estimation is reserve error.

We follow the measure of Anderson (1971), which is widely cited by the research regarding reserve error (e.g., Petroni, 1992; Petroni and Beasley, 1996; Gaver and Paterson, 2004; Hsu et al., 2019) to calculate reserve errors. Barth and Eckles (2018) point out that the calendar year development approach is more appropriate to measure reserve error in terms of solvency problems. The calendar year development measure is an aggregate concept that measures the difference between the aggregate loss reserve at time $t$ and the reestimated aggregate loss reserve at time $t + n$. The sign of the difference represents whether the aggregate loss reserve is overstated or understated at time $t$. According to the literature, a five-year window ($n = 5$) is appropriate to calculate reserve errors. Following the literature, the reserve error is scaled by total admitted assets (*TAssets*). The reserve error is defined as follows:

$$RESERROR_{i,t} = \left(Cumulative\ inccured\ loss_{i,t} - Cumulative\ inccured\ loss_{i,t+5}\right)/TAssets_{i,t}$$

A positive sign of reserve error, *RESERROR*, means that the initial estimation of loss reserve at time $t$ is greater than the reestimated loss reserve at time $t + 5$ , indicating the insurer overstated the loss reserve at time $t$.

### 1.3.2. Methodology

We find that the distribution of reserve errors is non-normal and positively skewed (Panels B–D of Table 1.1), indicating that the OLS approach, which assumes the distribution is normal and estimates the mean effect, is inappropriate. Instead, we use the quantile regression approach to address the non-normal and skewed distribution concerns. Please note that the incentive of estimating loss reserve may differ across different levels of reserve errors. At the median or lower quintile of reserve errors, managers have the incentive to be more conservative and reserve more, but at a very high quintile of positive reserve errors, managers may want to decrease reserve errors because reserving too much is costly. The quantile regression approach can measure the change in incentives across different quantile levels of reserve errors. In addition, the quantile regression is less sensitive to the distribution of the dependent variable and outliers, thus, helping us better understand conscientiousness's impact across different quantile levels of reserve errors.

The quantiles of the conditional distribution of the dependent variable are expressed as functions of independent variables (Koenker and Hallock, 2001). The quantile regression is based on minimizing the sum of asymmetric weighted absolute residuals to estimate the conditional quantile functions, providing a much more complete picture of the heterogeneity response of independent variables than would be offered by conditional mean models such as OLS (Koenker, 2005).

Our baseline quantile regression specification for the effect of CEO conscientiousness on reserve errors is as follows:[17]

$$Q_t\left(RESERROR_{i,t}|CONSC_i, Controls_{i,t}\right) \tag{1}$$
$$= \alpha_\tau + \beta_\tau CONSC_i + \lambda_\tau CEO\ Controls_{i,t-1} + \gamma_\tau Firm\ Controls_{i,t} + Year\ FE + e_{i,t},$$

where $RESERROR$ represents the reserve error, $CONSC$ represents CEO conscientiousness, $\beta_\tau$ represents the coefficient of conscientiousness, and $\lambda_\tau$ represents coefficients of control variables, all at $\tau^{\text{th}}$ percentile.

We include two types of control variables: CEO characteristics variables and firm characteristics variables, and year-fixed effects.[18] For CEO characteristics control variables, we include CEO vega and CEO delta to control for managers' risk-taking incentives (Coles et al., 2006). We also include CEO age and CEO tenure. Coles et al. (2006) suggest that CEO tenure is negatively related to firm risk and is used as a proxy for the level of risk aversion. Serfling (2014) argues that older CEOs adopt a less risky firm policy. Therefore, older CEOs are likely to estimate the reserve more cautiously.

We control for various firm characteristics that are likely to affect the reserve error. We use the natural log of net premium written (*LNNPW*), which can control the effect of risk pooling, as a proxy for firm size. A higher net premium growth rate (*NPWGROWTH*) may lead to higher income fluctuation, so insurers will reserve more to prepare for future loss claim uncertainties. Grace and Leverty (2012) point out that insurers manage reserve estimation for tax purposes because increasing the reserve estimation will reduce the current liability. Increasing the reserve estimation can postpone the tax payment until claims are ultimately resolved. Overestimating the

[17] The objective function of quantile regression is expressed as follows:

$$Q_N(\beta_\tau) = \min_{\beta_\tau} \sum_{i:y_i \geq x_i\beta_\tau}^{N} \tau|y_i - x_i\beta_\tau| + \sum_{i:y_i < x_i\beta_\tau}^{N} (1-\tau)|y_i - x_i\beta_\tau|.$$

[18] We do not include firm fixed effects as they would subsume variation in time-invariant conscientiousness measure.

loss reserve reduces the taxable income. Grace (1990) uses the tax shield to measure the incentive to overestimate loss reserve. The tax shield (*TAXSHIELD*) is calculated as net income plus estimated reserve divided by total assets. We use the natural logarithm of Tobin's Q (*LNQ*) to control for insurers' growth opportunities. Insurers with higher Tobin's Q, representing higher growth opportunities, would be more conservative in estimating reserves because they need to keep business operations steady and be prudent in supporting business expansion (Cummins et al., 2006).

According to Grace (1990), insurers are incentivized to smooth income for regulation concerns. Regulators are concerned about the high fluctuation of surplus from one year to the next. In addition, income stability is an indicator of firm risk. Thus, insurers may smooth income by estimating reserves. We use the previous 3 years' average ROA (*SMOOTH*) to measure income smoothing (Grace 1990).

Harrington and Danzon (1994) find that weak insurers mask the financial situation by underserving through reinsurance. Therefore, we control for reinsurance ceded to affiliated reinsurers (*REAFFILIATE*) and reinsurance ceded to nonaffiliated reinsurers (*RENONAFFILIATE*). We also control the loss ratio growth (*LRGROWTH*). A high loss ratio growth implies underwriting uncertainty, which impacts the reserve estimation.

The literature demonstrates that there is more uncertainty for long-tailed lines of business, which need more reserve discretion, resulting in overestimating loss reserves (Petroni and Beasley, 1996; Phillips et al., 1998; Beaver et al., 2003). Therefore, we control the percentage of the net loss incurred in long-tailed lines of business over the net loss incurred in whole business lines

(*LONGTAIL*).[19] We also control product diversification (*PRODHHI*) and geographical diversification (*GEOHHI*), which are calculated using Herfindahl Index.

Petroni (1992) and Gaver and Paterson (2004) suggest that weak insurers tend to under reserve to mask financial conditions to appear more solvent. We use an indicator variable, *WEAK*, to represent insurers' financial condition. Insurance regulators use IRIS ratios to analyze insurers' financial conditions and target those needing regulation attention. *WEAK* takes a value of 1 if the insurer has more than 3 out of the range IRIS ratios and 0 otherwise. In addition, we use the natural log of naive distance-to-default (*LNDD*), which is calculated following Bharath and Shumway (2008), to measure the default risk of the insurer. The default risk decreases as the distance-to-default increases. Appendix B provides the definitions of all the variables used in this study.

## 1.4. Summary Statistics and Empirical Results

### 1.4.1. Summary Statistics

Table 1.1 presents the summary statistics of the variables for the entire sample. The loss reserve error is scaled by the total admitted assets (*RESERROR*). The mean (median) reserve error is 0.009 (0.020), indicating that, on average, property-liability insurers overstate their loss reserves, which is consistent with the findings of the literature. The mean (median) of CEO conscientiousness score (*CONSC*) is -0.052 (-0.060). The average insurer has a 7.2% net premium growth rate (*NPWGROWTH*), a 4.2% three-year average ROA (*SMOOTH*), 0.8% loss ratio growth rate (*LRGROWTH*), and 71.3% loss incurred from the long-tail business lines (*LONGTAIL*). The minimum of reinsurance ceded to nonaffiliated reinsurers (*RENONAFFILIATE*) is greater than zero, indicating that all insurers in this sample transfer a portion of the insurance business to nonaffiliated reinsurers to diversify underwriting risk. The median of reinsurance ceded to affiliate

[19] Long-tailed lines of business are defined by Phillips et al. (1998).

reinsurers (*REAFFILIATE*) is zero, indicating that at least half of the insurers do not transfer underwriting risk to affiliated reinsurers. The average insurer has a product line Herfindahl Index (*PRODHHI*) of 0.361 and a geographical Herfindahl Index (*GEOHHI*) of 0.128, indicating that the insurer, on average, has approximately 3 business lines and operates in 8 states. The 75$^{th}$ quantile of *WEAK* is 0, representing that very few insurers have more than 3 unusual IRIS ratios.

### 1.4.2. CEO Conscientiousness and Reserve Error Baseline Result

Table 1.2 presents the results of the relation between CEO conscientiousness and reserve errors. In column (1), the coefficient on *CONSC* is insignificant. The OLS result shows no significant relation between CEO conscientiousness and reserve errors. One possible reason is that the OLS method focuses on the condition mean effect, which cannot capture the heterogeneous relation at different levels of reserve errors. A positive coefficient of conscientiousness indicates insurers reserve more, and a negative coefficient means reserve less.

Table 1.2 shows the coefficient of conscientiousness (*CONSC*) is significantly negative for the 75$^{th}$ quantile and higher (80$^{th}$, 85$^{th}$, 90$^{th}$, and 95$^{th}$), indicating that insurers with more conscientious CEOs reserve less than those with less conscientious CEOs.[20] One possible reason is that conscientious CEOs lower reserve errors to be more responsible for stockholders' wealth instead of conserveness at higher levels of reserve errors so that insurers do not over-reserve too much.[21] Because one important feature of conscientiousness is being responsible. In other words, at higher levels of reserve estimates, more conscientious CEOs decrease reserve errors to lower the cost of excess reserve. While over reserving can lower the probability of financial distress, there are disadvantages to over reserving. Holding excess reserves has opportunity costs.

[20] Panel A of Figure 1.2 demonstrates point estimates of the coefficients on *CONSC* from Table 1.2.
[21] Panel D of Table 1.1 shows that the median of reserve errors is positive.

Specifically, with excess loss reserve, insurers have less free cash flows to invest in positive NPV projects (financial or real assets). In other words, while conservatism in reserve estimates is important, reserving too much is not optimal. For economic significance, we find one standard deviation increase in conscientiousness will decrease the reserve by 1.317% of total assets (based on the result of $75^{th}$ quantile), which is about 2 million dollars on average. It should also be noted that even with the decrease in reserve estimation, insurers still over-reserve, indicating that risk-taking behavior does not compromise financial stability.

We next discuss the results of control variables. The natural log of net premium written (*LNNPW*) is negatively and significantly related to reserve errors at all quantile levels, implying larger insurers are less conservative in terms of reserve estimates. The coefficients of the natural log of Tobin's Q (*LNQ*) are positive and significant for most of the quantiles, indicating that CEOs of insurers with relatively stronger growth opportunities are more cautious and adopt a more conservative reserve policy to ensure solvency during business expansion. The reinsurance ceded to nonaffiliated reinsurers (*RENONAFFILIATE*) is negatively related to reserve errors, and the effect is significant in the upper tail of the conditional distribution ($75^{th}$, $80^{th}$, $85^{th}$, $90^{th}$, and $95^{th}$), implying that higher over reserved insurers transfer less underwritten risk to nonaffiliated reinsurers to save reinsurance costs. The percentage of the net loss incurred in long-tailed lines of business (*LONGTAIL*) are positively related to reserve error and significant at higher quantile level ($75^{th}$, $80^{th}$, $85^{th}$, $90^{th}$, and $95^{th}$), suggesting that insurers with high losses incurred from long-tail business lines have more conservative reserve estimations. One potential explanation for this result is that insurers with high losses incurred from long-tail business lines reserve more since the insurer needs to be able to pay future losses to hedge high uncertain losses. The estimated coefficients of the geographical Herfindahl Index (*GEOHHI*) are negatively related to reserve errors and significant at all quantile levels, suggesting that insurers operating in more states reserve more.[22]

The literature examines the relation between extraversion (one of the Big Five personality traits) and various corporate policies (e.g., Green et al., 2019; Lartey et al., 2020; Adebambo et al., 2022). The characteristics of extraversion include being talkative, energetic, and outgoing. Extraverts like to be a leader and often are the first to offer their opinion and suggestions. It is reasonable to suggest that the characteristics of extroverts are not relevant to reserve estimates. We thus use extraversion to perform a placebo test, replacing the CEO conscientiousness measure with CEO extraversion measure and rerunning our baseline specification. The results are in Table C.1 in Appendix C. The procedure we follow to generate CEOs' extraversion score (*EXTRA*) is similar to the procedure to form our conscientiousness score in Section 2.2. The mean (median) of *EXTRA* is -0.036 (0.120). The results show that the coefficients on CEO extraversion are generally insignificant, indicating extraversion personality traits are not associated with reserve errors. The evidence is consistent with our expectations.

#### 1.4.2.1 Propensity Score Matching

While we cannot completely rule out the possibility that our results in the previous section suffer from endogeneity issues. Roberts and Whited (2013) pointed out that the matching technique can alleviate asymptotic biases ascending from endogeneity or self-selection. Therefore, to mitigate self-selection-based endogeneity in our data, we use the widely known propensity score matching (PSM) technique (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 2002; Shipman et al., 2017).

[22] The negative sign of geographical Herfindahl Index is not consistent with the literature. We find that geographical Herfindahl Index is highly correlated with *LNQ*. We run the same regression dropping *LNQ* and find geographical Herfindahl is positively associated with reserve errors. This evidence is consistent with the literature.

To implement PSM, we form tercile groups of CEO conscientiousness score (*CONSC*) each year and define high conscientiousness, *HIGHCONSC*, as a dummy variable equal to one if a CEO's conscientiousness score falls under the top tercile group, otherwise set to zero. We report the results related to our PSM procedure in Table 1.3. In Panel A, we run logistic regression using *HIGHCONSC* as a dependent variable and on all the control variables from equation (1). We then estimate propensity scores as the predicted probabilities using the coefficients from this regression. CEOs in the high conscientiousness group (i.e., *HIGHCONSC* = 1) represent our treatment group. For each observation in our treatment group, we matched a sample from the low conscientiousness group (i.e., *HIGHCONSC* = 0) using the estimated propensity scores based on the nearest-neighbor matching approach with replacement and a caliper of 5%. Panel B reports the covariate balancing after the matching procedure. In addition, Panel B reports means and medians of all the covariates for the treated group (i.e., high conscientious CEOs) and the matched group (i.e., PSM matched group from low conscientious CEOs). We also report mean and median differences between these two groups for each covariate. As seen in Table 1.3, none of the means are different between these two groups, and the medians are also almost similar. This analysis ensures that the treated and the matched groups are statistically similar across all covariates except the dependent variable of interest, reserve error.

Using this PSM sample, we re-run our baseline specification in equation (1) and report the results in Panel C. In column (1), the OLS result shows the coefficient on *HIGHCONSC* is negative and statistically significant at 1% level. Additionally, using quantile regressions, this coefficient is significantly negative for the $90^{th}$ (at 10% level) and for the $95^{th}$ quantiles (at 1% level). These results confirm our baseline results using the PSM method and suggest that conscientious CEOs reserve less.

### 1.4.2.2 CEO versus CFO Conscientiousness and Reserve Error

Recent literature documents that the incentives of CFOs could be more dominant than those of CEOs for setting a firm's financial reporting and investment policies (Chava and Purnanandam, 2010; Jiang et al., 2010; Kim et al., 2011). This section assesses the differential impact of CFOs' versus CEOs' conscientiousness on the reserve error in insurance firms.

We report the results of CFO conscientiousness on the reserve error analysis in Table C.2 in Appendix C. Panel A shows the summary statistics of CFOs' conscientiousness measure, *CFOCONSC*, and the other CFO variables. The mean (median) of the conscientiousness measure for CFOs is higher than those of CEOs. In Panel B, we regress reserve error on CFO conscientiousness, controlling for CFO characteristic variables and firm controls. Examining CFOs conscientiousness solely, the coefficients on *CFOCONSC* are negative and statistically significant in the upper tail of the conditional distribution (75th and onwards). The evidence suggests that more conscientious CFOs reserve less, similar to the evidence for conscientious CEOs. However, when we include the CEO conscientiousness measure along with the CFO conscientiousness in Panel C, we find that the coefficients on CFOs conscientiousness (*CFOCONSC*) become insignificant with one exception, but the coefficients on CEOs conscientiousness (*CONSC*) remain significant.[23] The overall evidence implies that CEOs' conscientiousness prevails over CFOs' conscientiousness in deciding on the reserve estimate.

### 1.4.3. Channel of Conscientiousness and Reserve Errors

This section identifies the channel through which CEO conscientiousness affects reserve error. Specifically, we argue that insurers with more conscientious CEOs are likely to reserve more

[23] The coefficient of *CFOCONSC* is negative and statistically significant at the 95$^{th}$ percentile of reserve errors, implying CFOs have influence on the reserve error at the very high level of reserve errors.

than insurers with less conscientious CEOs when insurers have higher insolvency risk. The reason is that conscientious CEOs are more responsible for the insurers' financial health than less conscientious CEOs because being responsible and cautious are also the major characteristics of conscientiousness. In other words, more reserves can protect stakeholders from insolvency, especially when financial risk is high. Following the literature (e.g., Milidonis et al., 2019), we use Expected shortfall (ES) and Value at risk (VaR) at various confidence intervals to proxy the financial risk. Expected shortfall (ES) is defined as the conditional expected loss using 1 year of daily stock returns. Value at risk (VaR) is defined as the maximum expected loss that could occur using 1 year of daily stock returns at a specified confidence level (Milidonis et al., 2019).[24] It should be noted that ES contains more information than VaR, and the value of ES is beyond VaR. In addition, the 99.5 percent confidence level is consistent with the solvency capital requirement (Milidonis et al., 2019).

Table 1.4 shows the results of the impact of financial risks. In Panel A, coefficients of the interaction term between CEO conscientiousness and Expected shortfall with confidence levels of 99.5 (*CONSC×ES99.5*) are positive and significant in the upper tail of the conditional distribution ($75^{th}$, $80^{th}$, $85^{th}$, $90^{th}$, and $95^{th}$), implying that the negative relation between CEO conscientiousness and reserve errors is inverted to a positive relation when insurers face higher financial risk. For the upper tail of over-reserved insurers, conscientious CEOs pursue financial stability to avoid insolvency by reserving more when the financial risk is high. One possible explanation is that

---

[24] Both ES and VaR are based on stock price which reflects the value of the firm. The total value of a firm can be calculated by summing its equity value and debt value. The market value of a firm's debt is not directly observable; however, it can be estimated by equity value under the Merton DD model. Furthermore, what happens in the income statement is reflected in the stock price.
The equation of VaR is $VAR = \overline{R} + \sqrt{\sigma} z_c$. The equation of ES is. $ES = \overline{R} - \sqrt{\sigma}.\frac{1}{c}.\emptyset(z_c)$. $\overline{R}$ is the mean of 1 year of daily firm stock returns. σ is the variance of 1 year of daily firm stock returns. $z_c$ is the c-quantile of the standard normal distribution. $\emptyset$ is the density function.

conscientiousness's responsibility feature makes the CEO take a more conservative reserve policy when the financial risk is high. Panel B, Table 1.4 shows that the interaction term between CEO conscientiousness and VaR with a confidence level of 99.5 (*CONSC×VAR99.5*) is significantly positive in the upper tail of the conditional distribution (75th, 80th, 85th, 90th, and 95th).[25] The results of the interaction term between CEO conscientiousness and Expected shortfall (or Value at risk) with different confidence levels are qualitatively similar and presented in Appendix C (Table C.3).

### 1.4.4. Identification Strategy

We use the Sarbanes-Oxley Act (SOX) of 2002, an exogenous shock, as an identification strategy for our study. SOX requires CEOs to be responsible for the financial statements of their firms. For example, CEOs are required to certify the financial statement information according to Section 302 in SOX. Additionally, SOX increases penalties for violations of security acts in Sections 304, 807, 902, and 903. The passage of SOX may lead to some changes in CEOs' behaviors because it increases the liabilities of CEOs. For example, CEOs bear a higher risk of misreporting financial statement information and broader financial reporting responsibilities after SOX. It is well documented in psychology literature that conscientiousness is indicative of a propensity to follow social rules and norms (e.g., John et al., 2008; Roberts et al., 2009). We suggest that the increased legal exposure has a pronounced impact on conscientious CEOs, and they are more likely to follow the requirements of SOX because they tend to follow the rules. In other words, conscientious CEOs would be more responsible for the financial statement after SOX.

To examine the impact of SOX, we use *POSTSOX* as a dummy variable equals 0 if observations are during the implementation period of SOX (2002–2004) and 1 for 2005–2015. We follow Ho et al. (2013) to use a two-year lag because it takes time to revise the reserve policy.

---

[25] Panel B of Figure 1.2 shows point estimates of the coefficients on *CONSC×ES99.5* and *CONSC×VAR99.5*.

Table 1.5 presents the results of the impact of SOX on the relation between CEO conscientiousness and reserve error conditioning on financial risk. In Panel A (Panel B), we use Expected shortfall (Value at risk) with confidence levels of 99.5, *ES99.5* (*VAR99.5*), as a measure of financial risk. In Panel A, the coefficient of the three-way interaction term (*CONSC*×*POSTSOX*×*ES99.5*) is significantly negative, implying that more conscientious CEOs reserve less than less conscientious CEOs after SOX (compared with before SOX) when insurers face higher financial risk, possibly because they are more responsible for financial statements.

At higher levels of reserve errors, the cost of insolvency risk is relatively low. Facing insolvency risk, insurers with more conscientious CEOs reserve less than insurers with less conscientious CEOs to reduce the cost of holding liquid assets post SOX because more conscientious CEOs are more responsible for the financial statement and stockholders' wealth.

At lower levels of reserve errors, the marginal cost of insolvency is higher than the marginal cost of holding liquid assets, thus, more conscientious CEOs are likely to reserve more than less conscientious CEOs before SOX because more conscientious CEOs are more responsible. Post SOX, regardless of conscientiousness, CEOs need to reserve more to abide by the requirements of SOX. However, more conscientious CEOs would increase reserves relatively less than less conscientious CEOs post SOX because before SOX, more conscientious CEOs reserve relatively more than less conscientious CEOs. In other words, less conscientious CEOs are influenced more by SOX than more conscientious CEOs. Thus, after SOX, insurers with less conscientious CEOs would increase reserve more than more conscientious CEOs to abide by the requirements of SOX, indicating the coefficients of the interaction term (CONSC×POSTSOX×ES99.5) is significantly negative in the lower tail of reserve errors.[26]

---

[26] Here, we elaborate this concept with a numerical example. Consider a case where more conscientious CEOs initially under-reserve by $7,000,000 (reserve errors); after SOX, they revise their reserve estimates but still under-

For robustness, we also use VaR as a proxy for financial risk. The result of VaR is qualitatively similar, as shown in Panel B.[27] Overall, this evidence is consistent with one feature of conscientiousness: following the rules and norms.

### 1.4.5. Conscientiousness and Compensation

In this section, we investigate whether the value of conscientiousness of CEOs is rewarded through the compensation in the property-liability industry. Jung and Subramanian (2017) argue that CEOs get compensation for their talent and effort. In addition, higher pay is a reward for a CEO's unobservable talent, successful management, and firm performance (Barrick and Mount, 1991; Mount et al., 1998; Furnham et al., 1999; Barrick et al., 2001; Specht et al., 2011; Albuquerque et al., 2013; Bleidorn et al., 2018). The literature also shows that compensation and managerial style are influenced by the managers' latent traits, such as personality (Graham et al., 2012; Graham et al., 2013). The more narcissistic CEOs get higher compensation than the rest of the top management team (O'Reilly et al., 2014). Since pursuing accuracy and being responsible for reserve estimations are crucial for insurers' financial health, we suggest that conscientious CEOs are positively associated with compensation.

We use an OLS regression specification to examine the relationship between CEO conscientiousness and her compensation. We use CEO total compensation, *TDC1* variable in Execucomp, as the dependent variable in our model. The distribution of the natural log of total compensation is normal and not skewed, indicating that the OLS approach is appropriate for

reserve by $3,000,000 (reserve errors), thus increasing their reserve estimates by $4,000,000. In contrast, less conscientious CEOs, initially under-reserves by $10,000,000 (reserve errors), after SOX, under reserve by $5,000,000 (reserve errors), reflecting an increase of $5,000,000 in their reserve estimates. Consequently, more conscientious CEOs have a smaller increase in the reserve estimation compared to less conscientious CEOs, indicating a negative relationship in the lower tail of reserve errors.

[27] Additionally, the results using Expected shortfall and Value at risk with different confidence levels are qualitatively similar and presented in Appendix C (Table C.5).

depicting the relation between CEO conscientiousness and compensation. Due to data availability, the sample with CEO compensation reduces to 224 insurer-year observations.

We include CEO and firm controls in our model specification. For CEO characteristics-related variables, we include CEO vega and CEO delta. Guo et al. (2021) argue that "similar compensation levels do not mean equal compensation if compensation risk differs." Following the literature (e.g., Bebchuk and Fried, 2005; Yim, 2013), we also include CEO age, CEO tenure, and a dummy variable *CHAIRMAN* to control whether the CEO is also the board chairman. The firm controls are defined as follows. The natural log of net premium written (*LNNPW*) is a proxy for firm size. The business of larger insurers tends to be more complicated, and larger insurers tend to pay more. Profitability (*ROA*) is used to control firm performance. CEOs get higher compensation for better firm performance. Insurers can diversify underwriting risk through reinsurance to lower uncertainty. Thus, we control for reinsurance ratios (*RERATIO*). Shareholders encourage CEOs to bear risk. If a CEO invests a greater portion in low-risk projects, the CEO gets a lower reward. We use the tax-exempt ratio (*TAXEXEMPT*), which is measured by tax-exempt income divided by total investment income, to capture low-risk investment (D'Arcy and Garven, 1990). Firm risk affects CEO compensation (Core et al., 1999; Chang et al., 2016). We use five-year rolling data to calculate the standard deviation of ROA (*STDROA*), the standard deviation of ROI (*STDROI*), and the standard deviation of loss ratio (*STDLOSSRATIO*) to represent total risk, investment risk, and underwriting risk, respectively (Ho et al., 2013).

Table 1.6 presents the result of the relation between CEO conscientiousness and her compensation. The coefficient of CEO conscientiousness (*CONSC*) is positive and significant at the 5% level, implying that the conscientiousness trait is compensated by insurers. The evidence implies being responsible and following rules are rewarded by property-liability insurers.

Furthermore, we use the 2008 financial crisis as an external shock to examine how a financial crisis impacts the relation between CEO conscientiousness and compensation. *CRISIS* is a dummy variable that equals 1 if fiscal year observations are in 2007–2009 and 0 otherwise. Table 1.7 shows the coefficient of the interaction term between CEO conscientiousness and financial crisis (*CONSC*×*CRISIS*) is positive and significant at the 10% level, implying that more conscientious CEOs received higher compensation during the financial crisis. This result somewhat supports that insurers reward more managerial conscientiousness trait during the financial crisis.

## 1.5. Conclusion

This paper investigates the relation between CEO conscientiousness and reserve management in U.S. property-liability insurers. Our baseline results show that CEO conscientiousness is negatively associated with reserve error in the upper tail of the conditional distribution (i.e., at $75^{th}$ percentile and higher), indicating insurers with more conscientious CEOs reserve less than insurers with less conscientious CEOs at a higher level of reserve errors to lower the cost of excess reserve rather than conservatism when reserve errors are extremely conservative.

We also find that CEOs become more conservative when their insurers have higher financial risk. Furthermore, insurers with more conscientious CEOs reserve less than less conscientious CEOs after SOX (compared with before SOX) when insurers face higher financial risk, possibly because they are more responsible for financial statements. This evidence is consistent with one feature of conscientiousness: following the rules and norms. Finally, conscientious CEOs get higher compensation, suggesting that the conscientiousness trait is rewarded in the property-liability industry. The overall results of this paper are consistent with the features of conscientiousness: being responsible and following the rules.

REFERENCES


Abdel-Khalik, A.R. 2007. An empirical analysis of CEO risk aversion and the propensity to SMOOTH earnings volatility. *Journal of Accounting, Auditing & Finance*, 22(2), 201–235.

Adebambo, B., Bowen, R. M., Malhotra, S., and Zhu, P. 2022. CEO extraversion and the cost of equity capital. *Available at SSRN 3365155*.

Ahuja, G., C.M. Lampert, and V. Tandon. 2008. 1 moving beyond Schumpeter: management research on the determinants of technological innovation. *Academy of Management Annals*, 2(1), 1–98.

Albuquerque, A.M., G. De Franco, and R.S. Verdi. 2013. Peer choice in CEO compensation. *Journal of Financial Economics*, 108(1), 160–181.

Allport, G. W., and H. S. Odbert. 1936. Trait names: a psycho-lexical study. *Psychological Monographs*, 47(1), 171–220.

Almlund, M., A.L. Duckworth, J. Heckman, and T. Kautz. 2011. Personality psychology and economics. *In Handbook of the Economics of Education*, 4, 1–181

Anderson, D. 1971. Effects of Under- and Over-Evaluations in Loss Reserves. *Journal of Risk and Insurance*, 38(4): 585–600.

Banerjee, S., Humphery-Jenner, M., and Nanda, V. 2015. Restraining overconfident CEOs through improved governance: Evidence from the Sarbanes-Oxley Act. *Review of Financial Studies*, 28(10), 2812–2858.

Barrick, M. R., and M. K. Mount. 1991. The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1–26

Barrick, M. R., M. K. Mount, and T. A. Judge. 2001. Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9(1–2), 9–30.

Barrick, M.R., M. K. Mount, and J.P. Strauss. 1993. Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of applied psychology*, 78(5), 715.

Barth, M.M. and D.L. Eckles. 2018. Anatomy of an error: The case of loss reserve development error research. *Journal of Insurance Issues*, 41(2), 193–214.

Bebchuk, L.A., J.M. Fried. 2005. Pay without performance: overview of the issues. *Journal of Applied Corporate Finance*, 17(4), 8–23.

Berry-Stölzle, T. R., Eastman, E. M., and Xu, J. 2018. CEO overconfidence and earnings management: Evidence from property-liability insurers' loss reserves. *North American Actuarial Journal*, 22(3), 380–404.

Bertrand, M. and A. Schoar. 2003. Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics*, 118(4), 1169–1208.

Bharath, S.T. and T. Shumway. 2008. Forecasting default with the Merton distance to default model. *The Review of Financial Studies*, 21(3), 1339–1369.

Bleidorn, W., C. J. Hopwood, and R.E. Lucas. 2018. Life events and personality trait change. *Journal of Personality*, 86(1), 83–96.

Chang, W.J., R.M. Hayes, and S.A. Hillegeist. 2016. Financial distress risk and new CEO compensation. *Management Science*, 62(2), 479–501.

Chava, S. and Purnanandam, A., 2010. CEOs versus CFOs: Incentives and corporate policies. *Journal of Financial Economics*, 97(2), 263–278.

Coles, J.L., N.D. Daniel, and L. Naveen. 2006. Managerial incentives and risk-taking. *Journal of Financial Economics*, 79(2), 431–468.

Core, J.E., R.W. Holthausen, and D.F. Larcker. 1999. Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics*. 51(3), 371–406.

Costa Jr, P.T, and R. R. McCrae. 1992. Manual for the revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI). *Odessa, FL: Psychological Assessment Resources*.

Cummins, J.D., C.M. Lewis, and R. Wei. 2006. The market value impact of operational loss events for US banks and insurers. *Journal of Banking & Finance*, 30(10), 2605–2634.

Dah, M. A., Frye, M. B., and Hurst, M. 2014. Board changes and CEO turnover: The unanticipated effects of the Sarbanes–Oxley Act. *Journal of Banking & Finance*, 41, 97–108.

D'Arcy, S.P. and J.R. Garven. 1990. Property-liability insurance pricing models: An empirical evaluation. *Journal of Risk and Insurance*, 391–430.

Dehejia, R. H., and S., Wahba. 2002, Propensity score-matching methods for nonexperimental causal studies, *Review of Economics and Statistics*, 84, 151–161.

Furnham, A., C. J. Jackson, and T. Miller. 1999. Personality, learning style and work performance. *Personality and Individual Differences*, 27, 1113–1122.

Gabarro, J. J. 1987. The dynamics of taking charge. Boston: Harvard Business School Press.

Galasso, A. and T.S. Simcoe. 2011. CEO overconfidence and innovation. *Management Science*, 57(8), 1469–1484.

Gow, I.D., S.N. Kaplan, D.F. Larcker, and A.A. Zakolyukina. 2016. CEO personality and firm policies (No. w22435). *National Bureau of Economic Research*.

Grace, E.V., 1990. Property-liability insurer reserve errors: A theoretical and empirical analysis. *Journal of Risk and Insurance*, 28–46.

Grace, M.F. and J.T. Leverty. 2010. Political cost incentives for managing the property-liability insurer loss reserve. *Journal of Accounting Research*, 48(1), 21–49.

