APPLICATION OF HIDDEN MARKOV MODEL TO AUTO TELEMATICS DATA AND THE EFFECT OF UNIVERSAL DEMAND LAW CHANGE ON CORPORATE RISK TAKING IN THE U.S. PROPERTY CASUALTY INSURANCE INDUSTRY

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

---

by
Qiao Jiang
Diploma Date December 2022

Examining Committee Members:
Dr. Tianxiang Shi, Advisory Chair, Department of Risk, Actuarial Science, and Legal Studies
Dr. Martin Grace, Department of Risk, Actuarial Science, and Legal Studies
Dr. Krupa Viswanathan, Department of Risk, Actuarial Science, and Legal Studies
Dr. Sudipta Basu, External Member, Department of Accounting, Temple University

ABSTRACT

There are two themes in this dissertation, that is, the effect of universal demand law change on corporate risk-taking in the U.S. property & casualty insurance industry, and the application of hidden Markov model to auto telematics data. The first chapter presents my study in the first theme and the rest two chapters present the other theme.

In Chapter 1, "Does Shareholder Litigation Affect Corporate Risk-Taking? Evidence from the Property-Casualty Insurance Industry", I explore whether shareholder litigation affects corporate risk-taking differently depending on distinct organizational structures. I use a state's adoption of a Universal Demand (UD) Law as an exogenous shock and develop three risk-taking measures unique to the U.S. property-casualty insurance industry: leverage risk, asset risk, and underwriting risk. The insurance industry provides an interesting opportunity for the study as shareholders in mutual insurers are an ambiguous concept in the legal world, as opposed to the common argument in the insurance literature. The results show that along with UD law adoption, insurers increase their risk-taking. After taking organizational structure into account, the impact of the law's adoption differs among organizational forms. Stock insurers increase all three risk-taking measures while mutual insurers decrease their Leverage Risk and increase Asset Risk measures. For different time windows, stock insurers respond faster with respect to their Asset Risk compared to mutual insurers. In addition, I proceed to examine the main economic channel for the impact and find that the free cash flow argument is not the main channel.

Chapters 2 and 3 present the study in auto telematics data using a proprietary data source. Both studies are based on the application of hidden Markov model (HMM). Specifically, Chapter 2, "Auto Insurance Pricing Using Telematics Data: Application of a Hidden Markov Model", develops a HMM-based clustering framework to predict

auto insurance losses using driving characteristics extracted from telematics data. Through a simulation experiment based on a proprietary telematics data set, I show that HMM can effectively classify driving trips using model-implied hidden states, and HMM-based pricing methods provide better predictive power measured by both deviance statistics and mean squared error. Importantly, the proposed framework not only enables us to price usage-based insurances at a granular level, but it is also viable for estimating long-term insurance losses utilizing the limiting properties of HMM.

Chapter 3, "Theoretical Framework of a 3-Layer Hidden Markov Model for Auto Insurance Pricing", is a theoretical extension of the second chapter to improve the framework at a more granular level. I develop a 3-layer HMM for risk classification, which links driving behavior characteristics with risk classes and loss estimation. The proposed model presents a direct structure among all variables and utilizes time series data without aggregation. Furthermore, this study provides a theoretical framework to estimate the 3-layer HMM using the Expectation-Maximization (EM) algorithm. The parameters of Bernoulli distributed loss count (per unit of time) and Gamma distributed loss severity can be solved at least numerically, and the negative definite Hessian matrix indicates that the solution of the first-order condition of the log-likelihood function achieves its local maximum.

THIS DISSERTATION IS DEDICATED TO MY BELOVED WIFE AND SON.

ACKNOWLEDGMENTS

This dissertation is the result of an eight-year journey. During the last three years, I have thought about quitting the Ph.D. program hundreds of times. It is clear that without the help and support of my committee members and my family, I would never have completed this journey.

When I retrospect my thoughts of quitting, I find that the confidence of doing so is my belief in succeeding in my career without obtaining the Ph.D. degree. The belief rests on solid ground which is built on top of the type of person into which my Ph.D. program has shaped me: one who has passions for truth, perseverance, creative and critical thinking, and intellectual humility. Most of these qualities were gained through lessons from and interactions with Dr. Sudipta Basu. I want to express my most profound appreciation to him for being such an influential and inspirational role model in my life.

I am also extremely grateful to my committee chair, Dr. Tianxiang Shi, for his invaluable guidance, encouragement, and support during the last and most important three years of my doctoral studies at Temple. His patience and belief in me impelled me to continue the journey to its successful end.

I would also like to express my gratitude to my committee members, Dr. Martin Grace and Dr. Krupa Viswanathan, for their insightful comments, which contributed to the improvement of this dissertation.

Last but not least, I would like to thank my family members for their endless love and support. To my wife and son, your unconditional love was the main thing that kept me calm and confident during the most intense period of completing this dissertation. To my parents and sister, thank you for always encouraging me to pursue my dreams and for supporting me through all the difficult times. Live long and prosper.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1
DOES SHAREHOLDER LITIGATION AFFECT CORPORATE RISK-TAKING? EVIDENCE FROM THE PROPERTY-CASUALTY INSURANCE INDUSTRY

1.1 Introduction

Shareholder litigation is considered one of the three pillars - "vote, sell and sue" - for shareholders to exert corporate governance (Ferris et al., 2007; Pukthuanthong et al., 2017). Easterbrook and Fischel (1991) argue that shareholder litigation is necessary because there are voids in corporate governance when only enforcing contracts. First, contracts cannot be written in full detail since people are unable to foresee all contingencies. Second, although contracts could be written in enough detail to resolve all contingent problems in theory, the existence of unverifiable information makes it impossible for some contracts to be enforced effectively. Therefore, shareholder litigation has been considered as an effective tool for good corporate governance and as a last resort for shareholders to resolve conflicts.

This paper will investigate a specific type of shareholder litigation: shareholder derivative suits. A shareholder derivative suit is brought by a shareholder on behalf of the corporation for damage caused by a third party. Usually, the third party are managers or the board of directors of the corporation. For example, in 2004, a series of shareholder derivative suits were filed against AIG. The shareholders alleged that AIG insiders, such as Greenberg and Smith, misstated AIG's financial condition to intentionally deceive investors. As mentioned in Ferris et al. (2007), among derivative lawsuits filed between 1982 and 1999, 41% pertain to the duty of care, 26% pertain to the duty of loyalty, 16% pertain to mishandling corporate information, and 7% are associated with M&As. Any financial recovery from derivative lawsuits goes directly to corporations instead of to shareholders themselves.

There are two channels by which a shareholder derivative suit could affect managerial behavior. First, the suit can effectively impose liability on managers and the board of directors even when they have purchased Directors and Officers (D&O) liability insurance because this coverage does not cover misconduct. For example, if a breach of fiduciary duty is determined, the fiduciary is required to pay all profits he or she obtained to the corporation through the breach of his or her fiduciary duty. Moreover, breach of fiduciary duty causes reputation loss (Karpoff et al., 2008). The market would react accordingly, with managers or the board of directors suffering a penalty, such as reputational loss and personal liability from the market. This is an example of a direct and ex-post remedy from a derivative suit.

The other channel is indirect. That is, a derivative suit can represent a credible threat to managers and the board of directors and thereby affect managerial behavior ex ante. For instance, before a derivative suit, shareholders could initiate a threat of a suit (or a demand) that does not yet involve a court. Managers and the board of directors would realize that they are subject to a potential derivative suit and change their managerial behavior accordingly before a getting before a judge and jury.[1] Furthermore, even the fear of a derivative suit and the corresponding results, such as reputation impairment, could affect managerial behavior and limit misconduct by managers and the board of directors. At this point, it looks like a shareholder derivative suit is indeed an effective tool for better corporate governance. However, the real impact of a shareholder derivative suit on managerial behavior could be manifold.

A large volume of the extant literature proceeded from the premise that managers have control of the corporation and tend to exploit shareholders and other stakeholders. Meanwhile, it is well recognized that managers and the board of directors also possess more knowledge, especially inside information and expertise, enabling them

[1] Most allegations about damages caused by managers or the board of directors to the corporation stop at the pre-trial stage and never get to trial.

to make proper corporate decisions. It is noteworthy that corporate law is developing in a direction that favors the expert performance of directors and officers (Easterbrook and Fischel, 1991). Almost every state has an enabling statute that allows managers and investors to develop their own governance systems that are not subject to substantive scrutiny from regulation. In addition, courts apply a doctrine called the "business judgement rule" that recognizes that managers and the board of directors have enough information and expertise to make better corporate decisions for the benefit of investors. The implication of the business judgement rule is that a strict and routine judicial review or oversight of business decisions leads to worse decisions for wealth maximization. The main justifications is that judges are not competent to make good business decisions; that managers are too cautious and there will be fewer talented directors due to fear of personal liability; that the information market is more efficient than a judgement review in reflecting the performance of managers, etc.(Easterbrook and Fischel, 1991). Under the business judgement rule, less than optimal decisions can be tolerated and the board of directors are shielded from liability due to honest mistakes of judgement unless certain duties are breached. Therefore, the development of corporate law implies that whether a shareholder derivative suit could lead to a better result than decisions made by directors and officers is questionable.

In addition, the objective of a shareholder derivative suit is another concern since recovery from such suits goes to the corporation instead of directly to the shareholder. Since a majority of shareholders could change the board of directors through voting, they lack incentive to initiate a derivative suit. It is more likely for minority shareholders to file a derivative suit. In this case, although minority shareholders could benefit from the recovery to the corporation since they are owners, there exists a majority of free riders such that the cost to minority shareholders to pay for the derivative suit and return are imbalanced. In other words, there exists no direct link between stake and reward in a derivative suit. Therefore, it is not rational to launch a derivative suit based on a cost-benefit analysis. As a matter of fact, it is typically the attorneys who benefit more than shareholders in a derivative suit. As argued by

Easterbrook and Fischel (1991), "the combination of uninterested nominal plaintiffs and interested attorneys produces striking effects."

Due to a lack of proper incentives and the imbalance between the stakes and rewards, whether a derivative suit can lead to better corporate governance is uncertain and therefore constitutes an interesting empirical question. In this research, I conduct an empirical study to explore the impact of derivative suits on corporate decisions, especially risk-taking behavior, in the U.S. property-casualty insurance industry. I use the adoption of **universal demand (UD) laws** to determine the role of derivative suits on risk-taking behavior.

UD is a requirement of the Model Business Corporation Act (MBCA)[2] pertaining to shareholder derivative suits. The requirement is as follows:

No shareholder may commence a derivative proceeding until:

1. *a written demand has been made upon the corporation to take suitable action; and*

2. *90 days have expired from the date delivery of the demand was made unless the shareholder has earlier been notified that the demand has been rejected by the corporation or unless irreparable injury to the corporation would result by waiting for the expiration of the 90-day period*[3]

It is a state's discretion to include the requirement above in their statutes. When the UD requirement is included in a corporate statute, it becomes the UD law. Hereafter, I will use the term UD law to describe the requirement above since only laws can grant shareholder rights.

The UD law poses a procedural obstacle to shareholder derivative suits by requiring derivative plaintiffs (in our case, shareholders) to file a demand to the board of directors before launching the derivative suit (Bourveau et al., 2018), and it grants the board of directors the right to decide whether to accept or reject the demand.

[2]The Model Business Corporation Act (MBCA) is a standard or model law developed by the American Bar Association to provide enacting states with the law reflecting best practice for corporate chartering. Most states adopted the model law fully or partially.

[3]Model Business Corporation Act § 7.42

Although there is still a chance that shareholders can successfully initiate a derivative suit against the board of directors despite their rejection, the UD law hinders shareholders from exerting pressure on managers using derivative suits.

Risk-taking behavior has been discussed extensively in corporate governance because it embodies conflicts of interest between managers, shareholders, and debtholders. Several scholars have emphasized how such conflicts of interest lead to deviation from optimal risk-taking decisions.(Jensen and Meckling, 1976; Myers and Majluf, 1984) Shareholders, as the owners of residual profit, tend to prefer riskier projects while debtholders, whose return is contractually determined, prefer less risky projects. Managers, otherwise, may be risk averse or not, depending on their interest alignment with shareholders, job security concerns, possible wealth expropriation, and other reasons. Therefore, risk-taking decisions, although not merely a corporate decision, are at the core of corporate governance as they can reflect conflicts of interest among major stakeholders. the adoption of UD laws changes the bargaining powers for conflicts of interest between managers and shareholders. In turn, it should eventually change the risk-taking decisions of corporations.

How exactly shareholder derivative suits affect managerial behavior in the insurance industry has not yet been explored despite its potential advantages. First, there are various organizational forms in the industry. The interests of effective owners of insurers are different depending on the organizational forms. This difference provides a good opportunity to examine the channel through which shareholder derivative suits could affect managerial behavior. For example, shareholder derivative lawsuits are controversial in mutual companies. Policyholders, who are considered the owners of a mutual company, are not the same as normal shareholders in stock insurance companies. Although policyholders have the right to vote on the management of their mutual insurer, their interests are more like those of creditors instead of equity holders simply because their objective is to have the mutual company deliver the contractually determined payout when incidents occur rather than seeking to ensure the prosperity of the company. Since, traditionally, creditors are barred from initiating a derivative action on behalf of the company, the issue of policyholder standing arises as policyholders are more like creditors than owners for certain situations.

The distinct objectives of owners between mutual insurers and stock insurers provide an interesting quasi-natural experimental setting to study how managers respond to shareholder litigation risk. Shareholders are equity holders in stock companies, but policyholders in mutual companies have similar interests as creditors. In addition, mutual policyholders do not always realize they have an ownership interest in the company. Since the conflicts of interest between major stakeholders are largely different in mutual and stock insurance companies, managers' strategies to address such conflict may lead to different risk-taking behavior. (Jensen and Meckling, 1976; John et al., 2008)

Another benefit of investigating the topic in the property-casualty insurance industry is that insurance companies are more homogeneous than publicly traded firms in general. Therefore, the definition of risk-taking is consistent across all insurance companies. Also, it enables us to construct different risk-taking measures other than the standard deviation of accounting variables, such as underwriting risk. Another benefit of studying the insurance industry is that the insurance business is a business of risk. Therefore, risk consideration plays an important role in most of the decisions made in insurance. This permits us to find various risk measures that reveal different perspectives of business risk. For example, in this study, I assess three risks - leverage risk, asset risk and underwriting risk - that are associated with overall risk, asset side risk and liability side risk, respectively.

This paper is organized as follows: In Section 1.2, I introduce the history and current status of derivative suits and the UD law. Specifically, I focus on problems arising from derivative suits and law changes to address these problems. In Section 1.3, I discuss the literature on corporate risk-taking, UD law, and organizational structure. Especially, the UD law literature shows that the impacts of the law on corporate governance and other behavior are inconsistent. Some literature finds that the adoption of UD laws improves corporate governance and protects shareholders while other literature claims that the adoption of UD laws adversely affects firms and may eventually damage firm value. In Section 1.4, I provide an overview of the standing issue of derivative actions by policyholders. The discussion of this issue implies that derivative actions or lawsuits can still pose a credible threat to managers in mutual companies in some states. In the next section, I develop my hypotheses. The first hypothesis focuses on risk-taking and the adoption of UD laws in general. The second and third hypothesis introduces organizational structure because the impact of UD law could be mixed due to different mechanisms. I hypothesize that the effect of UD law on corporate risk-taking and the corresponding speed of adjustments varies depending on organizational structure. In Section 1.6, I present the data and methodology used for empirical analysis and outline the construction of variables. The next section presents the empirical results and discussions. Section 1.8 provides some robustness checks in which different risk-taking measures are used. Specifically, I present other regression results that determine the economic channels through which the adoption of UD laws affects insurers' risk-taking behavior, investigate how long-tail lines of business affect the speed of adjustment in risk-taking measures, and test the exogeneity of UD law adoption. Lastly, I draw some relevant conclusions in Section 1.9.

1.2 History and Current Status of Derivative Suits and UD Law

To understand the efficacy and limitations of derivative suits and UD law, it is helpful to know their histories. The concept of derivative suits in the U.S. was first applied in the 1940s, with the universal demand requirement being proposed in 1989. Both derivative suits and the UD requirement have relatively short histories and were introduced to reduce claim abuses. Therefore, whether their purposes are being satisfied has not been fully tested and remains controversial.

A shareholder derivative suit is a form of representation litigation in which the shareholders sue a third party on behalf of the corporation. As mentioned by Scarlett (2013), the U.S. common law uses the English common law's necessary party rule[4] and

[4]The necessary party rule (also called the indispensable party rule) decides which party or entity must be included in a lawsuit in order for the court to give a judgement. Traditionally, if a lawsuit

its exceptions from England during the 1700s. After the American Revolution, U.S. law started diverging from English law. Later, representative lawsuits were recognized and became legitimate, allowing plaintiffs to file lawsuits on behalf of other entities. Representative lawsuits did not differentiate between shareholder derivative suits and class action suits until the 1940s.

Unlike in England, shareholder derivative suits have been far easier to pursue in the U.S., even though certain restrictions are applied. In 1829, the first shareholder action in the U.S. was decided by the Louisiana Supreme Court in *Percy v. Millaudon.* This case was the earliest example in which the "directors' decisions are entitled to deference" (Scarlett, 2013). In 1832, *Robinson v. Smith*[5] became the first case to recognize that the corporation should be the one to bring a suit against its directors, which laid the foundation for shareholder derivative suits. If the corporation refuses to or is incapable of filing a suit, the shareholders could file a suit with the corporation as a defendant. It is worth noting that shareholder derivative suits were not formally developed since shareholders in previous cases could only sue on behalf of themselves or all shareholders, and corporations were brought to courts as defendants. Before shareholder derivative suits were named as such, people began observing the negative side of the suits. For instance, in *Hawes v. Oakland*,[6] the U.S. Supreme Court tried to impose limitations on shareholder derivative suits because the frequency by which they were brought by shareholders had placed an excessive burden on U.S. courts. In the late 1940s, the courts started to routinely describe shareholder derivative lawsuits as suits on behalf of the corporation, which became the current definition of shareholder derivative lawsuits. This might have been an implicit response to claim abuses as well as an attempt to limit frivolous claims.[7] In other words, the purpose

is brought to recover the damage to a property, all owners of the property must be named in the lawsuit. This limits representative litigation. Especially, as the number of shareholders increases over time, this rule becomes less practical for larger corporations.

[5]Robinson v. Smith, 3 Paige Ch. 222 (N.Y. Ch. 1832).

[6]Hawes v. Oakland, 104 U.S. 450 (1881).

[7]A frivolous lawsuit is a lawsuit filed with the intention of harassing the counterparty and extracting settlement fees without proper investigation and proof.

of introducing shareholder derivative suits may have been to alleviate the burden of claim abuses and frivolous claims.

What underlies the change of definition of shareholder derivative suits is a totally different normative foundation, which permits shareholders to bring a derivative suit in narrower situations (Scarlett, 2013). In an earlier age, when a shareholder derivative suit on behalf of shareholders themselves or all shareholders was brought, the focus of the suit emphasized the harm to shareholders. The courts' decisions reflected the balance between directors and shareholders. After the change, shareholder derivative suits are brought on behalf of the corporation and emphasize the harm to the corporation. The shift in emphasis significantly decreases shareholders' power against directors and leaves room for the business judgment rule, which presumes that directors are fulfilling their fiduciary duty.

The shift may also be explained by perceptions of corporations (Scarlett, 2013). The rise of public corporations might have caused a change in perception. That is, a public corporation should assume responsibility for all stakeholders, such as customers and employees. Therefore, shareholders are no longer at the center of the tension. However, the literature has shown that derivative suits hurt corporations instead of benefiting most stakeholders. For example, the Wood Report (Wood, 1944) found that the main beneficiaries of a derivative suit are the plaintiff's attorneys. The corporation pays for the plaintiff's attorneys and the legal fees of directors. A similar result was found by Romano (1991). In fact, later literature shows that the actual effect of a derivative suit is not obvious. The stock price of a corporation after a derivative suit demonstrates little marginal improvement and only part of the objectives of the derivative suit are achieved when it is lengthy and complicated. This further supports the argument that the main beneficiaries seem to be the attorneys (Kinney, 1994). If only attorneys benefit from derivative suits at the cost of all other stakeholders, this raises doubt about the efficacy of these suits for corporate governance. So, it would seem reasonable to impose certain limitations on initiating a derivative suit.

Due to the nature of derivative suits and possible abuses, obstacles were formally introduced so that shareholders must overcome a barrier to initiate their suit. One such obstacle is the demand requirement. The demand requirement was first introduced in Delaware, and it requires shareholders to file a demand to the board of directors before any derivative suit is filed. Specifically, the demand should ask the directors to pursue the corporate claim against the third party that hurt the company. There exists an exemption to the demand requirement. If shareholders prove demand futility[8] by showing that the majority of the board are disinterested, then the demand requirement would no longer be applicable. Later, Aronson[9] set a higher bar for derivative suits. However, imposing more restrictions on derivative suits leads to other problems, such as substantial judicial discretion, which works against the business judgement rule doctrine. This doctrine gives discretion to managers and presumes that they make corporate decisions honestly, fairly, and in good faith.