Graham, J.R., S. Li, and J. Qiu. 2012. Managerial attributes and executive compensation. *The Review of Financial Studies*, 25(1), 144–186.

Graham, J.R., C.R. Harvey, and M. Puri. 2013. Managerial attitudes and corporate actions. *Journal of Financial Economics*, 109(1), 103–121.

Green, T. C., R. Jame, and B. Lock. 2019. Executive extraversion: Career and firm outcomes. *The Accounting Review*, 94(3), 177–204.

Guo, X., V.K. Gupta, W.E. Jackson III, and S.C. Mortal. 2021. Is there a racial gap in CEO compensation?. *Journal of Corporate Finance*, 69, 102043.

Ho, C. L., Lai, G. C., and Lee, J. P. 2013. Organizational structure, board composition, and risk taking in the US property casualty insurance industry. *Journal of Risk and Insurance*, 80(1), 169–203.

Hsu, W., R. Huang, and G. Lai, 2019, Reserve Management and Audit Committee Characteristics: Evidence from U.S. Property-Liability Insurance Companies. *Journal of Risk and Insurance*, 86(4), 1019–1043.

Hirshleifer, D., and A.V. Thakor. 1992. Managerial conservatism, project choice, and debt. *The Review of Financial Studies*, 5(3), 437–470.

Ho, C.L., G.C. Lai, and J.P. Lee. 2013. Organizational structure, board composition, and risk taking in the US property casualty insurance industry. *Journal of Risk and Insurance*, 80(1), 169–203.

Holmström, B., 1999. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1), 169–182.

Hollander, S., Pronk, M. and Roelofsen, E., 2010. Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, 48(3), 531–563.

Jiang, J.X., Petroni, K.R. and Wang, I.Y., 2010. CFOs and CEOs: Who have the most influence on earnings management?. *Journal of Financial Economics*, 96(3), 513–526.

John, O. P., and S. Srivastava, 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin, L. A., & John, O. P. (Eds.), *Handbook of Personality Theory and Research*.

John, O. P., E. M. Donahue, and R. L. Kentle. 1991. The "Big Five" Inventory: Versions 4a and 5b. Tech. rep., Berkeley: University of California, Institute of Personality and Social Research.

John, O.P., L.P. Naumann, and C.J. Soto. 2008. Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues.

Judge, T. A., C. A. Higgins, C. J. Thoresen, and M. R. Barrick. 1999. The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621–652.

Jung, H.W.H. and A. Subramanian. 2017. CEO talent, CEO compensation, and product market competition. *Journal of Financial Economics*, 125(1), 48–71.

Kaplan, S.N., M. M. Klebanov, and M. Sorensen. 2012. Which CEO Characteristics and Abilities Matter? *The Journal of Finance*, 67, 973–1007.

Kern, M.L. and H.S. Friedman. 2008. Do conscientious individuals live longer? A quantitative review. *Health psychology*, 27(5), 505.

Kim, J.B., Li, Y. and Zhang, L., 2011. CFOs versus CEOs: Equity incentives and crashes. *Journal of Financial Economics*, 101(3), 713–730.

Koenker, R. 2005. Quantile regression [M]. *Journal of Economics Perspective*, 201(1), 15.

Koenker, R. and K.F. Hallock. 2001. Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156.

Kotter, J. P. 1982. The general managers. *New York: Free Press*.

Lartey, T., Kesse, K., and Danso, A. 2020. CEO extraversion and capital structure decisions: the role of firm dynamics, product market competition, and financial crisis. *Journal of Financial Research*, 43(4), 847–893.

Mairesse, F., M.A. Walker, M.R. Mehl, and R.K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30, 457–500.

Mehl, M. R., S. D. Gosling, and J. W. Pennebaker. 2006. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90, 862–877.

Mehl, M., J. Pennebaker, M. Crow, J. Dabbs, and J. Price. 2001. The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, and Computers*, 33, 517–523.

Milidonis, A., T. Nishikawa, and J. Shim. 2019. CEO inside debt and risk taking: Evidence from property–liability insurance firms. *Journal of Risk and Insurance*, 86(2), 451–477.

Mount, M. K., M. R. Barrick, and G. L. Stewart. 1998. Five-Factor Model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, 11, 145–165.

O'Reilly, Charles A., Bernadette Doerr, David F. Caldwell, and Jennifer A. Chatman, 2014b, Narcissistic CEOs and executive compensation. *The Leadership Quarterly*, 25, 218–231.

Pennebaker, J.W. and L.A. King. 1999. Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6), 1296.

Pennebaker, J.W., M.E. Francis, and R.J. Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001), 2001.

Peterson, S.J., F.O. Walumbwa, K. Byron, and J. Myrowitz. 2009. CEO positive psychological traits, transformational leadership, and firm performance in high-technology start-up and established firms. *Journal of Management*, 35(2), 348–368.

Petroni, K.R., 1992. Optimistic reporting in the property-casualty insurance industry. *Journal of Accounting and Economics*, 15(4), 485–508.

Petroni, K. and Beasley, M., 1996. Errors in accounting estimates and their relation to audit firm type. *Journal of Accounting Research*, 34(1), 151–171.

Phillips, R. D., J. D. Cummins, and F. Allen, 1998, Financial Pricing of Insurance in a Multiline Insurance Company, *Journal of Risk and Insurance*, 65(3): 597–636.

Roberts, B. W., Jackson, J. J., Fayard, J. V., Edmonds, G., & Meints, J., 2009, Conscientiousness. In M. Leary & R. Hoyle (Eds.), *Handbook of individual differences in social behavior*, 369–381.

Roberts, M.R. and Whited, T.M., 2013, Endogeneity in empirical corporate finance1, In *Handbook of the Economics of Finance* (Vol. 2, pp. 493-572), Elsevier.

Rosenbaum, P. R., and Rubin, D. B., 1983, The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, 41–55.

Selznick, P. 1957. Leadership in administration. New York: Harper & Row.

Serfling, M.A., 2014. CEO age and the riskiness of corporate policies. *Journal of Corporate Finance*, 25, 251–273.

Shen, W. and A.A. Cannella Jr, 2002, Power dynamics within top management and their impacts on CEO dismissal followed by inside succession. *Academy of Management Journal*, 45(6), 1195–1206.

Shipman, J. E., Swanquist, Q. T., and Whited, R. L., 2017, Propensity score matching in accounting research, *The Accounting Review*, 92(1), 213–244.

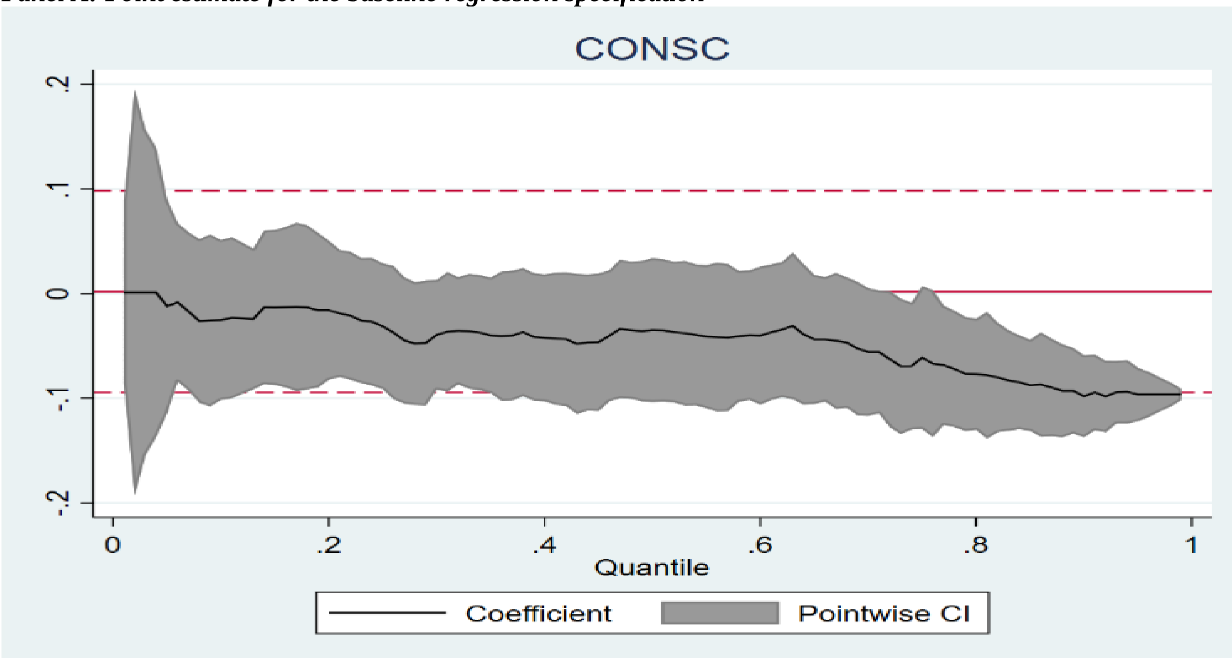
Specht, J., B. Egloff, and S.C. Schmukle. 2011. Stability and change of personality across the life course: the impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, 101(4), 862.

Vancil, R. M. 1987. Passing the baton. Boston: Harvard University Press.

Yim, S., 2013. The acquisitiveness of youth: CEO age and acquisition behavior. *Journal of Financial Economics*. 108(1), 250–273.

Zona, F., 2016. Agency models in different stages of CEO tenure: The effects of stock options and board independence on R&D investment. *Research Policy*, 45(2), 560–575

***Panel A: Point estimate for the baseline regression specification***




***Panel B shows quantile regression results for the conscientiousness interaction with financial risk***

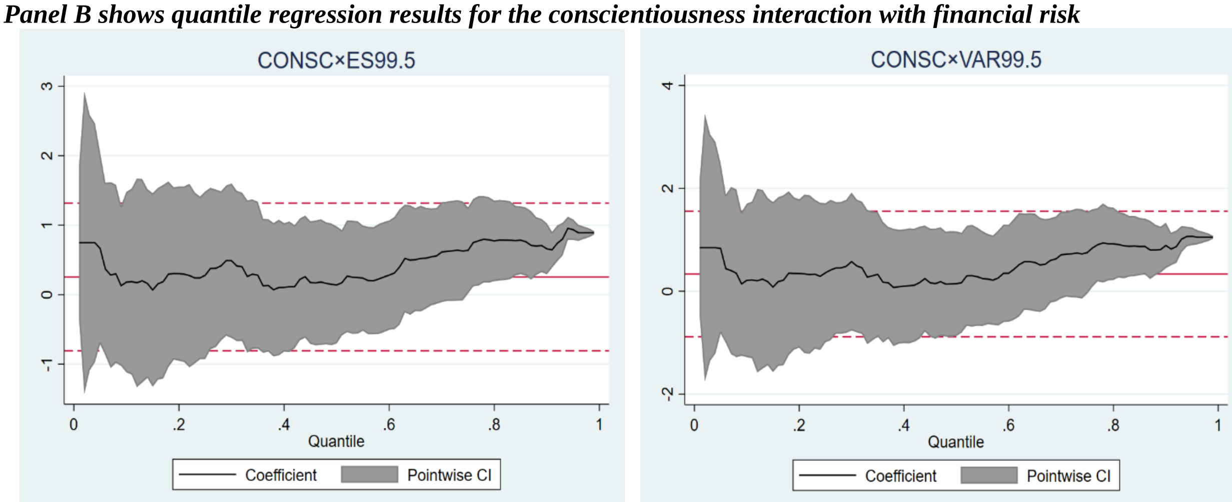



Figure 1.2: Point estimates for the effect of CEO conscientiousness on reserve error

Panel A demonstrates point estimates of the coefficients on *CONSC* from Table 1.2 for the effect of CEO conscientiousness on reserve errors. Panel B shows point estimates of the coefficients on *CONSC×ES99.5* and *CONSC×VAR99.5* from Table 1.3 for the effect of CEO conscientiousness on reserve error with financial risk mechanism. The solid dark curve represents point estimates of the coefficient for quantile regressions from the 1$^{th}$ percentile to the 95$^{th}$ percentile. The shaded area represents 95% pointwise confidence interval of quantile coefficients. The solid red straight line represents the OLS estimation, with two dashed lines depicting the 95% confidence level.

Table 1.1: Summary statistics

This table presents summary statistics of variables used in the regression model. The sample period is from 2002 to 2015. Expected shortfall (ES) and Value at risk (VAR) are computed at 99.5, 99, and 95 percent confidence levels. All the variables are defined in Appendix B.

Panel A: Summary Statistics

| Variable | N | MEAN | SD | MIN | P25 | P50 | P75 | MAX |
|---|---|---|---|---|---|---|---|---|
| *RESERROR* | 244 | 0.009 | 0.109 | -1.388 | -0.013 | 0.020 | 0.048 | 0.180 |
| *CONSC* | 244 | -0.052 | 0.216 | -0.607 | -0.214 | -0.060 | 0.108 | 0.361 |
| *VEGA* | 244 | 3.455 | 1.939 | 0 | 2.096 | 3.901 | 4.934 | 6.831 |
| *DELTA* | 244 | 5.345 | 1.546 | 0 | 4.310 | 5.318 | 6.364 | 9.330 |
| *AGE* | 244 | 4.028 | 0.124 | 3.738 | 3.951 | 4.025 | 4.094 | 4.443 |
| *TENURE* | 244 | 1.682 | 0.833 | -0.876 | 1.225 | 1.792 | 2.286 | 3.807 |
| *COMPENSATION* | 224 | 8.178 | 0.897 | 6.016 | 7.471 | 8.167 | 8.921 | 10.73 |
| *CHAIRMAN* | 224 | 0.442 | 0.498 | 0 | 0 | 0 | 1 | 1 |
| *FIRST4* | 155 | 0.516 | 0.501 | 0 | 0 | 1 | 1 | 1 |
| *LNNPW* | 244 | 14.400 | 1.428 | 10.500 | 13.310 | 14.240 | 15.260 | 17.380 |
| *NPWGROWTH* | 244 | 0.072 | 0.192 | -0.517 | -0.012 | 0.033 | 0.105 | 1.691 |
| *TAXSHIELD* | 244 | 0.280 | 0.143 | -0.068 | 0.179 | 0.243 | 0.395 | 0.631 |
| *LNQ* | 244 | 0.069 | 0.132 | -0.147 | -0.015 | 0.036 | 0.112 | 0.640 |
| *SMOOTH* | 244 | 0.042 | 0.038 | -0.104 | 0.024 | 0.041 | 0.066 | 0.144 |
| *REAFFILIATE* | 244 | 0.061 | 0.157 | -0.061 | 0 | 0 | 0.006 | 0.698 |
| *RENONAFFILIATE* | 244 | 0.135 | 0.116 | 0.001 | 0.038 | 0.096 | 0.234 | 0.506 |
| *RERATIO* | 224 | 0.173 | 0.206 | 0.00100 | 0.0390 | 0.089 | 0.256 | 0.915 |
| *LRGROWTH* | 244 | 0.008 | 0.277 | -2.233 | -0.066 | -0.002 | 0.070 | 2.691 |
| *LONGTAIL* | 244 | 0.713 | 0.212 | 0 | 0.666 | 0.742 | 0.828 | 1 |
| *PRODHHI* | 244 | 0.361 | 0.255 | 0.123 | 0.170 | 0.296 | 0.449 | 1 |
| *GEOHHI* | 244 | 0.128 | 0.162 | 0.036 | 0.049 | 0.070 | 0.099 | 0.889 |
| *WEAK* | 244 | 0.033 | 0.178 | 0 | 0 | 0 | 0 | 1 |
| *LNDD* | 244 | 2.629 | 0.758 | -0.554 | 2.324 | 2.778 | 3.139 | 4.163 |
| *ROA* | 224 | 0.034 | 0.047 | -0.303 | 0.016 | 0.033 | 0.055 | 0.179 |
| *TAXEXEMPT* | 224 | 0.462 | 0.300 | -0.596 | 0.271 | 0.455 | 0.686 | 1.372 |
| *ES99.5* | 244 | 0.057 | 0.040 | 0.021 | 0.034 | 0.043 | 0.065 | 0.304 |
| *ES99* | 244 | 0.053 | 0.037 | 0.019 | 0.031 | 0.040 | 0.060 | 0.281 |
| *ES95* | 244 | 0.041 | 0.028 | 0.015 | 0.024 | 0.031 | 0.046 | 0.219 |
| *VAR99.5* | 244 | 0.051 | 0.034 | 0.019 | 0.031 | 0.040 | 0.057 | 0.258 |
| *VAR99* | 244 | 0.046 | 0.030 | 0.018 | 0.028 | 0.036 | 0.052 | 0.233 |
| *VAR95* | 244 | 0.033 | 0.021 | 0.012 | 0.020 | 0.025 | 0.036 | 0.163 |
| *STDROA* | 224 | 0.022 | 0.017 | 0.001 | 0.010 | 0.019 | 0.028 | 0.094 |
| *STDROI* | 224 | 0.008 | 0.008 | 0.001 | 0.003 | 0.005 | 0.010 | 0.043 |
| *STDLOSSRATIO* | 224 | 0.084 | 0.127 | 0.010 | 0.030 | 0.048 | 0.080 | 0.858 |

Panel B: Shapiro-Wilk W Test for Normal Data

| Variable | N | W | V | z | Prob>z |
|---|---|---|---|---|---|
| *RESERROR* | 244 | 0.453 | 97.101 | 10.633 | 0.000 |

Panel C: Skewness/Kurtosis Tests for Normality

| Variable | N | Pr (Skewness) | Pr (Kurtosis) | Joint Prob>chi2 |
|---|---|---|---|---|
| *RESERROR* | 244 | 0.000 | 0.000 | 0.000 |

Panel D: The Distribution of Reserve Error

| Quantile | Value |
|---|---|
| 100% Max | 0.180 |
| 95% | 0.098 |
| 90% | 0.077 |
| 85% | 0.062 |
| 80% | 0.055 |
| 75% Q3 | 0.048 |
| Mean | 0.009 |
| 50% Median | 0.020 |
| 25% Q1 | -0.013 |
| 10% | -0.052 |
| 0% Min | -1.388 |

Table 1.2: CEO conscientiousness and reserve error

This table presents the main results of the baseline model, testing the relation between CEO conscientiousness and reserve error using the OLS and quantile regression methods. The dependent variable is reserve error (*RESERROR)*. All the variables are defined in Appendix B. The *t*-statistics are reported in parentheses. The standard deviations are clustered at firm level. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC*** | **0.002** | **-0.025** | **-0.031** | **-0.035** | **-0.061*** | **-0.077*** ** | **-0.088*** ** | **-0.098*** ** | **-0.096*** ** |
| | (0.039) | (-0.678) | (-1.068) | (-1.043) | (-1.776) | (-2.977) | (-4.130) | (-5.135) | (-7.727) |
| *VEGA* | -0.004 | 0.002 | 0.004 | -0.000 | -0.003 | -0.003 | -0.004 | -0.004 | -0.004** |
| | (-0.648) | (0.429) | (0.948) | (-0.070) | (-0.865) | (-0.862) | (-1.220) | (-1.339) | (-2.431) |
| *DELTA* | 0.001 | -0.008 | -0.008 | 0.001 | 0.005 | 0.004 | 0.005 | 0.003 | 0.001 |
| | (0.108) | (-1.692) | (-1.241) | (0.161) | (1.181) | (1.314) | (1.207) | (1.313) | (0.679) |
| *AGE* | -0.032 | 0.100 | 0.088 | -0.010 | -0.038 | -0.012 | -0.043 | -0.029 | -0.044* |
| | (-0.402) | (1.327) | (1.134) | (-0.109) | (-0.723) | (-0.250) | (-0.930) | (-0.886) | (-1.786) |
| *TENURE* | -0.003 | 0.002 | 0.002 | 0.001 | 0.003 | 0.003 | 0.007 | 0.007 | 0.012*** |
| | (-0.380) | (0.228) | (0.198) | (0.063) | (0.417) | (0.538) | (1.293) | (1.475) | (3.866) |
| *LNNPW* | -0.005 | -0.008 | -0.010** | -0.013** | -0.023*** | -0.024*** | -0.026*** | -0.026*** | -0.025*** |
| | (-0.478) | (-1.619) | (-2.140) | (-2.677) | (-5.083) | (-4.556) | (-5.337) | (-6.914) | (-9.903) |
| *NPWGROWTH* | 0.031 | 0.057 | 0.029 | 0.018 | -0.009 | -0.011 | -0.020 | -0.022 | -0.030** |
| | (1.147) | (1.346) | (0.852) | (0.670) | (-0.319) | (-0.414) | (-1.184) | (-1.501) | (-2.201) |
| *TAXSHIELD* | -0.180* | -0.154** | -0.053 | -0.035 | -0.069* | -0.080* | -0.096** | -0.090** | -0.095*** |
| | (-1.727) | (-2.521) | (-1.207) | (-0.860) | (-1.796) | (-1.888) | (-2.222) | (-2.498) | (-4.342) |
| *LNQ* | 0.174** | 0.176* | 0.181** | 0.094 | 0.115 | 0.136** | 0.149** | 0.173*** | 0.165*** |
| | (2.496) | (1.911) | (2.641) | (1.177) | (1.655) | (2.151) | (2.301) | (3.360) | (6.626) |
| *SMOOTH* | 0.199 | 0.225 | -0.099 | -0.075 | -0.002 | 0.090 | 0.125 | 0.104 | 0.106** |
| | (0.880) | (0.804) | (-0.426) | (-0.473) | (-0.010) | (0.638) | (0.981) | (1.215) | (2.403) |
| *REAFFILIATE* | -0.016 | -0.073 | -0.033 | -0.040 | 0.015 | 0.049 | 0.060 | 0.081* | 0.178*** |
| | (-0.281) | (-1.361) | (-0.713) | (-0.730) | (0.228) | (0.801) | (0.906) | (1.715) | (6.407) |
| *RENONAFFILIATE* | -0.059 | -0.157** | -0.121 | -0.083 | -0.131* | -0.158** | -0.197** | -0.203*** | -0.244*** |
| | (-0.675) | (-2.092) | (-1.411) | (-0.946) | (-1.757) | (-2.141) | (-2.644) | (-3.861) | (-8.495) |
| *LRGROWTH* | -0.005 | -0.022 | 0.013 | -0.012 | -0.004 | -0.009 | -0.002 | -0.003 | -0.003 |
| | (-0.253) | (-0.697) | (0.879) | (-0.736) | (-0.164) | (-0.753) | (-0.154) | (-0.294) | (-0.777) |

Table 1.2 CEO conscientiousness and reserve error (continued)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *LONGTAIL* | 0.053*** | 0.080*** | 0.048 | 0.017 | 0.061** | 0.066*** | 0.073*** | 0.077*** | 0.066*** |
| | (3.004) | (3.152) | (1.590) | (0.627) | (2.227) | (3.817) | (3.923) | (6.259) | (6.624) |
| *PRODHHI* | 0.008 | -0.005 | 0.007 | 0.012 | 0.009 | 0.009 | -0.003 | -0.002 | -0.013 |
| | (0.299) | (-0.168) | (0.272) | (0.406) | (0.295) | (0.357) | (-0.100) | (-0.083) | (-0.887) |
| *GEOHHI* | -0.079* | -0.086* | -0.094** | -0.122*** | -0.152*** | -0.154*** | -0.169*** | -0.167*** | -0.131*** |
| | (-1.887) | (-1.942) | (-2.695) | (-3.748) | (-5.422) | (-6.005) | (-6.793) | (-8.045) | (-6.471) |
| *WEAK* | -0.039 | -0.043 | -0.104** | -0.091** | -0.052 | -0.010 | -0.008 | -0.021 | -0.009 |
| | (-1.301) | (-1.413) | (-2.373) | (-2.063) | (-0.387) | (-0.158) | (-0.277) | (-0.854) | (-0.540) |
| *LNDD* | 0.013* | 0.017** | 0.014** | 0.005 | 0.003 | 0.002 | 0.000 | -0.003 | 0.006** |
| | (1.834) | (2.079) | (2.079) | (0.879) | (0.429) | (0.247) | (0.065) | (-0.616) | (2.724) |
| Intercept | 0.195 | -0.447 | -0.296 | 0.214 | 0.466** | 0.405** | 0.565*** | 0.515*** | 0.566*** |
| | (0.735) | (-1.515) | (-0.925) | (0.607) | (2.438) | (2.249) | (3.337) | (4.473) | (5.424) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table 1.3: CEO conscientiousness and reserve error using Propensity Score Matching

This table presents the results using Propensity Score Matching (PSM) method. Panel A reports the first-stage logit regression results for estimating propensity scores. Panel B reports the mean (median) difference between the treatment and matched sample using PSM method, and signs ***, **, * indicate the significance of these differences based on *t*-tests (Wilcoxon rank-sum test) for means (medians) at the 1%, 5%, and 10% levels, respectively. Panel C reports the results for the relation between high conscientiousness CEOs and reserve error using the OLS and quantile regression methods with the model specification in equation (1). The sample used in these regressions is the treatment and matched insurers using PSM procedure. All the variables are defined in Appendix B. The z(*t*)-statistics are reported in parentheses. The standard deviations are clustered at firm level. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

Panel A: First-stage logistic regression of PSM

| | Dep var = ***HIGHCONSC*** |
|---|---|
| *VEGA* | -0.382** |
| | (-2.465) |
| *DELTA* | 0.006 |
| | (0.026) |
| *AGE* | -3.321 |
| | (-1.316) |
| *TENURE* | 1.152*** |
| | (2.745) |
| *LNNPW* | 0.432** |
| | (2.082) |
| *NPWGROWTH* | 1.552 |
| | (0.876) |
| *TAXSHIELD* | -0.081 |
| | (-0.032) |
| *LNQ* | -0.387 |
| | (-0.141) |
| *SMOOTH* | 5.431 |
| | (0.504) |
| *REAFFILIATE* | 1.089 |
| | (0.408) |
| *RENONAFFILIATE* | -5.955* |
| | (-1.647) |
| *LRGROWTH* | -0.047 |
| | (-0.030) |
| *LONGTAIL* | -0.510 |
| | (-0.257) |
| *PRODHHI* | 7.581*** |
| | (5.020) |
| *GEOHHI* | -10.443*** |
| | (-3.394) |
| *WEAK* | -3.254 |
| | (-0.776) |
| *LNDD* | -1.263*** |
| | (-2.828) |
| Intercept | 8.462 |
| | (0.821) |
| Year FE | Yes |
| Observations | 244 |
| Pseudo R-sq. | 0.464 |

Panel B: Results of covariate balance checks after PSM procedure

| | High conscientious CEOs | | PSM matched group from low conscientious CEOs | | Differences in | |
|---|---|---|---|---|---|---|
| | Mean | Median | Mean | Median | Mean | Median |
| *VEGA* | 3.304 | 3.931 | 3.831 | 4.106 | -0.527 | -0.175 |
| *DELTA* | 5.321 | 5.323 | 5.629 | 5.876 | -0.308 | -0.553 |
| *AGE* | 4.033 | 4.025 | 4.074 | 4.043 | -0.041 | -0.018 |
| *TENURE* | 1.772 | 1.952 | 1.762 | 1.642 | 0.010 | 0.310 |
| *LNNPW* | 14.778 | 14.820 | 15.069 | 14.598 | -0.291 | 0.222 |
| *NPWGROWTH* | 0.059 | 0.029 | 0.051 | 0.041 | 0.008 | -0.012 |
| *TAXSHIELD* | 0.329 | 0.225 | 0.317 | 0.379 | 0.012 | -0.154 |
| *LNQ* | 0.107 | 0.033 | 0.056 | 0.036 | 0.051 | -0.003 |
| *SMOOTH* | 0.055 | 0.049 | 0.049 | 0.045 | 0.006 | 0.004 |
| *REAFFILIATE* | 0.023 | 0.000 | 0.020 | 0.000 | 0.003 | 0.000 |
| *RENONAFFILIATE* | 0.098 | 0.078 | 0.111 | 0.042 | -0.013 | 0.036 |
| *LRGROWTH* | 0.014 | 0.010 | 0.012 | 0.048 | 0.002 | -0.038 |
| *LONGTAIL* | 0.741 | 0.756 | 0.745 | 0.727 | -0.004 | 0.029 |
| *PRODHHI* | 0.424 | 0.325 | 0.352 | 0.316 | 0.072 | 0.009 |
| *GEOHHI* | 0.088 | 0.052 | 0.122 | 0.059 | -0.034 | -0.007** |
| *WEAK* | 0.020 | 0.000 | 0.000 | 0.000 | 0.020 | 0.000 |
| *LNDD* | 2.524 | 2.739 | 2.452 | 2.522 | 0.072 | 0.217 |

Panel C: Regression results using PSM procedure

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***HIGHCONSC*** | **-0.024*****| **-0.022** | **-0.016** | **-0.018** | **-0.018** | **-0.020** | **-0.020** | **-0.022*** | **-0.022***** |
| | (-3.244) | (-1.216) | (-0.793) | (-1.339) | (-1.229) | (-1.391) | (-1.495) | (-1.809) | (-2.825) |
| *VEGA* | 0.006** | -0.002 | 0.001 | 0.005 | 0.007** | 0.010** | 0.010*** | 0.010*** | 0.010*** |
| | (2.007) | (-0.330) | (0.124) | (1.359) | (2.366) | (2.543) | (2.726) | (3.032) | (3.704) |
| *DELTA* | -0.010* | -0.003 | 0.002 | -0.013* | -0.016** | -0.021** | -0.019** | -0.017* | -0.017*** |
| | (-1.894) | (-0.214) | (0.150) | (-1.799) | (-2.195) | (-2.371) | (-2.043) | (-1.846) | (-2.786) |
| *AGE* | 0.058 | 0.084 | -0.011 | -0.027 | -0.077 | -0.040 | -0.027 | -0.009 | -0.009 |
| | (1.164) | (0.633) | (-0.127) | (-0.345) | (-1.111) | (-0.438) | (-0.317) | (-0.117) | (-0.160) |
| *TENURE* | -0.008 | -0.008 | -0.004 | 0.008 | 0.014 | 0.013 | 0.007 | -0.002 | -0.002 |
| | (-0.778) | (-0.417) | (-0.300) | (0.478) | (0.887) | (0.761) | (0.343) | (-0.098) | (-0.122) |
| *LNNPW* | -0.016*** | -0.017 | -0.019** | -0.008 | -0.008 | -0.013 | -0.015 | -0.018 | -0.018* |
| | (-3.466) | (-1.574) | (-2.264) | (-1.319) | (-0.735) | (-1.150) | (-1.314) | (-1.648) | (-1.822) |
| *NPWGROWTH* | -0.066* | 0.015 | -0.078 | -0.063 | -0.070 | -0.098 | -0.097 | -0.101* | -0.101** |
| | (-1.803) | (0.128) | (-1.333) | (-1.633) | (-1.110) | (-1.522) | (-1.600) | (-1.718) | (-2.644) |
| *TAXSHIELD* | -0.013 | -0.100 | -0.062 | 0.008 | 0.015 | 0.042 | 0.040 | 0.021 | 0.021 |
| | (-0.334) | (-1.235) | (-0.877) | (0.069) | (0.236) | (0.661) | (0.604) | (0.345) | (0.598) |
| *LNQ* | 0.132*** | 0.167** | 0.123** | 0.067 | 0.137 | 0.196 | 0.195 | 0.211* | 0.211** |
| | (3.066) | (2.258) | (2.323) | (0.528) | (1.036) | (1.316) | (1.489) | (1.831) | (2.621) |
| *SMOOTH* | -0.006 | -0.048 | -0.102 | -0.082 | -0.168 | -0.034 | 0.080 | 0.242 | 0.242 |
| | (-0.035) | (-0.180) | (-0.509) | (-0.362) | (-0.437) | (-0.084) | (0.197) | (0.632) | (0.747) |
| *REAFFILIATE* | 0.115* | -0.033 | 0.087 | 0.077 | 0.240** | 0.235* | 0.222* | 0.203 | 0.203** |
| | (1.698) | (-0.135) | (0.514) | (1.000) | (2.043) | (1.957) | (1.771) | (1.638) | (2.671) |
| *RENONAFFILIATE* | -0.108 | -0.214 | -0.269* | -0.143 | -0.072 | -0.037 | -0.049 | -0.070 | -0.070 |
| | (-1.634) | (-1.515) | (-1.981) | (-1.128) | (-0.893) | (-0.395) | (-0.520) | (-0.801) | (-1.439) |
| *LRGROWTH* | -0.016 | -0.018 | -0.024 | -0.004 | -0.026 | -0.006 | -0.013 | -0.020 | -0.020 |
| | (-0.461) | (-0.381) | (-0.750) | (-0.087) | (-0.610) | (-0.174) | (-0.347) | (-0.512) | (-0.858) |
| *LONGTAIL* | 0.016 | 0.020 | 0.006 | 0.025 | 0.046* | 0.045 | 0.047 | 0.052 | 0.052*** |
| | (0.581) | (0.454) | (0.182) | (1.131) | (1.756) | (1.440) | (1.415) | (1.637) | (3.129) |
| *PRODHHI* | 0.006 | -0.011 | -0.032 | 0.026 | 0.012 | -0.022 | -0.033 | -0.059 | -0.059 |
| | (0.183) | (-0.189) | (-0.749) | (0.456) | (0.151) | (-0.279) | (-0.433) | (-0.839) | (-1.053) |