To alleviate these problems, the American Bar Association proposed a major change to derivative suits in 1989. Section 7.42 of the Model Business Corporation Act (MBCA)[10] pertaining to shareholder derivative suits added the UD requirement. This provision requires shareholders to file a written demand upon the corporation to take action before undertaking actions themselves. Then, the shareholders need to wait for 90 days unless the demand has been rejected or irreparable injury could occur to the corporation within 90 days. The same exemption applies to the UD requirement. Typically, if a demand is rejected by the board of directors, the court will dismiss the derivative case. However, courts always struggle with the eligibility of futility (Kinney, 1994), which leaves some uncertainty in the exemption. It is noteworthy that only if the UD requirement is included in state statutes, does the UD requirement become the UD law.

[8]Demand futility refers to case in which the majority of the board is not impartial or is acting unfairly in some way so that a demand filed to the board of directors is useless.

[9]Aronson, 473 A.2d at 814, https://law.justia.com/cases/delaware/supreme-court/1984/473-a-2d-805-4.html

[10]MBCA is the law model prepared by the American Bar Association for the purpose of guiding states about corporate law. The act is not mandatory for all states. Each state could either adopt certain sections of the MBCA or have its own statute. Currently, thirty-four states fully or partially follow the MBCA.

Thirty-four states and the District of Columbia adopted the MBCA in whole or in part. Among these states, twenty-two adopted the UD requirements; and these states are listed in Table 1.2. Although Pennsylvania, for example, does not follow the MBCA in general, it did adopt the UD requirement.

To conclude, starting from the introduction of shareholder derivative suits, problems prevailed such as high litigation costs and claim abuses. The demand requirement was proposed to resolve these problems. However, the efficacy of the demand requirement is dubious. The history of derivative suits and the development of demand requirements reveals a bias toward managers, which is consistent with the business judgment rule doctrine.

1.3 Literature Review

This section reviews the literature concerning three topics: corporate risk-taking, shareholder litigation and UD law, and organizational structures in the U.S. insurance industry. For corporate risk-taking, I start with the general argument for managers' incentive for risk-taking behavior. Then I review some of the risk-taking literature that focuses on the U.S. insurance industry. In the next subsection, I present a review of shareholder litigation and UD law. Studies find that shareholder litigation does not always have a positive effect on corporate governance but can instead impede good corporate governance. Some recent study also raises doubt on the validity of UD law adoption as an exogenous shock for derivative litigation. Lastly, a review of organizational structures in the U.S. insurance industry shows that mutual insurers and stock insurers have distinct practices in business due to their unique agency problems among stakeholders and access to capital, which also leads to their differential risk-taking behavior.

1.3.1 Corporate Risk-Taking

In their seminal paper, Jensen and Meckling (1976) clearly describe the inherent conflict of interest between shareholders and debtholders, arguing that shareholders prefer to make risky investments while debtholders prefer less risky investments. This is because shareholders only earn the residual profits and are not held accountable for the downside risk, while the opposite is true for debtholders. Galai and Masulis (1976) provide a theoretical framework that combines the capital asset pricing model (CAPM) and the option pricing model to show that the incentive of shareholders is for increased firm risk. Therefore, managers who are also shareholders have the same incentive, i.e., to take more risks. However, managers may also be risk-averse. Saunders et al. (1990) argue that the risk-taking incentives of managers depend on the degree to which managers' interests are aligned with those of shareholders. When managers' career concerns outweigh their aligned interest with shareholders, they may decrease their degree of risk-taking. That is, managers may take less risk than shareholders desire. As a result, managers' risk-taking decisions depend on the degree to which they are aligned with shareholders. This poses an empirical question concerning the extent of managers' risk-taking decisions given their alignment with shareholders and career concerns. For example, Agrawal and Mandelker (1987) find evidence supporting the hypothesis that executive holders of common stock and options in the firm reduces managerial incentive problems. This research examined the association between managers' compensation and riskiness of the firms.

Several insurance studies also address the risk-taking issue. For example, Mankaï and Belgacem (2016) posit reinsurance usage as a new endogenous decision variable and analyze its effect on risk-taking and capital using a sample of U.S. property–liability insurance firms. The authors find that capital, risk-taking, and reinsurance usage are all correlated. Similarly, Baranoff and Sager (2002) study the relationship between capital and two measures of risk: product risk and asset risk. They find that capital is positively associated with asset risk but negatively associated with product risk. The difference is explained by the hypothesized impact of guarantee funds. The study highlights the importance of considering the two risk measures separately.

Additional studies examine insurers' risk-taking behavior as it relates to other influential factors, such as regulation and organizational structure. Cheng et al.

(2011) investigate the relationship between the risk-taking behavior of life–health (LH) insurers and the stability of their institutional ownership using a simultaneous equation system model. Their main finding is that stable institutional ownership is associated with lower total risk. Lin et al. (2014) examine the risk-taking behavior of property–liability insurers in the presence of risk-based capital regulation using a simultaneous threshold regression. The authors find a nonlinear relationship between insurers' risk-taking behavior and regulatory pressure. For example, the efficacy of RBC regulation is only evident for poorly capitalized insurers and pertains to their asset risks and product risks. Ho et al. (2013) examine the impact of organizational structure and board composition on risk-taking in the U.S. property-casualty insurance industry, addressing different risk-taking behaviors from different perspectives. Their findings are manifold. For instance, the authors reveal that insurers with relatively more insiders have higher total risk. Also, they conclude that total risk is managed by an insurer through adjusting underwriting, investment, and leverage risks.

1.3.2 Shareholder Litigation and UD Law

Shareholder litigation has been identified by some literature as an effective tool for corporate governance, giving shareholders a key channel to enforce managers' fiduciary duties (Bourveau et al., 2018). For example, derivative lawsuits and securities class action suits are mechanisms employed to enforce state level law and federal level law, respectively. Although the 1995 Private Securities Litigation Reform Act (PSLRA)[11] imposes a hurdle for shareholders to file class action lawsuits, shareholders' ability to file derivative lawsuits remained unaffected by the act. After 1995, many shareholders diverted to filing derivative lawsuits instead of securities class action lawsuits(Erickson, 2010).

Another channel that affects managers' litigation risk is the emergence of directors' and officers' (D&O) liability insurance. D&O liability insurance can shield

[11]The Private Securities Litigation Reform Act is legislation passed by Congress in 1995 to reduce frivolous securities lawsuits by requiring more evidence from plaintiffs before suing.

managers from monetary losses from derivative lawsuits, but it is unable to cover losses caused by managers' wrongdoings, such as dishonesty and intentional misconduct (Cox, 1999). Other losses such as reputation damage and negative impacts on managers' future careers are also not covered by the insurance. Therefore, derivative lawsuits could still pose credible threats to managers to follow their fiduciary, loyalty, and care duties under state law, even though D&O insurance exists.

In addition to restricting manager behavior, derivative lawsuits often contribute to corporate governance reforms. Almost 80% of settlements of derivative lawsuits involves the improvement of governance practices regarding the composition of board members, provisions, and other terms. (Erickson, 2010) In other words, derivative lawsuits not only impose restrictions on some managerial behavior but also lead to certain corporate governance practices. Notably, however, the adoption of UD laws across twenty-three states from 1989 to 2005 has changed the landscape for utilizing derivative lawsuits for corporate governance reform.

UD law is a state level law that requires derivative plaintiffs - that is, shareholders in our case - to file a demand to the board of directors before launching a derivative lawsuit (Bourveau et al., 2018). However, generally, the board of directors comprise the defendants in the lawsuit (Appel, 2019). UD law imposes a significant barrier for shareholders to file a lawsuit against the board of directors. Although many states allow certain exemptions for shareholders to bypass the UD law, shareholders' ability to initiate a lawsuit has been largely restricted (Bourveau et al., 2018). Thus, shareholder litigation, which can be a tool for corporate governance, was significantly reduced following the adoption of UD laws.

Some studies demonstrate the favorable side of shareholder litigation. Pukthuanthong et al. (2017) show that for short-term institutional investors, shareholder litigation is an effective external monitoring device that substitutes for good internal corporate governance mechanisms. The result from Dalla Pellegrina and Saraceno (2016) supports the notion that shareholder litigation and public supervision are complementary in nature. By exploiting the relation between securities class action suits (SCAs) and CEO bonuses, the authors find that SCAs are likely to suppress the growth rate of CEO bonuses, and they claim that securities litigation works as a complementary tool of corporate governance. Similarly, McTier and Wald (2011) find evidence supporting the claim that shareholder class action suits lead to improvements in both governance and investment policy. The authors, examining sued firms, conclude that these firms on average decrease overinvestment as well as payouts and increase cash holdings.

Other studies demonstrate the dark side of shareholder litigation. Bourveau et al. (2018) explore the relation between shareholder litigation risk and corporate disclosure using the UD law as an exogenous shock. They find that lower shareholder litigation risk causes firms to release more information (earnings forecasts, 8-K filings, etc.). Lin et al. (2021) show that the pressure imposed by shareholder litigation limits firms' innovative activities. Specifically, when shareholder litigation risk decreased following the adoption of UD laws, firms explored and succeeded in more innovative activities such as more investment in R&D and more patents produced. Studies by Chu and Zhao (2021), Block et al. (1993), Kinney (1994), and Shaner (2014) further suggest that derivative litigation risk can discourage managers from taking necessary risks and experimenting with new ideas. Moreover, Nguyen et al. (2018) find that shareholder litigation incentivizes firms to pursue a conservative liquidity policy leading to a higher level but lower value of cash.[12]

Several studies in the law literature also criticize the positive impact of shareholder litigation on corporate governance. Erickson (2010) argues that the chief beneficiaries of shareholder derivative suits are law firms. The agreement of shareholders to settle shareholder derivative suits in exchange for corporate governance reforms is often unhelpful for both the corporations and their shareholders. Analyzing 535 public corporations randomly selected from different sources, Romano (1991) finds that shareholder litigation is a weak instrument for corporate governance. Moreover, Helland and Klick (2009) find a positive correlation between regulation and litigation in the insurance industry. The authors find no evidence that there is a trade-off between regulation and litigation but do find evidence that lawyers and regulators

[12] The value of cash is measured by how much the stock price increases after an extra dollar is spent.

actually "piggy-back" off each other. Erickson (2017) argues that shareholder litigation has its own agency costs, just like corporations. Such agency costs exist because plaintiffs' attorneys pursue their own interests instead of those of shareholders.

Following Appel (2019)'s study which introduces a state's UD law adoption as an exogenous shock on shareholder derivative litigation, there has been several papers applying the same identification strategy. Some of them conduct additional analysis on the exogeneity of UD law adoption from different perspectives (Manchiraju et al., 2020; Bourveau et al., 2018; Nguyen et al., 2018; Ni and Yin, 2018) while some other studies provide none (Huang et al., 2020; Houston et al., 2018). Donelson et al. (2022) is the first paper after Appel (2019) that directly investigate the effect of UD law adoption on shareholder derivative litigation risk using actual derivative litigation data. The authors conclude that no significant change in derivative litigation is found following the adoption of UD laws. They also find no impact of UD law adoption on several corporate behaviors such as aggressive accounting and corporate decisions, which raises doubt on the validity of UD law adoption as an exogenous shock. However, their result may not apply to our study for a few reasons. First, Donelson et al. (2022)'s study focuses on derivative litigation of publicly traded firms. Most of the sample are incorporated in Delaware where no UD law is adopted.[13] In contrast, the distribution of the U.S property-casualty insurance companies is more uniform that provides stronger testing power.[14] Second, the data in the study starts from 1996, where almost 50% of states that adopted UD laws are excluded from the sample. There may be a diminishing effect of UD law adoption as firms incorporated in states with late UD law adoption may undertake similar changes as the ones with earlier UD law adoption, which makes the effect of UD law adoption less impactful for later states.[15] Lastly, my study focuses on risk-taking behavior in the property-

[13] 54,900 out of 94,049 firm-year records are in Delaware and 24,951 are in states that never adopt UD laws.

[14] From 1988 to 2015, the average Herfindahl-Hirschman Index (HHI) of the direct written premium by states is 0.044. The standard deviation is 0.0012. It indicates that the businesses over states are very broadly spread.

[15] One possible explanation is that as UD law adoption introduces a procedural obstacle that decreases shareholder derivative litigation risk, managers are prone to take riskier but value-enhancing

casualty insurance industry, which is more homogeneous due to similar businesses and practices. In addition, I use additional time horizons prior to UD law adoption, as in Bertrand and Mullainathan (2003) and Ni and Yin (2018) to examine the exogeneity of the adoption of UD laws.

1.3.3 Organizational Structure

The insurance industry provides a unique research context to study organizational structure. Stock and mutual insurers are different in their ownership and control, where stock insurance companies are owned by stockholders while mutual companies are owned by policyholders. The different organizational structures lead to distinct agency problems and different survival strategies. (Fama and Jensen, 1983a,b) For instance, mutual insurers benefit from lower agency costs due to owner-policyholder conflict but such benefit is offset by larger owner-manager conflict. [16] Eventually, such difference implies distinct practices in business between stock and mutual insurers in various aspects.

A similar and more precise prediction is implied by the managerial discretion hypothesis proposed by Mayers and Smith (1981) and Mayers and Smith (1988). The authors argue that due to the high agency cost of owner-manager conflicts, mutual insurers will be more likely to succeed in the lines of business that require less managerial discretion. Additional empirical evidence is provided by several studies to confirm the argument. (Lamm-Tennant and Starks, 1993; Kleffner and Doherty, 1996; Lee et al., 1997; Dionne et al., 2007) Especially, the studies conclude that stock insurers tend to be involved in riskier business while mutual insurers are more successful in less risky lines of business.

In addition to agency problems, the other key difference between stock and mutual resides in differential access to capital. Stock insurers are able to access the

investments which makes firms incorporated in the corresponding states more competitive. Thus, firms in states without a UD law are forced to undertake similar changes to remain competitive.

[16]Policyholders have less control over managers due to multiple reasons, such as inability to align managers' interests using shares and a single vote for each policyholder.

capital market via issuing stocks when new capital is needed. Whereas, mutual insurers typically generate new capital through operations or investments. This is the capital-constraint argument in multiple studies. (Cummins and Doherty, 2002; Mayers and Smith, 2002) The argument also explains why mutual insurers have less risky businesses because during bad times, riskier lines require more capital which is not easily accessible to mutual insurers.

Moreover, there are several other hypotheses proposed to explain stock and mutual insurers' differential practices based on their key differences in agency costs and access to capital. For instance, the managerial entrenchment hypothesis predicts that the managers' turnover in mutual companies depends less on firm performance as policyholders have less mechanisms to affect changes in management. (He and Sommer, 2010) The maturity hypothesis predicts that mutual insurers are advantageous in long-tail lines over stock insurers, which Cummins et al. (1999) provides empirical evidence to support. Lastly, the expense preference hypothesis argues that due to the lack of mechanisms to oversight managers in mutual companies, the managers tend to spend money on perks instead of value-enhancing investments.

In addition to the general hypotheses surrounding the differential organizational structures, several studies examine the impact of organizational structure on corporate behavior related to our study, especially risk-taking in the insurance context. Baranoff and Sager (2003) investigate the relation between organizational structure and asset risk. They find that stock companies are associated with greater financial[17] and asset risk-taking. Cummins and Sommer (1996) uses agency theory for their argument about capital and risk-taking. The authors conclude that a larger separation between shareholders and managers results in lower risk-taking. The mutual company is an extreme case of separation between shareholders and managers since it is almost impossible to align managers' and shareholders' interests using stock shares. Thus, a mutual company should have lower risk-taking than a stock company. Ho et al. (2013) find evidence to support the argument above. The authors find that mutual

[17] Financial risk is defined as the ratio of capital to assets in Baranoff and Sager (2003).

insurers have lower total risk, underwriting risk, and investment risk compared to stock insurers.

1.4 Derivative Actions by Policyholders

The impact of UD law on mutual insurance companies is more complicated than on stock insurance companies. This complication originates in the standing of policyholders. For example, during 1993 and 1994, several derivative suits were filed against MetLife. The policyholders alleged that the MetLife managers failed to "adequately oversee, supervise, and maintain control over MetLife's sales agents."[18] MetLife responded to the suits by emphasizing that the plaintiff, as a policyholder, did not have proper standing to bring a derivative suit. Unlike shareholders, policyholders in mutual insurance companies have a creditor-like interest (Allegaert, 1996). More specifically, although policyholders have the right to vote, they only care about the receipt of a contractually determined payout instead of the growth and profitability of their insurer. As a result, pressure from policyholders may incentivize mutual insurers to forego long-term investment and take less risk, which would in turn adversely affect the growth of mutual insurers.

The creditor-like interest of policyholders is the key to differentiating policyholders and shareholders regarding their derivative standing. It raises the question of whether policyholders can sue a third party on behalf of their mutual insurer. Traditionally, the rule bars derivative actions by creditors on behalf of solvent companies. Therefore, policyholders' standing for derivative action is controversial such that derivative actions by policyholders may affect mutual insurers differently from stock insurers. Although the standing issue is pivotal for understanding the influence of policyholders on managerial behavior, this issue has not been resolved by judicial opinions. Generally, the current corporate law precedents simply assume standing in either way without discussing and litigating the standing issue; this encourages parties to settle for a strike suit, which is a lawsuit aimed at obtaining a private settlement.

[18] https://law.justia.com/cases/federal/district-courts/FSupp/935/286/2594288/

Allegaert (1996) proposes an explanation for the persistent ambiguity of policyholders' standing in derivative suits based on practical considerations. The author argued that both plaintiff's lawyers and D&O insurance carriers have not pressed judges to resolve the issue since both sides benefit from the uncertainty. For example, if the defendant, the mutual insurer in this case, presses the issue, it could risk granting the standing of policyholders to statute. As a result, it could encourage the next plaintiff to go to trial with clear standing. Next, I will introduce some legal background about policyholder derivative suits, including some cases, to shed light on how derivative actions could affect mutual insurers.

There are three bases on which a policyholder derivative suit can proceed: common law precedent, statutes, and the Federal Rule of Civil Procedure 23.1. It is noteworthy that statute and common law precedents suggest the opposite standing for policyholders' derivative actions. I will discuss each briefly below.

Using common law as a base, the trustee-like characteristics of certain activities undertaken by mutual insurers complicate the application of credit-like analysis to policyholders. As a result, courts softened the application of the debtor-creditor rule, which bars creditors from suing on behalf of the solvent company (Allegaert, 1996). Two cases in New York analyzed these mutual insurance issues.[19] In *Swan v. Mutual Reserve Fund Life Association*,[20] the court emphasized the differences between stock and mutual companies and explicitly considered the operative differences between policyholder and stockholder. In contrast, *Young v. Equitable Life Assurance Society*, addressing a similar issue, emphasized the similarities between policyholders and stockholders, stating that "the whole body of policy holders... have a quasi-ownership in all the assets of the corporation, and are, like stockholders of an ordinary corporation, in effect its cestui que trust...". Unlike in early cases, modern cases provide even less satisfying analyses regarding policyholder standing for derivative actions. Many courts tend to assume the standing and move on. *Koster v. (American) Lumbermens*

[19]There are about two dozen reported cases of policyholders derivative actions. This number does not reflect the actual frequency of derivative complaints since many derivative suits are settled privately.

[20]155 NY 9, 49 NE 258 (1898)

Mutual Casualty Co. is a noteworthy case because it did differentiate policyholder and stockholder actions, and it is the only derivative action by policyholders that reached the Supreme Court. Following this case, there has been a tendency among courts to assume standing analogous to stockholders' actions. Moreover, some cases are involved with statutes at the state level and with state insurance regulation since federal insurance regulation is absent.

Each state has different state laws for derivative actions. For example, according to Demott's practical manual on derivative actions, thirty-six[21] state laws allow derivative actions by "one or more shareholders."[22] Other states leave the law pertaining to derivative actions to the courts. Some states grant policyholders standing for derivative actions, such as Wisconsin and Pennsylvania, while other states bar derivative actions brought by policyholders, such as New York. It is notable that there may exist a division between the legislative and judicial treatment of policyholders' standing; this is highlighted by the New York framework since its law, passed in 1964, overrides precedent cases that supported policyholders' standing.

In addition, the Federal Rule of Civil Procedure 23.1 added confusion to the standing issue textually since it included "members of a corporation" in the prerequisites for derivative actions instead of "shareholders." However, even though Rule 23.1 creates ambiguity on the standing issue, state laws are the deciding factor in assuming a plaintiff's standing to sue.

Given the discussion and cases above, it is obvious that there exists uncertainty and controversy over derivative actions by policyholders of mutual insurers due to the standing issue. However, despite the complexity of derivative actions, policyholders in some states could pose a credible threat to managers through derivative actions just as shareholders can in stock companies.

---

[21] The discussion of state laws here emphasizes the number of shareholders allowed as the plaintiff(s) in derivative suits. They are not directly related to UD laws.

[22] DeMott, Shareholder Derivative Actions § 4:02 at 9-30

1.5 Hypothesis Development

In this section, I develop hypotheses to examine the effect of UD law adoption on risk-taking behavior. As in Ho et al. (2013), for simplicity, I use "risk-taking" to refer to all risk-taking measures. There are two possible channels through which the adoption of UD laws could affect risk-taking: the "Free Cashflow Reduction" argument and the"Managerial Myopia" argument.

As mentioned above, the plaintiffs' attorneys are the main beneficiaries of derivative suits. Moreover, most derivative cases are settled with cash because it is the optimal solution for both attorneys and managers. Specifically, self-interested attorneys can take a large share of cash settlements, and managers' litigation costs can be covered by their companies. Therefore, managers who are exposed to larger litigation risk tend to hold larger cash reserves for potential legal expenses and cash settlements in the future (Nguyen et al., 2018). When UD law adoption reduces litigation risk for managers, I expect managers to reduce their cash holdings[23] and make riskier but value-enhancing investments.

The other channel replies on managerial myopia (Stein, 1988, 1989). As Lin et al. (2021) argue, shareholder litigation distracts managers' attention and leads to high settlement fees as well as a deterioration of the company's reputation. All of these factors raise career concerns for managers. To justify their decisions and avoid litigation, self-interested managers should have more certain and short-term projects, thereby creating "managerial myopia." In other words, shareholder litigation motivates managers to overemphasize strategies that reduce risk and halt investment in risky but value-enhancing projects (Lin et al., 2021). Since the adoption of UD laws hinder shareholders from initiating derivative suits against managers, so they are more likely to make risky but value-enhancing investments. As a result, insurance

[23]The argument that the adoption of UD laws lowers the level of cash holdings has been supported by Nguyen et al. (2018). Re-examining the causal relationship between UD laws and cash holding is beyond the scope of this paper. A more efficient method for investigating economic channels would be to re-examine the causal relationship between UD laws and risk-taking after controlling for cash holdings. The results in this regard are shown in the Economic Channels section.

companies should tend to increase risk-taking behavior following the passage of UD laws. This discussion suggests the following hypothesis:

Hypothesis 1: Following the passage of UD laws, insurance companies will increase their risk-taking.

The first hypothesis posits the reactions of all insurance companies to UD law adoption as a general case. However, managers in stock insurance companies may react differently from those in mutual insurance companies.

It has been widely recognized that shareholders in stock insurance companies and policyholders in mutual insurance companies have opposite interests. The differences between these interests rest on how residual claims are received. Shareholders are residual claimants of stock insurers, whereas policyholders' return is a contractually pre-specified amount contingent on specific accidents. In other words, policyholders' interest is consistent with that of creditors (Cummins and Sommer, 1996). Therefore, shareholders have the desire to increase the risk of companies (Galai and Masulis, 1976). Policyholders, on the other hand, want managers to take less risk. If shareholders or policyholders can use shareholder derivative suits to exert pressure on managers to choose projects consistent with their interests, I should observe the opposite impacts of UD law adoption on risk-taking for both stock insurers and mutual insurers.

However, the argument presented above is questionable. Generally, there are three methods that shareholders could use to protect their rights: vote, sell, and sue. However, suing using derivative suits does not incentivize managers in the same way as voting and selling stocks since suing serves as a deterrent. More specifically, in voting and selling, shareholders can exercise their rights whenever their interests are not met by managers. However, for a derivative suit, shareholders are unable to initiate action unless managers cause damage to the corporation. As a result, to reduce their litigation risk, managers only need to be cautious and conservative; they should hold more cash for future legal costs and invest conservatively. These arguments apply to both stock and mutual insurers. After the passage of the UD law, as litigation risk is reduced, managers should make riskier corporate decisions. It is notable that the standing of policyholders in derivative suits remains an unresolved issue and is not included in the argument outlined above. However, I expect that the ambiguous standing of policyholders does not eliminate managers' exposure to litigation risk through derivative suits in mutual companies. The reason for this rests on the fact that both plaintiffs' attorneys and managers in mutual insurers are unwilling to clarify the standing issue. The ambiguity in the standing of policyholders is a source of litigation risk different from that of shareholders. Therefore, managers in mutual insurers are exposed to litigation risk as well.

Lastly, as stock insurers have greater risk-taking than mutual insurers in general (Baranoff and Sager, 2003; Cummins and Sommer, 1996; Ho et al., 2013), mutual insurers may undertake more conservative adjustments of risk-taking following the UD law adoption. That is, the overall impact of UD law adoption may be less dramatic on mutual insurers. Thus, the second hypothesis can be stated as follows:

Hypothesis 2: Following the UD law adoption, stock insurance companies and mutual insurance companies will change their risk-taking behavior in the same direction. Meanwhile, mutual insurers will undertake more conservative adjustments of risk-taking than stock insurers.

Stock and mutual insurance companies may respond to a UD law's adoption in different time horizons. Mutuals in the U.S. property-casualty insurance industry have more net premium written in a few long-tail business lines such as automobile liability and homeowners (Cummins et al., 1999; Powell, 2017). The argument above is based on the maturity hypothesis[24]. Also, in Table 1.1, I compare the average ratios of net premium written in long-tail business lines to that in all lines of stock and mutual insurers using a sample from 1988 to 2010. The t-statistics in column 4 are for a one-tailed test. The data also support the argument that mutual insurers have more net premium written in long-tail business lines. In asset liability management, insurers need to adjust their asset side based on their liability side. Since it takes more time for long-tail liabilities to change due to their longer duration, it is harder

[24]Maturity hypothesis argues that for long-tail business lines such as liability insurance, it is easier for stock managers to exploit policyholders' interests so that mutual insurers are more likely to succeed in long-tail business lines.

for insurers to adjust their assets when they have more net premium written in long-tail business lines, which could affect leverage risk and asset risk. Therefore, since mutual insurers have more long-tail business, it takes more time for mutual insurers to change their assets following the adoption of UD laws.

In addition, different speeds of risk-taking adjustment for stock and mutual insurers can also be expected due to differential abilities to raise capital. As explained in Xie et al. (2017), mutual insurers have no access to the private or public equity capital market. They mainly build up their surplus over operations and investments. Such build-up takes more time for mutual insurers than for stock insurers since the latter have more channels to raise capital. Both leverage risk and asset risk for mutual insurers are associated with a slower process. Thus, mutual insurers adjust their risk-taking slowly compared to stock insurers. These arguments lead to the third hypothesis:

Hypothesis 3: Following UD law adoption, it takes more time for mutual companies to adjust their risk-taking behavior than for stock companies.

1.6 Data and Methodology

Our sample consists of U.S. property-casualty insurance companies with positive net premium written, positive surplus, and assets larger than 1 million dollars from 1988 to 2010. Also, to retain only viable companies, companies that have very low surplus (Leverage Risk I > 1.00) and very high surplus (Leverage Risk I < 0.00) are removed. In addition to the NAIC data, I collect data from Appel (2019) for UD law adoption in each state and present it in Table 1.2. The final dataset includes 35,575 firm-year observations in total.

In this study, I use three risk-taking measures: leverage risk, asset risk, and underwriting risk. Notably, due to limited data, I use their ratios instead of standard deviations as in Ho et al. (2013). One limitation of the standard deviation as a risk measure is that the NAIC data are annual data. When calculating standard deviations of some variables right after the law change, I have to use some pre-

Table 1.1.: Long-Tail Business Lines for Stock and Mutual Insurers

	(1) Stock	(2) Mutual	(3) t-statistics
1988	0.555	0.567	-0.758
1989	0.557	0.570	-0.807
1990	0.562	0.577	-0.931
1991	0.578	0.580	-0.095
1992	0.575	0.585	-0.590
1993	0.575	0.577	-0.140
1994	0.580	0.566	0.806
1995	0.575	0.572	0.200
1996	0.577	0.594	-0.922
1997	0.584	0.599	-0.801
1998	0.581	0.607	-1.375*
1999	0.577	0.599	-1.204
2000	0.574	0.608	-1.861**
2001	0.581	0.600	-1.034
2002	0.587	0.609	-1.172
2003	0.597	0.615	-1.000
2004	0.600	0.620	-1.105
2005	0.603	0.631	-1.517*
2006	0.606	0.632	-1.363*
2007	0.600	0.633	-1.753**
2008	0.608	0.639	-1.661**
2009	0.608	0.634	-1.429*
2010	0.613	0.634	-1.137

Note: This table includes the average ratios of net premium written in long-tail business lines to those in all lines for both stock and mutual insurers from 1988 to 2010. Long-tail lines include homeowners, farmowners, private passenger auto liability, product liability, other liability, commercial multiple peril, ocean marine, medical malpractice, workers compensation, commercial auto liability, aircraft, boiler and machinery, international, and reinsurance.. The t-statistics of the difference in average ratios between stock and mutual insurers are reported in column 4 under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. It is worth noting that the null hypothesis for the t-test is $\mu_{mutual} - \mu_{stock} > 0$. Accordingly, one-tailed tests are performed.

Table 1.2.: Adoption of UD Laws

State	Year	Citation
Georgia	1989	Ga. Code Ann. § 14-2-742
Michigan	1989	Mich. Comp. Laws Ann. § 450.1493a
Florida	1990	Fla. Stat. Ann. § 607.07401
Wisconsin	1991	Wis. Stat. Ann. § 180.742
Montana	1992	Mont. Code. Ann. § 35-1-543
Virginia	1992	Va. Code Ann § 13.1-672.1B
Utah	1992	Utah Code. Ann. § 16-10a-740(3)
New Hampshire	1993	N.H. Rev. Stat. Ann. § 293-A:7.42
Mississippi	1993	Miss. Code Ann. § 79-4-7.42
North Carolina	1995	N.C. Gen. Stat. § 55-7-42
Arizona	1996	Ariz. Rev. Stat. Ann. § 10-742
Nebraska	1996	Neb. Rev. Stat. § 21-2072
Connecticut	1997	Conn. Gen. Stat. Ann. § 33-722
Maine	1997	Me. Rev. Stat. Ann. 13-C, § 753
Pennsylvania	1997	Cuker v. Mikalauskas (547 Pa. 600, 692 A.2d 1042)
Texas	1997	Tex. Bus. Org. Code. Ann. 607.07401
Wyoming	1997	Wyo. Stat. § 17-16-742
Idaho	1998	Idaho Code § 30-1-742
Hawaii	2001	Haw. Rev. Stat. § 414-173
Iowa	2003	Iowa Code Ann. § 490.742
Massachusetts	2004	Mass. Gen. Laws. Ann. Ch. 156D, § 7.42
Rhode Island	2005	R.I. Gen. Laws. § 7-1.2-710(C)
South Dakoda	2005	S.D. Codified Laws 47-1A-742

Source: Adopted from Appel (2019)

treatment variables. Consequently, the variables used to calculate standard deviations contain samples from two different distributions. These standard deviations from different adoption periods are not desired.

Secondly, the standard deviation used as a risk measure is the result of risk-taking behavior, which could be quite noisy. For example, the standard deviation of ROA is the result of insurers' risk-taking behavior. The standard deviation contains both the expected and unexpected parts of ROA, such as market shocks that managers did not consider when making risk-taking decisions. However, only the expected deviation can reflect managers' risk-taking decisions. For example, if managers choose the same ratio of risky investments to net assets, in 2005 and 2008, at the end of the fiscal years, they would have the same asset risks if the ratio is used for risk-taking behavior but very different asset risks if the standard deviation is chosen due to the financial crisis in 2008. Therefore, a ratio is a better risk-taking measure in this case.

The leverage risk measure used in the main results is denoted by Leverage Risk I and is defined as one minus the ratio of surplus to net admitted assets.[25] Later, for a robustness check, I use another leverage risk measure, Leverage Risk II, which replaces net admitted assets by total assets. Leverage risk measures are important for illustrating an insurer's insolvency probability because if an insurer has relatively more surplus, it is less likely to become insolvent in the future. The leverage risk measure captures the overall riskiness of an insurer. On average, Leverage Risk I is 0.555, as shown in Table 1.3 below. No significant difference can be observed between Leverage Risk I and Leverage Risk II. The asset risk measure, Asset Risk, is defined as the ratio of risky investments to total invested assets. The risky investments include common stock, preferred stock and real estate investments. The asset risk measure reflects the riskiness of the asset side of an insurer. There are two asset risk measures: one that without mortgages, called Asset Risk I, and another with mortgages, called Asset Risk II. The averages for Asset Risk I and Asset Risk II in the sample are 0.129 and 0.133, respectively. The underwriting risk measure is constructed as net premium

[25] Admitted assets generally include assets that are liquid and are available to pay claims when necessary.

written in "risky lines" divided by total net premium written.[26] This risk measure captures the riskiness of the liability side of an insurer. The average of Underwriting Risk I is 0.699 for the sample.

The summary statistics of other variables in the sample are also presented in Table 1.3. Concerning variables in UD law adoption, 28.2% of the observations are associated with UD law adoption. This is the size of the post-treatment group. Of the observations, 23.4% concern mutual firms and 41.4% are in states that enacted the MBCA.[27] The Herfindahl index of direct premium written in each state is 0.595, while the Herfindahl index of net premium written in different business lines is 0.503.

In Table 1.5, I compare variable values between stock and mutual insurers in the sample using t-statistics. For the leverage risk measures, Leverage Risk I in stock insurers is 0.436 while the risk in mutual insurers is 0.479. The difference is statistically significant at the 1% level. Stock companies hold relatively less surplus than mutual companies. The same pattern can be found in Leverage Risk II and in the asset risk measures. Surplus is not only a buffer for large financial loss for insurers but also an indicator of overall risk. Holding relatively more capital may imply that the risk to which stock insurers' business is exposed is larger than it is for mutual insurers. The underwriting risk reflects the risk in the core business. I find the same pattern for Underwriting Risk. Thus, stock insurers are riskier than mutual companies. In addition, I also find differences at the 1% significance level for all other variables except for the UD Law Dummy variable and the MBCA dummy variable. Concerning the MBCA, 41.2% of stock companies and 42.4% of mutual companies are domiciled in states that follow MBCA guidance. The distribution of stock and mutual insurers in MBCA states is statistically significantly different at the 5% level.

[26]These risky lines are aircraft, allied lines, boiler and machinery, burglary and theft, credit, credit accident and health, commercial auto physical damage, commercial auto liability, commercial multiple peril, fire, financial guaranty, fidelity, glass, group accident and health, international, inland marine, medical malpractice, mortgage, ocean marine, other liability, products liability, reinsurance, surety, and workers compensation.

[27]As mentioned previously, states that fully or partially follow MBCA guidance have the discretion to adopt UD laws.

Table 1.3.: Summary Statistics

	N	Mean	SD	Min	P25	P50	P90	Max
Risk-Taking								
Leverage Risk I	35,575	0.555	0.211	0.017	0.436	0.609	0.779	0.898
Leverage Risk II	35,575	0.557	0.212	0.017	0.439	0.613	0.782	0.900
Asset Risk I	35,575	0.129	0.162	0.000	0.000	0.071	0.351	0.751
Asset Risk II	35,575	0.133	0.165	0.000	0.000	0.074	0.361	0.765
Underwriting Risk	35,575	0.699	0.300	0.000	0.435	0.774	1.000	1.002
UD Law								
UD Law Dummy	35,575	0.282	0.450	0	0	0	1	1
UD Law Dummy (0)	35,575	0.020	0.139	0	0	0	0	1
UD Law Dummy (1)	35,575	0.019	0.138	0	0	0	0	1
UD Law Dummy (2+)	35,575	0.242	0.429	0	0	0	1	1
Other Variables								
M	35,575	0.234	0.423	0	0	0	1	1
MBCA	35,575	0.414	0.493	0	0	0	1	1
Affiliation	35,575	0.662	0.473	0	0	1	1	1
Log Total Asset	35,575	13.855	3.562	7.506	10.846	13.345	18.774	21.567
Log Age	35,575	3.240	1.125	0.000	2.565	3.296	4.691	5.136
Geo Mix	35,575	0.595	0.384	0.043	0.185	0.623	1.000	1.000
Busi Mix	35,575	0.503	0.307	0.112	0.240	0.421	1.000	1.090
Cash Ratio	35,575	0.067	0.170	-0.003	0.000	0.000	0.202	0.978

Note: This table presents summary statistics of the sample from property-casualty insurance companies from 1988–2010. For all variables, number of observations, mean, standard deviation and other percentiles are reported. Risk-taking variables include leverage risk, asset risk, and underwriting risk. Universal demand law variables include UD Law Dummy, UD Law Dummy (0), UD Law Dummy (1), and UD Law Dummy (2+). Variable definitions are given in Table 1.4.

Table 1.4.: Definition of Variables

Variable	Description
Leverage Risk I	$1 - \frac{\text{surplus}}{\text{net admitted assets}}$
Leverage Risk II	$1 - \frac{\text{surplus}}{\text{total assets}}$
Asset Risk I	Ratio of stock and real estate to total invested assets
Asset Risk II	Ratio of stock, real estate, and mortgage to total invested assets
Underwriting Risk	Ratio of net premium written in commercial lines to total net premium written
UD Law Dummy	Equals 1 if the state adopted a UD law
UD Law Dummy (0)	Equals 1 if the state adopted a UD law in the current year, and 0 otherwise
UD Law Dummy (1)	Equals 1 if the state adopted a UD law for one year, and 0 otherwise
UD Law Dummy (2+)	Equals 1 if the state adopted a UD law for more than one year, and 0 otherwise
Log Total Asset	Natural logarithm of total assets
M	Equals 1 if the company is mutual, and 0 otherwise
Log Age	Natural logarithm of the age of the company
MBCA	Equals 1 if the state follows MBCA guidance
Affiliation	Equals 1 if the insurance company is affliated
Geo Mix	Herfindahl index of direct premium written in each state
Busi Mix	Herfindahl index of net premium written in each line
Cash Ratio	$\frac{\text{cash,cash equivalent, and short term investments}}{\text{invested assets}}$

Overall, the table shows that firm characteristics are significantly different for stock and mutual insurers.

As mentioned in Section 1.3.2, an important concern pertaining to a state's UD law adoption is whether or not the adoption is exogenous. Most literature shows no evidence for the endogeneity problem in the adoption of UD laws. Bourveau et al. (2018) argue that under the setting of this natural experiment, managers do not choose if they will receive the treatment after the law change. Appel (2019) provides no evidence against this argument. Also, some literature finds no evidence that changes in firm behavior existed prior to the law change. Such studies decompose the adoption of UD laws into shorter time horizons (Bertrand and Mullainathan, 2003; Bourveau et al., 2018). Moreover, Lin et al. (2021) considers the possibility that firms may affect the adoption of UD laws using lobbyists. Based on the database of the Center for Responsive Politics, the authors find no association between any type of lobbying and UD laws. Lastly, Nguyen et al. (2018) used state GDP growth, GDP per capita, and both industry and state-year effects to control for possible unobserved factors that could affect both corporate decisions and the adoption of UD laws. Their findings reveal no evidence for possible unobserved factors. However, Donelson et al. (2022) raise doubt on whether UD law adoption has any substantial impact on real litigation risk using derivative litigation data. Therefore, even though UD law is not designed for the property-casualty insurance industry alone, it is possible that endogeneity still exists. In addition to the exogeneity test in Section 1.8.4, I compare variables in the sample using t-statistics in the treatment group and control group.

As evident in Table 1.6, the variables in the control groups are different from those in the treatment group. However, for almost all risk measures and all control variables, the magnitude of the differences is small. Such differences may be attributable to some factors associated with specific domiciles or time effects. Therefore, in the models below, I control for both year and state fixed effects.

I first use a linear regression model to examine the relationship between UD law adoption and risk-taking. The model is shown below:

Table 1.5.: Variable Statistics between Stock and Mutual Insurers

Variable Names	Stock	Mutual	t-statistics
Leverage Risk I	0.436	0.479	-15.781***
Leverage Risk II	0.434	0.483	-15.817***
Asset Risk I	0.112	0.197	-30.305***
Asset Risk II	0.117	0.201	-18.987***
Underwriting Risk	0.726	0.466	3.150***
UD Law Dummy	0.575	0.577	-0.140
Log Total Asset	13.854	13.379	10.920***
Log Age	2.980	4.139	-93.924***
Geo Mix	0.537	0.765	-50.139***
Busi Mix	0.504	0.466	10.306***
MBCA	0.412	0.424	-2.011**
Obser. Number	28,706	8,793	

Note: This table presents the average of variables in stock and mutual insurers in property-casualty insurance companies separately from 1988–2010. Also, a t-statistic is calculated for each variable. There are 26,706 stock insurers and 8,793 mutual insurers in the sample. Variable definitions are given in Table 1.4. p-values are reported in column 4 under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 1.6.: Variable Statistics between Treatment and Control Groups

Variable Names	Control	Treatment	t-statistics
Leverage Risk I	0.442	0.455	-4.979***
Leverage Risk II	0.442	0.459	-6.119**
Asset Risk I	0.129	0.138	-3.133***
Asset Risk II	0.135	0.141	-1.539
Underwriting Risk	0.653	0.695	-0.541
Log Total Asset	14.371	12.210	55.354***
Log Age	3.235	3.293	-4.530***
Geo Mix	0.588	0.597	-2.196**
Busi Mix	0.485	0.519	-9.806***
MBCA	0.316	0.656	-63.938***
Obser. Number	26,587	10,912	

Note: This table presents the average of variables in the treatment and control groups in property-casualty insurance companies separately from 1988–2010. Also, a t-statistic is calculated for each variable. There are 26,587 observations in the control group and 10,912 observations in the treatment group. Variable definitions are given in Table 1.4. The t-statistics are reported in column 4. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

$$RiskTaking_{ist} = \beta_0 + \beta_1 \text{UD Law}_{st} + \sum_j \theta_j Controls_{jist} + \alpha_i + \theta_{st} + \epsilon_{ist} \tag{1.1}$$

where j indicates the number of control variables, α_i is the firm level effect, θ_{st} is the year-state level effect, and ϵ_{ist} is the firm-year-state level error.