Table 1.3 CEO conscientiousness and reserve error using Propensity Score Matching Panel C (continued)

| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *GEOHHI* | -0.033 | -0.112 | -0.058 | 0.018 | 0.022 | 0.008 | 0.008 | 0.008 | 0.008 |
| | (-0.768) | (-0.992) | (-0.639) | (0.206) | (0.440) | (0.149) | (0.144) | (0.131) | (0.210) |
| *WEAK* | 0.092** | 0.066 | 0.113** | 0.047 | 0.034 | 0.075 | 0.060 | 0.076 | 0.076* |
| | (2.055) | (0.927) | (2.131) | (0.792) | (0.519) | (1.150) | (0.947) | (1.251) | (1.859) |
| *LNDD* | 0.019** | 0.014 | 0.008 | 0.011 | 0.012 | 0.022** | 0.022** | 0.022** | 0.022*** |
| | (2.580) | (1.208) | (0.945) | (0.796) | (1.531) | (2.089) | (2.279) | (2.675) | (4.082) |
| Intercept | 0.009 | -0.156 | 0.322 | 0.215 | 0.389 | 0.314 | 0.287 | 0.269 | 0.269 |
| | (0.042) | (-0.241) | (0.728) | (0.762) | (1.418) | (0.949) | (0.901) | (0.868) | (1.316) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Table 1.4: CEO conscientiousness and reserve error with financial risk mechanism

This table presents the results of the interaction term model, testing the relation between CEO conscientiousness and reserve error with financial risk mechanism using the OLS and quantile regression methods. The dependent variable is reserve error (*RESERROR)*. All the variables are defined in Appendix B. The *t*-statistics are shown in parentheses. The standard deviations are clustered at firm level. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

Table 1.4 Panel A: Financial risk measured by Expected shortfall with confidence levels of 99.5 (ES99.5)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *CONSC* | -0.008 | -0.035 | -0.037 | -0.039 | -0.100** | -0.112*** | -0.118*** | -0.112*** | -0.134*** |
| | (-0.118) | (-0.847) | (-0.748) | (-0.905) | (-2.713) | (-3.890) | (-4.394) | (-5.715) | (-10.977) |
| ***CONSC×ES99.5*** | **0.254** | **0.180** | **0.280** | **0.139** | **0.636*** | **0.773**** | **0.784***** | **0.659***** | **0.935***** |
| | (0.490) | (0.285) | (0.475) | (0.356) | (1.936) | (2.732) | (3.296) | (3.690) | (12.999) |
| *ES99.5* | -0.225* | -0.061 | -0.146 | -0.152 | -0.301** | -0.341*** | -0.315*** | -0.246*** | -0.342*** |
| | (-1.911) | (-0.365) | (-1.047) | (-1.247) | (-2.551) | (-3.738) | (-3.984) | (-3.283) | (-14.891) |
| *VEGA* | -0.004 | 0.001 | 0.005 | 0.000 | -0.004 | -0.003 | -0.002 | -0.002 | -0.002 |
| | (-0.590) | (0.305) | (1.236) | (0.073) | (-1.127) | (-1.014) | (-0.879) | (-0.826) | (-1.211) |
| *DELTA* | 0.001 | -0.008 | -0.007 | 0.001 | 0.004 | 0.003 | 0.003 | 0.003 | 0.002 |
| | (0.123) | (-1.486) | (-1.175) | (0.100) | (1.099) | (0.799) | (1.024) | (1.217) | (1.401) |
| *AGE* | -0.026 | 0.096 | 0.102 | -0.000 | -0.020 | -0.023 | -0.026 | -0.020 | -0.000 |
| | (-0.325) | (1.331) | (1.288) | (-0.000) | (-0.359) | (-0.499) | (-0.703) | (-0.601) | (-0.000) |
| *TENURE* | -0.004 | 0.002 | -0.005 | 0.001 | 0.005 | 0.004 | 0.005 | 0.005 | 0.004 |
| | (-0.503) | (0.240) | (-0.466) | (0.116) | (0.699) | (0.747) | (1.113) | (1.208) | (1.589) |
| *LNNPW* | -0.005 | -0.008 | -0.009* | -0.013*** | -0.022*** | -0.023*** | -0.025*** | -0.027*** | -0.028*** |
| | (-0.479) | (-1.462) | (-1.977) | (-2.932) | (-4.842) | (-4.699) | (-6.304) | (-7.248) | (-11.242) |
| *NPWGROWTH* | 0.038 | 0.058* | 0.057** | 0.039 | 0.010 | -0.006 | -0.014 | -0.027 | -0.031*** |
| | (1.547) | (1.821) | (2.072) | (1.215) | (0.579) | (-0.361) | (-0.878) | (-1.677) | (-2.954) |
| *TAXSHIELD* | -0.178* | -0.150** | -0.052 | -0.037 | -0.056 | -0.062 | -0.080** | -0.058 | -0.039* |
| | (-1.711) | (-2.226) | (-1.186) | (-0.914) | (-1.323) | (-1.405) | (-2.278) | (-1.638) | (-1.837) |
| *LNQ* | 0.184** | 0.169* | 0.168** | 0.089 | 0.104 | 0.156** | 0.170*** | 0.159*** | 0.167*** |
| | (2.678) | (1.821) | (2.632) | (1.318) | (1.639) | (2.487) | (3.063) | (3.133) | (6.588) |
| *SMOOTH* | 0.169 | 0.274 | -0.021 | -0.075 | -0.015 | -0.098 | 0.010 | 0.023 | 0.015 |
| | (0.730) | (0.855) | (-0.107) | (-0.495) | (-0.086) | (-0.784) | (0.102) | (0.267) | (0.309) |
| *REAFFILIATE* | -0.026 | -0.038 | -0.044 | -0.036 | 0.007 | 0.053 | 0.056 | 0.069 | 0.146*** |
| | (-0.440) | (-0.949) | (-0.951) | (-0.548) | (0.112) | (0.947) | (1.226) | (1.456) | (5.418) |

Table 1.4 CEO conscientiousness and reserve error with financial risk mechanism Panel A (continued)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *RENONAFFILIATE* | -0.052 | -0.162** | -0.076 | -0.084 | -0.139 | -0.143* | -0.161*** | -0.174*** | -0.185*** |
| | (-0.587) | (-2.320) | (-0.915) | (-0.982) | (-1.694) | (-1.903) | (-2.941) | (-3.722) | (-7.149) |
| *LRGROWTH* | -0.001 | -0.024 | 0.011 | -0.009 | -0.001 | 0.001 | -0.003 | -0.004 | -0.003 |
| | (-0.064) | (-0.734) | (0.674) | (-0.630) | (-0.057) | (0.087) | (-0.292) | (-0.463) | (-0.952) |
| *LONGTAIL* | 0.049*** | 0.074** | 0.047 | 0.018 | 0.066*** | 0.062*** | 0.069*** | 0.072*** | 0.075*** |
| | (3.055) | (2.322) | (1.479) | (0.699) | (3.120) | (3.651) | (4.072) | (5.316) | (8.827) |
| *PRODHHI* | 0.008 | -0.005 | 0.012 | 0.011 | 0.016 | 0.004 | 0.010 | 0.009 | 0.013 |
| | (0.287) | (-0.148) | (0.465) | (0.388) | (0.536) | (0.148) | (0.393) | (0.374) | (1.001) |
| *GEOHHI* | -0.077* | -0.090 | -0.075* | -0.116*** | -0.151*** | -0.166*** | -0.168*** | -0.162*** | -0.161*** |
| | (-2.017) | (-1.693) | (-1.780) | (-4.387) | (-4.494) | (-6.333) | (-7.417) | (-8.211) | (-12.102) |
| *WEAK* | -0.035 | -0.067 | -0.082** | -0.098** | -0.070 | -0.014 | -0.015 | -0.020 | -0.034*** |
| | (-1.179) | (-1.584) | (-2.521) | (-2.240) | (-1.081) | (-0.314) | (-0.501) | (-0.936) | (-3.340) |
| *LNDD* | 0.011 | 0.015* | 0.010 | 0.003 | 0.001 | 0.001 | 0.000 | -0.002 | -0.004* |
| | (1.521) | (1.981) | (1.598) | (0.655) | (0.082) | (0.193) | (0.045) | (-0.407) | (-1.814) |
| Intercept | 0.191 | -0.426 | -0.362 | 0.181 | 0.407** | 0.474*** | 0.493*** | 0.497*** | 0.446*** |
| | (0.700) | (-1.505) | (-1.109) | (0.536) | (2.055) | (2.839) | (3.529) | (4.536) | (4.918) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table 1.4 Panel B: Financial risk measured by Value at risk with confidence levels of 99.5 (VAR99.5)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *CONSC* | -0.011 | -0.035 | -0.038 | -0.038 | -0.102** | -0.117*** | -0.119*** | -0.119*** | -0.138*** |
| | (-0.161) | (-0.756) | (-0.745) | (-0.839) | (-2.702) | (-3.912) | (-4.396) | (-5.982) | (-11.328) |
| ***CONSC×VAR99.5*** | **0.335** | **0.211** | **0.357** | **0.145** | **0.751*** | **0.922**** | **0.869***** | **0.890***** | **1.067***** |
| | (0.561) | (0.290) | (0.545) | (0.293) | (1.883) | (2.691) | (3.213) | (4.139) | (12.758) |
| *VAR99.5* | -0.262* | -0.071 | -0.175 | -0.173 | -0.351** | -0.403*** | -0.359*** | -0.301*** | -0.413*** |
| | (-1.879) | (-0.346) | (-1.049) | (-1.127) | (-2.460) | (-3.473) | (-3.910) | (-3.281) | (-15.418) |
| *VEGA* | -0.004 | 0.001 | 0.005 | 0.000 | -0.003 | -0.003 | -0.002 | -0.002 | -0.002 |
| | (-0.589) | (0.297) | (1.197) | (0.092) | (-1.021) | (-1.048) | (-0.922) | (-0.850) | (-1.191) |
| *DELTA* | 0.001 | -0.008 | -0.007 | 0.001 | 0.004 | 0.003 | 0.003 | 0.003 | 0.001 |
| | (0.118) | (-1.534) | (-1.173) | (0.098) | (1.108) | (0.845) | (0.993) | (1.108) | (0.984) |
| *AGE* | -0.026 | 0.096 | 0.101 | 0.002 | -0.016 | -0.023 | -0.023 | -0.022 | 0.003 |
| | (-0.321) | (1.331) | (1.277) | (0.023) | (-0.296) | (-0.493) | (-0.618) | (-0.661) | (0.121) |
| *TENURE* | -0.004 | 0.002 | -0.005 | 0.001 | 0.003 | 0.004 | 0.005 | 0.006 | 0.006** |
| | (-0.503) | (0.227) | (-0.448) | (0.078) | (0.506) | (0.717) | (1.034) | (1.425) | (2.068) |
| *LNNPW* | -0.005 | -0.008 | -0.009* | -0.013*** | -0.023*** | -0.024*** | -0.025*** | -0.026*** | -0.027*** |
| | (-0.481) | (-1.495) | (-2.037) | (-2.977) | (-4.879) | (-4.949) | (-6.109) | (-7.435) | (-11.222) |
| *NPWGROWTH* | 0.038 | 0.058* | 0.059** | 0.039 | 0.008 | -0.005 | -0.016 | -0.025 | -0.029*** |
| | (1.546) | (1.962) | (2.085) | (1.154) | (0.490) | (-0.341) | (-0.964) | (-1.397) | (-2.846) |
| *TAXSHIELD* | -0.179* | -0.151** | -0.055 | -0.038 | -0.056 | -0.065 | -0.080** | -0.065* | -0.049** |
| | (-1.712) | (-2.425) | (-1.293) | (-0.904) | (-1.318) | (-1.504) | (-2.257) | (-1.883) | (-2.351) |
| *LNQ* | 0.185** | 0.169* | 0.170** | 0.091 | 0.111* | 0.160** | 0.171*** | 0.154*** | 0.184*** |
| | (2.692) | (1.781) | (2.668) | (1.327) | (1.727) | (2.520) | (3.017) | (3.018) | (7.161) |
| *SMOOTH* | 0.167 | 0.274 | -0.019 | -0.078 | -0.018 | -0.090 | 0.020 | 0.012 | -0.009 |
| | (0.722) | (0.876) | (-0.093) | (-0.527) | (-0.105) | (-0.734) | (0.197) | (0.140) | (-0.190) |
| *REAFFILIATE* | -0.026 | -0.038 | -0.045 | -0.036 | 0.004 | 0.052 | 0.058 | 0.069 | 0.142*** |
| | (-0.439) | (-0.800) | (-0.886) | (-0.578) | (0.063) | (0.936) | (1.239) | (1.453) | (5.568) |

Table 1.4 CEO conscientiousness and reserve error with financial risk mechanism Panel B (continued)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *RENONAFFILIATE* | -0.051 | -0.163** | -0.077 | -0.085 | -0.135 | -0.143* | -0.161*** | -0.179*** | -0.197*** |
| | (-0.584) | (-2.191) | (-0.932) | (-1.006) | (-1.652) | (-1.887) | (-2.935) | (-3.780) | (-7.722) |
| *LRGROWTH* | -0.001 | -0.024 | 0.011 | -0.009 | -0.001 | 0.002 | -0.003 | -0.005 | -0.001 |
| | (-0.073) | (-0.756) | (0.649) | (-0.587) | (-0.057) | (0.118) | (-0.357) | (-0.554) | (-0.370) |
| *LONGTAIL* | 0.050*** | 0.074** | 0.046 | 0.019 | 0.064*** | 0.063*** | 0.071*** | 0.066*** | 0.074*** |
| | (3.082) | (2.395) | (1.426) | (0.718) | (2.988) | (3.796) | (5.292) | (5.205) | (8.744) |
| *PRODHHI* | 0.008 | -0.005 | 0.011 | 0.010 | 0.014 | 0.005 | 0.013 | 0.003 | 0.008 |
| | (0.295) | (-0.155) | (0.411) | (0.335) | (0.453) | (0.157) | (0.499) | (0.147) | (0.638) |
| *GEOHHI* | -0.076* | -0.090* | -0.075* | -0.116*** | -0.152*** | -0.165*** | -0.170*** | -0.144*** | -0.154*** |
| | (-2.010) | (-1.776) | (-1.759) | (-4.173) | (-4.642) | (-6.340) | (-7.559) | (-5.815) | (-10.785) |
| *WEAK* | -0.035 | -0.066 | -0.068** | -0.098** | -0.067 | -0.018 | -0.018 | -0.015 | -0.034*** |
| | (-1.194) | (-1.442) | (-2.325) | (-2.394) | (-1.011) | (-0.423) | (-0.475) | (-0.677) | (-3.426) |
| *LNDD* | 0.011 | 0.015** | 0.010 | 0.003 | 0.001 | 0.000 | -0.000 | 0.000 | -0.004* |
| | (1.480) | (2.166) | (1.599) | (0.503) | (0.127) | (0.014) | (-0.016) | (0.124) | (-1.992) |
| Intercept | 0.191 | -0.425 | -0.354 | 0.177 | 0.397** | 0.480*** | 0.483*** | 0.503*** | 0.439*** |
| | (0.700) | (-1.505) | (-1.090) | (0.532) | (2.075) | (2.895) | (3.417) | (4.417) | (4.829) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table 1.5: CEO conscientiousness and reserve error using SOX as an exogenous shock with financial risk mechanism

This table presents the relation between CEO conscientiousness and reserve error with financial risk mechanism of the SOX period using the OLS and quantile regression methods. A dummy variable *POSTSOX* equals 1 from 2005 to 2015 and equals 0 from 2002 to 2004. The dependent variable is reserve error (*RESERROR*). All the variables are defined in Appendix B. The *t*-statistics are shown in parentheses. The standard deviations are clustered at firm level. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

Panel A: Financial risk measured by Expected shortfall with confidence levels of 99.5 (ES99.5)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *CONSC* | -0.258** | -0.611** | -0.333* | -0.125 | -0.243** | -0.272** | -0.271*** | -0.262*** | -0.261*** |
| | (-2.406) | (-2.397) | (-1.830) | (-1.150) | (-2.239) | (-2.732) | (-2.974) | (-3.609) | (-10.129) |
| ***CONSC×POSTSOX×ES99.5*** | **-3.947**** | **-10.644**** | **-7.759**** | **-2.086*** | **-2.508***** | **-2.346***** | **-1.834**** | **-1.403*** | **-1.829***** |
| | (-2.353) | (-2.707) | (-2.450) | (-1.769) | (-3.033) | (-3.394) | (-2.070) | (-2.020) | (-6.981) |
| *CONSC×POSTSOX* | 0.340** | 0.640** | 0.347* | 0.110 | 0.209* | 0.211** | 0.203** | 0.176** | 0.178*** |
| | (2.429) | (2.518) | (1.871) | (0.986) | (2.036) | (2.385) | (2.362) | (2.557) | (8.635) |
| *CONSC×ES99.5* | 3.277** | 10.114** | 7.246** | 1.950 | 2.492*** | 2.372*** | 1.886* | 1.613** | 2.282*** |
| | (2.739) | (2.755) | (2.369) | (1.661) | (2.876) | (3.124) | (2.034) | (2.445) | (8.520) |
| *POSTSOX×ES99.5* | 1.010** | 2.369** | 1.956** | 0.630** | 0.634** | 0.603*** | 0.472** | 0.350** | 0.506*** |
| | (2.279) | (2.325) | (2.423) | (2.148) | (2.698) | (3.480) | (2.118) | (2.066) | (5.212) |
| *POSTSOX* | -0.024 | -0.105** | -0.069 | -0.002 | -0.005 | -0.006 | -0.004 | 0.002 | -0.002 |
| | (-0.860) | (-2.144) | (-1.697) | (-0.122) | (-0.246) | (-0.408) | (-0.364) | (0.203) | (-0.504) |
| *ES99.5* | -0.898** | -2.291** | -1.952** | -0.539* | -0.642*** | -0.627*** | -0.494** | -0.421** | -0.622*** |
| | (-2.567) | (-2.397) | (-2.550) | (-1.899) | (-2.895) | (-3.713) | (-2.083) | (-2.622) | (-6.371) |
| *VEGA* | -0.001 | 0.001 | 0.001 | 0.000 | 0.001 | 0.001 | 0.001 | -0.001 | 0.001 |
| | (-0.091) | (0.353) | (0.359) | (0.040) | (0.223) | (0.556) | (0.339) | (-0.433) | (0.416) |
| *DELTA* | 0.002 | -0.006 | -0.002 | 0.004 | 0.005 | 0.003 | 0.002 | 0.005* | 0.004** |
| | (0.545) | (-1.160) | (-0.449) | (0.772) | (1.274) | (0.737) | (0.898) | (1.753) | (2.565) |
| *AGE* | -0.088 | 0.027 | 0.005 | -0.044 | -0.047 | -0.055 | -0.065 | -0.085** | -0.034 |
| | (-0.933) | (0.494) | (0.058) | (-0.583) | (-0.877) | (-1.257) | (-1.546) | (-2.338) | (-1.228) |
| *TENURE* | -0.005 | 0.004 | 0.004 | -0.001 | -0.002 | 0.002 | 0.003 | 0.003 | -0.001 |
| | (-0.459) | (0.330) | (0.302) | (-0.059) | (-0.248) | (0.235) | (0.451) | (0.465) | (-0.231) |
| *LNNPW* | -0.008 | -0.004 | -0.014*** | -0.018*** | -0.027*** | -0.027*** | -0.027*** | -0.028*** | -0.028*** |
| | (-0.878) | (-0.495) | (-3.234) | (-4.141) | (-6.235) | (-7.887) | (-7.587) | (-10.075) | (-14.325) |

Table 1.5 CEO conscientiousness and reserve error using SOX as an exogenous shock with financial risk mechanism Panel A (continued)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *NPWGROWTH* | 0.020 | 0.011 | 0.003 | -0.019 | -0.018 | -0.015 | -0.012 | -0.008 | -0.061*** |
| | (0.729) | (0.345) | (0.098) | (-0.838) | (-0.796) | (-0.832) | (-0.692) | (-0.435) | (-5.053) |
| *TAXSHIELD* | -0.151 | -0.112* | -0.020 | -0.021 | -0.063 | -0.088** | -0.083** | -0.082** | -0.029 |
| | (-1.523) | (-1.960) | (-0.413) | (-0.509) | (-1.390) | (-2.459) | (-2.112) | (-2.496) | (-1.025) |
| *LNQ* | 0.089 | 0.059 | 0.081 | 0.089 | 0.123* | 0.165** | 0.166** | 0.145** | 0.165*** |
| | (1.257) | (0.822) | (1.058) | (1.124) | (1.705) | (2.689) | (2.755) | (2.396) | (3.763) |
| *SMOOTH* | 0.562 | 0.715** | 0.223 | -0.014 | 0.201 | 0.249 | 0.301** | 0.323*** | 0.091 |
| | (1.620) | (2.624) | (0.960) | (-0.082) | (1.088) | (1.651) | (2.453) | (3.242) | (1.387) |
| *REAFFILIATE* | 0.014 | 0.083* | 0.019 | -0.029 | 0.006 | 0.038 | 0.042 | 0.017 | 0.123*** |
| | (0.227) | (1.747) | (0.403) | (-0.461) | (0.084) | (0.613) | (0.894) | (0.337) | (3.594) |
| *RENONAFFILIATE* | -0.011 | -0.125 | -0.144 | -0.048 | -0.096 | -0.151** | -0.134** | -0.135** | -0.064 |
| | (-0.113) | (-1.414) | (-1.447) | (-0.576) | (-1.325) | (-2.382) | (-2.292) | (-2.619) | (-1.628) |
| *LRGROWTH* | -0.008 | -0.015 | 0.009 | -0.010 | -0.011 | -0.005 | -0.007 | -0.008 | -0.008 |
| | (-0.404) | (-0.296) | (0.482) | (-0.536) | (-0.453) | (-0.426) | (-0.548) | (-0.780) | (-1.155) |
| *LONGTAIL* | 0.044* | 0.073*** | 0.050** | 0.018 | 0.043 | 0.045* | 0.045** | 0.046*** | 0.032** |
| | (1.964) | (3.016) | (2.293) | (0.602) | (1.266) | (2.010) | (2.682) | (3.261) | (2.284) |
| *PRODHHI* | 0.004 | 0.027 | -0.002 | 0.001 | 0.004 | -0.008 | -0.005 | -0.006 | -0.009 |
| | (0.157) | (0.993) | (-0.106) | (0.060) | (0.148) | (-0.352) | (-0.224) | (-0.292) | (-0.665) |
| *GEOHHI* | -0.070 | -0.048 | -0.111*** | -0.121*** | -0.146*** | -0.151*** | -0.161*** | -0.173*** | -0.043 |
| | (-1.392) | (-1.307) | (-3.637) | (-4.978) | (-5.916) | (-4.990) | (-6.799) | (-10.140) | (-0.577) |
| *WEAK* | -0.050** | -0.059* | -0.108** | -0.102** | -0.052 | -0.015 | -0.013 | 0.000 | 0.030 |
| | (-2.157) | (-1.984) | (-2.505) | (-2.421) | (-0.398) | (-0.295) | (-0.276) | (0.001) | (1.256) |
| *LNDD* | 0.001 | 0.006 | 0.003 | 0.004 | -0.002 | -0.002 | -0.001 | -0.002 | -0.001 |
| | (0.125) | (1.039) | (0.484) | (0.759) | (-0.336) | (-0.371) | (-0.177) | (-0.486) | (-0.509) |
| Intercept | 0.485 | -0.027 | 0.228 | 0.444 | 0.613*** | 0.665*** | 0.699*** | 0.789*** | 0.593*** |
| | (1.520) | (-0.109) | (0.709) | (1.602) | (3.213) | (3.721) | (4.083) | (4.988) | (4.940) |
| Year FE | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table 1.5 Panel B: Financial risk measured by Value at risk with confidence levels of 99.5 (VAR99.5)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *CONSC* | -0.275** | -0.628** | -0.359* | -0.119 | -0.251** | -0.279** | -0.267*** | -0.265*** | -0.263*** |
| | (-2.457) | (-2.318) | (-1.819) | (-1.037) | (-2.260) | (-2.732) | (-3.182) | (-3.566) | (-10.200) |
| ***CONSC×POSTSOX×VAR99.5*** | **-4.665**** | **-12.478**** | **-8.997**** | **-2.157*** | **-2.966***** | **-2.761***** | **-1.890**** | **-1.627*** | **-1.977***** |
| | (-2.373) | (-2.657) | (-2.378) | (-1.773) | (-2.980) | (-3.290) | (-2.142) | (-1.932) | (-6.283) |
| *CONSC×POSTSOX* | 0.355** | 0.662** | 0.378* | 0.105 | 0.216** | 0.218** | 0.197** | 0.178** | 0.176*** |
| | (2.456) | (2.477) | (1.882) | (0.896) | (2.062) | (2.403) | (2.500) | (2.514) | (8.504) |
| *CONSC×VAR99.5* | 3.959** | 11.790** | 8.343** | 2.013 | 2.946*** | 2.790*** | 1.972** | 1.872** | 2.504*** |
| | (2.763) | (2.563) | (2.266) | (1.637) | (2.836) | (3.054) | (2.071) | (2.319) | (7.781) |
| *POSTSOX×VAR99.5* | 1.193** | 2.768** | 2.222** | 0.661** | 0.747** | 0.709*** | 0.489** | 0.407* | 0.541*** |
| | (2.297) | (2.249) | (2.303) | (2.296) | (2.708) | (3.448) | (2.196) | (2.028) | (4.605) |
| *POSTSOX* | -0.027 | -0.112** | -0.070 | -0.001 | -0.007 | -0.008 | -0.003 | 0.001 | -0.001 |
| | (-0.949) | (-2.147) | (-1.663) | (-0.068) | (-0.332) | (-0.509) | (-0.237) | (0.072) | (-0.354) |
| *VAR99.5* | -1.069** | -2.666** | -2.212** | -0.558* | -0.757*** | -0.738*** | -0.518** | -0.490** | -0.678*** |
| | (-2.582) | (-2.230) | (-2.389) | (-1.965) | (-2.903) | (-3.709) | (-2.226) | (-2.540) | (-5.760) |
| *VEGA* | -0.001 | 0.000 | 0.002 | 0.000 | 0.001 | 0.002 | 0.001 | -0.001 | 0.000 |
| | (-0.094) | (0.004) | (0.475) | (0.025) | (0.220) | (0.608) | (0.298) | (-0.463) | (0.284) |
| *DELTA* | 0.002 | -0.005 | -0.002 | 0.004 | 0.005 | 0.003 | 0.003 | 0.005* | 0.004** |
| | (0.546) | (-1.127) | (-0.464) | (0.814) | (1.279) | (0.699) | (0.817) | (1.751) | (2.672) |
| *AGE* | -0.089 | 0.007 | 0.007 | -0.045 | -0.047 | -0.054 | -0.065 | -0.085** | -0.034 |
| | (-0.939) | (0.106) | (0.085) | (-0.556) | (-0.883) | (-1.215) | (-1.505) | (-2.687) | (-1.215) |
| *TENURE* | -0.005 | 0.006 | 0.004 | -0.001 | -0.002 | 0.002 | 0.003 | 0.003 | -0.000 |
| | (-0.454) | (0.533) | (0.249) | (-0.065) | (-0.245) | (0.258) | (0.407) | (0.545) | (-0.159) |
| *LNNPW* | -0.008 | -0.003 | -0.013*** | -0.018*** | -0.027*** | -0.027*** | -0.027*** | -0.028*** | -0.027*** |
| | (-0.878) | (-0.566) | (-3.117) | (-4.217) | (-6.233) | (-7.932) | (-8.361) | (-10.295) | (-13.819) |

Table 1.5 CEO conscientiousness and reserve error using SOX as an exogenous shock with financial risk mechanism Panel B (continued)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| *NPWGROWTH* | 0.020 | 0.008 | 0.002 | -0.018 | -0.018 | -0.015 | -0.012 | -0.009 | -0.059*** |
| | (0.739) | (0.235) | (0.061) | (-0.822) | (-0.783) | (-0.824) | (-0.689) | (-0.492) | (-4.982) |
| *TAXSHIELD* | -0.151 | -0.127** | -0.016 | -0.020 | -0.063 | -0.087** | -0.082** | -0.083** | -0.037 |
| | (-1.529) | (-2.401) | (-0.322) | (-0.439) | (-1.384) | (-2.446) | (-2.079) | (-2.577) | (-1.312) |
| *LNQ* | 0.088 | 0.063 | 0.072 | 0.089 | 0.122 | 0.167** | 0.167*** | 0.147** | 0.170*** |
| | (1.242) | (0.863) | (1.010) | (1.033) | (1.694) | (2.676) | (2.770) | (2.437) | (3.906) |
| *SMOOTH* | 0.565 | 0.714*** | 0.224 | -0.018 | 0.201 | 0.247 | 0.301** | 0.318*** | 0.098 |
| | (1.623) | (2.835) | (0.946) | (-0.108) | (1.088) | (1.641) | (2.488) | (3.228) | (1.510) |
| *REAFFILIATE* | 0.014 | 0.089* | 0.020 | -0.030 | 0.006 | 0.041 | 0.042 | 0.020 | 0.123*** |
| | (0.229) | (1.865) | (0.434) | (-0.466) | (0.084) | (0.661) | (0.892) | (0.397) | (3.604) |
| *RENONAFFILIATE* | -0.011 | -0.145 | -0.136 | -0.045 | -0.095 | -0.151** | -0.134** | -0.139*** | -0.067* |
| | (-0.110) | (-1.626) | (-1.370) | (-0.475) | (-1.315) | (-2.377) | (-2.296) | (-2.796) | (-1.712) |
| *LRGROWTH* | -0.008 | -0.021 | 0.006 | -0.010 | -0.011 | -0.005 | -0.007 | -0.008 | -0.006 |
| | (-0.400) | (-0.417) | (0.295) | (-0.548) | (-0.456) | (-0.402) | (-0.545) | (-0.750) | (-1.018) |
| *LONGTAIL* | 0.044* | 0.070*** | 0.042 | 0.018 | 0.043 | 0.045** | 0.046** | 0.046*** | 0.034** |
| | (1.958) | (2.902) | (1.465) | (0.635) | (1.259) | (2.055) | (2.742) | (3.390) | (2.467) |
| *PRODHHI* | 0.004 | 0.023 | 0.002 | 0.001 | 0.004 | -0.008 | -0.005 | -0.007 | -0.007 |
| | (0.157) | (0.841) | (0.104) | (0.034) | (0.159) | (-0.360) | (-0.222) | (-0.354) | (-0.550) |
| *GEOHHI* | -0.069 | -0.050 | -0.109*** | -0.121*** | -0.146*** | -0.151*** | -0.162*** | -0.173*** | -0.036 |
| | (-1.389) | (-1.361) | (-3.585) | (-4.984) | (-5.914) | (-6.797) | (-6.827) | (-10.434) | (-0.453) |
| *WEAK* | -0.050** | -0.062** | -0.116** | -0.103** | -0.053 | -0.015 | -0.013 | 0.001 | 0.026 |
| | (-2.179) | (-2.192) | (-2.557) | (-2.441) | (-0.395) | (-0.290) | (-0.272) | (0.048) | (1.192) |
| *LNDD* | 0.001 | 0.006 | 0.003 | 0.004 | -0.002 | -0.002 | -0.001 | -0.001 | -0.002 |
| | (0.115) | (1.167) | (0.583) | (0.722) | (-0.330) | (-0.366) | (-0.171) | (-0.417) | (-0.718) |
| Intercept | 0.491 | 0.059 | 0.214 | 0.451 | 0.615*** | 0.662*** | 0.699*** | 0.791*** | 0.580*** |
| | (1.530) | (0.192) | (0.643) | (1.511) | (3.222) | (3.662) | (3.978) | (5.938) | (4.727) |
| Year FE | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table 1.6: CEO conscientiousness and compensation

This table presents the results of CEO conscientiousness and compensation. The dependent variable is the natural log of CEO's total compensation (*COMPENSATION*). All the variables are defined in Appendix B. The *t*-statistics are shown in parentheses. The standard deviations are clustered at firm level. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| Variable | OLS |
|---|---|
| ***CONSC*** | **0.711**** |
| | (2.304) |
| *VEGA* | 0.030 |
| | (0.760) |
| *DELTA* | 0.170** |
| | (2.258) |
| *AGE* | -1.293 |
| | (-1.315) |
| *TENURE* | -0.332** |
| | (-2.689) |
| *CHAIRMAN* | 0.261* |
| | (1.936) |
| *LNNPW* | 0.353*** |
| | (4.940) |
| *ROA* | 0.872 |
| | (0.882) |
| *RERATIO* | 1.273*** |
| | (3.132) |
| *TAXEXEMPT* | -0.613*** |
| | (-3.154) |
| *STDROA* | 14.197*** |
| | (4.113) |
| *STDROI* | 4.303 |
| | (0.941) |
| *STDLOSSRATIO* | 0.592 |
| | (1.466) |
| Intercept | 7.465* |
| | (1.885) |
| Year FE | Yes |
| Observations | 224 |
| Adjusted R-squared | 0.641 |

Table 1.7: CEO conscientiousness and compensation using the financial crisis as an exogenous shock

This table presents the results of the interaction term between conscientiousness and the financial crisis of 2008. *CRISIS* equals 1 if observations are during 2007-2009 and 0 otherwise. The dependent variable is the natural log of CEO compensation (*COMPENSATION*). All the variables are defined in Appendix B. The *t*-statistics are shown in parentheses. The standard deviations are clustered at firm level. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| Variable | OLS |
|---|---|
| *CONSC* | 0.033 |
| | (0.090) |
| ***CONSC×CRISIS*** | **1.033*** |
| | (1.882) |
| *CRISIS* | -0.167 |
| | (-1.522) |
| *VEGA* | -0.011 |
| | (-0.249) |
| *DELTA* | 0.139* |
| | (1.901) |
| *AGE* | -1.295 |
| | (-1.327) |
| *TENURE* | -0.208* |
| | (-1.749) |
| *CHAIRMAN* | 0.223 |
| | (1.584) |
| *LNNPW* | 0.365*** |
| | (4.807) |
| *ROA* | 0.859 |
| | (0.744) |
| *RERATIO* | 1.385*** |
| | (3.026) |
| *TAXEXEMPT* | -0.595*** |
| | (-3.041) |
| *STDROA* | 13.494*** |
| | (3.637) |
| *STDROI* | 2.360 |
| | (0.467) |
| *STDLOSSRATIO* | 0.418 |
| | (1.162) |
| Intercept | 7.436* |
| | (1.915) |
| Year FE | NO |
| Observations | 224 |
| Adjusted R-squared | 0.582 |

# APPENDIX A: THE METHOD OF MAIRESSE ET AL. (2007)

Mairesse et al. (2007) develop four well-performed algorithms for scoring Big Five traits using continuous scales. The dependent variable of the algorithm is the individual trait score, and the independent variables are linguistic features. They train the model using two data samples. The first data sample is from Pennebaker and King (1999), containing 2,479 essays written by psychology students. Each essay is associated with a self-reported personality traits score.[28] The second data sample consists of conversations of 96 participants during a 2-day monitoring period and is recorded by the Electronically Activated Recorder (EAR) (Mehl et al., 2001 and 2006). Then, conversations are transcribed to text by well-trained research assistants. The participants also self-report their personality traits score. In addition, 18 independent observers also rated the participants' traits scores using a 7-point scale based on the description of Big Five traits from John and Srivastava (1999).