I then add an interaction term between the dummy variable for mutual companies and the dummy variable for the adoption of UD laws to the regression. The same is applied to the control variables as well. The model is shown below:

$$\begin{aligned} RiskTaking_{ist} =& \beta_0 + \lambda_1 \text{UD Law}_{st} + \lambda_2 Mutual_s + \lambda_3 \text{UD Law}_{st} \times Mutual_{ist} \\ &+ \sum_j \theta_j Controls_{jist} + \sum_j \theta_j Controls_{jist} \times Mutual_{ist} \\ &+ \alpha_i + \theta_{st} + \epsilon_{ist} \end{aligned} \tag{1.2}$$

The time span is relatively long, and thus the treatment effect may contain noise. To resolve this problem, I focus on shorter time horizons around the adoption of UD laws. Three dummy variables are introduced to examine the impact of UD law adoption in detail: UD Law Dummy (0), UD Law Dummy (1), and UD Law Dummy (2+). UD Law Dummy (0) equals 1 if the state in which the firm is domiciled adopted a UD law in the current year, and 0 otherwise. UD Law Dummy (1) equals 1 if the state in which the firm is domiciled adopted a UD law for exactly one year and 0 otherwise. UD Law Dummy (2+) equals 1 if the state in which the firm is domiciled adopted a UD law for more than one year and 0 otherwise. For instance, Wyoming adopted a UD law in 1997. UD Law Dummy (0) for insurers domiciled in Wyoming equals 1 in 1997 and 0 otherwise. Similarly, UD Law Dummy (1) for insurers domiciled in Wyoming equals 1 in 1997 and 0 otherwise. For UD Law Dummy (2+), it equals 1 for year larger than 1998 and 0 otherwise.

I regress risk-taking on the three UD law dummies to explore when insurers react to the passage of UD laws. This leads to the following model:

$$
\begin{aligned}
RiskTaking_{ist} = & \beta_0 + \phi_1 \text{UD Law } (0)_{st} + \phi_2 \text{UD Law } (1)_{st} + \phi_3 \text{UD Law } (2+)_{st} \\
& + \sum_j \theta_j Controls_{jist} + \alpha_i + \theta_{st} + \epsilon_{ist}
\end{aligned} \tag{1.3}
$$

I then interact the three dummies with the mutual dummy variable to test Hypothesis 3. This enables me to examine how quickly the firm responds to the law change and how long the effect lasts after the adoption.

In the next model, I include the dummy variables for UD law adoption at different time horizons.

$$
\begin{aligned}
RiskTaking_{ist} = & \beta_0 + \sum_i^3 \delta_i \text{UD Law (i)}_{st} + \delta_4 Mutual_s \\
& + \sum_i^3 \delta_{i+4} \text{UD Law (i)}_{st} \times Mutual_s + \sum_j \theta_j Controls_{jist} \\
& + \sum_j \theta_j Controls_{jist} \times Mutual_{ist} + \alpha_i + \theta_{st} + \epsilon_{ist}
\end{aligned} \tag{1.4}
$$

For all models, I use robust standard errors to account for any possible heteroscedasticity.

Table 1.7 shows the relevant coefficients and expected signs in the models for the hypotheses. Both model (1) and model (3) can be used to test Hypothesis 1, which claims that the adoption of UD laws increases risk-taking. I expect β_1 in model (1) and at least one of ϕ_1, ϕ_2 , and ϕ_3 in model (3) to be positive and statistically significant. Both model (2) and model (4) can be used to test Hypothesis 2. For the first part of Hypothesis 2, I expect the sign of λ_1 to be the same as the sign of $\lambda_1 + \lambda_3$ in model (2). For model (4), I expect $\delta_i > 0$ for i = 1,2, or 3, and $\delta_j + \delta_{j+4} > 0$ for j = 1, 2, or 3. That is, the effect of UD law adoption on risk-taking can be found for stock insurers and mutual insurers in some time window and such effect is to increase risk-taking. For the second part, the differential impacts of UD law adoption on risk-taking measures may be manifold. Mutual insurers may undertake less change in all risk-taking measures or have opposite changes in different risk-taking measures to maintain lower total risk than stock insurers.

Lastly, Hypothesis 3 concerns the different speeds of adjustment for stock and mutual insurers. Therefore, the hypothesis focuses on the duration of incremental changes that can be tested using model (4). For stock insurers, I expect a prompt response to UD law adoption. Thus, UD Law Dummy (0) or UD Law Dummy (1) should have statistically significant coefficients. In other words, in model (4), I expect $\delta_i > 0$ for i = 1 or 2 for stock insurers. Whereas for mutual insurers, the effect is expected to find after stock insurers since their speed of adjustment is slower. Thus, I expect to observe an incremental change for mutual insurers that is slower than that for stock insurers. In model (4), I expect $\delta_j + 4 - \delta_j > 0$ for $j > i$.

Table 1.7.: Hypotheses, Coefficients, and Expected Signs

Hypothesis	**Coefficients and Expected Signs**
Hypothesis 1	$\beta_1 > 0$ in model (1) and at least one of $\phi > 0$ in model (3)
Hypothesis 2	$sign(\lambda_1) = sign(\lambda_1 + \lambda_3)$ in model (2) and $\delta_i > 0$ for i = 1,2,3 and $\delta_j + \delta_{j+4} > 0$ for j = 1,2,3 in model (4)
Hypothesis 3	$\delta_i > 0$ for i = 1 or 2 and $\delta_{j+4} - \delta_j > 0$ for $j > i$ in model (4)

Difference-in-difference (DiD) is a common model used to analyze law changes. However, in this study, I could not use the model. DiD requires two dummy variables. The first variable indicates the time before and after a law change, while the second indicates when an observation is subject to a law change. Adoptions of UD law are staggered. The time point that divides the pre-treatment period and the post-treatment period cannot be identified. Actually, one advantage of using DID is that it differentiates the time effect and the treatment effect on the dependent variable. However, it also assumes that the time trend of the dependent variable is the same before and after the law change. In my model, I use year fixed effects to capture the shift of the dependent variables over time.

Size: Firm size affects an insurer's risk-taking decisions due to its effect on access to capital and investment opportunities (Mankaï and Belgacem, 2016). Baranoff and

Sager (2003) find that larger firms tend to have more risky asset portfolios and lower capital ratios. This variable is measured as the log of total assets.

Geographic Concentration: Geographic concentration is also associated with risk-taking, as found in Ho et al. (2013). This variable is measured by the Herfindahl index percentages of direct premium written in different areas.

Business Concentration: Business concentration is associated with risk-taking as well, as found in Ho et al. (2013). This variable is measured by the Herfindahl index percentages of net premium written in different lines.

Age: Age is related to the management and ownership structure. Younger firms are more likely to depend on the specialized services of entrepreneur-managers, while older firms are more likely to have greater separation between ownership and control (Easterbrook and Fischel, 1991). Such differences determine the governance mechanisms that suit firms best, which in turn determines the impact of a law on corporate governance.

1.7 Results

In Table 1.8, I run regressions with and without fixed effects and only use the UD Law Dummy variable as the shock to explore the impact of UD law adoption on all insurers' risk-taking. The R^2 is significantly larger when both year and state fixed effects are included compared to when they are not. Based on the results in column 4 through column 6, all three risk-taking measures increase and the effect is significant at the 5% level. That is, Leverage Risk, Asset Risk, and Underwriting Risk increase by 0.0082, 0.0125 and 0.0169, respectively, which is consistent with Hypothesis 1.

In Table 1.9, I highlight and differentiate stock companies and mutual companies to test Hypothesis 2. For stock companies, I find that all Leverage Risk, Asset Risk and Underwriting Risk are affected by UD law adoption at the 5% significance level. For mutual companies, their Asset Risk increase additionally by 0.0194 at the 1% significance level but the Underwriting Risk has no significant change. The reason that Underwriting Risk is not affected may be related to the business judgement

Table 1.8.: UD Law and Risk-Taking

Variables	(1) LR I	(2) AR I	(3) UWR	(4) LR I	(5) AR I	(6) UWR
UD Law Dummy	0.0335***	0.0152***	-0.0017	0.0082**	0.0125***	0.0169***
	(0.000)	(0.000)	(0.626)	(0.042)	(0.000)	(0.005)
Log Total Asset	0.0169***	0.0043***	0.0052***	0.0562***	0.0122***	0.0121***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log Age	-0.0077***	0.0440***	-0.0232***	-0.0135***	0.0428***	-0.0286***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Geo Mix	-0.0468***	0.0173***	-0.1840***	0.0299***	0.0408***	-0.1762***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Busi Mix	-0.1195***	-0.0066**	0.2412***	-0.0903***	-0.0038	0.2447***
	(0.000)	(0.025)	(0.000)	(0.000)	(0.204)	(0.000)
Observations	35,575	35,575	35,575	35,574	35,574	35,574
State Fixed Effect	NO	NO	NO	YES	YES	YES
Year Fixed Effect	NO	NO	NO	YES	YES	YES
R-sq	0.14	0.103	0.107	0.250	0.107	0.112

Note: This table presents the results of OLS regressions that estimate the relationship between the adoption of UD laws and corporate risk-taking. The sample consists of risk-taking measures and firm characteristics in property-casualty insurance firms during 1988 – 2010. I estimate pooled OLS regression in model (1). The dependent variables are corporate risk-taking measures. There are three risk-taking measures: Leverage Risk I, Asset Risk I and Underwriting Risk. In this table, the risk-taking measures are abbreviated as LR I, AR I, and UWR. The results without fixed effects are presented in column (1) – (3), respectively. The results with state and year fixed effects are presented in column (4) – (6), respectively. The R squared for column (4) - (6) is within R squared. Variable definitions are given in Table 1.4. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

rule. The business judgement rule presumes that managers possess the necessary knowledge, information and expertise to make corporate decisions in the best interest of the company. If plaintiffs want to successfully bring a derivative suit, they must rebut the presumption. It is more difficult for plaintiffs to rebut when it comes to the core business of an insurer since insurance products are more complicated and opaquer than asset investments.[28] Therefore, Underwriting Risk, which originates in the core business of insurance companies, should be less affected by shareholder derivative suits. In other words, the adoption of UD laws should not affect Underwriting Risk of mutual insurers.

In contrast to stock insurers, Leverage Risk of mutual insurers decreases aftera UD adoption. The overall effect of UD law adoption on Leverage Risk is negative -0.0183 (=0.0168-0.0351), which is statistically significant at the 1% level according to F statistics. Combining all results, it seems that, in response to UD law adoption, mutual insurers invest relatively more in risky assets while they also hold relatively more capital that offsets the effect on overall risk. However, stock insurers respond to the adoption of UD laws by increasing all risk-taking measures. This result partially supports Hypothesis 2.

As mentioned above, stock and mutual insurers may react to UD law in different time horizons. When all time horizons are combined, the effect of UD law adoption may be diluted due to different adjustment speeds. All of the following regression tables include results that consider different time horizons with both year and state fixed effects.

In Table 1.10, I examine shorter time horizons around UD law adoption. These three time horizons allow me to examine when insurers responded to the adoption of UD laws. Instead of comparing each coefficient of the UD law variables with the baseline, which corresponds to the control group only, the following discussion focuses on incremental changes in UD law variables as well. No risk-taking measures exhibit a significant change during the year of UD law adoption. One year later, only Asset Risk

[28]Additionally, it is harder for policyholders to learn details about businesses in mutual insurers as policyholders are atomic due to their voting power.

Table 1.9.: UD Law, Risk-Taking and Organizational Structure

Independent Variables	(1) Leverage Risk I	(2) Asset Risk I	(3) Underwriting Risk
UD Law Dummy	0.0168***	0.0068**	0.0153**
	(0.000)	(0.048)	(0.015)
UD Law Dummy × *Mutual*	-0.0351***	0.0194***	-0.0063
	(0.000)	(0.000)	(0.437)
Mutual	0.2482***	0.0301*	0.3194***
	(0.000)	(0.072)	(0.000)
Log Total Asset	0.0540***	0.0155***	0.0067***
	(0.000)	(0.000)	(0.000)
Log Age	-0.0031**	0.0273***	-0.0036*
	(0.012)	(0.000)	(0.062)
Geo Mix	0.0377***	0.0232***	-0.1463***
	(0.000)	(0.000)	(0.000)
Busi Mix	-0.0971***	0.0063**	0.1861***
	(0.000)	(0.049)	(0.000)
Log Total Asset × *Mutual*	-0.0033***	-0.0039***	-0.0077***
	(0.000)	(0.000)	(0.000)
Log Age × *Mutual*	-0.0445***	0.0247***	-0.0707***
	(0.000)	(0.000)	(0.000)
Geo Mix × *Mutual*	-0.0319***	-0.0165***	-0.1202***
	(0.000)	(0.008)	(0.000)
Busi Mix × *Mutual*	-0.0110	-0.0269***	0.2281***
	(0.230)	(0.000)	(0.000)
Observations	35,570	35,570	35,570
State Fixed Effect	YES	YES	YES
Year Fixed Effect	YES	YES	YES
within. R-sq	0.257	0.126	0.137

Note: This table presents the results of OLS regressions that estimate the relationship between adoption of UD law, corporate risk-taking, and organizational structure. The sample consists of risk-taking measures and firm characteristics in property-casualty insurance firms during 1988–2010. I estimate pooled OLS regression in model (2). The dependent variables are corporate risk-taking measures. There are three risk-taking measures: Leverage Risk I, Asset Risk I and Underwriting Risk. Results with year and state fixed effects are presented in columns (1)–(3), respectively. Variable definitions are in Table 1.4. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

increases by 0.0134 at the 5% significance level. After two years, all three risk-taking measures have statistically significant increases. In other words, insurance companies do not respond to UD law immediately after adoption. This finding supports the argument that insurance companies may not anticipate UD law adoption or take any action prior to adoption. The results in Table 1.10 support Hypothesis 1 for all risk-taking measures. Next, I separate stock insurers and mutual insurers to examine the relationship between UD law adoption and risk-taking at different time horizons.

Table 1.10.: UD Law and Risk-Taking in Detail

	(1)	(2)	(3)
Independent Variables	**Leverage Risk I**	**Asset Risk I**	**Underwriting Risk**
UD Law Dummy (0)	-0.0023	0.0084	0.0015
	(0.771)	(0.206)	(0.900)
UD Law Dummy (1)	0.0044	0.0134**	0.0064
	(0.567)	(0.040)	(0.593)
UD Law Dummy (2+)	0.0105**	0.0131***	0.0211***
	(0.013)	(0.000)	(0.001)
Log Total Asset	0.0562***	0.0122***	0.0121***
	(0.000)	(0.000)	(0.000)
Log Age	-0.0135***	0.0428***	-0.0286***
	(0.000)	(0.000)	(0.000)
Geo Mix	0.0299***	0.0407***	-0.1763***
	(0.000)	(0.000)	(0.000)
Busi Mix	-0.0903***	-0.0038	0.2447***
	(0.000)	(0.205)	(0.000)
Observations	35,574	35,574	35,574
within R-sq	0.250	0.107	0.120

Note: This table presents the results of OLS regressions that estimate the relationship between the adoption of UD laws and corporate risk-taking in different time horizons. There are three time horizons around the adoption. The sample consists of risk-taking measures and firm characteristics in property-casualty insurance firms during 1988–2010. I estimate pooled OLS regression in model (3). The dependent variables are corporate risk-taking measures. There are three risk-taking measures: Leverage Risk I, Asset Risk I and Underwriting Risk. The results with year and state fixed effects are presented in columns (1)–(3), respectively. Variable definitions are given in Table 1.4. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 1.11, like the results in Table 1.9, indicates that stock insurers change all risk-taking measures in response to UD law adoption. Leverage Risk and Underwriting Risk increase by 0.0193 and 0.0197 after two years of UD law adoption, respectively. Asset Risk increases by 0.0143 one year later and an additional 0.0063 thereafter. For Asset Risk, stock companies respond faster than mutual insurers. This could be attributable to the fact that the stock component in Asset Risk has higher liquidity in stock insurers than mutual insurers. For mutual insurers, Leverage Risk decreases by 0.0174 (-0.0367 + 0.0193) after two years of UD law adoption, and some change is observed during the first year with statistical significance at the 10% level. This indicates that mutual insurers adjust Leverage Risk faster than stock insurers. The Asset Risk of mutual insurers increase by 0.0251 at the 1% significance level when compared with stock insurers. These results partially support Hypothesis 1 and Hypothesis 2. In addition, for Asset Risk, it takes more time for mutual insurers to respond, which is consistent with Hypothesis 3. But mutual insurers adjust their Leverage Risk faster than stock insurers, which may be the preparation for a more dramatic change in Asset Risk than stock insurers.

Therefore, after investigating the relationship between UD law adoption and risk-taking for both stock insurers and mutual insurers within shorter time horizons, I find that both insurers increase Asset Risk following UD law adoption, which is consistent with Hypothesis 1 and Hypothesis 2. However, stock insurers increase Leverage Risk while mutual insurers decrease Leverage Risk, which does not support the first part of Hypothesis 2 but is consistent with the argument that mutual insurers tend to maintain lower total risk so that the decrease in Leverage Risk can be considered as an offset of the increase in Asset Risk. Moreover, significant differences in the speed of adjustments can be observed in Asset Risk and Leverage Risk. The former resides on the limited access to capital for mutual insurers and supports Hypothesis 3. The latter can be explained as the preparation for a more dramatic change in Asset Risk. Thus, the results show that organizational forms are indeed associated with the scale and the direction of change, as well as the speed of adjustment in risk-taking behavior in response to the adoption of UD laws.

Table 1.11.: UD Law, Risk-Taking, and Organizational Structure in Detail

Independent Variables	(1) Leverage Risk I	(2) Asset Risk I	(3) Underwriting Risk
UD Law Dummy (0)	0.0041	0.0104	-0.0029
	(0.636)	(0.147)	(0.824)
UD Law Dummy (1)	0.0115	0.0143**	0.0032
	(0.184)	(0.046)	(0.810)
UD Law Dummy (2+)	0.0193***	0.0063*	0.0197***
	(0.000)	(0.081)	(0.003)
Mutual	0.2490***	0.0271	0.3199***
	(0.000)	(0.106)	(0.000)
UD Law Dummy (0) × *Mutual*	-0.0242	-0.0121	0.0060
	(0.113)	(0.341)	(0.796)
UD Law Dummy (1) × *Mutual*	-0.0275*	-0.0067	-0.0028
	(0.077)	(0.604)	(0.905)
UD Law Dummy (2+) × *Mutual*	-0.0367***	0.0251***	-0.0075
	(0.000)	(0.000)	(0.386)
Log Total Asset	0.0540***	0.0155***	0.0067***
	(0.000)	(0.000)	(0.000)
Log Age	-0.0032**	0.0274***	-0.0036*
	(0.011)	(0.000)	(0.060)
Geo Mix	0.0377***	0.0232***	-0.1463***
	(0.000)	(0.000)	(0.000)
Busi Mix	-0.0971***	0.0064**	0.1861***
	(0.000)	(0.047)	(0.000)
Log Total Asset × *Mutual*	-0.0033***	-0.0037***	-0.0078***
	(0.000)	(0.000)	(0.000)
Log Age × *Mutual*	-0.0445***	0.0247***	-0.0706***
	(0.000)	(0.000)	(0.000)
Geo Mix × *Mutual*	-0.0320***	-0.0162***	-0.1203***
	(0.000)	(0.009)	(0.000)
Busi Mix × *Mutual*	-0.0110	-0.0270***	0.2282***
	(0.232)	(0.000)	(0.000)
Observations	35,570	35,570	35,570
State Fixed Effect	YES	YES	YES
Year Fixed Effect	YES	YES	YES
within R-sq	0.257	0.127	0.137

Note: This table presents the results of the pooled OLS regression estimated in model (4). The sample is the same as in Table 1.10. The dependent variables are corporate risk-taking measures. There are three risk-taking measures: Leverage Risk, Asset Risk and Underwriting Risk. The results with year and state fixed effects are presented in all columns. Variable definitions are given in Table 1.4. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

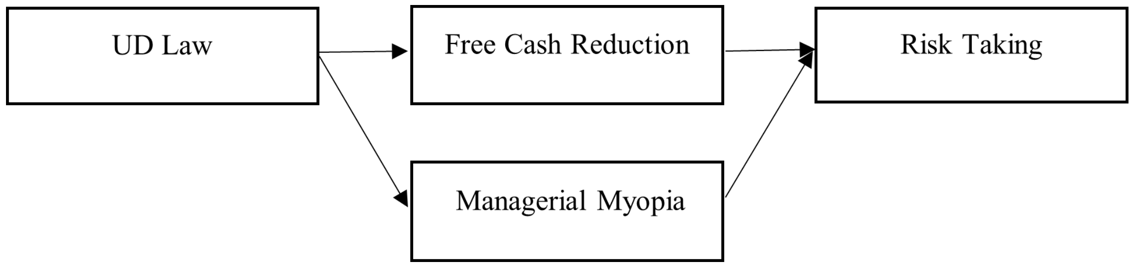

Figure 1.1.: Economic Channel Demonstration 1

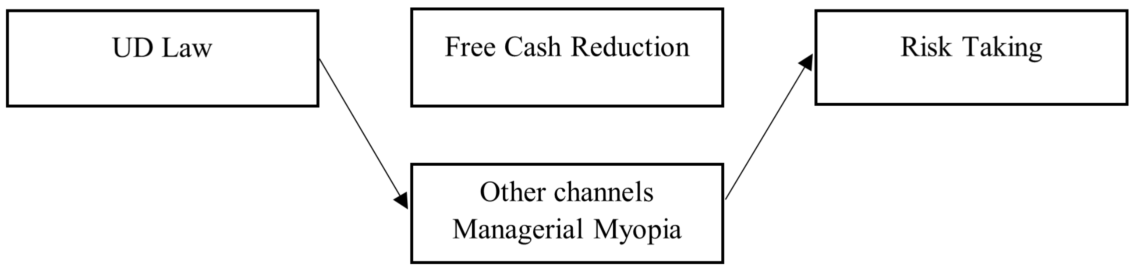

Figure 1.2.: Economic Channel Demonstration 2

1.8 Robustness Check

1.8.1 Alternative Risk-Taking Measures

I propose two additional methods to calculate Leverage Risk and Asset Risk. In Table 1.4, the alternative risk-taking measures are used to estimate the model in short time horizons to check the robustness of the model. The definitions are shown in Table 1.4.The model results are consistent with previous conclusions. That is, stock insurers have increased Leverage Risk, while mutual insurers decrease risk-taking behavior. Both types of insurers have a delayed response. For Asset Risk, stock insurers demonstrate a faster speed of adjustment compared with mutual insurers.