John and Srivastava (1999) describe conscientiousness as follows.

| Conscientiousness | |
|---|---|
| Low | High |
| Careless<br>Disorderly<br>Frivolous<br>Irresponsible<br>Slipshod<br>Undependable<br>Forgetful | Organized<br>Thorough<br>Planful<br>Efficient<br>Responsible<br>Reliable<br>Dependable<br>Conscientious<br>Precise<br>Practical<br>Deliberate<br>Painstaking<br>Cautious |

[28] The score reported is based on 5-point scale questionnaires from John et al. (1999).

These two data samples are trained by 4 algorithms: linear regression model, support vector regression, M5' model tree, and M5' regression tree. The features[29] used in each algorithm are: 88 Linguistic Inquiry and Word Count (LIWC) features (Pennebaker et al. 2001), 14 Machine Readable Cataloguing (MRC) features (Coltheart, 1981).[30] These features are related to the content and syntax. After the training of each algorithm, the personality scores are obtained. The final scores are the average scores from the 4 algorithms which apply 10-fold cross-validation to maximize out-of-sample prediction ability.[31] Mairesse et al. (2007) employ these algorithms to find out which LIWC and MRC features are significantly related to conscientiousness. They claim that the main feature of conscientiousness is the avoidance of using negative emotion words (e.g., fear, anger, depression, sadness). The other features of conscientiousness are described below. Conscientious people talk more about jobs and occupations, which are defined as content related to personal concerns in LIWC. They prefer to use longer words (e.g., words longer than six letters, number of syllables in the word), words related to communication (e.g., talk, listen, share), insight words capturing the sense of understanding or learning (e.g., think, know, consider), words acquired late by children, prompts (e.g., yeah, OK, huh), positive emotion words (e.g., happy, love, nice). They use fewer swear words and fewer pronouns (e.g., I, them, itself).

[29] Features mean independent variables used in the training algorithm.
[30] The list of features can be found in Table 6 of Mairesse et al. (2007)
[31] The 10-fold cross-validation means that the sample is randomly divided into 10 subsamples, using 9 subsamples as training dataset and the rest 1 subsample as test dataset.

# APPENDIX B: VARIABLE DEFINITIOS

| Variable | Definition |
|---|---|
| **Reserve management** | |
| *RESERROR* | The difference between the cumulative incurred loss at time *t* and the cumulative incurred loss at time *t* + 5, scaled by total admitted assets. (NAIC) |
| **CEO variables** | |
| *CONSC* | For each CEO and for a conference call, conscientiousness trait scores are generated using the four linguistic algorithms provided by *Personality Recognizer* application. These four scores are winsorized at 1st and 99th percentiles and averaged to get a call-level score. Then, a CEO-level conscientiousness trait score is estimated by taking a weighted average of all the call-level scores by the number of words spoken by the CEO in the Q&A section of each call. This CEO-level conscientiousness trait score is assigned to all the data points related to the CEO. Detailed discussion is in Section 2.2. (LexisNexis) |
| *HIGHCONSC* | Takes a value of 1 if a CEO's conscientiousness score falls under the top tercile group and 0 otherwise. |
| *VEGA* | The natural log of dollar changes in CEO wealth associated with a 0.01 change in the standard deviation of the firm's returns. (Execucomp) |
| *DELTA* | The natural log of dollar changes in CEO wealth associated with a 1% change in the firm's stock price. (Execucomp) |
| *AGE* | The natural log of CEO age. (Execucomp) |
| *TENURE* | The natural log of CEO tenure. (Execucomp) |
| *COMPENSATION* | The natural log of CEO total compensation. (Excucomp) |
| *CHAIRMAN* | The Chairman takes the value of 1 if CEO is also the board chairman, 0 otherwise. (Execucomp) |
| **Firm variables** | |
| *LNNPW* | The natural log of net premium written. (NAIC) |
| *NPWGROWTH* | 1-year increase of net premium written. (NAIC) |
| *TAXSHIELD* | The sum of net income and estimated loss reserve divided by total assets. (NAIC) |
| *LNQ* | The natural log of Tobin's Q. (Compustat) |
| *SMOOTH* | The previous 3 years' average ROA. (NAIC) |
| *REAFFILIATE* | The ratio of reinsurance ceded to affiliated reinsurers to total direct premium plus reinsurance assumed. (NAIC) |
| *RENONAFFILIATE* | The ratio of reinsurance ceded to nonaffiliated reinsurers to total direct premium plus reinsurance assumed. (NAIC) |
| *RERATIO* | The ratio of reinsurance ceded to reinsurers to total direct premium plus reinsurance assumed. (NAIC) |
| *LRGROWTH* | 1-year increase in loss ratio. (NAIC) |
| *LONGTAIL* | The net loss incurred in long-tailed lines of business over the net loss incurred in whole business lines (Phillips et al., 1998). (NAIC) |

APPENDIX B: VARIABLE DEFINITIOS (continued)

| | |
|---|---|
| *PRODHHI* | The line of business Herfindahl Index. (NAIC) |
| *GEOHHI* | The geographical Herfindahl Index. (NAIC) |
| *WEAK* | Takes a value of 1 if the insurer has more than 3 out of the range IRIS ratios and 0 otherwise. (NAIC) |
| *LNDD* | The natural log of naive distance-to-default (Bharath and Shumway 2008). (CRSP) |
| *ROA* | The operating income before depreciation divided by the book value of assets. (Compustat) |
| *TAXEXEMPT* | The tax-exempt income divided by total investment income, to capture low-risk investment (D'Arcy and Garven, 1990). (NAIC) |
| **Firm risk variables** | |
| *ES* | Expected shortfall (ES) is defined as the conditional expected loss using 1 year of daily firm stock returns, which is beyond the VaR. (CRSP) |
| *VAR* | Value at risk (VaR) is the maximum expected loss that could occur using 1 year of daily firm stock returns at a specified confidence level. (CRSP) |
| *STDROA* | The five-year standard deviation of returns on assets (ROA). (NAIC) |
| *STDROI* | The five-year standard deviation of returns on investment (ROI). (NAIC) |
| *STDLOSSRATIO* | The five-year standard deviation of loss ratio. (NAIC) |
| **External shock** | |
| *POSTSOX* | Equals 1 from 2005 to 2015 and equals 0 from 2002 to 2004. |
| *CRISIS* | Equals 1 if observations are during 2007-2009 and 0 otherwise. |

# APPENDIX C: ADDITIONAL RESULTS

Table C.1: CEO extraversion trait and reserve error

This table presents the results of the the relation between the CEO extraversion trait and reserve errors using the OLS and quantile regression methods. The dependent variable is reserve error (*RESERROR)*, and the main independent variable is CEO extraversion (*EXTRA*). The regressions include all the controls from Table 1.2. All the variables are defined in Appendix B. The *t*-statistics are shown in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***EXTRA*** | **0.052** | **0.031** | **0.009** | **0.007** | **0.017** | **0.026** | **0.020** | **0.012** | **0.007** |
| | (1.622) | (1.196) | (0.511) | (0.278) | (0.904) | (1.659) | (1.284) | (0.818) | (0.721) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.153 | -0.409 | -0.333 | 0.234 | 0.547*** | 0.740*** | 0.593*** | 0.498*** | 0.567*** |
| | (0.614) | (-1.438) | (-1.015) | (0.630) | (3.094) | (3.841) | (3.666) | (3.122) | (4.733) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table C.2: CEO versus CFO conscientiousness and reserve error

Panel A of this table presents summary statistics of CFO variables. Panel B reports the results of the relation between CFO conscientiousness and reserve error using the OLS and quantile regression methods. Panel C reports the results of the relation between CFO versus CEO conscientiousness and reserve error using the OLS and quantile regression methods. CFO variables are computed the same as CEO variables. All the other variables are defined in Appendix B. The *t*-statistics are reported in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

Panel A: Summary Statistics of CFO variables

| Variable | N | MEAN | SD | MIN | P25 | P50 | P75 | MAX |
|---|---|---|---|---|---|---|---|---|
| *CFOCONSC* | 113 | 0.041 | 0.182 | -0.443 | -0.091 | 0.045 | 0.224 | 0.399 |
| *CFOVEGA* | 113 | 2.216 | 2.232 | -15.455 | 1.593 | 2.488 | 3.367 | 4.640 |
| *CFODELTA* | 113 | 3.185 | 1.536 | -1.026 | 2.326 | 3.222 | 4.361 | 6.071 |
| *CFOAGE* | 113 | 3.933 | 0.119 | 3.611 | 3.850 | 3.932 | 4.025 | 4.190 |
| *CFOTENURE* | 113 | 1.475 | 0.714 | 0.000 | 1.099 | 1.609 | 2.079 | 2.833 |

Table C.2 Panel B: CFO conscientiousness and reserve error

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Variable | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CFOCONSC*** | **-0.008** | **-0.031** | **-0.007** | **-0.023** | **-0.033*** | **-0.037*** | **-0.039**** | **-0.035**** | **-0.028***** |
| | (-0.331) | (-0.897) | (-0.178) | (-0.947) | (-2.051) | (-1.913) | (-2.279) | (-2.772) | (-3.045) |
| *CFOVEGA* | 0.000 | 0.004 | 0.004 | -0.001 | -0.002 | -0.002 | -0.001 | -0.002 | -0.001 |
| | (0.060) | (0.538) | (0.718) | (-0.552) | (-1.094) | (-0.665) | (-0.821) | (-1.279) | (-1.684) |
| *CFODELTA* | 0.006 | -0.003 | -0.000 | 0.002 | 0.004 | 0.006 | 0.005 | 0.007*** | 0.009*** |
| | (1.449) | (-0.388) | (-0.030) | (0.544) | (1.312) | (1.626) | (1.636) | (3.650) | (6.353) |
| *CFOAGE* | -0.063 | 0.014 | -0.041 | -0.004 | 0.007 | -0.007 | 0.027 | -0.006 | -0.006 |
| | (-0.871) | (0.175) | (-0.480) | (-0.070) | (0.127) | (-0.095) | (0.521) | (-0.148) | (-0.243) |
| *CFOTENURE* | 0.004 | 0.002 | 0.006 | 0.010 | 0.009 | 0.009 | 0.010* | 0.008 | 0.008*** |
| | (0.465) | (0.176) | (0.452) | (0.942) | (1.194) | (1.543) | (2.068) | (1.677) | (3.430) |
| *Firm controls* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.567** | 0.264 | 0.501* | 0.394** | 0.444** | 0.560** | 0.473*** | 0.579*** | 0.570*** |
| | (2.466) | (0.942) | (1.900) | (2.372) | (2.434) | (2.543) | (3.166) | (4.113) | (7.109) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 113 | 113 | 113 | 113 | 113 | 113 | 113 | 113 | 113 |

Table C.2 Panel C: CEO versus CFO conscientiousness and reserve error

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC*** | **-0.073*** | **-0.083*** | **-0.087*** | **-0.072*** | **-0.101***** | **-0.094**** | **-0.117***** | **-0.120***** | **-0.120***** |
| | (-1.992) | (-1.811) | (-2.027) | (-1.954) | (-3.512) | (-2.503) | (-3.184) | (-5.325) | (-10.866) |
| *VEGA* | 0.006** | 0.004 | 0.007** | 0.001 | 0.001 | -0.002 | -0.004 | -0.004 | -0.004** |
| | (2.361) | (1.097) | (2.193) | (0.716) | (0.176) | (-0.481) | (-1.094) | (-1.268) | (-2.334) |
| *DELTA* | 0.001 | -0.004 | -0.004 | -0.004 | -0.008** | -0.004 | -0.005* | -0.005** | -0.005*** |
| | (0.297) | (-0.580) | (-0.488) | (-1.250) | (-2.616) | (-1.204) | (-1.782) | (-2.493) | (-4.489) |
| *AGE* | 0.043 | 0.058 | 0.078 | 0.039 | 0.047* | 0.012 | -0.001 | -0.000 | -0.000 |
| | (0.873) | (0.943) | (1.283) | (1.670) | (1.737) | (0.352) | (-0.034) | (-0.009) | (-0.018) |
| *TENURE* | 0.006 | 0.008 | 0.002 | 0.005 | 0.008* | 0.005 | 0.013* | 0.013*** | 0.013*** |
| | (0.860) | (0.702) | (0.243) | (1.254) | (1.971) | (0.501) | (2.019) | (3.074) | (5.900) |
| ***CFOCONSC*** | **-0.032** | **-0.007** | **-0.007** | **-0.013** | **-0.021** | **-0.030** | **-0.024** | **-0.022** | **-0.022**** |
| | (-1.621) | (-0.236) | (-0.284) | (-1.092) | (-1.237) | (-1.203) | (-0.888) | (-1.112) | (-2.370) |
| *CFOVEGA* | -0.003 | -0.002 | -0.002 | -0.001 | 0.001 | -0.001 | 0.003 | 0.003 | 0.003*** |
| | (-1.421) | (-0.697) | (-0.881) | (-0.578) | (0.395) | (-0.513) | (1.013) | (1.544) | (3.391) |
| *CFODELTA* | 0.009* | 0.008 | 0.010* | 0.009** | 0.011*** | 0.013** | 0.013*** | 0.013*** | 0.013*** |
| | (1.736) | (1.350) | (1.861) | (2.342) | (2.952) | (2.398) | (3.229) | (4.467) | (8.220) |
| *CFOAGE* | -0.120* | -0.080 | -0.122* | -0.036 | -0.038 | -0.024 | -0.012 | -0.014 | -0.014 |
| | (-1.862) | (-0.978) | (-1.886) | (-0.673) | (-0.761) | (-0.355) | (-0.166) | (-0.256) | (-0.468) |
| *CFOTENURE* | 0.007 | 0.010 | 0.014 | 0.008* | 0.010* | 0.001 | -0.005 | -0.006 | -0.006** |
| | (0.996) | (0.901) | (1.476) | (1.801) | (2.084) | (0.249) | (-0.712) | (-1.120) | (-2.143) |
| *Firm controls* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.542* | 0.215 | 0.341 | 0.246 | 0.249 | 0.502* | 0.491** | 0.496** | 0.496*** |
| | (2.036) | (0.694) | (0.956) | (1.261) | (1.440) | (1.962) | (2.654) | (2.559) | (4.759) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 110 | 110 | 110 | 110 | 110 | 110 | 110 | 110 | 110 |

Table C.3: CEO conscientiousness and reserve error with financial risk mechanism

This table presents the results of the interaction term model, testing the relation between CEO conscientiousness and reserve errors with financial risk mechanism using the OLS and quantile regression methods. The dependent variable is reserve error (*RESERROR)*. In Panels A and B, financial risk is measured by Expected shortfall with confidence levels of 99 and 95, respectively. In Panels C and D, financial risk is measured by Value at risk with confidence levels of 99 and 95, respectively. All the panels include respective all the other independent variables from Table 1.3. All the variables are defined in Appendix B. The *t*-statistics are shown in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

Panel A: Financial risk measured by Expected shortfall with confidence levels of 99 (ES99)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC* × *ES99*** | **0.274** | **0.195** | **0.302** | **0.151** | **0.690*** | **0.839**** | **0.852***** | **0.713***** | **1.016***** |
| | (0.488) | (0.285) | (0.473) | (0.340) | (1.939) | (2.733) | (3.304) | (3.684) | (13.038) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.191 | -0.426 | -0.362 | 0.181 | 0.407** | 0.474*** | 0.494*** | 0.497*** | 0.447*** |
| | (0.700) | (-1.505) | (-1.110) | (0.521) | (2.055) | (2.838) | (3.534) | (4.539) | (4.929) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

*Panel B: Financial risk measured by Expected shortfall with confidence levels of 95 (ES95)*

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC* × *ES95*** | **0.345** | **0.251** | **0.382** | **0.196** | **0.893*** | **1.074**** | **1.105***** | **0.910***** | **1.318***** |
| | (0.477) | (0.286) | (0.466) | (0.357) | (1.949) | (2.712) | (3.316) | (3.660) | (13.163) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.191 | -0.426 | -0.363 | 0.181 | 0.406** | 0.472** | 0.497*** | 0.498*** | 0.450*** |
| | (0.700) | (-1.506) | (-1.112) | (0.539) | (2.052) | (2.758) | (3.544) | (4.551) | (4.965) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table C.3 CEO conscientiousness and reserve error with financial risk mechanism (continued)

Panel C: Financial risk measured by Value at risk with confidence levels of 99 (VAR99)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC* × *VAR99*** | **0.374** | **0.234** | **0.399** | **0.159** | **0.830*** | **1.022**** | **0.963***** | **0.992***** | **1.184***** |
| | (0.566) | (0.290) | (0.550) | (0.293) | (1.876) | (2.690) | (3.217) | (4.144) | (12.754) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.191 | -0.425 | -0.353 | 0.176 | 0.399** | 0.480*** | 0.483*** | 0.501*** | 0.439*** |
| | (0.700) | (-1.505) | (-1.086) | (0.511) | (2.093) | (2.895) | (3.419) | (4.370) | (4.834) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Panel D: Financial risk measured by Value at risk with confidence levels of 95 (VAR95)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC* × *VAR95*** | **0.548** | **0.335** | **0.541** | **0.214** | **1.169*** | **1.424**** | **1.376***** | **1.422***** | **1.690***** |
| | (0.584) | (0.285) | (0.522) | (0.276) | (1.844) | (2.749) | (3.219) | (4.116) | (12.749) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.191 | -0.425 | -0.346 | 0.176 | 0.406** | 0.475*** | 0.484*** | 0.487*** | 0.442*** |
| | (0.700) | (-1.500) | (-1.073) | (0.549) | (2.227) | (2.868) | (3.236) | (4.195) | (4.858) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table C.4: CEO conscientiousness and reserve error using SOX as an exogenous shock with financial risk mechanism

This table presents the relation between CEO conscientiousness and reserve error with financial risk mechanism of the SOX period using the OLS and quantile regression methods. A dummy variable *POSTSOX* equals 1 from 2005 to 2015 and equals 0 from 2002 to 2004. The dependent variable is reserve error (*RESERROR*). In Panels A and B, financial risk is measured by Expected shortfall with confidence levels of 99 and 95, respectively. In Panels C and D, financial risk is measured by Value at risk with confidence levels of 99 and 95, respectively. All the panels include respective all the other independent variables from Table 1.5. All the variables are defined in Appendix B. The *t*-statistics are shown in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

Panel A: Financial risk measured by Expected shortfall with confidence levels of 99 (ES99)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC* × *POSTSOX* × *ES99*** | **-4.273**** | **-11.521**** | **-8.398**** | **-2.275*** | **-2.715***** | **-2.541***** | **-2.006**** | **-1.520*** | **-1.987***** |
| | (-2.353) | (-2.704) | (-2.452) | (-1.744) | (-3.035) | (-3.397) | (-2.129) | (-2.023) | (-7.006) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.485 | -0.030 | 0.229 | 0.444 | 0.613*** | 0.665*** | 0.699*** | 0.789*** | 0.593*** |
| | (1.520) | (-0.122) | (0.715) | (1.601) | (3.212) | (3.722) | (4.079) | (4.984) | (4.946) |
| Year FE | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Panel B: Financial risk measured by Expected shortfall with confidence levels of 95 (ES95)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC* × *POSTSOX* × *ES95*** | **-5.479**** | **-14.748**** | **-10.753**** | **-3.002** | **-3.482***** | **-3.262***** | **-2.317**** | **-1.951*** | **-2.580***** |
| | (-2.350) | (-2.691) | (-2.453) | (-1.632) | (-3.042) | (-3.404) | (-2.209) | (-2.032) | (-7.103) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.484 | -0.042 | 0.232 | 0.442 | 0.612*** | 0.665*** | 0.698*** | 0.789*** | 0.594*** |
| | (1.518) | (-0.171) | (0.735) | (1.600) | (3.211) | (3.721) | (3.920) | (4.972) | (4.963) |
| Year FE | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Table C.4 CEO conscientiousness and reserve error using SOX as an exogenous shock with financial risk mechanism (continued)

Panel C: Financial risk measured by Value at risk with confidence levels of 99 (VAR99)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC*** × ***POSTSOX*** × ***VAR99*** | **-5.181**** | **-13.853**** | **-9.972**** | **-2.375*** | **-3.294***** | **-3.066***** | **-2.089**** | **-1.802*** | **-2.183***** |
| | (-2.374) | (-2.656) | (-2.376) | (-1.767) | (-2.977) | (-3.286) | (-2.117) | (-1.922) | (-6.240) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.491 | 0.058 | 0.213 | 0.451 | 0.616*** | 0.661*** | 0.700*** | 0.791*** | 0.580*** |
| | (1.531) | (0.188) | (0.640) | (1.511) | (3.223) | (3.661) | (3.980) | (5.938) | (4.712) |
| Year FE | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

Panel D: Financial risk measured by Value at risk with confidence levels of 95 (VAR95)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | q0.10 | q0.25 | q0.50 | q0.75 | q0.80 | q0.85 | q0.90 | q0.95 |
| ***CONSC*** × ***POSTSOX*** × ***VAR95*** | **-7.418**** | **-19.822**** | **-14.169**** | **-3.288*** | **-4.719***** | **-4.395***** | **-2.935*** | **-2.557*** | **-3.053***** |
| | (-2.381) | (-2.653) | (-2.510) | (-1.746) | (-2.961) | (-3.391) | (-2.005) | (-1.876) | (-6.053) |
| *Other variables* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Intercept | 0.493 | 0.053 | 0.211 | 0.460 | 0.616*** | 0.661*** | 0.701*** | 0.789*** | 0.576*** |
| | (1.533) | (0.172) | (0.636) | (1.524) | (3.226) | (3.369) | (3.994) | (5.938) | (4.646) |
| Year FE | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 | 244 |

# CHAPTER 2: CORPORATE OPACITY AND NET PREMIUM WRITTEN FLOWS: EVIDENCE FROM U.S. PROPERTY-LIABILITY INSURERS

## 2.1. Introduction

The transparency of insurers refers to reliable and high information quality, which is valued by stakeholders. The transparency of financial institutions such as banks and insurers is an important issue because they are typically considered more opaque than non-financial corporations (Park, 2008). A growing literature suggests that corporates' transparency can improve the resource allocation process, lower transaction costs, reduce the cost of debt, and affect firms' valuation (Francis et al., 2009; Eckles et al., 2014; Lang et al., 2012). Han et al. (2018) find that transparency is positively associated with more conservative loss reserve for the property-liability insurers. Chen et al. (2022) argue that depositors are aware of information quality, which shapes their behavior, especially for uninsured deposits.

For the insurance industry, due to the nature of the complexity of liability structure and many different business lines, whether a typical policyholder's purchase behavior will be shaped by opacity is an open question. Policyholders are the main debtholders in the insurance industry. When policyholders buy insurance policies, insurers need to set up reserve to pay for future losses. The reason is that premiums are paid at the beginning of the policies, but the losses are paid during the policy period. In other words, reserves in the insurance industry are similar to the debt of non-financial industries and policyholders are like debtholders. The information quality (opacity) is valuable for current and prospective policyholders (debtholders) because policyholders would not be paid in full if insurers become insolvent. Yet, prospective policyholders may not have the incentives and expertise to understand the quality of information about insurers' financial health due to the complexity of the liability structure of insurers. Thus, whether the opacity of insurers has an impact on insurance purchase behavior is an empirical question.

This paper examines the relation between insurers' opacity and the net premium written flows, which reflect the insurance purchase behavior. The lack of transparency may lower policyholders' utility regarding the information risk of whether an insurer's information is reliable. A lower level of opacity means a higher level of outside monitoring, lowering information asymmetry. Everything else equal, insurers with a lower level of opacity indicate that their financial statement disclosures contain more reliable information. Transparency can enhance policyholders' belief that their future claims would be paid in full when they incurred losses. Thus, we expect a negative relationship between opacity and insurance purchases.

Policyholders are concerned about insurers' insolvency, especially when the information is accurate. Thus, we examine the interaction effect between insolvency risk and opacity on insurance purchase decisions. We argue that insurers' opacity will amplify the negative effect between insolvency risk and insurance purchase behavior.

Additionally, we are also interested in the interaction effect between the rating of insurers and opacity. As mentioned above, policyholders may not have interests in investing their resources to examine the financial health of insurers before they purchase insurance. Yet, it is not difficult to check the rating of an insurer. Since an insurer's rating reflects its transparency, we expect the rating can mitigate the negative effect between opacity and insurance purchases.

Policyholders suffer losses when insurers become insolvent even though there is a state-level guaranty association. Because the guaranty fund provides limited coverage and not all lines of business are protected. Thus, policyholders have more incentive to identify safer and more reliable insurers to avoid future losses due to the financial failure of insurers with the presence of the guaranty fund. In other words, policyholders are sensitive to opacity even though their policies are protected by the guaranty fund.

Property liability insurers can categorize their business lines into two main categories: commercial lines and personal lines. Commercial lines products provide protection against business-specific operational and liability risks, while personal lines products protect against personal property and liability risks. Due to the complexity of business risks, brokers and agents, who possess in-depth knowledge of insurers' financial situations and insurance product policies, play a significant role in the commercial lines insurance purchasing process. Consequently, brokers and agents are more likely to recommend products from less opaque insurers, thereby ensuring reliability and claim fulfillment. Thus, commercial lines product purchase behaviors are more likely to be influenced by opacity. In contrast, consumers of personal lines may not delve into insurers' financial statements to make their purchase decisions, making opacity a less significant factor. This paper uses a sample of publicly traded property-liability insurers because the opacity measure is only available for publicly traded insurers. The final sample consists of 499 insurer-year observations from 1997 to 2021. The empirical result shows that policyholders are willing to buy policies from less opaque insurers, and a one-standard-deviation decline in the opacity index is associated 16.83% increase in net premium written. A possible reason is that policyholders are aware of the information quality of insurers. We also find that policyholders are more sensitive to the information about insurers' financial risk when they are less opaque. For an insurer with an average Expected shortfall at a 99.5 confidence interval, a one-standard-deviation increase in opacity is associated with a decrease of 17.09% in net premium written. Furthermore, the empirical results also show that policyholders have incentives to buy policies from safer and less opaque insurers due to insufficient protection of the guaranty fund, especially for insurers with high financial risk.

This paper contributes to the literature that explores whether insurers' opacity has an impact on policyholders' purchase behavior. To our knowledge, this is the first paper to investigate policyholders' purchase behavior regarding insurers' opacity. We also provide evidence that policyholders are more sensitive to information about insurers' financial risk when they are less opaque. Additionally, our findings indicate that policyholders are aware of the insufficient protection offered by guaranty funds. Our research further suggests that opacity significantly influences the purchase behavior of commercial lines. This is due to the involvement of brokers and agents who possess in-depth knowledge of insurers' financial situations and product policies. We use personal lines as a placebo test to underscore this finding.

This paper is organized as follows. Section 2 reviews the related literature and outlines hypotheses. Section 3 describes the empirical methodology framework and data sources. Section 4 presents the summary statistics of the sample and empirical results. Section 5 concludes.

## 2.2. Hypothesis Development

### 2.2.1. Opacity and Net Premium Written

Due to the nature of insurers, such as the complicated liability structure, they are considered more opaque than non-financial corporations (Park, 2008). Thus, opaque firms are more likely to have a higher degree of information asymmetry, especially for property-liability insurers because of the uncertainty of loss estimation associated with natural disasters. We are interested in whether insurers' opacity has an impact on policyholders' purchase behavior.

Policyholders are less likely to devote time and resources to understanding annual or even quarterly financial statements thoroughly when they have a demand for policies. The reason is that most consumers can process a limited amount of information in the disclosure unless the information can be combined into a rating or ranking (Craswell, 2006; Sovern, 2010). In addition, consumers can also collect information on insurers through stock market performance and analysts' forecasting as they are not only policyholders but also potential debtholders. Policyholders pay the premium in exchange for the promise of indemnity when claims arise. In this sense, the risk involved (such as information risk) will be considered by policyholders because the claim might default if the insurer becomes insolvent (Eckles et al., 2014).

For the same type of risk (e.g., automobile, homeowners', etc.), consumers tend not to buy more than one policy. Thus, the information risk of the insurer is non-diversifiable (Eckles et al., 2014). Policyholders can perceive the information quality as a part of the utility function of default risk. Other things being equal, a lower quality of information implies a higher default risk because opaque firms are more likely to manipulate their financial statements, such as loss reserves (Han et al., 2018). Ertan et al. (2017) also argue that loans originated under a transparency regime have lower default probability.

Consumers are more willing to purchase from more transparent firms because greater transparency indicates outsider stakeholders can better monitor firms. Therefore, the information risk will be lower for more transparent insurers, benefiting policyholders due to lower default risk. As a result, opaque insurers provide lower utility to policyholders. Based on the above discussions, we suggest the following hypothesis:

*Hypothesis 1: Policyholders are less willing to buy policies from opaque insurers.*

### 2.2.2. The Interaction Effect between Opacity and Financial Risk on Net Premium Written

Poor financial health insurers might induce managers to manipulate financial statements to appear solvent (Petroni,1992; Gaver and Paterson, 2004). Additionally, under strong investor protection regimes, there are greater financial transparency and less earning management (Bhattacharya et al., 2003). Thus, policyholders are less likely to purchase insurance from insurers with the poor financial condition when the information is opaquer. The poor financial information serves as a signal which decreases the belief of policyholders that they can be paid in full when claims arise. In addition, Chen et al. (2022) suggest that uninsured depositors are sensitive to banks' performance when banks are more transparent. Thus, transparency plays an important role in conveying trustful financial information ex-ante that is valued by consumers (Dang et al., 2015). These arguments suggest the following hypothesis:

*Hypothesis 2: When insurers have higher financial risk, policyholders are likely to purchase more insurance from an insurer with lower opacity than insurers with higher opacity.*

### 2.2.3. Opacity and Insurance Guaranty Fund

If the guaranty fund provides complete protection to policyholders, the level of opacity may not play an important role in policyholder purchase behavior because policyholders would be paid in full for their claims when insurers are insolvent. However, in reality, the guaranty fund does not cover all lines of business. In addition, the guaranty fund provides only limited coverage to these lines of business that are covered. Thus, even with the protection of the guaranty fund, the safety and information quality shape policyholders' purchase behavior. Policyholders have the incentive to identify low financial risk insurers to avoid future losses due to the failure of insurers.