1.8.2 Economic Channels

The results in Table 1.10 reveal that, following the adoption of UD laws, insurers tend to increase their Asset Risk and Underwriting Risk. These results partially support Hypothesis 1. However, two channels are mentioned when developing Hy-

Table 1.12.: UD Law, Risk-Taking and Organizational Structure for Other Measures

Independent Variables	(1) Leverage Risk II	(2) Asset Risk II
UD Law Dummy (0)	0.0025	0.0102
	(0.775)	(0.165)
UD Law Dummy (1)	0.0091	0.0157**
	(0.296)	(0.033)
UD Law Dummy (2+)	0.0183***	0.0058
	(0.000)	(0.120)
Mutual	0.2463***	0.0420**
	(0.000)	(0.014)
UD Law Dummy (0) × *Mutual*	-0.0209	-0.0123
	(0.174)	(0.345)
UD Law Dummy (1) × *Mutual*	-0.0257	-0.0082
	(0.101)	(0.539)
UD Law Dummy (2+) × *Mutual*	-0.0337***	0.0253***
	(0.000)	(0.000)
Log Total Asset	0.0529***	0.0154***
	(0.000)	(0.000)
Log Age	-0.0024*	0.0280***
	(0.060)	(0.000)
Geo Mix	0.0373***	0.0227***
	(0.000)	(0.000)
Busi Mix	-0.0944***	0.0083**
	(0.000)	(0.011)
Log Total Asset × *Mutual*	-0.0027***	-0.0042***
	(0.000)	(0.000)
Log Age × *Mutual*	-0.0473***	0.0235***
	(0.000)	(0.000)
Geo Mix × *Mutual*	-0.0291***	-0.0188***
	(0.000)	(0.003)
Busi Mix × *Mutual*	-0.0160*	-0.0336***
	(0.082)	(0.000)
Observations	35,574	35,574
State Fixed Effect	YES	YES
Year Fixed Effect	YES	YES
within R-sq	0.249	0.122

Note: This table presents the results of the pooled OLS regression estimated in model (4). The sample is the same as in Table 1.10. The dependent variables are corporate risk-taking measures. There are two risk-taking measures: Leverage Risk II and Asset Risk II. The results with year and state fixed effects are presented in all columns. Variable definitions are given in Table 1.4. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

pothesis 1. First, it is possible that UD law adoption affects risk-taking behavior by alleviating the free cashflow problem. It is also possible that managerial myopia is reduced following the passage of UD law. These channels cannot be distinguished using previous results. In this section, I develop another model to determine which channel plays a more important role.

Figure 1.1 and Figure 1.2 demonstrate two possible structures for the economic channels. Figure 1.1 assumes that both a free cash flow problem and managerial myopia are channels through which UD law adoption could affect insurers' risk-taking. In contrast, Figure 1.2 assumes that the free cash flow problem is not one of these channels. If I add one more control variable, Cash_Ratio which is the ratio of cash, cash equivalents, and short-term investments to invested assets, the results of the model can provide support concerning which figure describes the appropriate economical channels.

The following model is used for this test. The model is very similar to model (3) except that it includes one more control variable: Cash_Ratio. After adding this control variable, I compare the results of model (3) and model (5), especially the magnitudes of coefficients for UD Law Dummy (0), UD Law Dummy (1), and UD Law Dummy (2+).

$$\begin{aligned} RiskTaking_{ist} =& \beta_0 + \phi_1 \text{UD Law } (0)_{st} + \phi_2 \text{UD Law } (1)_{st} + \phi_3 \text{UD Law } (2+)_{st} \\ &+ \gamma \text{Cash Ratio} + \sum_j \theta_j Controls_{jist} + \alpha_i + \theta_{st} + \epsilon_{ist} \end{aligned} \tag{1.5}$$

The different structures for the economic channels would lead to different results for model (5) when compared to model (3). When an independent variable affects a dependent variable, there should be some channels for the effect. The total effect is the sum of the effect of each channel if the channels are independent of each other. EC is used to denote the effect of an economic channel. In this study, I have the following equation:

$$EC_{total} = EC_{cf} + EC_{myopia} + EC_{others}, \tag{1.6}$$

where EC_{total} is the total effect of UD law adoption on risk-taking behavior, EC_{cf} is the effect of the cash flow channel, EC_{myopia} is the effect of the managerial myopia channel, and EC_{others} is the effect of all other possible channels. The coefficients of UD Law Dummy variables of model (3) correspond to the total effect of UD law adoption, that is, EC_{total}. Moreover, the coefficients of the UD Law Dummy variables of model (5) correspond to the sum of EC_{myopia} and EC_{others} because the model controls for the cash flow ratio. More specifically, after controlling for the cash flow ratio, the effect of UD law adoption in model (5) cannot come from the cash flow ratio. Instead, the sources of the total effect of UD law adoption can only contain managerial myopia and other possible channels.

If Figure 1.1 illustrates an appropriate economic channel structure, EC_{cf} is not 0 since the free cash flow problem is an economic channel. Therefore, the following relationship can be observed:

$$EC_{total} > EC_{myopia} + EC_{others} \tag{1.7}$$

This implies that the results of model (3) would have larger coefficients for the UD Law Dummy variables than those in model (5). If Figure 1.2 is an appropriate model, EC_{cf} is 0 since the figure assumes that the free cash flow problem is not an economic channel. Therefore, the following relationship can be observed:

$$EC_{total} = EC_{myopia} + EC_{others} \tag{1.8}$$

In other words, the coefficients of the UD law dummies of model (3) are the same as the coefficients of UD Law Dummy variables compared to the results of model (5).

Table 1.13 shows that when Cash Ratio is controlled, the coefficients of the UD Law Dummy variables in model (5) do not change significantly, compared to corresponding coefficients in model (3). For Leverage Risk, the coefficients of UD Law Dummy (0) and UD Law Dummy (1) in both models are insignificant. The coefficients of UD Law Dummy (2+) are both statistically significant, which is 0.0105 in model (3) and 0.0108 in model (5). Based on 2-sample t-statistics, the difference is not statistically significant. The same conclusion can be drawn for Underwriting Risk for which the coefficient of UD Law Dummy (2+) in model (3) is 0.0211, while that in model (5) is 0.0230. Moreover, for Asset Risk, the coefficients for UD Law

Dummy (1) and UD Law Dummy (2+) in model (3) and model (5) show no statistical difference.

The results in Table 1.13 support the structure for the economic channel described in Figure 1.2, where the free cash flow problem is not the main economic channel. In other words, "managerial myopia" may be a more important economic channel for the effect of UD law adoption on the risk-taking behaviors of insurance companies.

Table 1.13.: Determine Economic Channels

	(1)	(2)	(3)
Independent Variables	**Leverage Risk I**	**Asset Risk I**	**Underwriting Risk**
UD Law Dummy (0)	-0.0022	0.0073	0.0023
	(0.783)	(0.273)	(0.850)
UD Law Dummy (1)	0.0045	0.0124*	0.0071
	(0.558)	(0.058)	(0.555)
UD Law Dummy (2+)	0.0108**	0.0104***	0.0230***
	(0.011)	(0.005)	(0.000)
Log Total Asset	0.0564***	0.0106***	0.0132***
	(0.000)	(0.000)	(0.000)
Log Age	-0.0134***	0.0421***	-0.0281***
	(0.000)	(0.000)	(0.000)
Geo Mix	0.0298***	0.0411***	-0.1765***
	(0.000)	(0.000)	(0.000)
Busi Mix	-0.0904***	-0.0027	0.2440***
	(0.000)	(0.356)	(0.000)
Cash Ratio	0.0088	-0.0826***	0.0574***
	(0.325)	(0.000)	(0.000)
Observations	35,574	35,574	35,574
within R-sq	0.250	0.112	0.113

Note: This table presents the results of OLS regressions that determine the economic channels through which UD law adoption affects risk-taking behavior. There are three time horizons around the adoption of UD laws. The sample consists of risk-taking measures and firm characteristics in property-casualty insurance firms during 1988–2010. I estimate pooled OLS regression in model (3) with the cash ratio variable as an extra control. The dependent variables are corporate risk-taking measures. There are three risk-taking measures: Leverage Risk I, Asset Risk I and Underwriting Risk. The results with state and year fixed effects are presented in columns (1)–(3), respectively. Variable definitions are given in Table 1.4. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

1.8.3 Additional Control Variables

In hypothesis development, I argue that mutual insurers tend to respond to UD law adoptions slower than stock insurers due to more net premium written in long-tail lines of business such as homeowners and automobile, which impedes quick adjustment of the liability side and asset side. In this section, a new variable, Long-Tail, is added to model (4) to explore whether insurers respond to UD law adoption differently when long-tail businesses are controlled.

Compared to the previous result in model (4), the new result in Table 1.14 shows no significant difference for all three risk-taking measures. The conclusion remains the same. It implies that long-tail lines of business may not be the factor to explain the differential speed of adjustments between mutual and stock insurers.

1.8.4 Additional Time Horizons

To investigate the validity of using UD law adoption as an exogenous shock, I introduce two more variables that reveal the effect of UD laws in earlier time periods, that is, UD Law Dummy (-1) and UD Law Dummy (-2) which corresponds to one year before UD law adoption and two years before UD law adoption, respectively. When all states are included, the coefficients of UD Dummy (-1) for Asset Risk and UD Dummy (-2) for Underwriting Risk are slightly significant, which might imply some pre-treatment effect. On the other hand, the significant coefficients could also indicate the ripple effect of earlier UD law adopted states on later states, as explained in Section 1.3.2. Then, I include states that adopted UD laws before 1996 and re-run the model. Table 1.15 shows the result, where no significant coefficients are found for UD Law Dummy (-1) and UD Law Dummy (-2). The result supports the ripple effect argument and validates the exogeneity of UD law for earlier time.

Table 1.14.: UD Law, Risk-Taking, and Organizational Structure in Detail

Independent Variables	(1) Leverage Risk I	(2) Asset Risk I	(3) Underwriting Risk
UD Law Dummy (0)	0.0048	0.0102	-0.0034
	(0.569)	(0.152)	(0.793)
UD Law Dummy (1)	0.0130	0.0141**	0.0019
	(0.123)	(0.050)	(0.885)
UD Law Dummy (2+)	0.0203***	0.0063*	0.0178***
	(0.000)	(0.083)	(0.007)
Mutual	0.1486***	0.0287	0.5320***
	(0.000)	(0.104)	(0.000)
UD Law Dummy (0) × *Mutual*	-0.0289*	-0.0115	0.0112
	(0.053)	(0.365)	(0.627)
UD Law Dummy (1) × *Mutual*	-0.0333**	-0.0060	0.0041
	(0.029)	(0.642)	(0.861)
UD Law Dummy (2+) × *Mutual*	-0.0488***	0.0260***	0.0116
	(0.000)	(0.000)	(0.176)
Log Total Asset	0.0484***	0.0163***	0.0118***
	(0.000)	(0.000)	(0.000)
Log Age	0.0019	0.0265***	-0.0074***
	(0.127)	(0.000)	(0.000)
Geo Mix	0.0195***	0.0264***	-0.1331***
	(0.000)	(0.000)	(0.000)
Busi Mix	-0.0875***	0.0043	0.1825***
	(0.000)	(0.185)	(0.000)
Log Total Asset × *Mutual*	-0.0048***	-0.0036***	-0.0053***
	(0.000)	(0.000)	(0.000)
Log Age × *Mutual*	-0.0305***	0.0238***	-0.0947***
	(0.000)	(0.000)	(0.000)
Geo Mix × *Mutual*	-0.0350***	-0.0163***	-0.1124***
	(0.000)	(0.009)	(0.000)
Busi Mix × *Mutual*	-0.0057	-0.0263***	0.2108***
	(0.525)	(0.001)	(0.000)
Long-Tail	0.1080***	-0.0217***	-0.0579***
	(0.000)	(0.000)	(0.000)
Long-Tail × *Mutual*	0.0991***	0.0019	-0.2394***
	(0.000)	(0.783)	(0.000)
Observations	35,570	35,570	35,570
State Fixed Effect	YES	YES	YES
Year Fixed Effect	YES	YES	YES
within R-sq	0.290	0.128	0.156

Note: This table presents the results of the pooled OLS regression estimated in model (4) with the long-tail variable as an extra control. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 1.15.: UD Law and Risk-Taking in Additional Time Horizons

Independent Variables	(1) Leverage Risk I	(2) Asset Risk I	(3) Underwriting Risk
UD Law Dummy (-1)	-0.0020	0.0234	-0.0037
	(0.915)	(0.133)	(0.888)
UD Law Dummy (-2)	-0.0197	0.0119	0.0299
	(0.312)	(0.455)	(0.284)
UD Law Dummy (0)	-0.0028	0.0278	0.0086
	(0.891)	(0.116)	(0.770)
UD Law Dummy (1)	0.0066	0.0400**	0.0132
	(0.772)	(0.040)	(0.687)
UD Law Dummy (2+)	-0.0006	0.0492**	0.0031
	(0.981)	(0.023)	(0.931)
Log Total Asset	0.0602***	-0.0062***	-0.0026
	(0.000)	(0.000)	(0.359)
Log Age	-0.0170***	0.0458***	-0.0424***
	(0.000)	(0.000)	(0.000)
Geo Mix	0.0507***	0.0062	-0.2792***
	(0.000)	(0.268)	(0.000)
Busi Mix	-0.1139***	-0.0133**	0.1927***
	(0.000)	(0.041)	(0.000)
Observations	5673	5673	5673
State Fixed Effect	YES	YES	YES
Year Fixed Effect	YES	YES	YES
within R-sq	0.271	0.110	0.131

Note: This table presents the results of OLS regressions that estimate the relationship between the adoption of UD laws and corporate risk-taking in different time horizons. There are five time horizons around the adoption of UD laws. The sample consists of risk-taking measures and firm characteristics in property-casualty insurance firms for states that adopted UD laws before 1996. I estimate pooled OLS regression in model (3) with extra time horizons. The dependent variables are corporate risk-taking measures. There are three risk-taking measures: Leverage Risk, Asset Risk and Underwriting Risk. The results with year and state fixed effects are presented in columns (1)–(3), respectively. Variable definitions are given in Table 1.4. p-values are reported in parentheses under the corresponding estimated coefficients. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

1.9 Conclusion

I examined the effect of shareholder derivative lawsuits on corporate risk-taking behavior for both stock and mutual insurers in the property-casualty insurance industry using UD law adoption as an exogenous shock. The different organizational structures provide a great opportunity to examine corporate risk-taking behavior since conflicts within stock insurance companies and mutual companies result in distinct managers' objectives, implying possible distinct responses to UD law adoption. Using NAIC data and UD law adoption at the state level, I constructed three risk-taking measures, Leverage Risk, Asset Risk, and Underwriting Risk.

I found that both stock and mutual insurers increased their Asset Risk after UD law adoption. Also, mutual insurers needed more time than stock insurers to respond to UD law adoption for Asset Risk. The robustness check finds that restricted access to capital is the main reason to explain the differential speed of adjustments. Moreover, only stock insurers increase Underwriting Risk where no significant change is observed for mutual insurers. One explanation could be that it is more difficult for policyholders to question managers' decisions about the core business since it requires more expertise and information. As policyholders are more atomic than shareholders in stock insurers, it is harder for them to acquire necessary business information. Therefore, managers have an advantage in this situation due to the business judgement rule. For Leverage Risk, stock and mutual insurers undertake opposite adjustments. It could be explained by the differential overall risk level between stock and mutual insurers. That is, mutual insurers may decrease Leverage Risk to offset the effect of increased Asset Risk to maintain a low overall risk level. In the robustness check, after controlling for free cash holdings, I found that the free cash flow argument is not the main channel by which UD law adoption affects risk-taking behavior.In other words, managerial myopia is a more important channel. lastly, to examine the exogeneity of UD law adoption, two more time horizons are introduced in model 3. The result shows that at least for states that adopted UD laws in the early years, there is no evidence for endogeneity issue.

CHAPTER 2
AUTO INSURANCE PRICING USING TELEMATICS DATA: APPLICATION OF A HIDDEN MARKOV MODEL

2.1 Introduction

Risk classification is one of the most important topics in insurance pricing and risk management. In auto insurance, policyholders' self-reported information (e.g., age, gender, driving record, zip code, education) and vehicle characteristics (e.g., car make and model) are traditionally used as rating factors to identify the riskiness of an insured. While the approach has proven to be effective, these traditional rating factors at best serve as static proxy measurements for actual risk (Henckaerts and Antonio, 2022). Consequently, insurers have been facing premium and/or revenue loss due to misclassification (e.g., caused from omitted or misstated underwriting information). According to Verisk Insurance Solutions (2016), personal line automobile insurers face at least $29 billion in premium leakage annually, for which unrecognized drivers ($10.3 billion) and underestimated mileage ($5.4 billion) are the two largest sources.

Fortunately, in the era of digitalization and big data, telematics and related technology advancements have enabled insurers to develop novel risk classification methods that could significantly mitigate premium leakage, directly linking insurance losses with driving behavior and vehicle usage. Telematics data are the computerized information collected from devices and can be transmitted over long distances; these data record driving information, including GPS coordinates, acceleration in each direction, speed, and many others at high frequency. Auto insurances that rely on the use of telematics data for risk classification and pricing are generally called Usage-Based Insurance (UBI), with multiple variations such as Pay-As-You-Drive (PAYD), Pay-

How-You-Drive (PHYD), and Pay-As-You-Go (PAYG) (NAIC, 2021). The global UBI market has been growing steadily, valued at $18.9 billion in 2021 and projected to reach $78.9 billion by 2028 at a compound annual growth rate of 26.9% (Vantage Market Research, 2022). In the U.S., UBI insurances are not only offered by major personal line players (e.g., Progressive's Snapshot, Travelers' IntelliDrive) but also by emerging insurtech companies, such as Metromile and Root. For example, Metromile, located in San Francisco and founded in 2011, sells pay-per-mile car insurance as its only product and collects telematics data using its own onboard diagnostic (OBD) devices, namely Metromile Pulse.

There is a growing body of literature on extracting useful information from telematics data for risk assessment and management. Earlier work has mainly focused on the risk exposures of duration and driving distance under different categories (e.g., speed intervals, road types, daytime vs. nighttime) and their predictive power for accident rates (Paefgen et al., 2014; Ayuso et al., 2014, 2016; Boucher et al., 2017). Using the generalized additive models (GAM) with compositional predictors, Verbelen et al. (2018) find that descriptive statistics of telematic variables increase predictive power and render gender a redundant rating variable. Except for driving habits, another set of studies have examined various indicators for driving behavior, such as speeding, hard braking, accelerations, and sudden turns. Huang and Meng (2019) incorporate these variables in claim frequency models using logistic regression and machine learning techniques. So et al. (2021) introduce a cost-sensitive multi-class adaptive boosting (AdaBoost) algorithm to deal with the accident imbalance problem in predicting claim frequencies based on driving behavior variables. Guillen et al. (2020, 2021) study driving behavior in near-miss events and their economic implications. Besides, telematics covariates extracted from speed-acceleration heatmaps are used for classifying driving styles and predicting claims (Wüthrich, 2017; Gao and Wüthrich, 2018; Gao et al., 2019, 2022). It is worth noting that roughly three months or 4,000 km of observation are sufficient to obtain stable telematics features (Duval et al., 2022), and hence, incumbent insurers may easily incorporate telematics insights into existing pricing structures using updating mechanisms (e.g., Denuit et al., 2019; Henckaerts and Antonio, 2022). Overall, telematics technology not only provides better predictive power than traditional rating methods but also helps mitigate moral hazard and adverse selection (Filipova-Neumann and Welzel, 2010), change driving patterns and behavior (Parry, 2005; Reimers and Shiller, 2019; Soleymanian et al., 2019; Guillen et al., 2021), and enhance UBI insurers' market share and overall financial performance (Che et al., 2021).

Along with the aforementioned literature, another stream of studies directly explores telematics information at the trip level. Weidner et al. (2016) extract covariates from time series of telematics data using discrete Fourier transforms. Gao and Wüthrich (2019) propose to allocate individual car driving trips to selected drivers using convolutional neural networks (CNNs) based on feature information from high-frequency GPS location data. Similarly, Meng et al. (2022) propose a supervised driving risk scoring neural network model to identify risky or safe drivers. Typically, dealing with trip-level telematics data requires new statistical and machine learning approaches that can interpret raw features into new representations for predictive modeling (Gao et al., 2022). Readers are also referred to Blier-Wong et al. (2021) for a review of machine learning applications in P&C insurances. While machine learning models can automate various decision making processes and are powerful in identifying risk patterns and trends, their complexity and lacking of interpretability may limit their usage in practice under the current regulatory environment (Virgilis et al., 2022).

In this paper, we propose a hidden-Markov-model (HMM)-based clustering method for auto insurance pricing. In this setting, driving behavior is inferred by various measurable factors from telematics data, and a hidden state is assigned to describe driving status at a given moment. The riskiness of a driver can be then quantified based on the transitions and distributions of hidden states. There is a branch of IEEE studies on the applications of HMM for driving behavior recognition and prediction, including but not limited to aggressive driving (Derbel and Landry, 2015), fatigued driving (Qin et al., 2012; Craye et al., 2016), driving skills (Deng et al., 2018; Liu et al., 2020), and driving behavior in different environments (highway vs. local) (Sathyanarayana et al., 2008; Amsalu and Homaifar, 2016; Deng and Söffker, 2019; Deng et al., 2020). Compared with other algorithms, such as neural networks and Gaussian mixed models, HMM demonstrates high accuracy and outstanding performance in predicting real-time driving behavior through time series data (Deng and Söffker, 2022). However, as far as we know, no existing papers have examined the application of the powerful HMM tool in auto insurance pricing.

To fill this gap, we develop an HMM framework to predict driving status and its associated probability of insurance losses. Specifically, we take telematics information for each trip as input to find the implied driving status (hidden states) using HMM. We then cluster driving trips based on the similarities of their hidden state distributions and transitions and use Poisson GAM models to predict loss frequencies by cluster. Our results show that the proposed HMM approach effectively captures driving characteristics and outperforms the examined purely driving-behavior-based classification models in predicting accident rates. There are several advantages to use the proposed HMM framework. First, it allows insurers to reveal the dynamics of driving behavior and determine rates by weighing driving status (hidden states). In the meantime, the stationary states of the HMM may also be used to obtain a long-term loss prediction, which reduces the sensitivity of individual trips. This is especially useful for pricing new policies with limited data. Second, HMM can be flexibly applied to any time series data with different lengths, which is particularly convenient for investigating individual trips. Many machine learning models, such as CNN and long short term memory, require fixed lengths of input, although some techniques (e.g., padding, truncating, or splitting) exist that can resolve the problem with some costs. Third, HMM is a well-visualized graphical model with great interpretability. It provides a clear model structure that associates risks with prices and could be easier for regulatory rate filing.

The rest of the paper is organized as follows. The general HMM framework and associated clustering and prediction methods are discussed in Section 2.2. Section 2.3 introduces the data and simulation. Section 2.4 compares several HMM-related models and illustrates how the proposed models can be used for auto insurance pricing. Section 2.5 concludes the paper.

2.2 Methodology

In this section, we propose an HMM approach that extracts driving behavior information from telematics data for accident prediction. Under this framework, drivers' riskiness is inferred through hidden states. The distribution of hidden states is then used for clustering trips into different risk classes through the so-called K-means algorithm (Hastie et al., 2009). The resulting classifications can be incorporated into existing predictive models, such as the Poisson GAM model, for auto insurance pricing.

2.2.1 Hidden Markov Model

HMMs are Bayesian networks proposed by Baum and Petrie (1966) and have been applied in numerous fields, such as speech recognition, pattern recognition, and time series clustering (e.g., Rabiner and Juang, 1986; Smyth, 1996; Ghahramani, 2001). There are two components in an HMM: a Markov sequence of hidden states Z_t (i.e., not directly observable) and observations X_t (also known as emission) that only depend on the hidden state at time t. That is,

$$\Pr\left[Z_t \,|Z_1, Z_2, ..., Z_{t-1}\right] = \Pr\left[Z_t \,|Z_{t-1}\right], \tag{2.1}$$

$$\Pr\left[X_t \,|Z_1, Z_2, ..., Z_t, X_1, X_2, ..., X_{t-1}\right] = \Pr\left[X_t \,|Z_t\right], \tag{2.2}$$

for $t = 1, 2, \cdots$. Thus, the joint distribution of the sequence of hidden states and observations can be written as

$$\Pr\left[Z_1, Z_2, ..., Z_t, X_1, X_2, ..., X_t\right] = \Pr\left[Z_1\right] \prod_{k=2}^{t} \Pr\left[Z_k \,|Z_{k-1}\right] \prod_{k=1}^{t} \Pr\left[X_k \,|Z_k\right]. \tag{2.3}$$

Figure 2.1 graphically shows how observations and hidden states are generated over time. The HMM is fully specified by the following parameter space $\{\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}\}$:

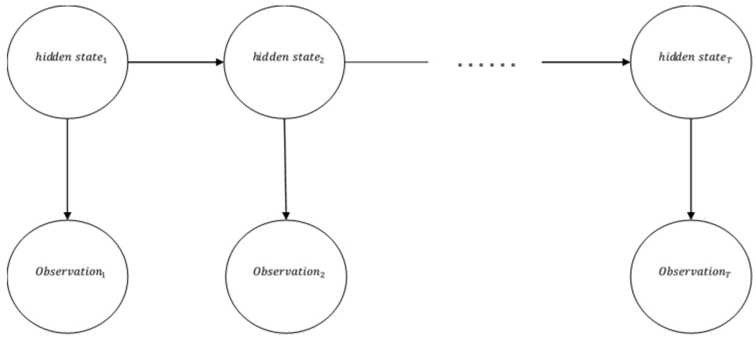

Figure 2.1.: Graphic Illustration of Hidden Markov Model

- Initial state distribution $\boldsymbol{\pi}$: we denote $\pi_i = \Pr(Z_1 = i)$ as the probability of being in hidden state i at time 1 for $i = 1, 2, ..., n$, where n is the total number of hidden states.

- State transition model $\mathbf{A}$: $a_{ij} = \Pr(Z_t = j | Z_{t-1} = i)$, for $i, j = 1, 2, ..., n$, denotes the time-independent transition probability of hidden state from i to j between time $t - 1$ and t.

- Observation (emission) model $\mathbf{B}$: the probability of making observation X_t given state Z_t is denoted as $b_{kj} = \Pr(X_t = k | Z_t = j)$, for $j = 1, 2, .., n$ and $k = 1, 2, ..., m$, where m is the number of distinct observations.

In our application, we assume that the risk status of a driver is not directly observable but follows a Markov process. However, various measures describing driving behavior are observable, such as speed, hard braking (negative acceleration), big vehicle angle changes, and trip duration. For example, high speed will give drivers less time to react to accidents, hard braking can be an indicator of following too closely or aggressive driving, big angle change can be associated with emergency corrections or drowsy driving (e.g., McDonald et al., 2012), and long time spent driving may cause driver fatigue and affect driving behavior. We illustrate how drivers' risk status is probabilistically inferred by driving behavior variables through the associated hidden states under the HMM framework.

To fit a discrete HMM with telematics data, we first convert the numerical values of driving behavior variables per observation into categorical values. Specifically, we divide speed into four categories based on the speed limits of different road types: less than 35 mph, between 35 and 50, between 50 and 80, and higher than 80. Vehicle angle changes are divided into two categories based on a cutoff of 7.5 degrees per second, or equivalent to 0.131 radian per second, which is approximately the 90th percentile in the data. Accelerations are divided into two categories based on -7 mph per second as in Desai et al. (2021). A reduction in speed greater than 7 mph per second is often considered hard braking.[1] Finally, trip duration is divided into two categories, within or over two hours, to capture potential long-time driving effects. As a result, there are $m = 4 \times 2 \times 2 \times 2 = 32$ distinct driving behavior observations. Each observation X_t is a tuple of the categorical variables defined above, i.e.

$$X_t = [speed_i,\ hard\_braking_j,\ big\_angle\_change_s,\ time\_duration_r]. \tag{2.4}$$

The parameters $\{\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}\}$ can be estimated using the Expectation-Maximization (EM) algorithm, which is also known as the Baum-Welch algorithm (Baum et al., 1970). While the emission model given hidden states can follow any distributions, we adopt the multinomial distribution, which is a natural choice for discrete multi-variate observations. The detailed EM algorithm for estimating HMM can be found in, for example, Ghahramani (2001), and the Python package *hmmlearn* is used for model fitting. It is worth noting that the EM algorithm does not estimate the number of hidden states in an HMM, which should be specified beforehand. To find the optimal number of hidden states, we perform parameter estimation for all cases when the number of states varies from 2 to 9 and select the number of states with the minimum AIC or BIC score. After obtaining the model parameters, the optimal sequence of hidden states can then be assigned to the observations in each trip and further used for clustering and risk classification.

[1]For example, this threshold is adopted by Progressive's Snapshot. https://www.progressive.com/auto/discounts/snapshot/snapshot-details/

2.2.2 K-means Algorithm

Different from classification by drivers, the HMM framework enables us to cluster observations at the more granular trip level. Specifically, the K-means algorithm (Hastie et al., 2009) is applied to classify individual trips into K groups using the distribution of hidden states for each trip. The K-means algorithm is a very popular unsupervised clustering method that divides observations based on dissimilarity, which is often measured by the Euclidean distance between observation vectors. It has been used by researchers to classify car drivers in the actuarial literature (Wüthrich, 2017; Gao et al., 2019).

Despite its popularity, the K-means algorithm has some limitations (Ahmed et al., 2020). First, the random initialization of centroids (i.e., the center of each cluster) may lead to unexpected convergence and hence impact the clustering results. To overcome this drawback, we use multiple random initializations and select the most popular centroids. Over 90% of the time, the same group of centroids are selected. Second, the number of clusters needs to be specified in advance. Following Gao et al. (2019), we determine the optimal number of clusters based on the performance of predictive models that incorporate the resulting clusters.[2] Specifically, the clustering results based on different numbers of clusters are tested in the Poisson GAM models (to be introduced later) for car accident prediction. We use the deviance statistics of Poisson distribution to assess relative performance. During this process, stratified k-fold cross-validation is also used to test performance robustness.

The clusters identified by the K-means algorithm are considered as our risk classes with varying driving behavior. It is important to point out that the risk classes based on the distribution of hidden states are different from those based on aggregate driving behavior information, such as the heatmap (Wüthrich, 2017; Gao et al., 2019). One advantage of using hidden-state-based clustering is that it not only reflects the heterogeneity of overall driving behavior but also captures the transition of driving

[2]There are multiple ways to select the optimal number of clusters. For example, the elbow method is a popular one, in which the explained variation is plotted as a function of the number of clusters, and the elbow of the curve is picked as the optimal number of clusters (e.g., Ketchen and Shook, 1996; Bholowalia and Kumar, 2014).

behavior. Consider a simple example where there are two different trips with the same speed distribution, e.g., 50% of the time at 20 mph and 50% of the time at 80 mph, as shown in Figure 2.2. We can see that the speed of the first trip changes much more frequently and dramatically than that of the second trip. Such information is lost in the aggregate speed distribution but can be identified by the hidden-state approach as speed changes are properly captured by the transitions of hidden states.

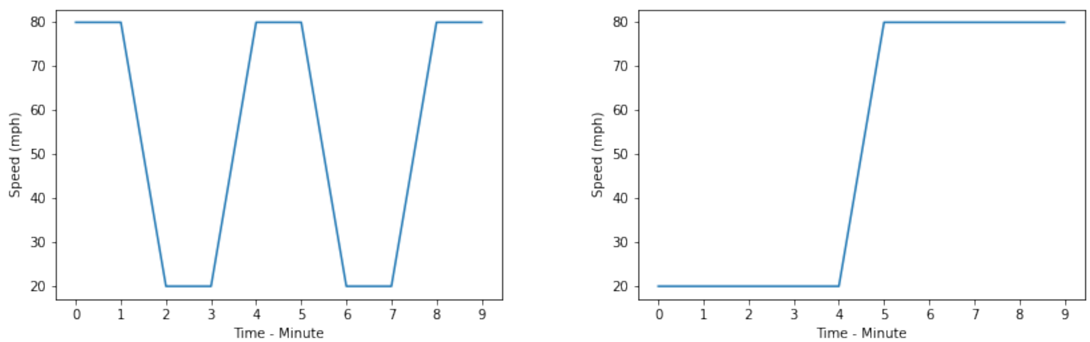

Figure 2.2.: Different Driving Behaviors with the Same Speed Distribution

2.2.3 Accident Prediction

In this subsection, we propose some Poisson GAM models that incorporate HMM-based classifications. The Poisson GAM model has been widely applied along with telematics data to predict the number of accidents for auto insurances (e.g., Boucher et al., 2017; Verbelen et al., 2018; Gao et al., 2019). It can incorporates exposure naturally so that no covariates on trip duration are needed. The general form of a Poisson GAM model based on driving behavior can be defined as

$$y_i \sim \text{Poisson}(\lambda_i \cdot e_i) \qquad \text{with}$$
$$\ln(\lambda_i) = \beta_0 + \beta_1 \cdot DB_1 + ... + \beta_n \cdot DB_n + \theta \cdot cluster, \tag{2.5}$$

where y_i is the number of accident occurrences, λ_i is the Poisson parameter, e_i is the exposure for the ith unit (e.g., trip or driver), DB_j ($j = 1, 2, ..., n$) stands for the information vector of the jth driving behavior (e.g., speed, big angle change, hard braking), β_j is the coefficients vector for the corresponding driving behavior variable, *cluster* represents the risk class that the ith unit belongs to, and θ is the coefficient for the risk class. Typically, risk classes are classified based on certain driving behavior characteristics.

In this paper, we focus on speed, big angle change, and hard braking. Thus, Equation (2.5) is specified at the trip level as follows:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i \cdot e_i) \qquad \text{with} \\ \ln(\lambda_i) &= \beta_0 + \beta_1 \cdot speed + \beta_2 \cdot hard\_braking + \beta_3 \cdot big\_angle\_change \\ &\quad + \beta_4 \cdot hard\_braking \cdot big\_angle\_change + \theta \cdot cluster, \end{aligned} \tag{2.6}$$

where β_4 aims to capture the interdependence between hard braking and big angle change. To compare the performance of the HMM approach with existing classification methods, we first regress (2.6) using the driving-behavior-based clusters (Model 1). To be more Specific, we calculate the frequency distributions of three driving behavior variables for each trip as input for trip classification using the K-means algorithm. Our Model 2 is the same as Model 1 except that the hidden-state-based clusters are used in Equation (2.6).

In addition to clustering trips, hidden states may also be considered as a natural classification of driving status/environment when we further explore each observation of a trip. Therefore, we also examine Model 3 replacing clusters by hidden states:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i \cdot e_i) \qquad \text{with} \\ \ln(\lambda_i) &= \beta_0 + \beta_1 \cdot speed + \beta_2 \cdot hard\_braking + \beta_3 \cdot big\_angle\_change \\ &\quad + \beta_4 \cdot hard\_braking \cdot big\_angle\_change + \gamma \cdot hidden\_state, \end{aligned} \tag{2.7}$$

where γ is the coefficients vector for hidden states. In this model, we group observations based on hidden states and estimate the average Poisson parameter per unit time for each hidden state j.

Next, we incorporate both hidden states and hidden-state-based clusters into analysis, which is our Model 4:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i \cdot e_i) \qquad \text{with} \\ \ln(\lambda_i) &= \beta_0 + \beta_1 \cdot speed + \beta_2 \cdot hard\_braking + \beta_3 \cdot big\_angle\_change \\ &\quad + \beta_4 \cdot hard\_braking \cdot big\_angle\_change \\ &\quad + \theta \cdot cluster + \gamma \cdot hidden\_state. \end{aligned} \tag{2.8}$$

In this model, we aim to obtain the average Poisson parameter λ'_{jk} for each state j in each cluster k.

Finally, a model that incorporates no risk class and hidden state is included as a benchmark for model performance comparison, called Model 5:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i \cdot e_i) \qquad \text{with} \\ \ln(\lambda_i) &= \beta_0 + \beta_1 \cdot speed + \beta_2 \cdot hard\_braking + \beta_3 \cdot big\_angle\_change \\ &\quad + \beta_4 \cdot hard\_braking \cdot big\_angle\_change. \end{aligned} \tag{2.9}$$

Two different metrics are used to evaluate the performance of the aforementioned models. The first one is deviance statistics for Poisson distribution (e.g., Gao et al., 2019):

$$D = 2\sum_{i=1}^{N}\{Y_i log(\frac{Y_i}{\lambda_i e_i}) - (Y_i - \lambda_i e_i)\}, \tag{2.10}$$

where N is the number of trips and Y_i is the actual number of accidents for the ith trip. Lower values of D indicate relatively better performance. Second, we aggregate the predicted number of accidents from trips and calculate the mean squared error (MSE) of accidents by device. It measures how accurately a model predicts the number of accidents at the device level.

2.2.4 Limiting Behavior and Insurance Pricing

The proposed Model 4 also allows us to calculate a long-term accident probability for each risk class per time unit. This can be done by examining the limiting behavior of the HMM.

First, for each cluster k, we calculate the sample transition matrix $\hat{\mathbf{A}}$ with elements

$$\hat{a}_{ij} = \frac{\sum_{\text{trips in cluster } k} \sum_t I_{\{Z_t=j, Z_{t-1}=i\}}}{\sum_{\text{trips in cluster } k} \sum_t I_{\{Z_{t-1}=i\}}}. \tag{2.11}$$

The sample transition matrix is further used to determine the stationary distribution $\boldsymbol{\pi}^L = \left(p_1^L, ..., p_n^L\right)^\top$ of the HMM (if it exists), i.e.,

$$\boldsymbol{\pi}^L \cdot \hat{\mathbf{A}} = \boldsymbol{\pi}^L. \tag{2.12}$$

The long-term expected Poisson parameter λ'_{*k} for the kth cluster can then be expressed as

$$\lambda'_{*k} = \sum_{j=1}^{n} \lambda'_{jk} \cdot p_j^L. \tag{2.13}$$

Finally, the average Poisson parameter for the total duration λ_{total} satisfies

$$\lambda_{total} = \sum_k T_k \cdot \lambda'_{*k}, \tag{2.14}$$

where T_k is the trip duration of a driver in the kth cluster. The accident probability predicted by (2.14) reflects the cost of insurance if a driver keeps the same driving pattern for a long period of time, and it can be adjusted gradually as more telematics data are collected.

2.3 Data and Simulation

2.3.1 Data and HMM Estimation

Our telematics data comprise a proprietary data set from Verisk. Driving behavior information, such as GPS locations, speed, and average accelerometer in each direction, was recorded every second and collected by OBD devices. There are 12 variables in total with definitions provided in Table 2.1. The data contain 10,232,749 observations from 79 OBD devices between 05/14/2019 and 07/09/2019. The number of observations for each device varies from 9,947 to 318,147 with an average of 129,528. According to the GPS information, the records were scattered in different states of the U.S., including California, Wisconsin, and Ohio. The overall distribution of GPS locations is shown in Figure 2.3.

Table 2.1.: Variable Definitions of the Telematics Data

Variable Name	Variable Definition
GD_ID	ID for each observation
GD_DEVICE_ID	ID for each OBD device
GD_INGESTION_DATE	Timestamp when each observation is transmitted
GD_RAW_TYPE	Type of data
GD_DATE	Timestamp when each observation is recorded
GD_FUEL_LEVEL	Fuel level at each moment
GD_LATITUDE	Latitude at the moment
GD_LONGITUDE	Longitude at the moment
GD_SPEED	Speed at the moment
GD_AVG_X	Average accelerometer at x-direction
GD_AVG_Y	Average accelerometer at y-direction
GD_AVG_Z	Average accelerometer at z-direction

The raw data are pre-processed in several steps. First, trips are identified based on GD_DATE. While each device stores the telematics data every second, it transmits data every five minutes. Thus, we split records into trips whenever the gap between two consecutive observations is more than five minutes. Second, to capture the relevant driving behavior, we remove records with consecutive zero speed at the beginning and end of each trip. If consecutive zero speed in the middle of trip lasts more than two minutes, we further split the trip into smaller trips. Third, we remove any duplicated records and estimate missing records using linear interpolation. Finally, trips with a duration of fewer than five minutes are removed.