Insurers with less opacity are expected to be better monitored by outsider stakeholders, which motivates insurers not to manipulate financial statements and results in better meeting the obligations to pay claims to policyholders. In addition, with incomplete protection of the guaranty fund, policyholders perceive higher utility from insurers with higher information quality (transparency). The higher opacity amplifies the concerns of the probability of financial failure when insurers face high financial risk. Thus, policyholders will be attracted to less opaque insurers to purchase insurance because less opacity enhances their belief in getting their claims paid. Theses argument leads to the following hypothesis:

*Hypothesis 3-1: Even with the guaranty fund, policyholders are likely to purchase more insurance from insurers with less opacity, especially when insurers' financial risk is high.*

For lines of business not protected by the guaranty fund, these are riskier for policyholders.[32] This is because these lines of business have, in general, high risk and are covered by limited insurers. Since policyholders are aware of the risk of not being protected by the guaranty fund, they would be more sensitive to the opacity when insurers face high financial risk. Therefore, policyholders of policies that are not protected by the guaranty fund will be more sensitive to financial risk when opacity changes. Thus, we propose the following hypothesis:

*Hypothesis 3-2: Without the guaranty fund, policyholders are likely to purchase insurance from insurers with less opacity, especially when insurers' financial risk is high.*

#### 2.2.4. Opacity and Purchase Behavior: Commercial Lines and Personal Lines

Property liability insurers typically classify business lines into two major categories: personal lines and commercial lines. Personal lines include products designed to protect individuals or families from risks and losses associated with personal property and liability, such as homeowners insurance and auto insurance. In contrast, commercial lines offer protection for businesses against operational and liability risks, such as malpractice, commercial auto, and workers' compensation insurance. Given the varied and specific needs of businesses, commercial insurance products require more sophisticated and tailored policies. For example, technology firms, due to their unique set of operation scopes and risks, have several essential insurance products to

---

[32] The detailed lines of business which are not covered by guaranty fund can be found in *PROPERTY AND CASUALTY INSURANCE GUARANTY ASSOCIATION MODEL ACT.*

meet their needs: cyber liability insurance, which can protect against data breaches and cyber-attacks. Professional liability insurance can protect against the errors and omissions of the software, which may lead to financial loss for clients.

Therefore, commercial lines are inherently more complex than personal lines, reflecting different business types’ diverse risks and requirements. This complexity often necessitates the involvement of brokers and agents, who can assess all possible unique insurance situations and have useful and in-depth knowledge about the insurers’ financial situation and their products. Opacity can play an important role when brokers and agents make recommendations for policies. For less opaque insurers, there will be less earning management, which results in better meeting the obligations to pay claims to policyholders. Thus, commercial lines purchase behaviors are more likely to be influenced by opacity because brokers and agents know more about insurers, so they are willing to choose products from more transparent insurers.

In contrast, individual policyholders purchasing personal lines products may not delve into detailed financial analyses, opting instead to rely on straightforward rating information. This preference is because ratings are direct and easier to understand, suggesting that opacity may not play a significant role when individuals purchase personal lines products.

Overall, the role of brokers and agents is crucial in navigating the complexities of commercial insurance, where transparency and the insurer's financial health significantly impact policy selection. In the personal lines market, however, simplicity and accessibility of information allow individuals to make choices with less reliance on financial analyses. These arguments suggest the following hypothesis:

*Hypothesis 4-1: For commercial lines, policyholders are more likely to purchase insurance from insurers with less opacity, under the guidance of brokers and agents.*

*Hypothesis 4-2: For personal lines, policyholders' purchase behaviors are less influenced by opacity.*

## 2.3. Data and Methodology

The initial sample consists of US publicly traded property-liability insurers from 1996-2021. Following the literature (Han et al. 2018), the opacity index is calculated based on four factors: trading volume, bid-ask spread, the number of analysts following, and analysts' forecasting error. The trading volume and bis-ask spread data are from the Center for Research in Security Prices (CRPS). The number of analysts following and analysts' forecasting error data are from the Thomson Financial Institutional Brokers Estimate System (I/B/E/S). The insurers' financial data are from the National Association of Insurance Commissioners (NAIC). Value at risk and Expected shortfall, proxies for firm risks, are calculated from CRSP. The rating data is from A.M. Best. After merging all datasets, the final sample consists of 499 insurer-year observations from 1997 to 2021. [33]

The baseline regression specification to test our argument is as follows:

$$Ln\left(Net\ Premium\ Written_{i,t}\right) = \beta_0 + \beta_1 Opacity_{i,t-1} + BX_{i,t-1} + f_i + \varepsilon_{i,t}$$

where $Net\ Premium\ Written_{i,t}$ represents the net premium written by insurer i in year t from 1997 to 2021. $Opacity_{i,t-1}$ is the transparency measure of insurer i in year t-1. A lower opacity index means the insurer has more transparent public information. $X_{i,t-1}$ is a set of control variables. $f_i$ is firm fixed effect. Insurance demand is highly subjected to the underwriting cycle due to macroeconomic developments and uncertainties. Thus, for year fixed effect, we use historical hard market episodes

[33] NAIC dataset starts from 1996. The return of average equity is available from 1997 and we lag one year so the data sample starts from 1998.

year dummies to control. We lag one year for all independent variables because policyholders make the decision based on last year's information.

For policyholders who do not have access to insurers' private information, insurers' public information can come from the stock market, the rating agency, analysts' reporting, etc. Based on the literature (Anderson et al., 2009; Wang, 2011; Han et al., 2018), the insurers' opacity index combines the information of four factors: trading volume, bid-ask spread, the number of analysts following, and analysts' forecasting error. The trading volume is correlated with information asymmetry (Chae, 2005) and captures firm-specific information (Bessembinder et al., 1996). Therefore, we take the natural log of average daily trading volume during the fiscal year as the measure of the trading volume. The second factor is the bid-ask spread, which is widely used as a proxy for information asymmetry among investors (Nagar et al., 2019), and higher bid-ask spreads imply lower transparency of individual firms (Leuz and Verrecchia, 2000). To calculate the bid-ask spreads, the first step is to calculate daily bid-ask spreads as the daily ask price minus the daily bid price and scaled by the average of daily ask and bid prices. The second step is calculating annual bid-ask spreads by averaging the daily bid-ask spreads during the fiscal year.

Analysts play the role of intermediaries between firms and investors. Analysts' reports provide financial information to investors. If more analysts follow a firm, investors can get more comprehensive information about the firm, indicating lower opacity. Thus, we include the analysts following into the opacity measure. The analysts following is calculated as the natural log of the number of analysts who provide earnings forecasts nine months before the end of the fiscal year. The last but not the least factor in the opacity index is analysts' forecasting error. Hope (2003) suggests that analysts' forecast accuracy is positively related to financial disclosure quality. Anderson et al. (2009) use analysts' forecast errors to proxy information availability. The analysts' forecast errors are calculated as the square of the difference between the mean earnings forecast of all analysts (nine months before the end of the fiscal year) and actual earnings, then scaled by the stock price.

After calculating these four proxies of the firm's opacity, we rank each proxy to deciles from ten to one. For each rank of the proxy, the rank value of ten means the opaquest information, and the value of one means the most transparent information. Then, we sum up the four rank values of the proxy and scaled by 40 to get the opacity index from 0.1 to 1.0. The lower value of the opacity index means higher transparency. Anderson et al. (2009) argue that this opacity index provides a comprehensive and robust measure of a firm's opacity, which includes market trades and analyst coverage information.

We include a set of control variables in the regression analysis. The price of a policy (*PRICE*) is an essential factor influencing policyholders' purchase behavior. There is no available data for unit price; thus, we use loss incurred divided by net premium earned (the inverse of loss ratio) to proxy the price. Policyholders can get public information not only from the stock market and analyst reporting but also from the rating agency. A.M. Best, a leading rating agency, provides rating services specializing in the insurance industry. A.M. Best's financial strength rating incorporates detailed public and proprietary financial information such as liquidity, asset, certified actuarial and loss-reserve reports, investment detail, annual business plans, etc. Once the rating is published, A.M. Best still monitors and updates the rating, which reflects the agency's opinion of whether the insurer can meet the obligation to policyholders and still remain solvent. Therefore, the A.M. Best rating provides comprehensive information to policyholders. Thus, we include the A.M. Best rating as one of the control variables.

We also control the effect of default risk, demand, and service quality on policyholders' purchase behavior. Following the literature (e.g., Milidonis et al., 2019), the Expected shortfall

(ES) at the 99.5 percent confidence level is used to proxy the financial risk.[34] We use 1 year of daily stock returns to calculate Expected shortfall (*ES*), of which the definition is the conditional expected loss. An indicator variable *WEAK* is also used to reflect insurers' financial condition. The insurance industry is highly regulated, and regulators use Insurance Regulatory Information System (IRIS) ratios to target insurers needing regulation attention. *WEAK* takes a value of 1 if the insurer has more than 3 out of the range IRIS ratios and 0 otherwise. Policyholders respond to macroeconomic changes, which result in changes in insurers' performance. This effect is called underwriting cycle. Thus, we control time dummies of hard market episodes as a proxy for demand. We include insurer fixed effects and various insurer characteristics to control time-invariant and time-variant components of service quality. Other control variables are return on assets (*ROA*), product diversification (*PRODHHI*), geographical diversification (*GEOHHI*), which are calculated using the Herfindahl Index, and the percentage of net premium written from coastal states (*COASTAL*).

## 2.4. Summary Statistics and Empirical Results

### 2.4.1. Summary Statistics

Table 2.1 presents summary statistics on financial variables for the whole sample. The insurer opacity index (OPACITY) has a mean (median) of 0.564 (0.550) and a standard deviation is 0.206, which are comparable to Han et al. (2018). The average insurer in our sample has A.M. Best financial strength rating (Mean of Rating=3.371) between "Superior" (A++ or A+) and "Excellent" (A or A-). The median rating (Median Rating=3) is "Excellent" (A or A-). The 75$^{th}$ quantile of *WEAK* is 0, representing that very few insurers have more than 3 unusual IRIS ratios. The average insurer has a product line Herfindahl Index (*PRODHHI*) of 0.415 and a geographical

[34] The 99.5 percent confidence level is consistent with the solvency capital requirement (Milidonis et al., 2019).

Herfindahl Index (*GEOHHI*) of 0.196, indicating that the insurer, on average, has approximately 3 business lines and operates in 5 states.

### 2.4.2. Opacity and Net Premium Written Baseline Result

Table 2.2 presents the relation between insurers' opacity index and the natural log of net premium written. Recall that the lower opacity index means the insurer is more transparent. The coefficient of the insurer opacity index is negative and significant at a 5% level, supporting hypothesis 1 that policyholders are willing to buy policies from more transparent insurers. This indicates that policyholders take information quality into consideration when they purchase policies, suggesting that more transparent insurers provide more utility to policyholders. A one-standard-deviation decline in opacity index is associated 17.92% (= -0.206*(-0.870)) increase in net premium written. This result is consistent with Han et al. (2018), who argue that more transparent insurers provide more conservative loss reserve estimation and prevent earning management through manipulating reserve estimation. Thus, policyholders get higher utility from less opaque insurers because of high information quality.

The results of the control variable are consistent with expectations. A.B. Best financial strength rating (*RATING*) is positive and significant at a 5 % level, indicating that insurers with higher ratings attract more policyholders if they use rating to assess insolvent risk. This result is consistent with Halek and Eckles (2010) that information provided by the rating agency is valuable to policyholders. Consumers of financial products are sensitive to insolvency risk, so they demand lower prices of products to compensate for higher risk (Epermanis and Harrington 2006). Financial strength ratings try to provide comprehensive and unbiased opinions regards insurers' insolvency risk and reduce opacity. Thus, the influence of price might be incorporated into financial strength ratings, so the price is insignificant. The time dummy of hard market episodes (*HARD*) is negative and significant at a 1% level, indicating that policyholders respond to macroeconomic shocks. The insurance industry's underwriting cycle is influenced by the hardening market. Geographical diversification (*GEOHHI*) is negative and significant at a 1% level, suggesting that geographical diversification enriches coverage area and diversifies underwriting risk.

### 2.4.3. Opacity and Financial Risk

This section identifies whether policyholders are more sensitive to financial risk at insurers with lower opacity index. We use Expected shortfall (ES) at a 99.5 confidence interval to proxy the financial risk (Milidonis et al., 2019). The main focus is the interaction term of opacity and Expected shortfall at a 99.5 confidence interval. The coefficient of interaction term measures the change in net premium written-financial risk sensitivity as opacity changes. Table 2.3 presents the relation between opacity and NPW-financial risk sensitivity of policyholders.

The interaction term between opacity and Expected shortfall at a 99.5 confidence interval is negative and significant at a 5% level, indicating that policyholders are more sensitive to insurers' financial risk with a lower opacity index: a one-standard-deviation decrease in opacity amplifies the average sensitivity by 71.21% in net premium written (= -0.206*(-4.708)/1.362). The coefficient of opacity is negative and significant, suggesting that less opaque insurers have higher net premium written growth rate: for an insurer with an average Expected shortfall at a 99.5 confidence interval, a one-standard-deviation decrease in opacity is associated with an increase of 18.30% (= ( -0.639+(-4.708)*0.0530) * (-0.206)) in net premium written. Policyholders are more sensitive to information about insurers' financial risk when they are less opaque. It makes sense because policyholders buy promises from insurers; thus, they are more risk-averse. The trustful information provided by transparent insurers increases the utility perceived by policyholders, especially when financial risk is high.

### 2.4.4. Opacity and Net Premium Written Protected by Guaranty Fund

If policyholders buy policies from solvent insurers, they can receive the full coverage listed in the policy if a claim happens and also receive services such as roadside assistance. What if insurers become insolvent? The insurance guaranty association is nonprofit and operated at a state level. The stated purpose of the guaranty fund is to provide a backup fund for policyholders to cover policies when the insurer is insolvent. However, the coverage provided by the guaranty fund is limited and varies for different states. Thus, the opacity and financial condition would influence the policyholders' purchase behavior.

Table 2.4 presents the relation between insurers' opacity index and the net premium written from lines of business protected by the guaranty fund. The coefficient of opacity is negative and significant at a 5% level, indicating that policyholders recognize that the guaranty fund doesn't provide them with complete protection; otherwise, the opacity doesn't impact purchase behavior. This result is consistent with Sommer (1996); there is no reason for purchasers to pay higher prices to buy products from safer firms if the guaranty fund protection is complete. The opacity index conveys the information quality to policyholders willing to buy policies from less opaque insurers, which provide higher benefits to policyholders. The economic magnitude is also significant: a one-standard-deviation decline in the opacity index is associated 19.32% (= -0.204*(-0.947)) increase in net premium written from lines of business protected by the guaranty fund. The coefficient of price is negative but insignificant, and the rating is significant, indicating that the policyholders are aware that the guaranty fund doesn't provide full protection, so, on average, policyholders will be attracted by insurers with lower opacity and higher ratings.

### 2.4.5. Opacity and Net Premium Written Protected by Guaranty Fund with Financial Risk

As discussed in the previous section, due to insufficient protection of the guaranty fund, policyholders have incentives to buy policies from safer and less opaque insurers. This incentive will be amplified when they perceive the financial risk of insurers. Therefore, the influence of financial risk on demand will be greater if the insurer is less opaque.

In Table 2.5, we explore whether policyholders will be more sensitive to the net premium written from lines of business protected by guarantee fund with financial risk when opacity changes. The coefficient of the interaction term between opacity and Expected shortfall is negative and significant at a 5 % level, indicating that policyholders are sensitive to opacity, especially when the insurer's financial risk is high. Moreover, the economic magnitude is significant: for an insurer with an average Expected shortfall at a 99.5 confidence interval, a one-standard-deviation decrease in opacity is associated with an increase of 19.86% (= ( -0.687+(-5.404)*0.0530) * (-0.204)) in net premium written. The opacity index indicates lower information quality which harms policyholders when their policies are not fully protected by the guaranty fund.

### 2.4.6. Opacity and Net Premium Written Not Protected by Guaranty Fund

The guaranty fund doesn't cover all lines of business; for example, insurance products that offer protection against investment or financial risks are excluded from the protection offered by the guarantee fund. For these products, policyholders are aware of the high risk and concern about the insurers' financial strength. Therefore, they have fewer incentives to buy products with the lowest price regardless of safety. Not all insurers provide products for lines of business that are not protected by the guaranty fund. As a result, policyholders have limited choices of insurers for these products. Based on the initial risks of policies, policyholders are more sensitive to opacity when insurers face high financial risk. Because once insurers go bankrupt, policyholders get nothing and lose the premium they paid. For these policyholders, net premium written-financial risk sensitivity as opacity changes is pronounced. The opacity can provide some information to policyholders, especially when financial risk is high. Furthermore, the comprehensive evaluation of insurers can also be informed by financial strength ratings.

In Table 2.6, the baseline result of the coefficient of opacity is not significant. In column 2, the coefficient of opacity is negative and significant at a 5% level, and the coefficient of the interaction term between opacity and rating is positive and significant at a 10% level. This result indicates that the negative effect of opacity is mitigated by rating. The impact of opacity is mitigated by a higher rating because insurers are more trustful with a higher rating. In column 3, the coefficient of the interaction term between opacity and Expected shortfall is negative and significant at a 5% level, indicating that when the insurer's financial risk is high, policyholders are sensitive to opacity: a one-standard-deviation decrease in opacity amplifies the average sensitivity by 40% in net premium written (= -0.202*(-13.315)/6.724). In addition, the financial strength rating is significant at a 1% level, indicating that policyholders trust the information provided by the rating agency.

### 2.4.7. Opacity and Purchase Behavior: Commercial Lines and Personal Lines

Commercial lines insurance products protect against business-specific operational and liability risks, playing an important role in the risk management strategies of business entities. The complexity of business entities' risk situations arises from the diverse risks and requirements of different business types, necessitating the involvement of brokers and agents. Brokers and agents, who can assess all possible unique insurance situations and possess useful and in-depth knowledge about insurers' financial situations and their products, provide the most accurate and relevant advice to meet clients' unique needs. Consequently, brokers and agents not only ensure that policies fit clients' needs but also verify whether insurers can fulfill their obligations to pay claims to policyholders. Thus, brokers and agents are more likely to recommend products from less opaque insurers, as they can access more information. In Table 2.7, we explore whether business entities' purchase behavior is influenced by opacity. In column 3, the coefficient of opacity is negative and significant at a 5% level, indicating that opacity will shape policyholders' purchase behavior and they are more willing to purchase products from more transparent insurers. A one-standard-deviation decline in opacity index is associated 21.6% (= -0.204*(-1.059)) increase in net premium written from commercial lines. This evidence suggests that opacity is an important factor for brokers and agents to make insurance purchase recommendations.

In contrast, individual policyholders purchasing personal lines more rely on straightforward rating information. Thus, opacity may not play an important role when individuals purchase personal lines products.

In Table 2.8, the coefficients of opacity are insignificant for all columns but the coefficients of rating are positive and significant at 5% level, indicating that personal lines policyholders may look at ratings directly because rating information is more straightforward and easier to understand.

## 2.5. Conclusion

This paper investigates whether insurers' opacity has an impact on policyholders' purchase behavior. The results indicate that policyholders take information quality into consideration, which means opacity plays an important role when they make purchase decisions.

We explore that opacity negatively affects policyholders' purchase behavior, especially when insurers' financial risk is high. In addition, the guaranty fund doesn't provide complete protection to policies; thus, policyholders care about the safety and opacity of insurers. For less opaque insurers, policyholders perceive more utility regarding lower default risk because of higher information quality. Our research further suggests that opacity significantly influences the purchase behavior of commercial lines. This is due to the involvement of brokers and agents who possess in-depth knowledge of insurers' financial situations and product policies.

# REFERENCES

AM Best. 2009. An Explanation of Best's Credit Rating System and Procedures, Oldwick, NJ.

Bessembinder, H., K. Chan, and P. Seguin. 1996. An empirical examination of information, differences of opinion, and trading activity. *Journal of Financial Economics*, 40, 105–134.

Bhattacharya, U., H., Daouk, and M. Welker. 2003. The World Price of Earnings Opacity. The Accounting Review, 78(3), 641–678.

Chae, J. 2005. Trading Volume, Information Asymmetry, and Timing Information. *The Journal of Finance*, 60: 413-442.

Chen, Q., I., Goldstein, Z. Huang, and R. Vashishtha. 2022 Bank transparency and deposit flows. *Journal of Financial Economics*, 146(2), 475–501.

Craswell, R. 2006. Taking information seriously: Misrepresentation and nondisclosure in contract law and elsewhere. *Va. L. Rev.*, 92, p.565.

Dang, T.V., G., Gorton, and B., Holmström 2015. Ignorance, debt and financial crises, Unpublished working paper.

Dang, T.V., G., Gorton, B., Holmström, and G., Ordoñez. 2017. Banks as secret keepers. *The American Economic Review*. 107, 1005–1029.

Eastman, E., D. L. Eckles, and M. Halek. 2021. Asymmetry in Earnings Management Surrounding Targeted Ratings. *Available at SSRN*: https://ssrn.com/abstract=3846968

Eckles, D.L., M. Halek, and R. Zhang. 2014. Information risk and the cost of capital. *Journal of Risk and Insurance*, 81(4), pp.861-882.

Epermanis, K., and S. E. Harrington. 2006. Market discipline in property/casualty insurance: Evidence from premium growth surrounding changes in financial strength ratings. *Journal of Money, Credit & Banking*, 38 (6): 1515–1544.

Ertan, A., M. Loumioti, and R. Wittenberg-Moerman. 2017. Enhancing loan quality through transparency: Evidence from the European Central Bank loan level reporting initiative. *Journal of Accounting Research*, 55(4), pp.877-918.

Francis, J., S., Huang, I., Khurana, and R. Pereira. 2009. Does corporate transparency contribute to efficient resource allocation? *Journal of Accounting Research*, 47, 943–989.

Gaver, J.J. and J.S., Paterson. 2004. Do insurers manipulate loss reserves to mask solvency problems? *Journal of Accounting and Economics*, 37(3), pp.393-416.

Halek, M. and Eckles, D.L. 2010. Effects of Analysts' Ratings on Insurer Stock Returns: Evidence of Asymmetric Responses. *Journal of Risk and Insurance*, 77: 801-827.

Han, S.Y., G.C. Lai, and C.L. Ho. 2018. Corporate transparency and reserve management: Evidence from US property-liability insurance companies. *Journal of Banking & Finance*, Elsevier, vol. 96(C), pages 379-392.

Hope, O.K. 2003. Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. Journal of accounting research, 41(2), pp.235-272.

Sovern, J., 2010. Preventing future economic crises through consumer protection law or how the truth in lending act failed the subprime borrowers. *Ohio St. LJ*, 71, p.761.

Lang, M.H., K.V. Lins, and M.G. Maffett. 2012. Transparency, liquidity, and valuation: international evidence on when transparency matters most. . *Journal of Accounting Research*,50, 729–774 .

Leuz, C., and R. E. Verrecchia. 2000. The Economic Consequences of Increased Disclosure. *Journal of Accounting Research*, 38, 91–124.

Milidonis, A., T. Nishikawa, and J. Shim. 2019. CEO inside debt and risk taking: Evidence from property–liability insurance firms. *Journal of Risk and Insurance*, 86(2), 451–477.

Nagar, V., J. Schoenfeld, and L. Wellman. 2019. The effect of economic policy uncertainty on investor information asymmetry and management disclosures. *Journal of Accounting and Economics*, 67(1), pp.36-57.

Park, S., 2008, The Opacity of Insurance Companies, Working Paper, University of Pennsylvania.

Sommer, D. W. 1996. The Impact of Firm Risk on Property-Liability Insurance Prices. *The Journal of Risk and Insurance*, 63(3), 501–514.

Petroni, K.R., 1992. Optimistic reporting in the property-casualty insurance industry. *Journal of Accounting and Economics*, 15(4), pp.485-508.

Table 2.1: Summary statistics

This table presents summary statistics of variables used in the regression model. The sample period is from 1997 to 2021. All the variables are defined in Appendix A.

| Variable | N | MEAN | SD | MIN | P25 | P50 | P75 | MAX |
|---|---|---|---|---|---|---|---|---|
| *LNNPW* | 499 | 14.050 | 1.265 | 11.630 | 13.110 | 13.890 | 14.750 | 17.080 |
| *LNNPW_GRT* | 497 | 13.970 | 1.285 | 11.460 | 13.020 | 13.830 | 14.720 | 17.010 |
| *LNNPW_NGRT* | 398 | 10.61 | 2.905 | 1.744 | 9.817 | 11.330 | 12.220 | 15.120 |
| *LNNPW_PSNL* | 448 | 12.850 | 1.890 | 4.940 | 11.830 | 12.850 | 13.970 | 16.98 |
| *LNNPW_COMML* | 496 | 12.490 | 2.200 | 1.616 | 11.720 | 12.870 | 13.810 | 16.31 |
| *OPACITY* | 499 | 0.564 | 0.206 | 0.125 | 0.400 | 0.550 | 0.725 | 0.925 |
| *PRICE* | 499 | 1.957 | 0.629 | 1.197 | 1.621 | 1.808 | 2.059 | 5.184 |
| *RATING* | 499 | 3.371 | 0.527 | 1 | 3 | 3 | 4 | 4 |
| *ES99.5* | 499 | 0.053 | 0.030 | 0.021 | 0.035 | 0.045 | 0.064 | 0.296 |
| *WEAK* | 499 | 0.112 | 0.316 | 0 | 0 | 0 | 0 | 1 |
| *HARD* | 499 | 0.493 | 0.500 | 0 | 0 | 0 | 1 | 1 |
| *PRODHHI* | 499 | 0.415 | 0.295 | 0.118 | 0.167 | 0.259 | 0.647 | 1 |
| *GEOHHI* | 499 | 0.196 | 0.269 | 0.039 | 0.055 | 0.073 | 0.146 | 0.997 |
| *ROA* | 499 | 0.048 | 0.038 | -0.060 | 0.023 | 0.046 | 0.070 | 0.156 |
| *COASTAL* | 499 | 0.487 | 0.197 | 0.0240 | 0.377 | 0.499 | 0.551 | 1 |

Table 2.2: Opacity and net premium written

This table presents the main results of the baseline model, testing the relation between insurers' opacity index and net premium written using the OLS regression method. The dependent variable is the natural log of net premium written (*LNNPW*). All the variables are defined in Appendix A. Standard deviations are clustered at firm level. The *t*-statistics are reported in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| | (1) | (2) | (3) |
|---|---|---|---|
| Variable | LNNPW | LNNPW | LNNPW |
| *OPACITY* | -1.163*** | -0.836** | -0.870** |
| | (-3.23) | (-2.516) | (-2.556) |
| *PRICE* | | -0.021 | -0.013 |
| | | (-0.486) | (-0.267) |
| *RATING* | | 0.203** | 0.222** |
| | | (2.193) | (2.474) |
| *ES99.5* | | -1.141 | -0.736 |
| | | (-1.678) | (-1.146) |
| *WEAK* | | -0.105 | -0.105 |
| | | (-1.544) | (-1.526) |
| *HARD* | | | -0.082*** |
| | | | (-3.604) |
| *PRODHHI* | | -0.107 | -0.035 |
| | | (-0.409) | (-0.128) |
| *GEOHHI* | | -1.479*** | -1.629*** |
| | | (-2.787) | (-3.071) |
| *ROA* | | -0.318 | -0.609 |
| | | (-0.630) | (-1.188) |
| *COASTAL* | | 14.303*** | -0.284 |
| | | (37.004) | (-0.478) |
| Intercept | | (-0.409) | 14.414*** |
| | | -1.479*** | (32.648) |
| Observations | 501 | 499 | 499 |
| Adjusted R-squared | 0.940 | 0.950 | 0.951 |
| Firm FE | Yes | Yes | Yes |

Table 2.3: Opacity and net premium written with financial risk mechanism

This table presents the results of the interaction term model, testing the relation between insurers' opacity index and net premium written with the financial risk mechanism. The dependent variable is the natural log of net premium written (*LNNPW*). All the variables are defined in Appendix A. Standard deviations are clustered at the firm level. The *t*-statistics are shown in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| Variable | LNNPW |
|---|---|
| *OPACITY* | -0.639* |
| | (-1.716) |
| *OPACITY ×ES99.5* | -4.708** |
| | (-2.360) |
| *ES99.5* | 1.362 |
| | (1.401) |
| *PRICE* | -0.010 |
| | (-0.206) |
| *RATING* | 0.220** |
| | (2.427) |
| *WEAK* | -0.103 |
| | (-1.500) |
| *HARD* | -0.073*** |
| | (-3.636) |
| *PRODHHI* | -0.037 |
| | (-0.133) |
| *GEOHHI* | -1.601*** |
| | (-3.092) |
| *ROA* | -0.659 |
| | (-1.244) |
| *COASTAL* | -0.276 |
| | (-0.471) |
| Intercept | 14.304*** |
| | (31.665) |
| Observations | 499 |
| Adjusted R-squared | 0.951 |
| Firm FE | Yes |

Table 2.4: Opacity and net premium written protected by guaranty fund

This table presents the main results of the baseline model, testing the relation between insurers' opacity index and net premium written of business lines protected by guaranty fund using the OLS regression method. The dependent variable is the natural log of net premium written of business lines protected by the guaranty fund (*LNNPW_GRT*). All the variables are defined in Appendix A. Standard deviations are clustered at the firm level. The *t*-statistics are reported in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| Variable | LNNPW_GRT |
|---|---|
| *OPACITY* | -0.947** |
| | (-2.724) |
| *PRICE_GRT* | -0.043 |
| | (-0.906) |
| *RATING* | 0.206** |
| | (2.152) |
| *ES99.5* | -0.932 |
| | (-1.539) |
| *WEAK* | -0.109 |
| | (-1.560) |
| *HARD* | -0.092*** |
| | (-3.763) |
| *PRODHHI* | 0.394 |
| | (1.465) |
| *GEOHHI* | -1.569** |
| | (-2.559) |
| *ROA* | -0.618 |
| | (-1.177) |
| *COASTAL* | -0.314 |
| | (-0.463) |
| Intercept | 14.326*** |
| | (30.207) |
| Observations | 497 |
| Adjusted R-squared | 0.950 |
| Firm FE | Yes |

Table 2.5: Opacity and net premium written protected by guaranty fund with financial risk mechanism

This table presents the results of the interaction term model, testing the relation between insurers' opacity index and net premium written of business lines protected by guaranty fund with the financial risk mechanism. The dependent variable is the natural log of net premium written of business lines protected by the guaranty fund (*LNNPW_GRT*). All the variables are defined in Appendix A. Standard deviations are clustered at the firm level. The *t*-statistics are shown in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| Variable | LNNPW_GRT |
|---|---|
| *OPACITY* | -0.687* |
| | (-1.828) |
| *OPACITY ×ES99.5* | -5.404** |
| | (-2.518) |
| *ES99.5* | 1.512 |
| | (1.532) |
| *PRICE_GRT* | -0.047 |
| | (-0.984) |
| *RATING* | 0.204** |
| | (2.094) |
| *WEAK* | -0.106 |
| | (-1.532) |
| *HARD* | -0.084*** |
| | (-3.820) |
| *PRODHHI* | 0.400 |
| | (1.525) |
| *GEOHHI* | -1.539** |
| | (-2.591) |
| *ROA* | -0.635 |
| | (-1.187) |
| *COASTAL* | -0.305 |
| | (-0.457) |
| Intercept | 14.210*** |
| | (29.688) |
| Observations | 497 |
| Adjusted R-squared | 0.951 |
| Firm FE | Yes |