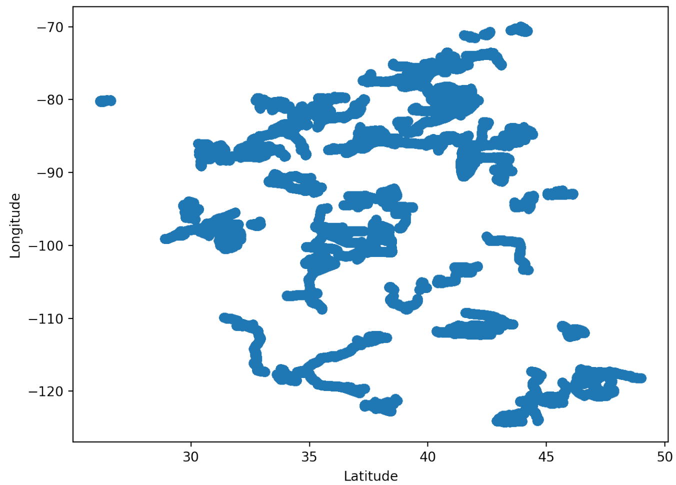

Figure 2.3.: The Geographic Distribution of Trips

After data cleaning, we extract the relevant variables for our analysis. The speed information and the time duration of each trip can be directly obtained from the data. The acceleration for each second can be calculated as the difference in speed between the current and previous observations. The vehicle angle change θ is calculated as the angle between two consecutive accelerometer vectors, i.e.,

$$\theta_t = \arccos \frac{\vec{v}_t \cdot \vec{v}_{t-1}}{|\vec{v}_t||\vec{v}_{t-1}|}, \tag{2.15}$$

where $\vec{v}_t$ is the accelerometer vector at time t with three elements (i.e., GD_AVG_X, GD_AVG_Y, and GD_AVG_Z). The summary statistics of these variables are reported in Table 2.2, and their overall distributions are shown in Figure 2.4.

There is a large number of records in the low speed range from 0 to 10 mph, which might be associated with events such as traffic, parking, or waiting at traffic lights. We also observe that many drivers keep their speed within 5 mph or 10 mph above the speed limit (see the local peaks in Figure 2.4). The acceleration distribution is relatively symmetric with most observations staying around 0. Vehicle angle changes are generally within 20 degrees or equivalent to 0.349 in radian. Sudden large angle changes are likely for situations such as parking, sudden turns, and car accidents.

Table 2.2.: Summary Statistics of Key Variables

				Percentile			
Variable	**Mean**	**Min**	**Max**	**25th**	**50th**	**75th**	**99th**
Speed (mph)	39.74	0.00	103.75	16.88	41.25	62.50	81.88
Acceleration (mph/s)	0.00	-16.25	12.5	-0.63	0.00	0.63	3.75
Angle change (degrees/s)	3.67	0.00	65.37	1.78	3.09	4.93	12.26
Duration (mins)	22.62	5	293.30	8.67	14.48	26.80	112.20

Lastly, the majority of the trips in the data are within 50 minutes, with the average being 22.62 minutes.

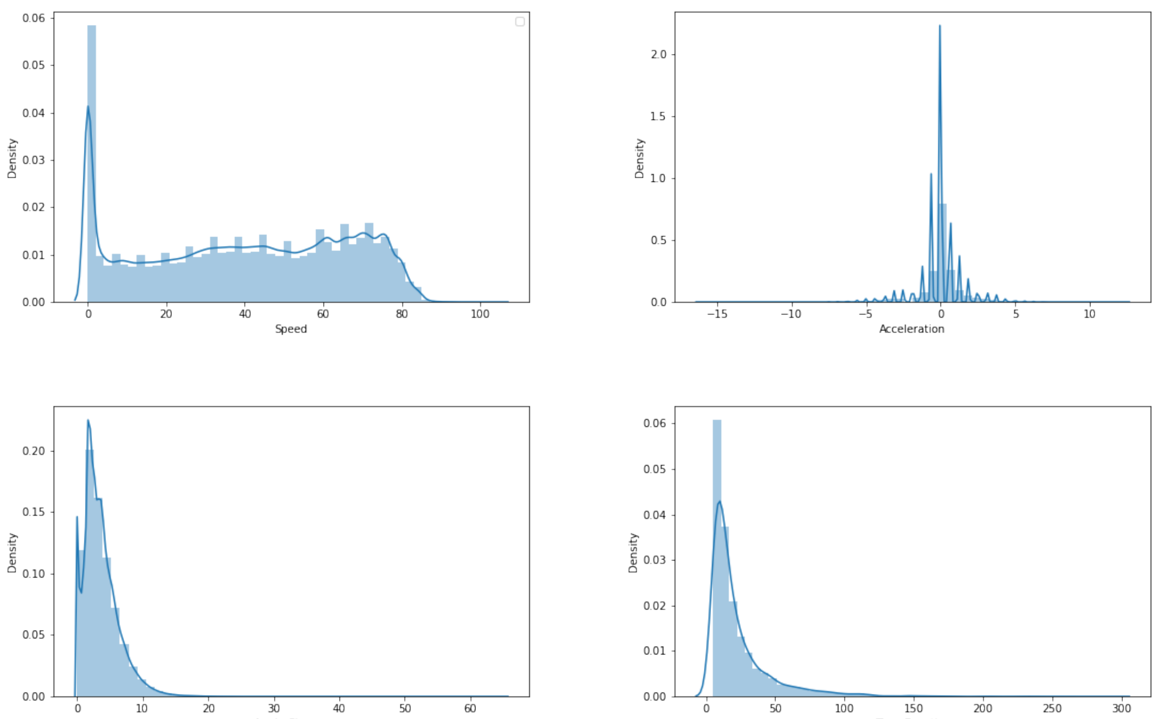

Figure 2.4.: Distributions of Key Variables

Using the categorical variables (2.4) as input, we fit the HMM through the EM algorithm when the number of hidden states changes from 2 to 9. The AIC and BIC scores of the fitted models by number of hidden states are presented in Figure 2.5. Both indicators suggest that the optimal number of hidden states is 7.

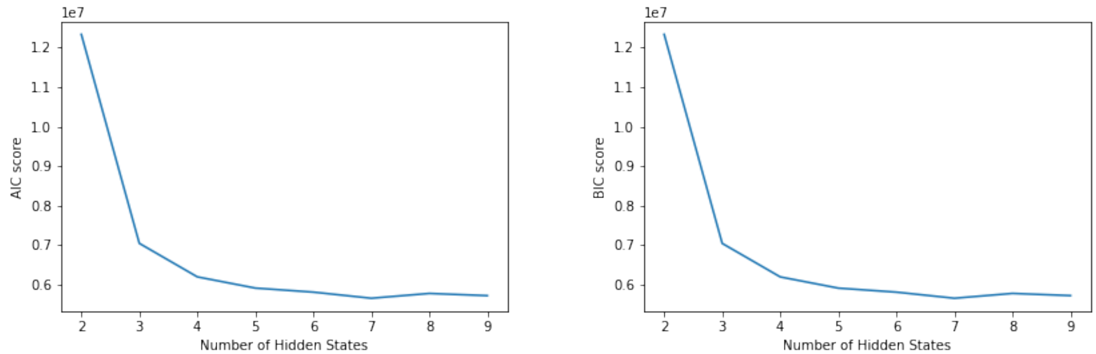

Figure 2.5.: AIC and BIC Scores by Number of Hidden States

2.3.2 Data Simulation

The provided data do not include accident information. Thus, we simulate accident occurrences based on existing findings between driving behavior and accident likelihood and then illustrate the proposed methods using the simulated data.

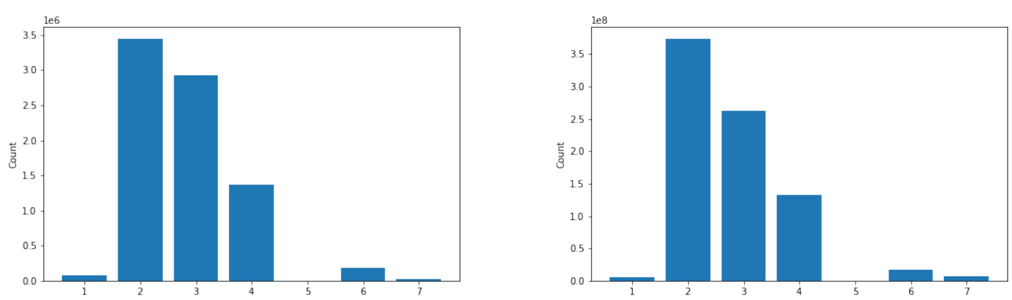

Figure 2.6.: Number of Simulated Observations by Hidden State: Original Data vs. Simulated Data

The simulation contains two steps. The first step is to increase the number of trips. For each trip in the original data, we use the seven-state HMM to generate

100 similar trips with the same time length based on the sample transition matrix **A** and emission matrix **B** of the trip, which yields 591,500 simulated trips in this step. The fitted HMM is then applied to assign hidden states to all records in the simulated data. The number of records by hidden state for both the original and simulated data are compared in Figure 2.6. We can see that the characteristics of the original trips (implied by hidden states) are largely kept in their simulated ones. Table 2.3 reports the mean and standard deviation of the (percentage) weight of each categorical variable in the simulated data. Most records are in hidden states 2, 3, and 4 (i.e., 63.4%, 17.8%, and 17.1%, respectively). Hard braking is rare and only accounts for 0.0536% of the total observations. About 8.31% of the records have angle changes more than 7.5 degrees per second, which confirms our categorical converting criteria in Section 2.1.

Table 2.3.: Mean and Standard Deviation of the Percentage Weight of Variables in the Simulated Data

Variables	Mean	Standard Deviation
speed_1	0.644	0.301
speed_2	0.171	0.165
speed_3	0.179	0.259
speed_4	0.00597	0.0441
hard_braking	0.000536	0.00184
big_angle_change	0.0831	0.0378
hidden_state_1	0.00868	0.00769
hidden_state_2	0.634	0.304
hidden_state_3	0.178	0.257
hidden_state_4	0.171	0.165
hidden_state_5	0.000017	0.00017
hidden_state_6	0.00693	0.0485
hidden_state_7	0.00133	0.0276

Figure 2.7 further explores the distributions of driving behavior by hidden state in the simulated data. The two upper graphs show that hidden state 2 has the highest number but low frequencies of hard braking and big angle change, while hidden state

5 has the highest frequency but very low counts for both driving behaviors. It seems that hidden state 5 could be a major state to capture high-risk driving patterns. The implied hidden states are also able to effectively differentiate speed intervals (see the two lower graphs in Figure 2.7). The majority of low-speed observations (between 0 to 35 mph) appear in hidden states 1, 2 and 7, and middle-range speeds are mostly in hidden states 3, 4, and 5. Hidden state 6 mainly consists of high-speed observations, especially those above 80 mph.

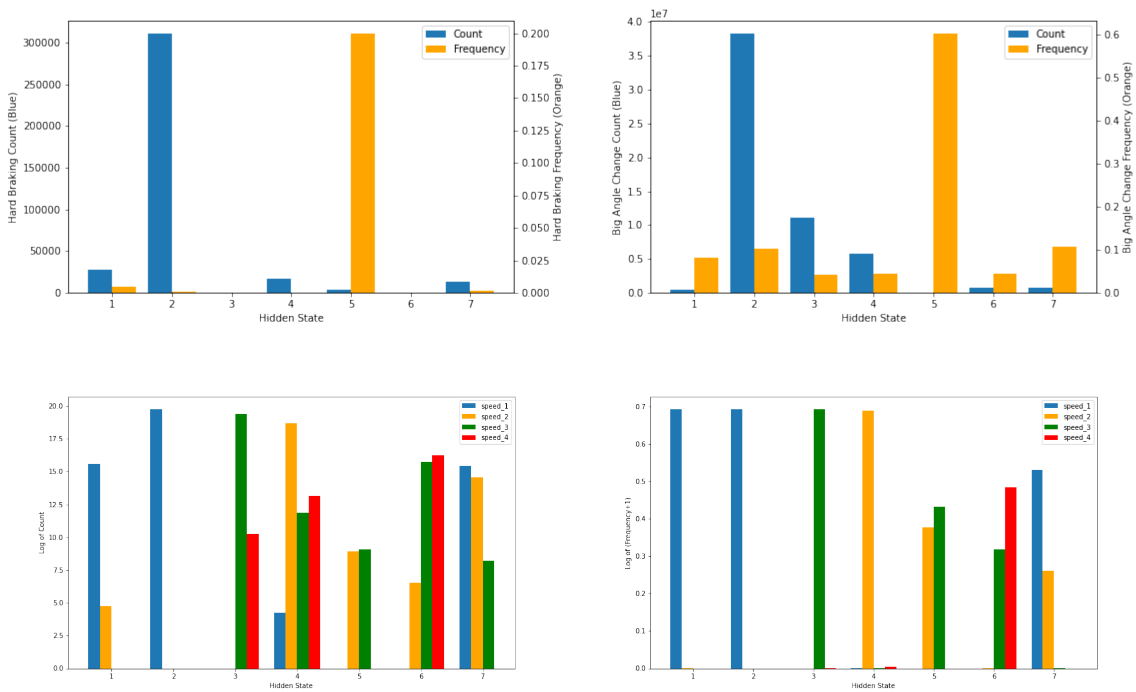

Figure 2.7.: Distributions of Driving Behaviors by Hidden State

The second step is to independently simulate accidents based on different driving behaviors. Existing studies have shown that hard braking and big angle change, among other behaviors, play important roles in predicting accident rates (e.g., Huang and Meng, 2019). Since the purpose of the simulation is to provide illustrative data for our model application, we simplify the real situation and mainly associate accident probabilities with hard braking and big angle changes as follows:

$$\Pr[\text{accident}|\text{hard braking}] = 0.0008, \quad (2.16)$$

$$\Pr[\text{accident}|\text{big angle change}] = 0.00001, \quad (2.17)$$

$$\Pr[\text{accident}|\text{no hard braking or big angle change}] = 0.0000005. \quad (2.18)$$

The above scale is consistent with the accident statistics found in Boucher et al. (2017), Verbelen et al. (2018), Huang and Meng (2019), and Desai et al. (2021). For example, Boucher et al. (2017) found that the average claim frequency of a Spanish telematics data set is around 0.15 for 5,000 km, which is translated into approximately 5×10^{-7} per second using an average speed of 39.74 mph in our data, similar to the values assigned in our simulation.

For each observation of a trip, we can calculate the accident probability based on whether hard braking and/or big angle changes are observed. Starting from the first observation, we draw a Bernoulli number with the calculated accident probability. If the random number is 0, we move to the next observation and repeat the process. Once an accident (i.e., a Bernoulli number of 1) is observed, the trip is assumed to be terminated, and later observations in the trip are removed. The overall distributions for trip duration and accident frequency by hidden state are shown in Figure 2.8. Most of the trips are shorter than 30 minutes, and the trip duration decreases exponentially when the time interval increases every 30 minutes. Hidden state 2 has the highest number of accidents, which is explained by the fact that it has the highest number of total observations, as well as the highest number of hard braking instances and big angle changes. However, hidden state 5 has much higher accident probability than all other states. This is associated with the significantly higher frequencies of hard braking and big angle changes observed in this state (see Figure 2.7).

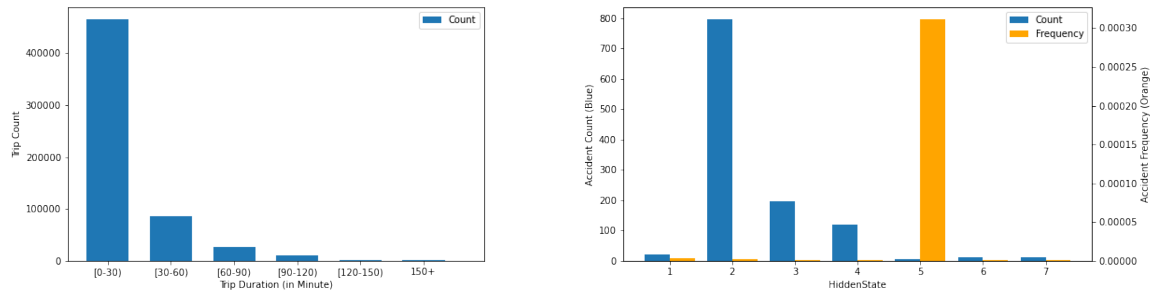

Figure 2.8.: Distributions of Trip Duration and Accident by Hidden State

2.4 Results

2.4.1 Clustering and Model Comparison

In this section, we present the classification results based on hidden states using the K-means algorithm. We also compare the proposed method with a driving-behavior-based classification method. To do so, we first calculate the percentage weights of hidden states in each trip. These vectors are then used to measure the dissimilarity or closeness between trips based on their Euclidean distances. For a given number of clusters K, the K-means algorithm randomly initializes K centroids and divides all trips into K groups based on their closeness to each centroid. The algorithm next re-calculates the centroid of each group by averaging the weight vectors within the group. Once the new centroids are determined, trips are reassigned into different groups based on their closeness to the new centroids. The process is repeated until the centroids converge. The obtained clusters are used as a categorical covariate in Poisson GAM Models 1 (driving-behavior-based) and 2 (hidden-state-based) for claim frequency prediction.

As previously mentioned, the optimal number of clusters is determined by minimizing the deviance statistics (2.10) of the predictive models. To assess model performance, a five-fold cross-validation technique is adopted for calculating the (average) deviance statistics. First, the simulated data set is divided into five equal subsets.[3] In each cross-validation, one subset is picked as the test set, and the remaining subsets are combined into a training set. The training set is then used to train the model (i.e., estimate the model parameters), and the test set is used to evaluate the performance (i.e., calculate the deviance statistics). The process is repeated until all the remaining subsets are used test sets. Finally, we calculate the average of the deviance statistics from all cross-validations. The cross-validation technique significantly mitigates the random errors caused from data splitting.

Figure 2.9 depicts the average deviance statistics of Models 1 and 2, respectively, when the number of clusters varies from 2 to 6. We see that the optimal number of clusters is five for both clustering methods. Table 2.4 reports the total number of observations of each (hidden-state-based) cluster in the optimal case, while the percentage of hard braking and big angle change observations in each hidden state and cluster are presented in Tables 2.5 and 2.6, respectively. Consistent with previous findings, we observe that hidden state 5 has significantly higher percentages of hard braking and big angle change than others across all the clusters. The only exception is hidden state 7, in which 100% of the observations are identified as hard braking in clusters 1, 2, and 5. However, the total number of these observations is very small. For example, only 9 observations are in cluster 2 and hidden state 7. Generally speaking, hidden states 3, 4, and 6 have the lowest percentages of hard braking and big angle changes.

The fitted parameters for Models 1 and 2 in the optimal clustering scenarios are reported in Table 2.7. As expected, the coefficients for both hard braking and big angle change are positive and statistically significant. The coefficients for the speed vector are mostly insignificant, which confirms that accident probability is not dependent on speed in our experiment. While we do not specifically impose a dependence between hard braking and angle change in predicting accident rates, the interaction term on hard braking and angle change still shows statistical significance for both

[3]Since the accident distribution is highly skewed, uneven accident frequencies are possible in these subsets. To avoid such a situation, stratified sampling is implemented so that all subsets have relatively consistent accident frequencies.