Table 2.6: Opacity and net premium written not protected by guaranty fund

This table presents the main results of the baseline model, testing the relation between insurers' opacity index and net premium written of business lines not protected by guaranty fund using the OLS regression method and the interaction term between opacity and financial risk. The dependent variable is the natural log of net premium written of business lines not protected by the guaranty fund (*LNNPW_NGRT*). All the variables are defined in Appendix A. Standard deviations are clustered at the firm level. The *t*-statistics are reported in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| | (1) | (2) | (3) |
|---|---|---|---|
| Variable | LNNPW_NGRT | LNNPW_NGRT | LNNPW_NGRT |
| *OPACITY* | 1.524 | -4.814** | 2.142 |
| | (0.923) | (-2.500) | (1.261) |
| *OPACITY ×RATING* | | 1.788* | |
| | | (1.984) | |
| *RATING* | 0.610*** | -0.360 | 0.607*** |
| | (2.955) | (-0.704) | (3.027) |
| *OPACITY ×ES99.5* | | | -13.315** |
| | | | (-2.239) |
| *ES99.5* | 1.191 | 0.833 | 6.724*** |
| | (0.879) | (0.682) | (2.812) |
| *PRICE_NGRT* | 0.017 | 0.013 | 0.021 |
| | (1.113) | (0.934) | (1.347) |
| *WEAK* | -0.190 | -0.197 | -0.177 |
| | (-0.958) | (-0.989) | (-0.895) |
| *HARD* | -0.106 | -0.098 | -0.087 |
| | (-1.266) | (-1.203) | (-1.035) |
| *PRODHHI* | -2.181 | -1.942 | -2.160 |
| | (-1.664) | (-1.554) | (-1.671) |
| *GEOHHI* | -3.710** | -3.525*** | -3.658** |
| | (-2.554) | (-3.037) | (-2.477) |
| *ROA* | -0.760 | -1.600 | -1.099 |
| | (-0.316) | (-0.644) | (-0.451) |
| *COASTAL* | 0.943 | 0.579 | 0.967 |
| | (0.737) | (0.465) | (0.737) |
| Intercept | 8.349*** | 11.997*** | 8.084*** |
| | (5.542) | (11.583) | (5.394) |
| Observations | 398 | 398 | 398 |
| Adjusted R-squared | 0.907 | 0.909 | 0.908 |
| Firm FE | Yes | Yes | Yes |

Table 2.7: Opacity and net premium written from commercial lines

This table presents the main results of the baseline model, testing the relation between insurers' opacity index and net premium written of commercial lines using the OLS regression method. The dependent variable is the natural log of net premium written of commercial business lines (*LNNPW_COMML*). All the variables are defined in Appendix A. Standard deviations are clustered at the firm level. The *t*-statistics are reported in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| | (1) | (2) | (3) |
|---|---|---|---|
| Variable | LNNPW_COMML | LNNPW_COMML | LNNPW_COMML |
| OPACITY | -1.326*** | -1.055** | -1.059** |
| | (-3.119) | (-2.075) | (-2.065) |
| PRICE | | -0.067 | -0.061 |
| | | (-0.508) | (-0.450) |
| RATING | | 0.289** | 0.300** |
| | | (2.459) | (2.604) |
| ES99.5 | | -1.611* | -1.189 |
| | | (-1.862) | (-1.400) |
| WEAK | | -0.197 | -0.201 |
| | | (-1.439) | (-1.441) |
| HARD | | | -0.084** |
| | | | (-2.442) |
| PRODHHI | | 0.182 | 0.231 |
| | | (0.277) | (0.353) |
| GEOHHI | | -3.276* | -3.327* |
| | | (-1.704) | (-1.749) |
| ROA | | -1.671** | -1.953** |
| | | (-2.135) | (-2.418) |
| COASTAL | | -2.213 | -2.196 |
| | | (-1.641) | (-1.648) |
| Intercept | 13.240*** | 14.076*** | 14.044*** |
| | (55.007) | (14.023) | (14.048) |
| | | | |
| Observations | 498 | 496 | 496 |
| Adjusted R-squared | 0.947 | 0.957 | 0.957 |
| Firm FE | YES | YES | YES |

Table 2.8: Opacity and net premium written from personal lines

This table presents the main results of the baseline model, testing the relation between insurers' opacity index and net premium written of personal lines using the OLS regression method. The dependent variable is the natural log of net premium written of personal business lines (LNNPW_PSNL). All the variables are defined in Appendix A. Standard deviations are clustered at the firm level. The t-statistics are reported in parentheses. Significant levels at the 1%, 5%, and 10% are indicated by ***, **, and *, respectively.

| | (1) | (2) | (3) |
|---|---|---|---|
| Variable | LNNPW_PSNL | LNNPW_PSNL | LNNPW_PSNL |
| OPACITY | 0.176 | -0.050 | -0.057 |
| | (0.189) | (-0.106) | (-0.120) |
| PRICE | | -0.634* | -0.633* |
| | | (-1.728) | (-1.721) |
| RATING | | 0.257** | 0.260** |
| | | (2.569) | (2.628) |
| ES99.5 | | -1.474* | -1.317* |
| | | (-1.786) | (-1.942) |
| WEAK | | 0.044 | 0.041 |
| | | (0.467) | (0.441) |
| HARD | | | -0.032 |
| | | | (-0.688) |
| PRODHHI | | -3.551*** | -3.530*** |
| | | (-3.800) | (-3.795) |
| GEOHHI | | -3.175** | -3.181** |
| | | (-2.465) | (-2.484) |
| ROA | | 3.025* | 2.925* |
| | | (1.848) | (1.829) |
| COASTAL | | 0.609 | 0.611 |
| | | (0.445) | (0.449) |
| Intercept | 12.742*** | 14.680*** | 14.679*** |
| | (24.546) | (10.646) | (10.672) |
| Observations | 450 | 448 | 448 |
| Adjusted R-squared | 0.909 | 0.942 | 0.942 |
| Firm FE | YES | YES | YES |

# APPENDIX A: VARIABLE DEFINITIONS

| Variable | Definition |
|---|---|
| **Dependent variable** | |
| *LNNPW* | The natural log of net premium written. (NAIC) |
| *LNNPW_GRT* | The natural log of net premium written of business lines protected by the guaranty fund. (NAIC) |
| *LNNPW_NGRT* | The natural log of net premium written of business lines not protected by the guaranty fund. (NAIC) |
| *LNNPW_PSNL* | The natural log of net premium written of personal lines. (NAIC) |
| *LNNPW_COMML* | The natural log of net premium written of commercial lines. (NAIC) |
| **Independent variable** | |
| *OPACITY* | An index that sums up the four rank values of the proxy (trading volume, bid-ask spread, the number of analysts following, and analysts' forecasting error) and scaled by 40 to get the opacity index from 1.0 to 0.1. The lower value of the opacity index means higher transparency. (CRSP and I/B/E/S) |
| *PRICE* | The loss incurred divided by net premium earned (the inverse of loss ratio). (NAIC) |
| *RATING* | A.M. Best's financial strength rating: "Superior" (A++ or A+) equals 4; "Excellent" (A or A-) equals 3; "GOOD" (B++ or B+) equals to 2; "Vulnerable Ratings" (B and below) equals 1. (A.M. Best) |
| *ES_99.5* | Expected shortfall (ES) is defined as the conditional expected loss using 1 year of daily firm stock returns at 99.5% confidence level. (CRSP) $ES = \overline{R} - \sqrt{\sigma}.\frac{1}{c}.\emptyset(z_c)$. $\overline{R}$ is the mean of 1 year of daily firm stock returns. σ is the variance of 1 year of daily firm stock returns. Zc is the c-quantile of the standard normal distribution. ∅ is the density function. |
| WEAK | The WEAK takes a value of 1 if the insurer has more than 3 out of the range IRIS ratios and 0 otherwise. (NAIC) |
| *HARD* | Time dummies of hard market episodes. Equals 1 if observations are during 2000-2003, 2008-2012, and 2018-2020, and 0 otherwise. (Swiss Re Institute) |
| *PRODHHI* | The line of business Herfindahl Index. (NAIC) |
| *GEOHHI* | The geographical Herfindahl Index. (NAIC) |
| *ROA* | The return on assets. (NAIC) |
| *COASTAL* | The percentage of net premium written from coastal states. (NAIC) |

# CHAPTER 3: PHYSICIAN FRAUD DETECTION USING MACHINE LEARNING METHODS

## 3.1. Introduction

The study uses machine learning methods to detect physician frauds and estimates the healthcare cost savings through detecting frauds. Healthcare costs have become a major expenditure in the U.S. since 1980 (Li et al., 2008) and are expected to continue growing due to an aging population and advancing health technology (Organisation for Economic Co-operation, 2008 and 2009). According to the Centers for Medicare & Medicaid Services (CMS), healthcare spending grew 4.6% in 2018 and reached $3.6 trillion, accounting for 17.7% of the nation's Gross Domestic Product (GDP). National health spending is projected to grow at an average annual rate of 5.4% for the next decade and to reach $6.2 trillion by 2028, without the consideration of COVID-19. However, a conservative estimation by the National Health Care Anti-Fraud Association (NHCAA) shows that 3% of total healthcare expenditures were lost due to frauds each year. Other estimations of fraud by some government and law enforcement agencies reach as high as 10% of the annual health outlay. Healthcare fraud is not only in the U.S. but in the rest of the world (e.g., He et al., 1997; Yamanish et al., 2004; Ortega et al., 2006; Aral et al., 2012; Shin et al., 2012). Besides the direct financial losses, frauds also severely hinder the healthcare system from providing quality services because frauds reduce the funds available to the healthcare system. Therefore, effective fraud detection is vital in reducing the cost and improving the quality of healthcare services.

Healthcare frauds come from many sources: service providers, insurance policyholders, and insurance carriers (Li et al., 2008). Among the sources, frauds from service providers are the most severe (Pflaum and Rivers, 1991; He, Wang, Graco and Hawkins, 1997; Yang and Hwang, 2006; and Li et al., 2008). Physicians play the most critical role among service providers because physicians determine the type of treatments (surgery vs. non-surgery) and length of hospital stays. The purpose of this paper is to construct features to detect fraudulent physicians using supervised learning and analyze the importance of features. Supervised learning is critical in understanding fraud since models are trained on labeled datasets where fraudulent behaviors are known. This training provides a better understanding of the characteristics that differentiate legitimate claims from fraudulent ones, thereby improving the accuracy of predictions when applied to real data (Dua and Bais, 2014). Detecting fraudulent physicians has important implications for insurers: saving claim costs, speeding up the claim review process, narrowing down the fraud investigation range, and excluding suspicious physicians as external reviewers for the insurers.

Our dataset consists of information from long-term health insurance policies issued by a Taiwanese life insurer. The reasons we use this dataset are stated below. First, the dataset contains confidential information that is not available to most researchers in other countries/regions (e.g., the U.S.). Specifically, the dataset contains the following information: physicians' characteristics, length of hospital stays, specific surgery descriptions, claim amount and frequency, actual claim payments, zip code of the physician and the policyholder, whether the insurer investigates a claim, whether a patient is a returned patient, and insurance agents' characteristics. Second, the Taiwanese health insurance market consists of both a public and private long-term health insurance market. Our results will have implications in both public and private insurance markets.

The dataset originally contains 922,154 claims covering in Taiwan from 1995 to 2018. After cleaning the data and transforming from claim level to physician level, the final sample is composed of 34,832 physicians from 2010 to 2018. Among them, 21,839 physicians file multiple claims (the multi-claim sample) while the rest file single claim (the single-claim sample). We first construct 32 features to describe the claim basics, physician characteristics, fraud strategies, early signals, insured characteristics, and agent characteristics, and these features are used for supervised learning. We find our sample is extremely imbalanced. Specifically, only 0.63% and 0.15% physicians in the multi-claim sample and the single-claim sample are labeled fraudulent, respectively. To address the imbalanced data issue, we choose two data sampling methods: the random under sampling method for RUSBoot model and the class weights method for the neural network model. Both error-based and cost-based measures are used for model performance.

We present the results of the neural network method and RUSBoot model below.[35] For the multi-claim sample, the neural network model performs slightly better than RUSBoost method; however, the difference is not statistically significant. We find AUROC score is 0.781 for neural network model with class weights method, indicating the model with class weights method performs reasonably well.[36] To estimate the savings from the identification of fraudulent physicians, we need to assume the fraud rate of fraudulent physicians' total claim amount. If we assume that the fraud rate of fraudulent physicians' total claim amount ranges between 30% and 70%, then the percentage saved by our models is approximately between 16.3% and 36.9% under the best-case scenario. For the single-claim sample, the cost of investigating fraudulent physicians outweighs the potential cost savings, possibly due to the extremely low percentage of fraudulent physicians and their small claim amounts. We also use the permutation importance method to identify the feature importance in distinguishing fraudulent physicians from legitimate ones and find that fraudulent physicians are more likely to be associated with those in the eastern area of Taiwan, the percentage of insureds whose age are less than 18, the percentage of surgeries due to illness, whether the physician can perform any surgery with higher severity, and the "steal a little, all the time" strategy because large

[35] The results of other classification method can be found in Appendix D.
[36] AUCROC is a performance metric used in machine learning to evaluate binary classification models.

claim amounts may arouse attention of the life insurer. Early signals from the life insurer are useful in detecting fraudulent physicians.[37]

This study contributes to the literature in several ways. First, this study uses data that contains unique, confidential, and valuable information. Without confidential data such as patients' and physician characteristics, it is difficult to detect fraudulent activities. All personal identification details, including names and addresses, were encoded prior to sharing the data with us. Second, our data is from a private insurer. According to Li et al. (2008), only 14% of research on healthcare fraud detection focuses on the private insurance area while the rest on the governmental health insurance area. Using data from a private insurer can provide additional insight since physicians and patients may have different perceptions about filing fraudulent claims against private insurers versus governmental health departments. On the one hand, they may be more likely to file fraudulent claims against governmental health departments because it is funded by tax revenues and private insurers are more likely to investigate suspicious claims to maximize profits. On the other hand, they may be less likely to file fraudulent claims against governmental health departments because the penalty can be more severe if they are caught. Third, our labels of fraudulent physicians are accurate and reliable since they are from one of three sources: the life insurer, the Supreme Court, and the Bureau of National Health Insurance (NHI) of Taiwan. Accurate labels are important for supervised machine learning methods, which provide a better understanding of the characteristics that differentiate legitimate claims from fraudulent ones (Dua and Bais, 2014). If the labels of fraudulent physicians are wrong, the prediction based on the wrong label would be incorrect as well. Fourth, focusing on the physician level is more efficient because one medical professional can submit multiple present- and future-claim submissions. Ortega et al. (2006) state that the detection of sources of fraudulent and abusive behavior, such as medical

[37] Claims associated with the physician that have been investigated or denied are considered early signals.

professional, is a far more efficient strategy than the analysis of individual medical claims. Fifth, we report the construction of 32 important features, while very few papers in the healthcare fraud detection literature disclose the features used. To our best knowledge, we are the first to analyze the importance of features in detecting fraudulent physicians. Sixth, we are the only paper to use cost-based performance measure in healthcare fraud detection literature except for Ortega et al. (2006). They unfortunately did not report the cost estimation in detail. Finally, we address the imbalanced data problem using random under sampling method and class weights method. Except for Herland et al. (2019), we are the only paper to address the imbalanced data issue in healthcare fraud detection literature, even though imbalanced data is the norm (Phua et al., 2004).

The remainder of the paper is organized as follows. Section 2 presents the background and literature. Section 3 describes the data and data pre-processing steps. Section 4 provides machine learning methodologies and performance measures. Section 5 presents results and Section 6 analyzes feature importance. Section 7 concludes.

## 3.2. Background and Literature Review

### 3.2.1. Background of Physician Frauds

We first provide the background of physician frauds. According to the National Health Care Anti-Fraud Association, healthcare fraud is defined as “an intentional deception or misrepresentation made by a person, or an entity, with the knowledge that the deception could result in some unauthorized benefit to him or some other entities” (Guidelines to healthcare fraud, 1991; Yang and Hwang, 2006)[38]. According to Li et al., (2008), healthcare frauds can typically be divided into three types based on the parties who commit the fraud: service provider (such as physicians, hospitals and ambulance companies); policyholders; and

[38] We are not distinguishing healthcare fraud and abuse in our paper.

insurance carriers (including governmental health departments and private insurers). Recently, conspiracy frauds, a fraud type involving more than one party, makes healthcare fraud even more complicated (Li et al., 2008; Liu et al., 2013).

Among all types, frauds from service providers account for the greatest proportion (Pflaum and Rivers, 1991; He, Wang, Graco and Hawkins, 1997; Yang and Hwang, 2006; and Li et al., 2008) because service providers can commit fraud in various possible ways and on a broad scale. Physicians, for example, engage in fraudulent behaviors that include, but are not limited to, the following: billing for services not rendered, unbundling, upcoding, providing unnecessary services or procedures, misrepresenting non-covered treatments as medically necessary, falsifying patients' diagnoses and/or treatment histories to justify unnecessary treatments, waiving deductibles, co-payments, or co-insurance, and engaging in corruption such as kickbacks and bribery (Li et al., 2008; Dua and Bais, 2014). Additionally, physicians can submit years-long fraudulent claims which can sum up to a massive amount. For example, Michael J. Ligotti, a Florida doctor, fraudulently billed approximately $121 million and $560 million for laboratory testing claims and other services to private insurers and Medicare from May 2011 to March 2020, respectively.

Our analysis focuses on physicians further because they make the major decisions on treatments including whether the surgery is necessary, and the length of hospital stays. According to Wynia et al. (2000), physicians decide on the services offering to their patients. It affects 80% of all healthcare expenditures and enormously affects healthcare quality. Additionally, focusing on physician fraud detection can generate a higher saving potential (Bayerstadler, Dijk and Winter, 2016). Besides financial costs, some types of fraud schemes such as surgeries will cause significant physical risk to the patients (Yang and Hwang, 2006). Thus, it is of high importance to construct features of physicians and build models to detect fraudulent physicians from the legitimate ones.

Traditionally, human experts review claims and identify healthcare frauds. This task is both time consuming and expensive (Matheus, Piatetsky-Shapiro, and McNeill, 1996; Li et al., 2008; Waghade and Karandikar, 2018). In recent years, electronic claim systems have been increasingly implemented to automatically perform audits and review claims data. However, these systems can only detect certain types of fraud according to pre-defined simple rules specified by experts (Liu and Vasarhelyi, 2013). To meet the needs of fraud detection, more sophisticated anti-fraud systems incorporating data mining, machine learning, and other methods are needed. These antifraud systems are developing to automatically learn fraud patterns from data and specify "fraud likelihood" of each case.

### 3.2.2. Literature Review

Classification algorithms such as decision tree, support vector machine (SVM), multi-layer perceptron (MLP), *k*-Nearest Neighbor, and random forest models are widely used in the literature to detect healthcare frauds at different levels. A systematic review by Ai et al. (2019) shows that 22%, 44%, 19%, and 15% of the analyses are on the physician level, claim level, facility level, and other levels, respectively.

We first review papers detecting frauds at the physician level. Based on the data from a Chilean private health insurer, Ortega, Figueroa, and Ruz (2006) propose a detection system that uses a committee of MLP neural networks for each entity involved in the fraud/abuse problem, including medical claims, affiliates, medical professionals, and employers. Their detection system discovers approximately 75 fraudulent and abusive cases per month, making the detection process 6.6 months earlier. They consider the costs of different types of errors while evaluating the model performance. However, they cannot list the features chosen and only report the results of the employer model due to a disclosure contract. He, Wang, Graco and Hawkins (1997) train a MLP neural network to classify the practice profiles of 1,500 general practitioners (G.P.) from the Health Insurance Commission of Australia. They use 28 features to summarize the G.P.'s practice over a year; however, the features are not reported due to legal and professional reasons. The agreement rate is 63% initially and reaches 70% while applying a proper probability cutoff. Using the same data and features, He et al. (2000) apply a genetic algorithm to the *k*-Nearest Neighbor technique to optimize the weights of all features and the agreement rate improves to 78%. Branting et al. (2016) apply a decision tree algorithm and propose graph analytics to generate 15 features as inputs. Twelve thousand healthcare providers excluded from the Office of Inspector General's List of Excluded Individuals and Entities (LEIE) are labeled as fraudulent and matched by the same number of randomly selected non-excluded providers. Their models generate mean F1-measure 0.919 and Area Under Receiver Operating Characteristics (AUROC) 0.960. The prior analyses neither do not reflect the imbalanced nature of fraudulent physicians in their datasets or do not address the imbalanced data issue except for Herland et al. (2019). They use the LEIE to generate fraud labels and apply the logistic regression, random forest, and gradient tree boosting methods. They use under sampling to deal with the imbalance data issue and AUROC is used as the performance measure.

Another stream of literature detects healthcare frauds at the claim level. For example, Yang and Hwang (2006) propose a data-mining framework using the clinical pathways concept to detect service providers' fraud, i.e., detecting abnormal procedures for a certain disease using data from National Health Insurance (NHI) of Taiwan. Lee, Kim, and Shin (2012) apply genetic algorithm to detect bill claim frauds. Using 600 claims for medical expenses in Seoul area, they find that genetic algorithm generates the highest AUROC comparing with decision tree, neural network, and regression analysis. Some papers label outliers as frauds with the assumption that outliers engage in frauds. For example, Pandy, Saroliya and Kumar (2018) propose a rules-based scoring system and the claim frequencies lying in the third quantile are considered to have higher probability of being fraudulent and are labeled as frauds. Using the database of a

Turkish insurer, Kirlidog and Asuk (2012) choose SVM and mark the claim as fraud if the probability of anomaly is larger than 50%.

There are few papers that focus on the facility level. Shin et al. (2012) propose a scoring model to detect outpatient clinics with abusive utilization patterns based on profiling information. Using the NHI database in Taiwan, Liou, Tang and Chen (2008) use logistic regression, neural networks, and classification trees to detect fraudulent hospitals whose contracts are terminated.

Most papers discussed above use relatively small datasets from governmental health departments while a few of them use data from private insurers. Thus, using a large dataset from a private insurer adds valuable information to the literature. Except for Ortega, Figueroa, and Ruz (2006), the literature uses error-based methods to evaluate model performance which do not consider the cost differences between different types of error. Ortega, Figueroa, and Ruz (2006), unfortunately, cannot report their estimation of the costs of different errors in detail.

Although various algorithms have been applied to physician fraud detection, very few papers emphasize feature construction. Additionally, to our best knowledge, no paper mentions the importance of features in detecting fraudulent physicians. Table 3.1 summarizes the most relevant studies on feature construction at the physician level. Panel A reports papers using supervised learning while Panel B also includes unsupervised learning methods to illustrate as many features as possible[39]. Columns 2-5 present the number of features, feature set, data sources, and algorithms examined in each paper, respectively. A few observations can be implied. First, the number of features ranges from 5 to 125. Second, many papers do not report features or briefly mention them due to either data confidentiality or legal restrictions. Third, for papers report features, features constructed vary widely due to the sources of data and physician’s segmentation of operation. Unlike other fraud detection area, e.g., credit card fraud

[39] Supervised learning requires labels on historical fraud data while unsupervised learning does not need labels.

detection, the raw fields of each transaction are standardized while reporting and the literature/industry can form a framework for feature construction. Thus, constructing a feature set and analyzing their importance in detecting fraudulent physicians from the legitimate ones are useful not only in physician fraud detection area but also providing insights to form more standardized claim report format.

## 3.3. Data and Data Preprocessing

### 3.3.1. Data

In this study, we use a unique and confidential claim dataset from one of the top five life insurers in Taiwan. The claim dataset is from a surgery endorsement added to a primary life insurance policy. The surgery endorsement pays out to the insured if she undergoes a covered surgery. The dataset originally contains 922,154 claims covering in Taiwan from 1995 to 2018[40]. The dataset starts in 1995 but with limited records on claims (4,365 claims) from 1995 to 2009; thus, we exclude these claims. Our final sample period is from 2010 to 2018. For each claim record, the dataset provides information on the claim date and amount, the surgery endorsement policy, the primary life insurance policy, the characteristics (e.g., gender, age, education, marriage, and zip code) of the insured, agent, claim agent, and investigator if applicable. The dataset also provides information related to the surgery such as the diagnosis code, diagnosis description, surgery code, surgery description, length of stay, physician, hospital location, etc. Given the confidential nature of the data and consumer privacy protection, all individual identification data elements such as names and addresses were coded by the life insurer before sharing the data with us. We clean the dataset by deleting claims without information needed later such as actual claim amount paid by the insurer, length of stay, hospital zip code, etc. This step reduces the number of claims to 899,983.

[40] We got partial claims for year 2018.

The life insurer labels the fraudulent physicians in the dataset based on the claim history and other confidential information. Additionally, we hand collect data from the judgments of the Supreme Court of Taiwan and the blacklist of social insurance (National Pension Program) to label other fraudulent physicians. In summary, the fraudulent physicians are labeled from three sources: the life insurer, the Supreme Court, and the social insurance. Although the physicians labeled as "fraudulent" are clean due to the cumulated claim experience and the governmental resources, it should be noted that the physicians labeled as "legitimate" may contain noises. The reasons are as follows. Some physicians are too cunning to be detected and still marked as legitimate. A physician can behave in a fraudulent manner some of the time but not at other times (Bolton and Hand, 2002); thus, they are hard to be detected. Another limitation of our dataset is that we cannot determine the exact claims from which the fraudulent physicians are labeled.

### 3.3.2. Data Transformation

Because our goal is to detect fraudulent physicians, we reconstruct the data from the claim level to the physician level. This leads to 34,832 physicians in our sample. As observed in Figure 3.1, the number of claims from physicians varies widely. Specifically, 37.3% of physicians correspond to a single claim, while the remaining 62.7% of physicians have more than one claim. Among physicians with multiple claims, 53% file between 2 and 10 claims, 26% file between 11 and 50 claims, and the rest of them file more than 50 claims. The average number of claims is 39 for those with multiple claims. The broad fraud detection literature (e.g., Whitrow et al., 2009; Wu et al., 2019) suggests that the aggregated information leads to better performance than a single claim/transaction. We separate physicians into two groups: those with multiple claims and those with single claim (multi-claim sample and single-claim sample later for short).

Table 3.2 summarizes the number of physicians, the percentage of fraudulent physicians, the number of claims, total claim amounts and other variables for the multi-claim sample and the single-claim sample. Our analyses focus on the multi-claim sample for the following reasons. First, the percentage of labeled fraudulent physicians is 0.63% for the multi-claim sample while it is 0.16% for the single-claim sample. Second, physicians with single claim have lower total number of claims, the percentage of suspicious claims, total claim amount, and the average claim amount compared with physicians with multiple claims. Thus, the life insurer should be less worried about physicians with single claim and the benefits from detecting these physicians are lower. Third, aggregated information can be more useful in physician fraud detection.

### 3.3.3. Feature Construction

Given the data availability, we construct features to describe the physicians' practice in consultation with the industry experts and by referring to the literature. The features listed below are based on aggregated claim information. Table 3.3 reports the summary statistics for the multi-claim sample, including the means and standard deviations of all features for the fraudulent and legitimate physicians and the *p*-value for the t-tests of mean differences. The corresponding features for physicians with single claim are presented in Appendix B. We categorize the features into six groups: claim basics, physician characteristics, fraud strategies, early signals, insured characteristics, and agent characteristics. These features within the six groups are discussed in detail below.

1) Claim characteristics

Following the physician fraud detection literature (Lin, Lin, Li and Kuo, 2008; Musal, 2010; Bauder and Khoshgoftaar, 2016; Johnson and Nagarur, 2016; Herland, Bauder and Khoshgoftaar, 2019), we use the natural logarithm of the number of claims of the physician (*lnn_claim*) and the natural logarithm of the average claim amount paid by the insurer to the physician (*lnavg_claim_actual_amt*) to represent the claim characteristics.

2) Physician characteristics

Physician characteristics are represented by the physician's surgery practice, the percentage of returning patients of the physician, surgery types, the affiliation of physicians, and the location of the physician's affiliation.

The number of unique surgery categories (*n_surgery_type*) and the diversification of surgery categories (*hhi_surgery_type*) are used to characterize the physician's surgery practice. The surgery category system is the classification system that provides a detailed description of surgeries, shared by all physicians. If the physician performs *I* unique types of surgeries, let *n_surgery_type* = *I* denote the count for $i$ unique surgery and $n_i$ times for each surgery category $i$ and then define diversification of surgery categories as *hhi_surgery_type* = $\sum_{i=1}^{I}(\frac{n_i}{\sum_{j=1}^{I} n_j})^2$. We argue that specialized physicians are less inclined to commit fraud due to the higher opportunity cost because specialists are paid more. Thus, physicians who perform more unique surgery categories and more diversified are more likely to commit frauds. The life insurer ranks surgery complexities from 1 to 10 with 10 being the most complicated. Physicians who can perform complicated surgeries are more skilled, paid more, and have a higher opportunity cost if being caught. Thus, they have lower incentives to commit fraud. We measure the complexity with a dummy variable (*d_complexity6*) which equals 1 if the physician can perform surgeries with a complexity 6 or higher.

Following Lin, Lin, Li and Kuo (2008), we include the percentage of returning patients of the physician (*pct_retpatient*) as one characteristic. While the percentage of returning patients is not directly associated with the characteristics of physicians, it is an indirect measure because one way for a physician to commit fraud is to ask patients to come again and again so that the physician can charge more. In other words, physicians are more likely to commit conspiracy frauds with returning patients than with new patients. We also measure the percentage of surgeries performed by the physician due to illness rather than accidents (*pct_sur_illness*). We argue that it's less likely for physicians to commit fraud on surgeries due to accidents because accidents are typically associated with some evidence of police reports.

We categorize the affiliation of physicians into four types of hospitals and create a dummy variable for each type. Based on the size and prestige of the affiliations, they are medical centers (*d_medical_center*), regional hospitals (*d_regional_hospital*), district hospitals (*d_district_hospital*), and clinic hospitals (*d_clinic_hospital*). Physicians in medical centers relatively have less incentives to commit frauds since they usually receive fixed salary. Additionally, they have more reputation concern and the higher opportunity cost while being caught. Smaller medical institutions are more inclined to commit frauds. Clinic hospitals, however, are less likely to commit frauds in our case since they have fewer chances to operate surgeries. Thus, we argue that fraudulent physicians are more likely to be in regional hospitals and district hospitals.

We further catergorize the location of the physician's affiliation into five regions according to the geographical location and the economic development of the region. *Region0* is the southeast of Taiwan including Keelung, Yilan, Hualien, Quemoy, and Lienchiang; *Region1* is the northeast including Taipei, New Taipei and Taoyuan; *Region2* is the northwest including Hsinchu and Miaoli; *Region3* is the west including Taichung, Changhua, and Nantou; *Region4* is the southwest including Yunlin, Chiayi and Tainan; and *Region5* is the east including Kaohsiung, Pingtung, Taitung, or Penghu. Among them, *Region1* is the most developed in terms of economy while *Region5* is the least developed.

3) Fraud strategies

Sparrow (2000) and Travaille et al. (2011) describe the "hit-and-run" and the "steal a little, all the time" as two extremes in fraud strategy spectrum. The former strategy acquires large amounts of money quickly and disappear while the later one uses bulk of seemingly legitimate claims to hide the incremental stealing.

The "hit-and-run" strategy is measured by the maximum claim amount to total claim amounts (*pct_max_sum_amt*). The "steal a little, all the time" strategy is proxied by whether the physician charges an abnormally high price per surgery or performs an abnormally high frequency of surgeries. Specifically, we use *abvp75_claim_amt* to measure the abnormally high price. Given each surgery category $i$ and year $t$, we aggregate the amount differences if the surgery's claim amount of the physician is larger than the 75$^{th}$ percentile claim amount in the sample and 0 otherwise, and scale it by the physician's total claim amount[41]. One method to charge a high price unnoticed is through filing a longer length of stay for the same surgery category. For each $i$ and $t$, we count the number of claims with the length of hospital stays greater than the 75$^{th}$ percentile in the sample, and scale it by the physician's total number of claims (*d75_hosp_days*). We generate a dummy variable, *abnm_freq_surgery_10*, to measure the abnormally high frequency of surgeries. We use this dummy variable if a physician performs a certain surgery category more frequently than other physicians and the surgery can be performed by many other physicians, then it is a red flag indicating the physician is more likely to be fraudulent. For any $i$ and $t$, if the physician performs more than the average number of surgeries in the sample and the surgery can be performed by more than 10% of all physicians that year, we set the dummy variable equal to 1; 0 otherwise.

4) Early signals

Following Hillerman, Souza, Reis, and Carvalho (2017), e treat claims from the physician that have been investigated or denied as early signals. Investigations are usually initiated by home office consultants and finalized by regional investigators. A physician under investigation is indicative of a higher likelihood of committing fraud in the future. It is

---

[41] All dollar amounts are in 2016 value.

measured by the natural logarithm of the number of claims of the physician being investigated by the life insurer scaled by the length of service (*lninv_seryr*). Claim denials are measured by the percentage of claims related to the physician with amount paid by the life insurer less than the filed amount (*pct_per_benefit*).

5) Insured characteristics

Some studies show that the importance of geo-location in healthcare fraud detection (Musal, 2010; Liu and Vasarhelyi, 2013; Branting et al., 2016; Johnson and Nagarur, 2016). We argue that if the insured travels an abnormally long distance to visit a physician, the physician is more likely to be fraudulent. Specifically, we calculate the driving distance between the insured and the physician (i.e., the affiliated hospital) using the centroids of zip codes.[42] If any insured's driving distance is longer than the 75$^{th}$ percentile driving distance of all patients in the insured's zip code and she is not going to a medical center. We set the dummy variable *distp75_ins_hosp* equal to 1, and 0 otherwise. Another measure is abnormal visit (*abnm_visit*). We set the dummy variable, *abnm_visit*, to equal 1 if the insured travels to an abnormal place to visit a physician, and 0 otherwise. If an insured person lives in the metropolitan area but travels to a non-metropolitan area to see a physician, this visit is considered abnormal because the hospitals in the metropolitan area are better. A physician with an abnormal visit pattern from patients is more likely to be fraudulent.

Other characteristics of the insureds are also included. Following Hillerman, Souza, Reis and Carvalho (2017), we include the percentage of insureds whose age are less than 18 (*pct_evtins_age_lt18*) or greater than 65 (*pct_evtins_age_gt65*). These two groups are less likely to question physicians for additional services. Younger patients, however, are usually dependents of the primary insurance holders who can monitor the services needed.

[42] We would like to thank Bing map for the free access to academy.

The insured can initiate or collaborate on frauds with the physician, a phenomenon known as conspiracy fraud. We generate a dummy variable (*d_blacklist_insured*) that equals 1 if insured is on the blacklist of the life insurer in the physician's claims, and 0 otherwise. We also include the natural logarithm of the average endorsement face amount to the primary life insurance face amount (*lnavg_rider_prmy*) as a feature because a higher face amount may indicate a higher probability of fraud. Finally, we add the percentage of insureds who are male (*pct_insured_male*) as a feature.

6) Agent characteristics

The agent and the claim agent are the gatekeepers while underwriting policies and filing claims. We measure the percentage of agents (*pct_agent_punish*) and claim agents (*pct_claim_agent_punish*) who have been punished by the life insurer in the claims related to physician fraud since sales agents and claim agents who have a history of fraudulent claims are more likely to commit fraud again.

## 3.4 Methodology

The steps for our machine learning approach are as follows. First, we split the multi-claim sample[43] into the training set and the test set. The training set is used to fit the model, while the test set is used to evaluate the model's performance. Specifically, we use 80% of the physicians in the multi-claim sample for the training set and the rest 20% as the test set. Several models are chosen including RUSBoost, and neural network. Other methods' results can be found in Appendix D. We address the imbalanced data issue in the training algorithm. Both error-based and cost-based methods are used to evaluate model performance. We repeat the procedure 10 times and compute the mean and standard deviation of the performance measures of the 10 trials.

[43] The same for the single-claim sample.

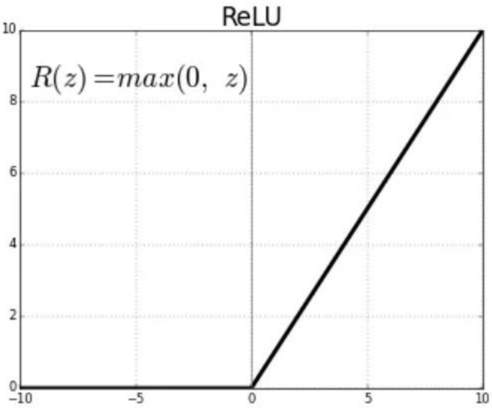
### 3.4.1. Address Imbalanced Data Issue

Imbalance data refers to a dataset within which the number of minority class observations is far less than the majority class. Imbalanced data issue is common in the rare events detection problems such as fraudulent credit card transactions (Panigrahi et al., 2009), accounting fraud in U.S. publicly traded firms (Bao et al., 2020), and healthcare fraud detection (Herland et al., 2019). Rare events detection problems can be viewed as classification problems and are difficult to observe because of infrequency. For example, in our dataset, only 0.63% and 0.15% of physicians in the multi-claim and single-claim samples are labeled fraudulent, respectively. Imbalanced data will cause models to perform well for the majority class but underperform for the minority class (López et al., 2013). However, the main purpose of the model is to detect rare events, which are minority observations. Thus, it's essential to take the imbalance data issue into consideration.

We use two methods to address the imbalance data issue. The first method is random under sampling, which is used to rebalance the data by randomly removing observations from the majority class. The second method is class weights, where the model assigns the class weights inversely proportional to their respective frequencies. A higher class weight means that the model emphasizes this class and penalizes mistakes in it more heavily. For the imbalanced data, a higher weight is given to the algorithm's cost function of minority observations during the training, providing a higher penalty for the mistakes in the minority class to reduce errors.

### 3.4.2. Model Specification

The neural network is a method that processes data in a way that mimics the human brain in which the biological neurons signal to one another and can learn from data and make intelligence predictions with limited human assistance. The neural network comprises an input layer, one or more hidden layers, and an output layer. The input layer contains information from the real world, which is captured by features. The features are the input variables. The hidden layers are between the input layer and the output layer. The input of hidden layers is the output from the previous layer, and then processes information further and passes the information to the next layer. The output layer gives the final results of the whole processing progress. There are artificial neurons on each layer, and every neuron is connected to every other neuron in the next layer. If the output of an individual neuron is above a threshold, which is determined by the activation function, then the information of this neuron will be passed to the next layer of the neural network. Otherwise, the information of the neuron will not be analyzed by the next layer. The ReLU non-linear activation function is the most commonly used in neural networks.




A loss function of the neural network is used to assess how well the coefficients are doing at the classification problem. The mean squared error (MSE) is an example of the loss function. In our neural network model, we use the focal loss function introduced by Lin et al., 2017. The focal loss function can address imbalanced data issues by penalizing misclassification observations heavily. The weight parameters in the focal loss function are the class weights, which are inversely proportional to the class's respective frequencies. The focal loss function is defined as

$$L(y, \hat{p}) = -\alpha y(1-\hat{p})^{\gamma} \log(\hat{p}) - (1-y)\hat{p}^{\gamma} \log(1-\hat{p})$$

Where y is a binary class label, $\hat{p}$ is the estimated class probability, $\alpha$ is class weight, and $\gamma$ is the focusing parameter. The higher $\gamma$ indicates a higher penalty for misclassified observations. The goal is to minimize the loss function to get better estimated coefficients. We use FastAI deep learning library to train the model.

RUSBoost is a hybrid method of random under-sampling and AdaBoost algorithm, alleviating the imbalanced data issue and improving model performance. Random under-sampling (RUS) removes observations of the majority class randomly until the imbalance data is resampled to balanced data. Thus, the minority class will not be treated as a noise of the sample and will improve the prediction accuracy of the minority class. The AdaBoost algorithm is one of the most common and efficient ensemble learning methods. An ensemble learning method is an approach that combines multiple models to provide better predictive performance. The AdaBoost algorithm aims to train a sequence of classifiers, such as Decision Trees, and reweights misclassified observations in the next iteration. For each iteration, the weight of misclassified observations will increase in the subsequent iteration, forcing the future classification to concentrate more on the observations that are misclassified in the previous iterations. Therefore, in subsequent iterations, the model is more likely to correctly classify observations that were misclassified in previous classifications. When iterations are completed, the final classification result will be the majority vote of models. This method is particularly effective for imbalanced data because the minority class is more likely to be misclassified. In summary, the RUSBoost method performs random under sampling in each iteration to address the issue of imbalanced data. Then, based on AdaBoost, it reweights the misclassified observations to achieve better prediction performance. We use the Python scikit-learn package's "RUSBoostClassifier" to train the model. The algorithm of RUSBoost can be found in Appendix C.

### 3.4.3. Performance Measure

Both error-based and cost-based methods are used to measure the model's performance.

We first use error-based method to evaluate the model's performance according to the confusion matrix. Figure 3.2 is a confusion matrix where positive and negative stand for fraudulent and legitimate, respectively. True positive, TP, represents the number of fraudulent physicians correctly predicted by the model while false negative, FN, represents the number of fraudulent physicians neglected by the model. Neglecting fraudulent physicians can lead to a severe problem since a physician can submit hundreds of claims over years before being caught. False positive, FP, represents the number of legitimate physicians who are incorrectly marked as fraudulent by the model while true negative, TN, represents the number of legitimate physicians who are correctly predicted. Following the literature (Hanley and McNeil, 1982; Ferri et al., 2011; Müller and Guido, 2016), AUROC is used as the indicator for the discriminating power of the classifier. It is the area under the receiver operating characteristic curve which plots the true positive rate (Equation (1)) against the false positive rate (Equation (2)) at different decision-making thresholds. AUROC equals 1, indicating the classifier performs perfectly, while 0.5 indicates randomly. The more AUROC is closer to 1, the better the classifier is.

$$True\ positive\ rate = TP/(TP + FN) \quad (1)$$

$$False\ positive\ rate = FP/(TN + FP) \quad (2)$$

We use the cost-based method to estimate the cost savings achieved by using machine learning to predict fraudulent claims. The cost-based method calculates the savings by considering the difference between the cost of investigating fraudulent physicians and the cost incurred from neglecting such fraud. The cost-based performance evaluation method is preferred since the class size is uneven and the cost for different types of misclassifications is different (Bolton and Hand, 2002). Table 3.4 summarizes the costs to the insurer based on the outcomes in the confusion matrix, where *C_investigation*, *C_physician_legit*, and *C_physician_fraud* stand for the average cost of investigation and the average total claim amounts related to legitimate and fraudulent physicians, respectively.

Following Phua et al. (2004), the cost savings of the model can be calculated using Equation (3).

$$\begin{aligned}\text{Model cost savings} &= \text{No action} - (\text{Misses costs} + \text{False alarms costs} + \text{Normal costs} \\ &\quad + \text{Hits costs}) \\ &= \text{TP} * (C\_physician\_fraud \ * Fraud\ rate - \ C\_investigation\ ) - \text{FP} \\ &\quad * C\_investigation\end{aligned} \tag{3}$$

where No action means the life insurer does not use any model to detect frauds and the fraud rate is defined as the percentage of total claim amounts that are assumed to be frauds. It is hard to estimate the fraud rate and we assume it ranges from 30% to 70%. The higher the fraud rate, the larger the benefits of detecting fraudulent physicians. *C_investigation*, and *C_physician_fraud* are costs of investigation and costs of physician fraud, respectively.

Equation (4) restates the model's cost savings as a percentage by scaling it to the best-case scenario, where all physicians are correctly predicted by the model as legitimate or fraudulent. Since the cost-based method is threshold-dependent, we choose the threshold giving the optimal cost savings.

$$\begin{aligned}\text{Percentage saved} &= \frac{\text{Model cost savings}}{\text{Best} - \text{case scenario cost savings}} * 100\% \\ &= \frac{Model\ cost\ savings}{(TP \ + FN) * \left(C_{physician_{fraud}} * Fraud\ rate \ - \ C_{investigation}\right)} * 100\%\end{aligned} \tag{4}$$

To investigate the physician, investigators of the life insurer examine claims related to the physician, overview the claims including the certification of diagnosis, and request detailed medical records from hospitals for the most suspicious claims. Thus, the cost of investigation is composed of two parts: the labor costs of the investigator and the cost of medical records from hospitals. We assume the investigator requests medical records for 10% of all claims related to the physician for two reasons. It is estimated by the U.S. Department of Health and Human Services that 10% of annual healthcare spending is fraud. Since requesting medical records needs permission from the insured, a high percentage may lead to customer dissatisfaction. According to the life insurer, the average annual salary for investigators is NT$ 1-1.2 million. There are approximately 30 investigators in the company and around 37,000 cases are investigated each year. Thus, the labor cost per investigation is around NT$ 811-973. We round it to NT$ 1000 to adjust for the bonuses they may receive occasionally while dealing with complicated cases. We use the average cost of medical records per claim in our dataset, NT$ 1,000, as the proximation for the cost of each medical record. Since the average number of claims is 105 for labeled fraudulent physicians with multi-claims; further investigating 10% of claims leads to approximately 10.5 medical records from hospitals, costing NT$ 10,500. As a result, the estimated average cost of investigation equals NT$ 11,500 for physicians with multi-claims and NT$ 2,000 for physicians with single claim.

We use the average total claim amounts related to labeled fraudulent physician in our sample as the estimation for *C_physician_fraud*. For the multi-claim sample, the average total claim amount is NT$ 1,841,428 while for the single-claim sample, the average total claim amount is NT$ 7,501.

### 3.5. Results

We first report results for physicians with multi-claims since they can cause larger damage to the life insurer. Table 3.5 reports the AUROC scores for the neural network, RUSBoost, and Logit models, respectively. The Logit model is used as a baseline to compare with the neural network and RUSBoost models. Overall, AUROC is 0.78 with a standard deviation of nearly 0.04, indicating the models all perform reasonably well. It can be observed that the neural network model performs better than the RUSBoost models, but the advantages are not obvious. The baseline model Logit performs worse than the neural network and

RUSBoost models. The performances of the different methods addressing the imbalanced data issue are similar while the class weight method slightly outperforms the others. We also try other conventional classification methods with different resampling methods. The results of other methods can be found in Appendix D.

Table 3.6 presents the precision and recall scores for the neural network, RUSBoost, and Logit models. The precision and recall tradeoff is a common issue in imbalanced data when evaluating the performance of classification models. In general, increasing the recall will decrease the precision, and vice versa because they are inversely related. To detect fraudulent physicians, high recall scores are more important than precision scores. Recall is defined as the proportion of actual positives (true fraudulent physicians) that are correctly identified as such by the model. The precision score measures how many of the observations predicted as positive (fraudulent physicians) are actually positive. The recall score of the neural network is 0.911, which is significantly higher than the other two models, indicating that the neural network method performs well in detecting fraud physicians.

Table 3.7 reports the percentage of cost savings for the neural network and RUSBoost models. The neural network model performs persistently better than RUSBoost model. We assume different fraud rates for submitted claims to conduct sensitivity analyses. As the fraud rate increases from 30% to 70%, the percentage saved by the neural network model increases from approximately 16.3% to 36.9%. The evidence suggests that our neural network model can achieve 16.3% in cost savings when scaling to the best-case scenario, where all physicians are correctly predicted, assuming that 30% of the total claim amounts from fraudulent physicians are fraudulent. This ratio increases to 36.9% if 70% of their total claim amounts are fraudulent. Note that the actual cost savings of our models should be greater than these numbers since it is highly likely that those physicians labeled as legitimate but predicted by our models as fraudulent are, in fact, fraudulent.

Using the neural network with class weight method as an example, we further illustrate the dollar amount saved by our model. Figure 3.3 shows that the amount saved is monotonically increasing to the fraud rate. The amount saved is NT$ 2.37 million when 30% of the total claim amounts are frauds of the fraudulent physicians. It enlarges to NT$ 6.84 million when the fraud rate is 50% and to NT$ 12.74 million when the fraud rate is 70%. The amount saved can be even larger if the fraud rate is greater than 70%.

For physicians with single claim (not tabulated), our models do not perform better than randomly assigning a physician as fraudulent or legitimate since the average AUROC is nearly 0.5. Additionally, the percentage saved is almost 0 since the cost of investigating suspicious physicians offsets the cost savings from detecting fraudulent physicians. The results make sense since there are only 20 fraudulent physicians out of 12,993 physicians and their average total claim amount is just 0.41% of that of fraudulent physicians with multi-claims. In short, due to the low frequency as well as low severity, it would be more economical to  physicians with single claim until they file more claims. As discussed by Bolton and Hand (2002) that our machine learning models do not intend to reduce frauds to the lowest level but rather getting a tradeoff between the cost of detecting a fraud and the savings to be made by detecting it.

Besides model cost savings, our results also generate additional economic value for the insurer by predicting which physicians are more likely to be fraudulent. First, it can accelerate the claim review process since the reviewers only need to pay attention to claims from those marked physicians. Policies in the surgery endorsement require the insurer to pay within 15 days of receiving the claim; otherwise, there will be interest on the delayed amount paid. Accelerating the review process without omitting many suspicious claims generates economic value. Additionally, it can largely narrow down the range of physicians to be investigated and improve the efficiency of the investigation. According to the experts, no more than 3% of all claims can be investigated due to client satisfaction and investigation costs. Finally, if the insurer needs external reviewers for some claims, the life insurer can avoid engaging those who are more likely to commit fraud.

### 3.6. Feature Importance

To understand the features that play the most important roles in physician fraud detection, we employ the permutation importance method. Permutation importance is a method used to calculate the contribution of each feature to the prediction accuracy of a model, providing insight into how much the prediction accuracy depends on the feature. It is calculated in five steps. The first step involves training the model and obtaining the performance accuracy metrics. The second step requires randomly shuffling one feature in the dataset and then making predictions using the shuffled dataset. In the third step, the accuracy metrics are recalculated based on the predictions made in the second step. The fourth step involves calculating the decrease in model performance due to the shuffled feature. A significant decrease indicates that the original value of the feature is crucial for making accurate predictions. The last step is to repeat the process for each feature in the dataset. The interpretation of permutation importance is that if the model relies heavily on a feature for making accurate predictions, shuffling the feature's values will significantly worsen the model's performance. Based on the multi-claim sample, Figure 3.4 ranks in descending order the features that contribute most to the neural network with the class weight model's prediction of fraudulent physicians, measured by the average permutation importance of ten random states.

We first analyze the features related to physician characteristics. The most important feature is *Region5*, which is in the eastern area of Taiwan and is the least developed area. It suggests that insurers need to pay more attention to the physicians in the least developed area. For the hospital location, we observe a strong geo-clustering effect for *Region3* and *Region1*, which are the sixth and fourteenth most important features, respectively. Meanwhile *Region0, Region2*, and *Region4* are less informative features. *d_complexity6* is the third most important feature, indicating that more skilled physicians have a higher opportunity cost, and this feature is important for detecting fraudulent activities. The fourth most important feature is *pct_sur_illness,* which means the percentage of surgeries due to illness. It indicates that life insurers need to pay more attention to surgeries due to illness. One possible explanation is that it might be more difficult for physicians to cheat on surgeries due to accidents. The fifth most important feature is *d_medical_center,* suggesting that physicians in the medical center are an important factor in detecting fraudulent activity. *pct_retpatient* is the thirteenth most important feature, suggesting that returning patients, compared to new ones, are more likely to be associated with fraud. For other features related to physician characteristics, *hhi_surgery_type* is the twelfth most important feature, indicating that the diversification of surgery categories is an important feature in detecting fraud. For physician affiliation, life insurers need to pay more attention to physicians who are in district hospitals and regional hospitals than in clinic hospitals and medical centers. The least important feature, *n_surgery_type,* further indicates that fraudulent physicians are less likely to be associated with performing a variety of different surgery categories.

We next analyze features on claim characteristics. *lnn_claim* is the tenth most important feature, indicating fraudulent physicians are more likely to associate with more claims. Additionally, *lnavg_claim_actual_amt* is ranked as the twenty-eighth most important feature, indicating that fraudulent physicians are not significantly different from legitimate physicians in terms of the average claim amounts paid by insurers, this suggests the effectiveness of claim investigators. The larger claim number is a red flag to detect fraudulent physicians.

We next analyze the features of fraud strategies. The eighth most important feature is *d75_hosp_days,* indicating that fraudulent physicians are more likely to require patients to stay longer in the hospital. *pct_max_sum_amt* is the thirtieth most important feature showing that fraudulent physicians are less likely to file a huge claim amount. Taking these two features together, it illustrates that fraudulent physicians prefer the “steal a little, all the time” strategy rather than the “hit and run” strategy. Filing a huge amount may arouse the attention of the life insurer, and it is more likely to be caught. Filing a moderate amount through longer hospital stays and multiple times is a “safer card” to play for fraudulent physicians but makes fraud detection harder for the life insurer. Additionally, *abvp75_claim_amt* is the twenty-fourth most important feature. Thus, detecting fraudulent physicians based on abnormally high prices may also be effective.

Early signals from the life insurer are useful in detecting fraudulent physicians. The eleventh most important feature is *lninv_seryr*, which shows that fraudulent physicians are more likely to be associated with arousing attention from the life insurer and get more investigations during their service years.

Among the features of insured characteristics, the age of insureds is useful in detecting fraudulent physicians. *pct_evtins_age_lt18* is the second most important feature. One possible explanation is that patients less than 18 are in general the dependents of their parents and parents are more care about children so that it may cause overtreatment. *distp75_ins_hos* is the seventeenth most important feature related to fraudulent physicians, which indicates if the insured travels to an abnormal place to visit, the physician may be fraudulent.

Agent characteristics also play an important role in detecting fraudulent physicians. The ninth most important feature, *pct_claim_agent_punish*, shows that a higher percentage of punished claim agents is associated with fraudulent physicians. Meanwhile, the fifteenth most important feature, pct_agent_punish, illustrates that a higher percentage of punished agents is also an important feature in detect fraudulent physicians.

In summary, the cluster in the eastern area of Taiwan, the percentage of insureds whose age are less than 18, the percentage of surgeries due to illness (vs. accident), and whether the physician can perform any surgery with high complexity are important features to detect fraudulent physicians. They prefer the “steal a little, all the time” strategy and try to avoid large claim amounts, which may arouse the attention of the life insurer. Early signals from the life insurer are also useful in detecting fraudulent physicians.

### 3.7. Conclusion

Physician fraud detection is an important part of healthcare fraud detection. Physicians can file claims costing millions of dollars and quickly adjust fraudulent strategies according to the changes in detection methods. Thus, it is vital to continuously update detection methods and find new patterns of fraud.

Using the healthcare claim data from a life insurer in Taiwan, we first transform the data to the physician level and generate 32 features, including claim basics, physician characteristics, fraud strategies, early signals, insured characteristics, and agent characteristics. Then, we apply the neural network and RUSBoost models to detect fraudulent physicians. The class weight and random under sampling methods are used to address the imbalanced data issue.

For the multi-claim sample, the neural network model performs reasonably well, with the AUROC scores around 0.781. As the fraud rate of fraudulent physicians’ total claim amount increases from 30% to 70%, the percentage saved from our models increases from approximately 16.3% to 36.9% of the best-case scenario cost savings. The cost savings calculated are the lowest since some fraudulent physicians correctly predicted may still be labeled as legitimate. For the single-claim sample, the cost of investigation outweighs the cost savings from detecting fraudulent physicians due to the extremely low percentage of fraudulent physicians and their small claim amounts.

We use the permutation importance method to understand the feature importance in distinguishing fraudulent physicians from legitimate ones. We find that fraudulent physicians are associated with the cluster in the eastern area of Taiwan, the percentage of insureds whose age are less than 18, the percentage of surgeries due to illness, and whether the physician can perform any surgery with high complexity. Our evidence suggests that physicians prefer the "steal a little, all the time" strategy and try to avoid large claim amounts, which may arouse the attention of the life insurer. Early signals from the life insurer are also useful in detecting fraudulent physicians.

Besides the cost savings discussed above, our results generate other economic value for the insurers in practice. First, it can accelerate the claim review process since the reviewers only need to pay attention to claims from those marked physicians. Second, it can largely narrow down the range of physicians who need to be investigated and improve the efficiency of the investigation. Finally, if the insurer needs external reviewers for some claims, the life insurer can avoid engaging those who are more likely to commit fraud.

Future works include, but are not limited to, the following: It would be better to address the noises in the labels since it is highly likely that some fraudulent physicians are not detected and still marked as legitimate. We can consider the hybrid machine learning technique, which combines supervised learning and unsupervised learning to improve the classifier's performance. Detecting fraudulent claims in addition to fraudulent physicians is another interesting question.

# REFERENCES

Aral, K. D., Güvenir, H. A., Sabuncuoğlu, İ., and Akar, A. R. 2012. A prescription fraud detection model. *Computer methods and programs in biomedicine*, *106*(1), 37-46.

Bao, Y., Ke, B., Li, B., Yu, Y.J. and Zhang, J. 2020. Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research*, 58(1), pp.199-235.

Bayerstadler, A., van Dijk, L., and Winter, F. 2016. Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance. *Insurance: Mathematics and Economics*, 71, 244-252.

Bolton, R. J., and Hand, D. J. 2002. Statistical fraud detection: A review. *Statistical Science*, 235-249.

Branting, L. K., Reeder, F., Gold, J., and Champney, T. 2016. Graph analytics for healthcare fraud risk estimation. *In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 845-851. IEEE.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.

Copeland, L., Edberg, D., Panorska, A. K., and Wendel, J. 2012. Applying business intelligence concepts to Medicaid claim fraud detection. *Journal of Information Systems Applied Research*, 5(1), 51.

Dua, P. and Bais, S. 2014. Supervised learning methods for fraud detection in healthcare insurance. *Machine learning in healthcare informatics*, pp.261-285.

He, H., Hawkins, S., Graco, W. and Yao, X. 2000. Application of genetic algorithm and k-nearest neighbour method in real world medical fraud detection problem. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 4(2), pp.130-137.

He, H., Wang, J., Graco, W., and Hawkins, S. 1997. Application of neural networks to detection of medical fraud. *Expert systems with applications*, 13(4), 329-336.

Herland, M., Bauder, R. A., and Khoshgoftaar, T. M. 2019. The effects of class rarity on the evaluation of supervised healthcare fraud detection models. *Journal of Big Data*, 6(1), 21.

Johnson, M. E., and Nagarur, N. 2016. Multi-stage methodology to detect health insurance claim fraud. *Health care management science*, 19(3), 249-260.

Khoshgoftaar, T.M., Van Hulse, J. and Napolitano, A. 2010. Supervised neural network modeling: an empirical investigation into learning from imbalanced data with labeling errors. *IEEE Transactions on Neural Networks*, 21(5), pp.813-830.

Kirlidog, M., and Asuk, C. 2012. A fraud detection approach with data mining in health insurance. *Procedia-Social and Behavioral Sciences*, 62, 989-994.

Li, J., Huang, K. Y., Jin, J., and Shi, J. 2008. A survey on statistical methods for health care fraud detection. *Health care management science*, 11(3), 275-287.

Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P. 2017. Focal loss for dense object detection. *In Proceedings of the IEEE international conference on computer vision*, pp. 2980-2988.

Liou, F. M., Tang, Y. C., and Chen, J. Y. 2008. Detecting hospital fraud and claim abuse through diabetic outpatient services. *Health care management science*, 11(4), 353-358.

Liu, Q., and Vasarhelyi, M. 2013, November. Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information. *In 29th world continuous auditing and reporting symposium (29WCARS), Brisbane, Australia*.

López, V., Fernández, A., García, S., Palade, V., and Herrera, F. 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 25078, 113–141.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., and Lee, S. I. 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1), 56-67.

Lundberg, S. M., and Lee, S. I. 2017, December. A unified approach to interpreting model predictions. *In Proceedings of the 31st international conference on neural information processing systems*, pp. 4768-4777.

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

Mullick, S.S., Datta, S., Dhekane, S.G. and Das, S. 2020. Appropriateness of performance indices for imbalanced data classification: An analysis. *Pattern Recognition*, 102, p.107197.

Musal, R. M. 2010. Two models to investigate Medicare fraud within unsupervised databases. *Expert Systems with Applications*, 37(12), 8628-8633.

Ortega, P.A., Figueroa, C.J. and Ruz, G.A. 2006. A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile. *DMIN*, 6, pp.26-29.

Pandey, P., Saroliya, A. and Kumar, R., 2018. Analyses and detection of health insurance fraud using data mining and predictive modeling techniques. In *Soft Computing: Theories and Applications: Proceedings of SoCTA 2016, Volume 2*, pp. 41-49. Springer Singapore.

Panigrahi, S., Kundu, A., Sural, S., and Majumdar, A. K. 2009. Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. *Information Fusion*, 10 (4), 354–363.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.

Pflaum, B. B., and Rivers, J. S. 1991. Employer strategies to combat health care plan fraud. *Benefits quarterly*, 7(1), 6.

Phua, C., Alahakoon, D., and Lee, V. 2004. Minority report in fraud detection: classification of skewed data. *Acm sigkdd explorations newsletter*, 6(1), 50-59.

Salcedo-Sanz, S., Cuadra, L., Portilla-Figueras, A., Jiménez-Fernández, S., and Alexandre, E. 2012. A review of computational intelligence algorithms in insurance applications. *Statistical and Soft Computing Approaches in Insurance Problems*, 1-50.

Gu, S., Kelly, B. and Xiu, D. 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), pp.2223-2273.

Shin, H., Park, H., Lee, J., and Jhee, W. C. 2012. A scoring model to detect abusive billing patterns in health insurance claims. *Expert Systems with Applications*, 39(8), 7441-7450.

Sparrow, M.K., 2019. *License to steal: how fraud bleeds America's health care system*. Routledge.

Thornton, D., Mueller, R. M., Schoutsen, P., and Van Hillegersberg, J. 2013. Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection. *Procedia technology*, 9, 1252-1264.

Van Capelleveen, G., Poel, M., Mueller, R. M., Thornton, D., and Van Hillegersberg, J. 2016. Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. *International journal of accounting information systems*, 21, 18-31.

Viveros, M. S., Nearhos, J. P., and Rothman, M. J. 1996, September. Applying data mining techniques to a health insurance information system. In *VLDB* (Vol. 96, pp. 286-294).

Waghade, S. S., and Karandikar, A. M. 2018. A comprehensive study of healthcare fraud detection based on machine learning. *International Journal of Applied Engineering Research*, 13(6), 4175-4178.

Williams, G.J. and Huang, Z., 1997. Mining the knowledge mine: The hot spots methodology for mining large real world databases. *In Advanced Topics in Artificial Intelligence: 10th*

*Australian Joint Conference on Artificial Intelligence, AI'97 Perth, Australia, November 30–December 4, 1997 Proceedings 10* (pp. 340-348). Springer Berlin Heidelberg.

Wu, Y., Xu, Y., and Li, J. 2019. Feature construction for fraudulent credit card cash-out detection. *Decision Support Systems*, 127, 113155.

Wynia, M. K., Cummins, D. S., VanGeest, J. B., and Wilson, I. B. 2000. Physician manipulation of reimbursement rules for patients: between a rock and a hard place. *Jama*, 283(14), 1858-1865.

Yamanishi, K., Takeuchi, J. I., Williams, G., and Milne, P. 2004. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Mining and Knowledge Discovery*, 8(3), 275-300.

Yang, W. S., and Hwang, S. Y. 2006. A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications*, 31(1), 56-68.

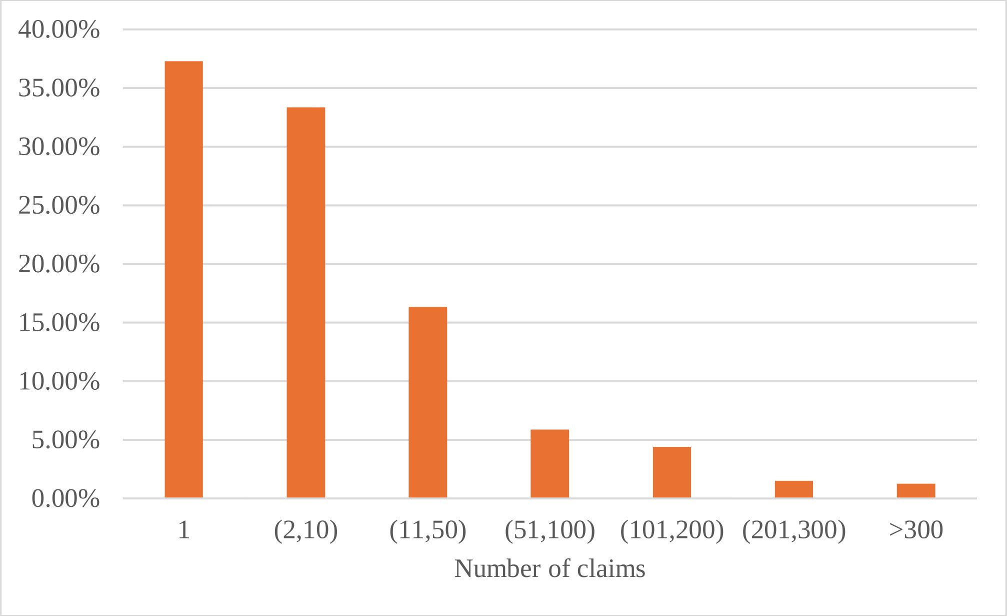



Figure 3.1: Frequency on physicians’ number of claims

This figure presents the number of claims from physicians varies widely. Specifically, 37.3% physicians correspond to single claim while the rest 62.7% physicians relate to more than one claim. For physicians with multiple claims, 53% of them file 2 to 10 claims, 26% file 11 to 50 claims, and the rest file more than 50 claims. The average is 39 for those with multiple claims.

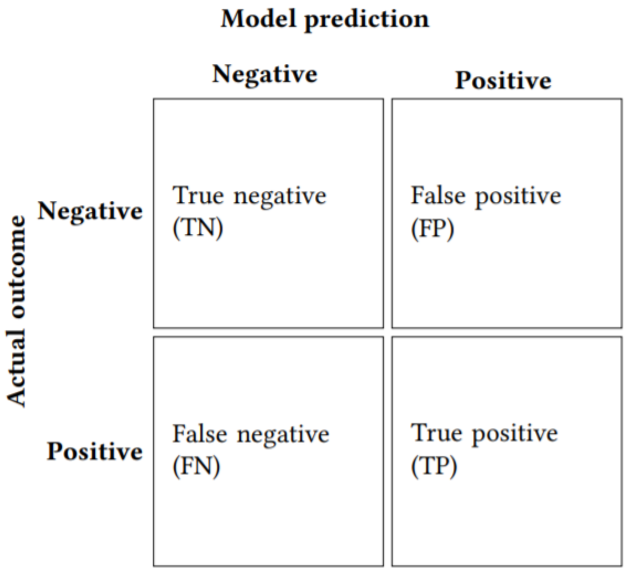



Figure 3.2: Confusion matrix

This figure presents a confusion matrix where positive and negative stand for fraudulent and legitimate, respectively. True positive, TP, represents the number of fraudulent physicians correctly predicted by the model while false negative, FN, represents the number of fraudulent physicians neglected by the model. False positive, FP, represents the number of legitimate physicians who are incorrectly marked as fraudulent by the model while true negative, TN, represents the number of legitimate physicians who are correctly predicted.

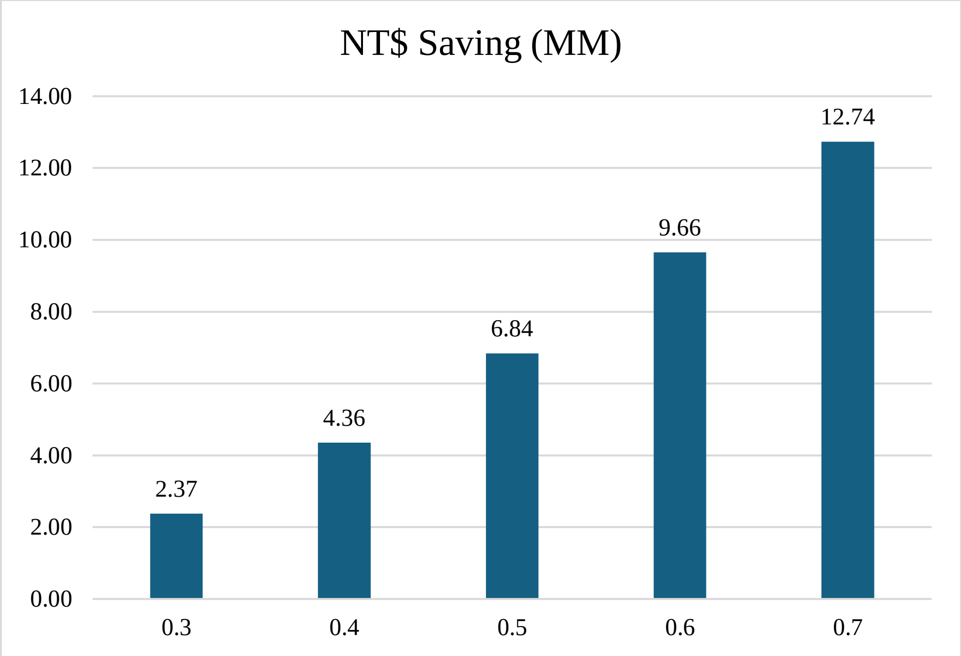



Figure 3.3: Amount saving by neural network with class weights method

This figure shows that the amount saved is monotonically increasing to the fraud rate. The amount saved is NT$ 1.65 million in the test set when 30% of the total claim amounts are frauds of the fraudulent physicians. It enlarges to NT$ 5.01 million when the fraud rate is 50% and to NT$ 9.94 million when the fraud rate is 70%.

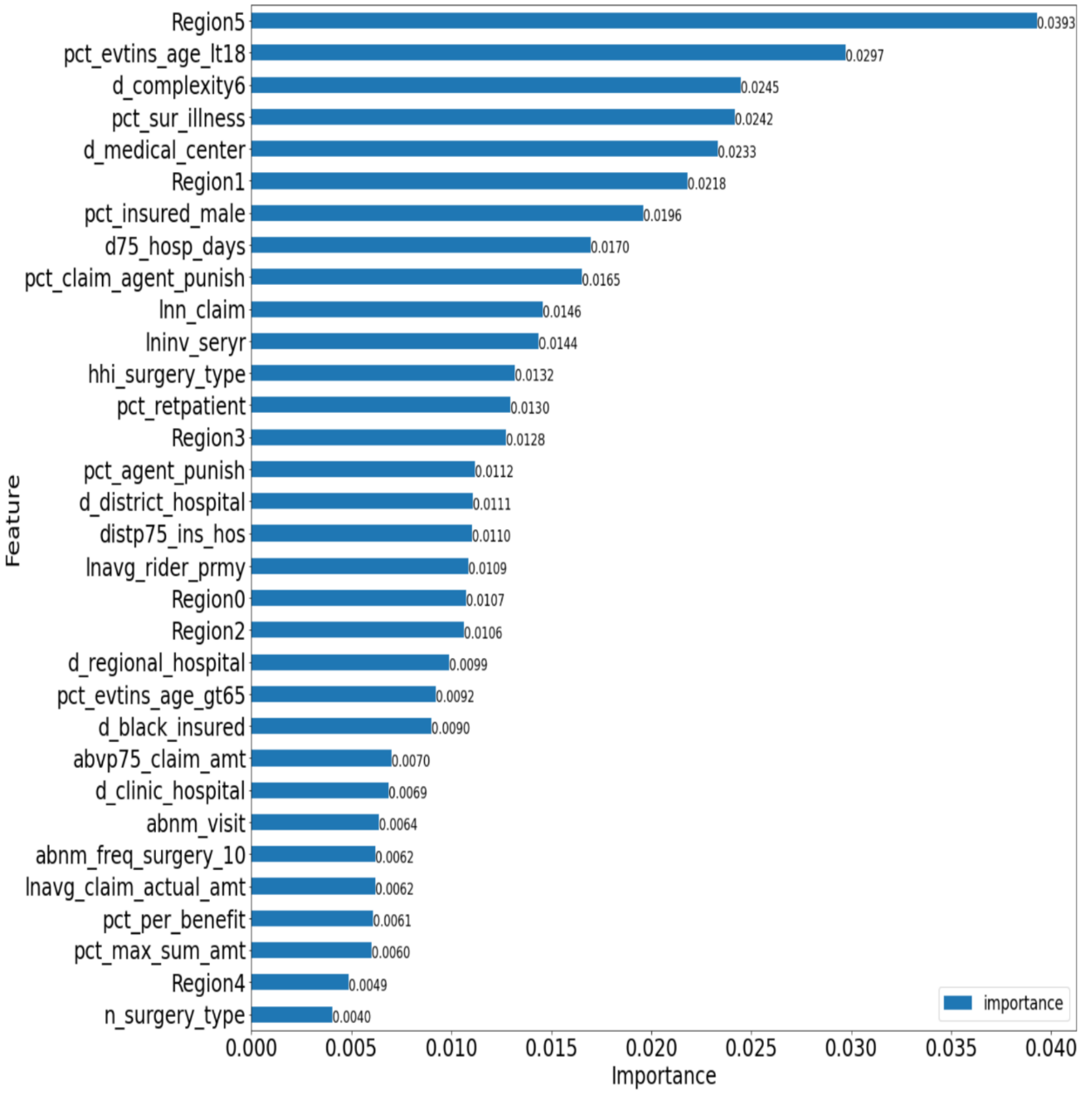



Figure 3.4: Feature importance based on permutation importance

This figure ranks in descending order the features that contribute most to the neural network with the class weight model's prediction of fraudulent physicians, measured by the average permutation importance of ten random states.

Table 3.1: Main papers related to feature construction at physician level

This table summarizes the most relevant studies on feature construction at the physician level. Panel A reports papers using supervised learning while Panel B also includes unsupervised learning methods to illustrate as many features as possible. Supervised learning requires labels on historical fraud data while unsupervised learning does not need labels.

| Reference | No | Main features | Data sources | Algorithms |
|---|---|---|---|---|
| *Panel A: Supervised learning* | | | | |
| Johnson and Nagarur (2016) | 5 | the length of visits | Claims from a private U.S. insurer | A new multistage methodology |
| | | the number of visits | | |
| | | the number of diagnoses | | |
| | | the number of services | | |
| | | the number of medications | | |
| Herland, Bauder and Khoshgoftaar (2019) | 18 | medical provider's specialty or practice | U.S. Medicare datasets, list of Excluded Individuals/Entities (LEIE) | Logistic regression<br>Random forest<br>Gradient tree boosting |
| | | provider's gender | | |
| | | number of procedures/services the provider performed | | |
| | | number of distinct Medicare beneficiaries receiving the service | | |
| | | number of distinct Medicare beneficiary/per day services performed | | |
| | | average of the charges that the provider submitted for the service | | |
| | | average payment made to a provider per claim for the service performed | | |
| | | number of distinct Medicare beneficiaries receiving the drug | | |
| | | number of drugs the provider administered | | |
| | | number of standardized 30-day fills | | |
| | | cost paid for all associated claims | | |
| He, Wang, Graco and Hawkins (1997) | 28 | NA | 1500 G.P.s of Australia | MLP; SOM |
| He, Hawkins, Graco and Yao (2000) | 28 | briefly mentioned | 1500 G.P.s of Australia | Genetic algorithm & k-Nearest Neighbor |

Table 3.1 Main papers related to feature construction at physician level (continued)

| | | | | |
|---|---|---|---|---|
| Ortega, Figueroa & Ruz (2006) | 125 | NA | A Chilean private health insurer | MLP |
| Branting, Reeder, Gold and Champney (2016) | 15 | 15 derived features on structural-similarity features, drug-type behavior-vector features, providers' attributes, and risk-propagation features | U.S. Medicare datasets; LEIE | behavioral similarity and geospatial colocation |
| *Panel B: Unsupervised learning* | | | | |
| Lin, Lin, Li and Kuo (2008) | 10 | the average claim fee submission to Bureau of NHI per month | Medical-claim database of the NHI, Taiwan | SOM; PCA |
| | | the average claim cases per month in claim data | | |
| | | the average prescription days per month in claim data | | |
| | | the average visits number in per case per month | | |
| | | total consultation fee divided by total cases | | |
| | | total treatment fee divided by total cases | | |
| | | total drug fee divided by total cases | | |
| | | total claim fee divided by total cases | | |
| | | total cases of antibiotic prescription divided by total cases | | |
| | | total cases of injection prescription divided by total cases | | |
| Musal (2010) | 7 | provider zip codes | Medicare claims on infusion therapy drugs | Clustering<br>Regression analysis<br>Descriptive statistics |
| | | beneficiary zip codes | | |
| | | provider paid amount | | |
| | | provider identification keys | | |
| | | beneficiary identification keys | | |
| | | provider physical address | | |
| | | socio-demographic information | | |

Table 3.1 Main papers related to feature construction at physician level (continued)

| | | | | |
|---|---|---|---|---|
| auder and Khoshgoftaar (2016) | 7 | medical provider's specialty, e.g., Cardiology | Medicare dataset on the dermatology and optometry provider specialties in Florida | A new general outlier detection model, based on Bayesian inference, using probabilistic programming |
| | | code for specific medical service furnished by the provider | | |
| | | unique provider identification number | | |
| | | provider's first name | | |
| | | provider's last name | | |
| | | number of services provided/procedures performed | | |
| | | amount Medicare paid the provider for services performed | | |
| Hillerman, Souza, Reis and Carvalho (2017) | 8 | growth rate of expenses | Claims related to psychotherapy sessions, health insurance provider in Brazil | clustering |
| | | previous investigation | | |
| | | accreditation time | | |
| | | percent of out-of-state patients | | |
| | | percent of young or elderly patient | | |
| | | percent of denied billings | | |
| | | average weekly visitation frequency | | |
| | | percent of visits in weekends or holidays | | |
| Major and Riedinger (2002) | 27 | 27 heuristics in five categories: financial, medical logic, abuse, logistics and identification | Travelers Insurance Company | A hybrid knowledge/statistical-based system |
| Copeland, Edberg, Panorska, Wendel and Reno (2012) | 12 | NA | Medicaid claims from Nevada | statistical methods |

Table 3.2: Basics for the multi-claim and single-claim samples

This table summarizes the number of physicians, the percentage of fraudulent physicians, the number of claims, total claim amounts, and the average claim amount.

| | | Multi-claim sample | Single-claim sample |
|---|---|---|---|
| No. of physicians | | 21,839 | 12,993 |
| Pct. of fraudulent physicians | Supreme Court | 0.15% | 0.04% |
| | Life insurer | 0.44% | 0.11% |
| | Social insurance | 0.04% | 0.01% |
| No. of claims | | 886,990 | 12,993 |
| Pct. of suspicious claims | | 1.62% | 0.15% |
| Total claim amount (NT$ B) | | 15.58 | 0.19 |
| Average claim amount (NT$ K) | | 17.56 | 14.28 |

Note: all amounts are in 2016 value.

Table 3.3: Summary statistics

This table presents the means and standard deviations of all features for the fraudulent and legitimate physicians for the multi-claim sample. p-value for the t-tests of mean differences is reported.

| Features | Multi-claim subsample | | |
|---|---|---|---|
| | Fraudulent physicians | Legitimate physicians | *p*-value |
| *lnn_claim* | 3.526 | 2.466 | 0.000*** |
| *lnavg_claim_actual_amt* | 9.358 | 9.045 | 0.017** |
| *d_medical_center* | 0.183 | 0.289 | 0.002*** |
| *d_regional_hospital* | 0.285 | 0.231 | 0.139 |
| *d_district_hospital* | 0.292 | 0.108 | 0.000*** |
| *d_clinic_hospital* | 0.212 | 0.348 | 0.000*** |
| *Region0* | 0.263 | 0.387 | 0.003*** |
| *Region1* | 0.022 | 0.060 | 0.003*** |
| *Region2* | 0.051 | 0.039 | 0.528 |
| *Region3* | 0.175 | 0.199 | 0.489 |
| *Region4* | 0.168 | 0.137 | 0.287 |
| *Region5* | 0.321 | 0.178 | 0.001*** |
| *n_surgery_type* | 4.168 | 2.888 | 0.000*** |
| *hhi_surgery_type* | 0.546 | 0.657 | 0.000*** |
| *d_complexity6* | 0.540 | 0.340 | 0.000*** |
| *pct_retpatient* | 0.198 | 0.222 | 0.159 |
| *pct_sur_illness* | 0.636 | 0.759 | 0.000*** |
| *pct_max_sum_amt* | 0.193 | 0.302 | 0.000*** |
| *abvp75_claim_amt* | 0.065 | 0.063 | 0.865 |
| *d75_hosp_days* | 0.273 | 0.213 | 0.003*** |
| *abnm_freq_surgery_10* | 0.365 | 0.235 | 0.002*** |
| *lninv_seryr* | -2.394 | -3.601 | 0.000*** |
| *pct_per_benefit* | 0.061 | 0.084 | 0.031** |
| *d_black_insured* | 0.102 | 0.015 | 0.001*** |
| *distp75_ins_hosp* | 0.737 | 0.529 | 0.000*** |
| *abnm_visit* | 0.263 | 0.175 | 0.021** |
| *pct_evtins_age_lt18* | 0.036 | 0.049 | 0.003*** |
| *pct_evtins_age_gt65* | 0.082 | 0.077 | 0.595 |
| *lnavg_rider_prmy* | 1.231 | 1.081 | 0.019** |
| *pct_insured_male* | 0.493 | 0.497 | 0.836 |
| *pct_agent_punish* | 0.348 | 0.336 | 0.476 |
| *pct_claim_agent_punish* | 0.612 | 0.641 | 0.082* |

Table 3.4: Costs of outcomes

This table summarizes the costs to the insurer based on the outcomes in the confusion matrix, where C_investigation, C_physician_legit, and C_physician_fraud stand for the average cost of investigation and the average total claim amounts related to legitimate and fraudulent physicians, respectively.

| Outcomes | Costs |
|---|---|
| TP (Hits) | TP * *C_investigation* |
| FP (False alarms) | FP * (*C_investigation* + *C_ physician_legit*) |
| FN (Misses) | FN * *C_ physician_fraud* |
| TN (Normal) | TN * *C_ physician_legit* |

Table 3.5: AUCROC scores

AUROC is used as the indicator for the discriminating power of the classifier. It is the area under the receiver operating characteristic curve which plots the true positive rate ($True\ positive\ rate = TP/(TP + FN)$ Equation (1)) against the false positive rate ($False\ positive\ rate = FP/(TN + FP)$ Equation (2)) at different decision-making thresholds. AUROC equals 1, indicating the classifier performs perfectly, while 0.5 indicates randomly. The more AUROC is closer to 1, the better the classifier is.

| Training algorithm | Methods deal with imbalanced data | Mean | Std. Dev. |
|---|---|---|---|
| Neural Network | Class weight | 0.781 | 0.043 |
| RUSBoost | Random under sampling | 0.777 | 0.035 |
| Logit | Class weight | 0.751 | 0.031 |

Table 3.6: Precision and Recall scores

Precision score is calculated by $Precision = TP/(TP + FP)$ . Recall score is calculated by $Recall = TP/(TP + FN)$ Precision score and recall score are inversely related.

| Training algorithm | Methods deal with imbalanced data | Precision | Recall |
|---|---|---|---|
| Neural Network | Class weight | 0.00901 | 0.91111 |
| RUSBoost | Random under sampling | 0.01499 | 0.69630 |
| Logit | Class weight | 0.01437 | 0.68148 |

Table 3.7: Model cost saving in decimal for multi-claims

| | | Fraud rate = 30% | | Fraud rate = 40% | | Fraud rate = 50% | | Fraud rate = 60% | | Fraud rate = 70% | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| Neural Network | Class weight | 0.163 | 0.060 | 0.223 | 0.058 | 0.278 | 0.062 | 0.327 | 0.066 | 0.369 | 0.070 |
| RUSBoost | Random under sampling | 0.047 | 0.099 | 0.180 | 0.093 | 0.264 | 0.091 | 0.326 | 0.086 | 0.374 | 0.082 |

Model cost savings $=$ No action– (Misses costs + False alarms costs + Normal costs + Hits costs) $=$ TP $*$ $(C\_physician\_fraud * Fraud\ rate - C\_investigation) -$ FP $* C\_investigation$ (3)

Equation (3), where No action means the life insurer does not use any model to detect frauds, and Fraud rate is defined as the percentage of total claim amounts that are, in fact, frauds. It is hard to estimate the fraud rate and we assume it is from 30% to 70%.

Percentage saved $=$ (Model cost savings)/(Best – case scenario cost savings) $* 100\% = (Model\ cost\ savings)/((TP + FN) * (C\_physician\_fraud * Fraud\ rate - C\_investigation)) * 100\%$ (4)

Equation (4) restates the model cost savings in percentage by scaling it to the best-case scenario cost savings when all physicians are correctly predicted as legitimate or fraudulent by the model. Since the cost-based method is threshold-dependent, we choose the threshold, giving the optimal cost savings.

It is with high importance to estimate C_investigation, and C_physician_fraud reasonably while using cost-based performance evaluation method. The estimated average cost of investigation (C_investigation) equals NT$ 11,500 for physicians with multi-claims. The estimated average total claim amount (C_physician_fraud) is NT$ 1,841,428 for physicians with multi-claims.

This table presents the model cost saving in decimal by scaling it to the best-case scenario cost savings when all physicians are correctly predicted as legitimate or fraudulent by the model.

# APPENDIX A: DEFINITION OF FEATURES

| Features | Definition |
|---|---|
| 1) Claim basics | |
| *lnn_claim* | the natural logarithm of the number of claims related to the physician |
| *lnavg_claim_actual_amt* | the natural logarithm of the average claim amount paid by the insurer |
| 2) Physician characteristics | |
| *d_medical_center* | dummy variable equals 1 if the hospital is a medical center; 0 otherwise |
| *d_regional_hospital* | dummy variable equals 1 if the hospital is a regional hospital; 0 otherwise |
| *d_district_hospital* | dummy variable equals 1 if the hospital is a district hospital; 0 otherwise |
| *d_clinic_hospital* | dummy variable equals 1 if the hospital is a clinic hospital; 0 otherwise |
| *Region0* | the southeast of Taiwan including Keelung, Yilan, Hualien, Quemoy, and Lienchiang |
| *Region1* | the northeast of Taiwan including Taipei, New Taipei and Taoyuan |
| *Region2* | the northwest of Taiwan including Hsinchu and Miaoli |
| *Region3* | the west of Taiwan including Taichung, Changhua, and Nantou |
| *Region4* | the southwest of Taiwan including Yunlin, Chiayi and Tainan |
| *Region5* | the east of Taiwan including Kaohsiung, Pingtung, Taitung, or Penghu |
| *n_surgery_type* | the number of unique surgery categories |
| *hhi_surgery_type* | $= \sum_{i=1}^{I}(\frac{n_i}{\sum_{j=1}^{I} n_j})^2$, where $I$ is the unique surgery categories and $n_i$ is the performance times for surgery category $i$ |
| *d_complexity6* | dummy variable equals 1 if the physician can perform any surgery with complexity 6 or higher |
| *pct_retpatient* | the percentage of return patients |
| *pct_sur_illness* | the percentage of surgeries due to illness |
| 3) Fraud strategies | |
| *pct_max_sum_amt* | the maximum claim amount to total claim amounts |
| *abvp75_claim_amt* | given each surgery category and year, the aggregated amount differences if the surgery's claim amount is larger than the 75th percentile claim amount or 0 then scaled it by the physician's total claim |
| *d75_hosp_days* | for all surgery category and year, count the claims with length of hospital stays greater than the 75th percentile, and scale it by the physician's total number of claims |

APPENDIX A: DEFINITION OF FEATURES (continued)

| | |
|---|---|
| *abnm_freq_surgery_10* | for any surgery category and year, if the physician performs more than the average number of surgeries and the surgery can be performed by more than 10% of all physicians in that year, this dummy variable equal to 1; 0 otherwise. |
| 4) Early signals | |
| *lninv_seryr* | the natural logarithm of the number of claims of the physician being investigated by the life insurer and scaled by the length of service year |
| *pct_per_benefit* | the percentage of claims related to the physician with claim amount received less than the filed amount |
| 5) Insured characteristics | |
| *d_black_insured* | dummy variable equals 1 if there are insureds on the blacklist of the life insurer in the physician's claims; 0 otherwise |
| *distp75_ins_hosp* | dummy variable equals to 1 if any insured's driving distance is longer than the 75th percentile driving distance of all patients in the insured's zip code, and she is not going to a medical center; 0 otherwise |
| *abnm_visit* | dummy variable equals 1 if there is any insured who lives in the metropolitan area but travels to a non-metropolitan area to visit a physician; 0 otherwise. |
| *pct_evtins_age_lt18* | the percentage of insureds whose age are less than 18 |
| *pct_evtins_age_gt65* | the percentage of insureds whose age are greater than 65 |
| *lnavg_rider_prmy* | the natural logarithm of the average endorsement face amount to the primary life insurance face amount |
| *pct_insured_male* | the percentage of insureds who are male |
| 6) Agent characteristics | |
| *pct_agent_punish* | the percentage of agents been punished by the life insurer in the claims related to the physician |
| *pct_claim_agent_punish* | the percentage of claim agents been punished by the life insurer in the claims related to the physician |

# APPENDIX B: SUMMARY STATISTICS FOR THE SINGLE-CLAIM SAMPLE

This table presents the means and standard deviations of all features for the fraudulent and legitimate physicians for the single-claim sample. p-value for the t-tests of mean differences is reported.

| | Single claim subsample | | | |
|---|---|---|---|---|
| Features | Fraudulent physicians | Legitimate physicians | *p*-value | Note: corresponding features of multi-claim physicians |
| | | | | *lnn_claim* |
| *lnclaim_actual_amt* | 6.630 | 8.031 | 0.218 | *lnavg_claim_actual_amt* |
| *d_medical_center* | 0.050 | 0.278 | 0.000*** | *d_medical_center* |
| *d_regional_hospital* | 0.400 | 0.231 | 0.073* | *d_regional_hospital* |
| *d_district_hospital* | 0.150 | 0.119 | 0.666 | *d_district_hospital* |
| *d_clinic_hospital* | 0.400 | 0.372 | 0.798 | *d_clinic_hospital* |
| *Region0* | 0.150 | 0.422 | 0.014** | *Region0* |
| *Region1* | 0.100 | 0.062 | 0.484 | *Region1* |
| *Region2* | 0.100 | 0.042 | 0.406 | *Region2* |
| *Region3* | 0.250 | 0.175 | 0.377 | *Region3* |
| *Region4* | 0.200 | 0.128 | 0.336 | *Region4* |
| *Region5* | 0.200 | 0.172 | 0.737 | *Region5* |
| *surgery_type* | 3.400 | 4.327 | 0.234 | *n_surgery_type* |
| | | | | *hhi_surgery_type* |
| *d_complexity6* | 0.150 | 0.076 | 0.376 | *d_complexity6* |
| | | | | *pct_retpatient* |
| *sur_illness* | 0.550 | 0.755 | 0.033** | *pct_sur_illness* |
| | | | | *pct_max_sum_amt* |
| *abvp75_claim_amt* | 0.005 | 0.014 | 0.048** | *abvp75_claim_amt* |
| *d75_hosp_days* | 0.300 | 0.178 | 0.156 | *d75_hosp_days* |
| | | | | *abnm_freq_surgery_10* |
| *investigate* | 0.100 | 0.019 | 0.256 | *lninv_seryr* |
| *per_benefit* | 0.805 | 0.914 | 0.240 | *pct_per_benefit* |
| *d_black_insured* | 0.000 | 0.002 | 0.000*** | *d_black_insured* |
| *distp75_ins_hosp* | 0.300 | 0.166 | 0.108 | *distp75_ins_hosp* |
| *abnm_visit* | 0.050 | 0.033 | 0.664 | *abnm_visit* |
| *evtins_age_lt18* | 0.000 | 0.060 | 0.000*** | *pct_evtins_age_lt18* |
| *evtins_age_gt65* | 0.050 | 0.071 | 0.713 | *pct_evtins_age_gt65* |
| *lnrider_prmy* | 0.481 | 0.488 | 0.982 | *lnavg_rider_prmy* |
| *insured_male* | 0.400 | 0.489 | 0.427 | *pct_insured_male* |
| *agent_punish* | 0.450 | 0.340 | 0.298 | *pct_agent_punish* |
| *claim_agent_punish* | 0.700 | 0.631 | 0.524 | *pct_claim_agent_punish* |

APPENDIX C: ALGORITHM OF RUSBOOST

**Algorithm RUSBoost**

**Given:** Set $S$ of examples $(x_1, y_1), ..., (x_m, y_m)$ with minority class $y^r \in Y, |Y| = 2$

Weak learner, $WeakLearn$

Number of iterations, $T$

Desired percentage of total instances to be represented by the minority class, $N$

1 Initialize $D_1(i) = \frac{1}{m}$ for all $i$.

2 Do for $t = 1, 2, ..., T$

a Create temporary training dataset $S'_t$ with distribution $D'_t$ using random undersampling

b Call $WeakLearn$, providing it with examples $S'_t$ and their weights $D'_t$.

c Get back a hypothesis $h_t : X \times Y \rightarrow [0, 1]$.

d Calculate the pseudo-loss (for $S$ and $D_t$):

$$\epsilon_t = \sum_{(i,y):y_i \neq y} D_t(i)(1 - h_t(x_i, y_i) + h_t(x_i, y)).$$

e Calculate the weight update parameter:

$$\alpha_t = \frac{\epsilon_t}{1 - \epsilon_t}.$$

f Update $D_t$:

$$D_{t+1}(i) = D_t(i)\alpha_t^{\frac{1}{2}(1+h_t(x_i,y_i)-h_t(x_i,y:y\neq y_i))}.$$

g Normalize $D_{t+1}$: Let $Z_t = \sum_i D_{t+1}(i)$.

$$D_{t+1}(i) = \frac{D_{t+1}(i)}{Z_t}.$$

3 Output the final hypothesis:

$$H(x) = \underset{y\in Y}{\operatorname{argmax}} \sum_{t=1}^{T} h_t(x, y) \log \frac{1}{\alpha_t}.$$

# APPENDIX D: OTHER CLASSIFICATION METHOD RESULTS WITH OTHER METHODS DEAL WITH IMBALANCED DATA

These methods are at the data level and rebalance the data during data pre-processing step. The first method is over sampling (Over sampling method later). By duplicating the fraudulent physicians, their number in the training set is increased to the same as the legitimate ones, increasing the chances of correct predictions by the algorithms. The second method is SMOTE that is to synthesize new "fraudulent" physicians from the existing observations. SMOTE works by selecting observations close in the feature space, drawing a line between them, and generating a new sample at a point along that line. Hybrid method is combining random under sampling to trim the number of legitimate physicians and SMOTE to over sample the minority class to balance the data (Hybrid method later). Specifically, we synthesize fraudulent physicians to be 10% of the legitimate physicians and set the fraudulent and legitimate physicians to be one-to-two during training. The last method is class weights where the model assigns the class weights inversely proportional to their respective frequencies (Class weights method later).

| Training algorithm | Methods deal with imbalanced data | AUCROC Mean |
|---|---|---|
| XGboost | Over sampling | 0.735 |
| | SMOTE | 0.741 |
| | Hybrid | 0.746 |
| | Class weight | 0.758 |
| RidgeClassifier | Over sampling | 0.771 |
| | SMOTE | 0.724 |
| | Hybrid | 0.720 |
| | Class weight | 0.772 |
| BaggingClassifier | Over sampling | 0.720 |
| | SMOTE | 0.670 |
| | Hybrid | 0.696 |
| | Class weight | 0.768 |
| LGBMClassifier | Over sampling | 0.656 |
| | SMOTE | 0.690 |
| | Hybrid | 0.690 |
| | Class weight | 0.736 |
| Linear SVC | Over sampling | 0.770 |
| | SMOTE | 0.720 |
| | Hybrid | 0.718 |
| | Class weight | 0.770 |
| Logistic | Over sampling | 0.769 |
| | SMOTE | 0.720 |
| | Hybrid | 0.728 |
| | Class weight | 0.771 |
| | | |
| LDA | Over sampling | 0.774 |
| | SMOTE | 0.725 |
| | Hybrid | 0.723 |
| | Class weight | 0.774 |
| SGD | Over sampling | 0.769 |
| | SMOTE | 0.723 |
| | Hybrid | 0.730 |
| | Class weight | 0.766 |
| Random Forest | Over sampling | 0.749 |
| | SMOTE | 0.745 |
| | Hybrid | 0.759 |
| | Class weight | 0.760 |

# CONCLUSIONS

This dissertation consists of three essays on corporate finance and machine learning. The first chapter investigates the relationship between the CEOs' conscientiousness trait and reserve management within U.S. property-liability insurance insurers. The baseline results show that CEO conscientiousness is negatively associated with reserve error in the upper tail of the conditional distribution (i.e., at 75th percentile and higher), indicating insurers with more conscientious CEOs reserve less than insurers with less conscientious CEOs at a higher level of reserve errors to lower the cost of excess reserve rather than conservatism when reserve errors are extremely conservative. I also find that CEOs become more conservative when their insurers have higher financial risk. Furthermore, insurers with more conscientious CEOs reserve less than less conscientious CEOs after SOX (compared with before SOX) when insurers face higher financial risk, possibly because they are more responsible for financial statements. This evidence is consistent with one feature of conscientiousness: following the rules and norms. Finally, conscientious CEOs get higher compensation, suggesting that the conscientiousness trait is rewarded in the property-liability industry. The overall results of this paper are consistent with the features of conscientiousness: being responsible and following the rules.

The second chapter studies whether the opacity of insurers shapes a typical policyholder's purchase behavior. The results indicate that policyholders take information quality into consideration, which means opacity plays an important role when they make purchase decisions. I explore that opacity negatively affects policyholders' purchase behavior, especially when insurers' financial risk is high. In addition, the guaranty fund doesn't provide complete protection to policies; thus, policyholders care about the safety and opacity of insurers. For less opaque insurers, policyholders perceive more utility regarding lower default risk because of higher information quality. Our research further suggests that opacity significantly influences the purchase behavior of commercial lines. This is due to the involvement of brokers and agents who possess in-depth knowledge of insurers' financial situations and product policies.

The third chapter studies how to detect fraudulent physicians using supervised machine learning algorithms and analyzes the importance of features. I first transform the data to the physician level and generate 32 features, including claim basics, physician characteristics, fraud strategies, early signals, insured characteristics, and agent characteristics. Then, I apply the neural network and RUSBoost models to detect fraudulent physicians. The class weight and random under sampling methods are used to address the imbalanced data issue. For the multi-claim sample, the neural network model performs reasonably well, with the AUROC scores around 0.781. As the fraud rate of fraudulent physicians' total claim amount increases from 30% to 70%, the percentage saved from our models increases from approximately 16.3% to 36.9% of the best-case scenario cost savings. The cost savings calculated are the lowest since some fraudulent physicians correctly predicted may still be labeled as legitimate. I use the permutation importance method to understand the feature importance in distinguishing fraudulent physicians from legitimate ones. I find that fraudulent physicians are associated with the cluster in the eastern area of Taiwan, the percentage of insureds whose age are less than 18, the percentage of surgeries due to illness, and whether the physician can perform any surgery with high complexity. The evidence suggests that physicians prefer the "steal a little, all the time" strategy and try to avoid large claim amounts, which may arouse the attention of the life insurer. Early signals from the life insurer are also useful in detecting fraudulent physicians.