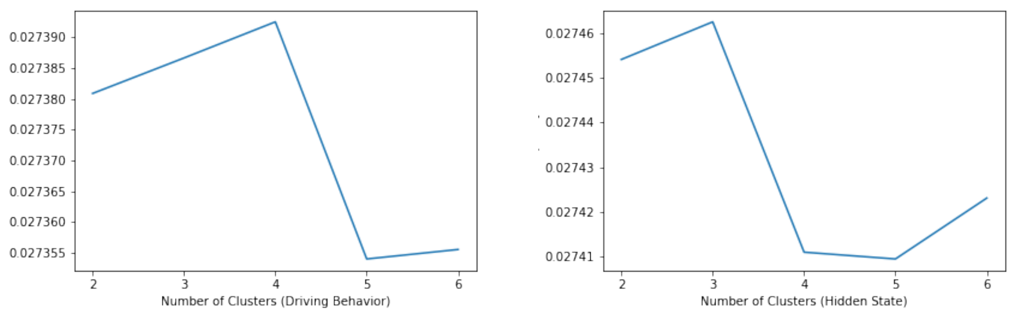

Figure 2.9.: Average Deviance Statistics for Driving-Behavior-Based Clusters (Left) and Hidden-State-Based Clusters (Right)

Table 2.4.: Total Number of Observations by Cluster and Hidden State

	cluster_1	**cluster_2**	**cluster_3**	**cluster_4**	**cluster_5**
hidden_state_1	656,070	1,152,126	1,476,690	851,833	1,851,507
hidden_state_2	137,901,391	30,541,163	69,443,276	30,443,640	105,115,727
hidden_state_3	1,147,126	7,859,285	77,288,346	164,686,812	12,132,586
hidden_state_4	6,035,826	37,316,980	34,090,516	22,437,304	33,017,596
hidden_state_5	527	1,190	5,662	7,018	1,674
hidden_state_6	10,579	31,472	9,876,942	7,632,811	310,883
hidden_state_7	521	9	5,339,720	1,641,842	164

Table 2.5.: Percentage of Hard Braking in Hidden States by Cluster

	cluster_1	**cluster_2**	**cluster_3**	**cluster_4**	**cluster_5**
hidden_state_1	0.069089	0.079480	0.090404	0.091472	0.078026
hidden_state_2	0.098096	0.110385	0.104266	0.108329	0.103001
hidden_state_3	0.046298	0.035636	0.042931	0.041727	0.042441
hidden_state_4	0.040420	0.038820	0.048976	0.049128	0.040415
hidden_state_5	0.362429	0.773109	0.636524	0.581932	0.527479
hidden_state_6	0.037055	0.035714	0.045289	0.043635	0.044846
hidden_state_7	0.094050	0.000000	0.113048	0.090210	0.225610

Table 2.6.: Percentage of Big Angle Change in Hidden States by Cluster

	cluster_1	**cluster_2**	**cluster_3**	**cluster_4**	**cluster_5**
hidden_state_1	0.069089	0.079480	0.090404	0.091472	0.078026
hidden_state_2	0.098096	0.110385	0.104266	0.108329	0.103001
hidden_state_3	0.046298	0.035636	0.042931	0.041727	0.042441
hidden_state_4	0.040420	0.038820	0.048976	0.049128	0.040415
hidden_state_5	0.362429	0.773109	0.636524	0.581932	0.527479
hidden_state_6	0.037055	0.035714	0.045289	0.043635	0.044846
hidden_state_7	0.094050	0.000000	0.113048	0.090210	0.225610

models. This implies that some inner structures among driving behaviors exist in the original data, and the first step of our simulation using the HMM framework to generate trips is able to capture these structures. Furthermore, some coefficients of the five clusters are statistically significant, which indicates that both Models 1 and 2 are able to capture additional layers of information from risk classification, other than from individual driving behavior variables. We also observe that Model 2 has more statistically significant risk class coefficients than Model 1. It appears that purely driving-behavior-based clusters include relatively less effective information than hidden-state-based clusters, partly due to the fact that individual driving behavior variables themselves are also regressed in the model.

The performance measures of Models 1 and 2 are summarized in Table 2.8. Model 2 has slightly better deviance statistics but higher mean squared errors by device than Model 1. Overall, Models 1 and 2 provide comparable results, indicating that risk classification based on the HMM method can deliver results with similar quality as those driving-behavior-based clustering methods.

To examine the performance of Models 3 and 4, we further group observations in a trip by their hidden states and calculate the percentage weights of different driving behavior variables for each state. They are the input for estimating Model 3 and Model 4. Specifically, the percentage weights of the ith driving behavior vector in the hidden state k of jth trip are given as

$$frequency_{\{DB_i, Trip_j, State_k\}} = \frac{\text{Number of } DB_i \text{ observed in } Trip_j \text{ and } State_k}{\text{Length of } Trip_j \text{ in } State_k}. \quad (2.19)$$

The estimated parameters are presented in Table 2.7. Similar to the parameters in Models 1 and 2, the coefficients for hard braking, big angle change, and their interaction term are all statistically significant. Some coefficients for the hidden states are also statistically significant, which shows that hidden states help capture additional driving performance information that is not fully reflected by individual driving behavior variables. In particular, by incorporating hidden states as a categorical vector, the coefficients for hidden-state-based clusters become more statistically significant in Model 4 than in Model 2. It appears that regressing on both hidden-state-based

Table 2.7.: Estimated Parameters for Models 1-5

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
speed_2	0.070	-0.158	0.019	-0.035	-0.902^{***}
speed_3	-0.529^{**}	-0.489	0.102	0.203	-0.991^{***}
speed_4	-0.138	-0.158	-0.529	-0.447	-0.546
$hard\_braking(HB)$	5.766^{***}	5.765^{***}	7.324^{***}	7.237^{***}	5.726^{***}
$big\_angle\_change(BAC)$	8.913^{***}	8.671^{***}	6.125^{***}	6.210^{***}	8.499^{***}
$HB \cdot BAC$	0.784^{***}	0.784^{***}	-3.885^{***}	-3.941^{***}	0.777^{***}
cluster_1	-0.064	-0.214^{**}		0.508^{***}	
cluster_2	0.367^{***}	-0.124^{*}		0.118	
cluster_3	0.011	0.376^{***}		-0.172^{***}	
cluster_4	-0.108	0.086		-0.301^{***}	
cluster_5	-0.206^{**}	-0.123		-0.152^{**}	
hidden_state_1			0.611^{***}	0.503^{**}	
hidden_state_2			0.520^{***}	0.294	
hidden_state_3			-0.233	-0.177	
hidden_state_4			0.0142	0.049	
hidden_state_5			0.903^{**}	-0.805^{*}	
hidden_state_6			0.057	0.110	
hidden_state_7			-0.0667	0.025	

Note: Stars *, **, and *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

clusters and hidden states can better extract driving behavior characteristics and their underlying structures from the data. This is also verified by the performance measures in Table 2.8. Model 4 outperforms the other four models by both deviance statistics and MSE by device except for Model 5, which has larger deviance statistics but smaller MSE by device than Model 4. The better MSE by device for Model 5 may be driven by a single device. After removing the device, the new MSE by device for Model 4 becomes 25.34, which is smaller than that for Model 5, that is, 27.10.

Table 2.8.: Performance Measures for Models 1 - 5

Model	Deviance Statistics	MSE by Device	Number of Trips
Model 1	0.02730	38.47	591,500
Model 2	0.02729	39.86	591,500
Model 3	0.02743	39.89	591,500
Model 4	0.02729	38.30	591,500
Model 5	0.02736	37.70	591,500

2.4.2 Insurance Pricing

The predicted mean accident rates λ in the proposed models can be used to determine the premiums for various UBI policies, such as PAYD and PHYD. In particular, Model 4 provides the Poisson accident rate per second by cluster and hidden state. Utilizing the stationary properties of the HMM, we are also able to provide long-term prediction of accident rates.

Specifically, we calculate the mean accident rate λ for each risk class and each hidden state using both the simple average and exposure-weighted average approaches. The results are reported in Tables 2.9 and 2.10, respectively. To obtain the long-term insurance rates, the stationary distributions of hidden states for each risk class is determined using Equations (2.11) and (2.12) (Table 2.11). As expected, the mean accident rates for hidden states 3, 4, and 6 are among the lowest, while those for hidden state 5 are among the highest across all clusters. While different clusters generally present similar accident rates in a given hidden state, we do observe distinct accident likelihoods in hidden state 7. The mean accident rates for hidden state 7 in clusters 1, 2, and 5 are much higher than in all other cases. This is because hidden state 7 in these clusters has observations with hard braking only (see Table 2.5). Since the number of observations for hidden state 7 in these clusters is extremely low, the transition probabilities to hidden state 7 are almost negligible for these cases (see Table 2.11).

Table 2.9.: Simple Average Mean Accident Rate (λ) by Risk Class and Hidden State (in 10^{-6})

	cluster_1	**cluster_2**	**cluster_3**	**cluster_4**	**cluster_5**
hidden_state_1	18.30	3.96	5.17	9.46	4.17
hidden_state_2	3.06	2.36	1.74	1.66	1.68
hidden_state_3	2.22	1.19	0.80	0.69	0.97
hidden_state_4	2.19	1.02	0.88	0.82	0.81
hidden_state_5	154.19	220.32	294.10	333.36	151.89
hidden_state_6	3.24	3.27	0.86	0.92	0.89
hidden_state_7	2168.49	1107.61	35.00	47.66	1646.25

Table 2.10.: Exposure-Weighted Average Mean Accident Rate (λ) by Risk Class and Hidden State (in 10^{-6})

	cluster_1	**cluster_2**	**cluster_3**	**cluster_4**	**cluster_5**
hidden_state_1	9.78	3.01	2.81	3.27	2.64
hidden_state_2	2.96	2.21	1.58	1.43	1.59
hidden_state_3	1.70	1.03	0.80	0.70	0.83
hidden_state_4	1.62	1.04	0.84	0.73	0.80
hidden_state_5	166.61	236.46	343.25	410.57	172.62
hidden_state_6	1.17	0.77	0.76	0.69	0.59
hidden_state_7	2188.14	1107.61	1.52	1.50	1514.91

Finally, using Equation (2.13), we obtain the long-term accident rate vector $\vec{\lambda}$ from the stationary distribution of HMM and values in Table 2.9, i.e.,

$$\vec{\lambda} = [3.1, 1.6, 2.2, 1.2, 1.4] \cdot 10^{-6}, \tag{2.20}$$

or values in Table 2.10, i.e.,

$$\vec{\lambda} = [2.9, 1.5, 1.1, 0.8, 1.3] \cdot 10^{-6}. \tag{2.21}$$

Once the trip duration of a policy in each cluster is estimated, the total expected Poisson accident rate λ_{total} can be found by substituting (2.20) or (2.21) into (2.14). This

Table 2.11.: Stationary Distribution of HMM by Risk Class

	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
hidden_state_1	0.004798	0.015144	0.007227	0.003246	0.012617
hidden_state_2	0.939423	0.358528	0.300911	0.092016	0.660268
hidden_state_3	0.009763	0.110802	0.426383	0.765730	0.090475
hidden_state_4	0.045912	0.515042	0.177196	0.095390	0.234190
hidden_state_5	0.000001	0.000013	0.000024	0.000025	0.000007
hidden_state_6	0.000100	0.000470	0.056525	0.035813	0.002442
hidden_state_7	0.000002	0.000000	0.031735	0.007780	0.000001

pricing approach not only digs into trip-level driving characteristics and their inner structures but also provides a balanced solution to aggregate extracted information from a long-term perspective.

2.5 Conclusions

This study proposes an HMM framework to explore telematics data and perform auto insurance risk classification and pricing. The implied hidden states are used for trip clustering via the K-means algorithm. To test the predictive power of the proposed method, the Poisson GAM model is used to estimate accident rates and compare the performance of different classification methods. We find that the hidden-state-based classification method appears to outperform the examined driving-behavior-based method in terms of both deviance statistics and MSE of accidents by OBD device. This suggests that the HMM framework not only captures the information from individual driving behavior variables and their relations but also the transitions of driving behavior under different environments. Overall, the proposed HMM framework provides an effective pricing approach for UBI policies and allows insurers to develop rates based on driving behavior and their transitions.

As an initial step to investigate auto insurance pricing through an HMM framework, there are several questions that we leave for future research. First, our analysis is based on simulated data. The experiment tests whether the proposed model can capture some assumed relations between driving behavior and accident rates. It would be interesting to further explore on raw telematics data whether the proposed model is more powerful in extracting information than other models. Second, there are many directions to combine the traditional HMM with other algorithms (e.g., artificial neural network) to complete multi-purpose tasks and improve model performance. We expect that these more advanced HMM frameworks would further increase the predictive power for accident rates. Finally, an extra layer may also be added into the existing HMM framework so that the loss severity and its association with driving behavior and hidden states can be studied simultaneously. This three-layer HMM framework might provide a more integrated way to investigate the interdependence of claim frequency and severity and the impact from driving behavior.

CHAPTER 3
THEORETICAL FRAMEWORK OF A 3-LAYER HIDDEN MARKOV MODEL FOR AUTO INSURANCE PRICING

3.1 Introduction

Usage-based insurance (UBI) has gained increasing popularity in the auto insurance industry in recent years. Its success is propelled by the technological advancement named telematics technology, which enables insurance companies to collect a tremendous amount of driving behavior data (e.g., videos, GPS information, speed information, car status, and accelerometer and magnetometer sensor data). Many emerging insurtech companies and major personal auto insurers apply state-of-the-art models to telematics data to develop competitive UBI products (e.g., Root Car Insurance, Metromile, Progressive's Snapshot, State Farm's Drive Safe & Save).

The emergence of telematics data has not only helped insurance companies understand how policyholders drive but also led to a new insurance pricing paradigm. Unlike traditional rate making approaches, where insurance premiums are determined based on policyholders' self-reported rating variables, such as vehicle information and personal information, telematics-data-based pricing methods rely on real-time driving behavior information and directly link loss information with policyholders' riskiness inferred from driving behavior. These novel methodologies allow insurers to more accurately classify risks and estimate losses at different granularity levels (e.g., policyholder level, trip level, or the level of each second).

In Jiang and Shi (2022), a hidden-Markov-model (HMM)-based framework is developed to predict claim frequencies using telematics data. The study shows that hidden-state-based risk classes would provide more predictive power than purely driving-behavior-based ones for loss prediction. As an initial step to utilizing HMM for auto insurance pricing, that study left a few interesting problems for further investigation. In particular, loss severity was not discussed. In this paper, we revisit the problem and propose a 3-layer HMM framework that takes both telematics data and loss data as input to 1) classify risks using driving behavior information and 2) estimate loss frequency and severity simultaneously. A unique advantage of this approach is that time series data can be utilized without aggregation. This is different from traditional two-step methods, in which driving behavior features are typically first extracted and aggregated from telematics data and then incorporated in predictive models (e.g., Amsalu and Homaifar, 2016; Wüthrich, 2017; Huang and Meng, 2019; Gao et al., 2022).

There has been a series of studies on developing sophisticated HMM-related models for driving behavior prediction. In addition to direct applications of the classic HMM (e.g., Berndt et al., 2008; Al-Sultan et al., 2013; Craye et al., 2016; Deng et al., 2018), many researchers have proposed combining HMM with other well-established learning algorithms to achieve better model performance. For example, Boyraz et al. (2007) and Zong et al. (2009) utilized an artificial neural network and HMM to identify and predict the manoeuvring behavior of drivers.[1] The support vector machine (SVM) is another popular choice for combination with HMM to classify driving patterns (Aoude et al., 2012), predict driving status (Tadesse et al., 2014), and predict accidents (Xiong et al., 2018). More organic and complex combinations, such as modeling some parts of HMM by deep neural network (DNN), random forest, and Gaussian mixture models, can be found in Lefèvre et al. (2015), Zhang et al. (2015), and Wang et al. (2018), for example.

Another strand of research seeks to modify the classic HMM through additional assumptions or structures. For example, auto-regressive HMMs, in which observations are generated by auto-regressive time series while the time series transitions are governed by HMM, have been used in various driving behavior predictions, such as steering angle (Hamada et al., 2016). Hierarchical Dirichlet process (HDP) HMMs

[1] Manoeuvring behavior is the certain driving behavior that requires skill and care. Typically, driving manoeuvring includes left/right turning, lane changing, slowing down, etc.

do not require a pre-specified number of hidden states and are also widely used to predict driving behavior, such as lane changes (Zhang et al., 2022) and driving style (Wang et al., 2019). Gadepally et al. (2011) utilized hierarchical HMM (HHMM), a nested HMM in which states can be emitted by other states, for pattern recognition (e.g., left/right turns, stop and straight through interactions). The same model was also used in Zhu et al. (2015) to study the driving conditions of heavy-duty vehicles. Finally, Deng and Söffker (2019) proposed a novel multi-layer HMM, in which three HMMs are cascaded, to predict driving behavior.[2]

While all these studies have demonstrated that classic HMMs and their variants provide excellent performance in driving behavior prediction, few have explored the application of these models in auto insurance pricing. The proposed 3-layer HMM provides a one-step solution to incorporating driving behavior information for auto insurance pricing. This model also offers long-term insurance rates that are derived from the limiting properties of HMM. Specifically, the average loss per second for each risk class can be determined using the stationary distribution of HMM, and the expected insurance losses can be considered as the exposure-weighted average. We also show that numerical solutions can be easily obtained for some parametric loss models (e.g., Bernoulli distributed loss frequency per unit of time and Gamma distributed loss severity).

The rest of the paper is organized as follows. Section 3.2 illustrates the basic set-up of the model. Section 3.3 introduces an auto insurance pricing framework based on the proposed HMM model. The next section develops formulas for evaluation and decoding, giving important definitions and steps for model estimation. Section 3.5 shows the application of the expectation-maximization (EM) algorithm for model estimation. The last section concludes the paper.

[2]It is noteworthy that the term "layer" has different meanings; in our model, it represents variables (e.g., risk class, driving behavior, and loss).

3.2 3-Layer HMM

In this section, we introduce a 3-layer HMM tailored for auto insurance pricing. The proposed 3-layer HMM is a special case of Bayesian networks. Appendix A provides a brief introduction to Bayesian networks and describes how to decompose multi-variable probabilities based on Bayesian networks.

There are three important variables in our analysis. We first define them as follows:

X_t : The risk class of the driver at time t **(unobservable)**

Y_t : The driver's behavior at time t **(observable)**

L_t : The driver's loss at time t **(observable)**

The first variable X_t represents the risk class of the driver at time t, which is unobservable, and the output we want to obtain from the model. Risk classes are directly associated with drivers' expected losses. Ideally, we want policyholders with the same expected loss to pay the same premium for sharing the same risk. The risk class could change over time and is inferred based on driving behavior. Y_t represents the driving behavior at time t. This variable is observable and serves as the model input. Driving behavior can be characterized by multiple indicators, such as hard braking, sudden turns, and accelerations. L_t denotes the loss size due to an potential accident at time t. This variable is also observable and serves as the input of the model. L_t follows a highly skewed distribution during a trip, since only the last moment of the trip can have non-zero loss. Meanwhile, some L_t cannot be collected instantaneously when the insured is driving. Moreover, $X_t \in \{1, 2, ..., N\}$ and $Y_t \in \{1, 2, ..., M\}$ are discrete variables, and L_t is continuous.

A 3-layer HMM is graphically depicted by Figure 3.1, in which directed arrows indicate the direction of information propagation between nodes. In cross-section, the risk class X_t provides information about drivers' driving behavior. Therefore, there is a directed arrow from X_t to Y_t. Since driving behavior may impact losses, a directed arrow from Y_t to L_t is also added. Moreover, the risk class of the driver at time t

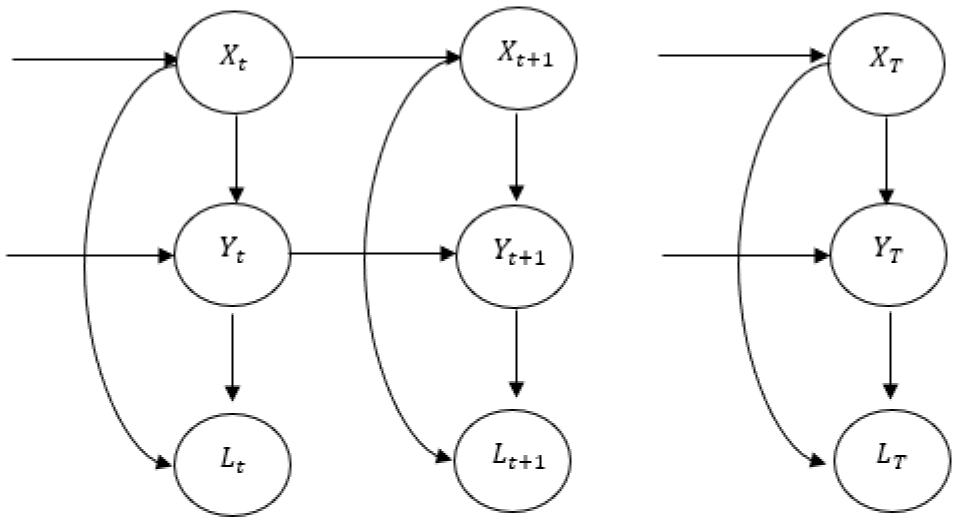

Figure 3.1.: 3-Layer HMM

may also help predict the probability of loss at the same moment. Thus, a directed arrow is placed from X_t to L_t. In time series, we assume that both X_t and Y_t follow the Markov process. That is, the values at time t+1 only depend on those at time t. Directed arrows also appear from X_t to X_{t+1} and Y_t to Y_{t+1}. Importantly, according to Pearl (1988), the graphical relationship shown in Figure 3.1 can be decomposed as follows:

$$\begin{aligned} & P(X_{t-1}, Y_{t-1}, L_{t-1}, X_t, Y_t, L_t) \\ = \; & P(X_{t-1})P(Y_{t-1}|X_{t-1})P(L_{t-1}|X_{t-1}, Y_{t-1})P(X_t|X_{t-1})P(Y_t|X_t, Y_{t-1})P(L_t|X_t, Y_t). \end{aligned} \tag{3.1}$$

Using Equation (3.1) and the properties of Bayesian networks, we reach the following proposition:

Proposition 1

$$P(X_t, Y_t, L_t|X_{1:t-1}, Y_{1:t-1}, L_{1:t-1}) = P(X_t|X_{t-1})P(Y_t|X_t, Y_{t-1})P(L_t|X_t, Y_t) \text{ for } t > 1. \tag{3.2}$$

The proof for this proposition is given in Appendix B. Proposition 1 provides an important decomposition used in estimation, evaluation, and decoding in later sections.

Next, we introduce the parameter space that fully defines the 3-layer HMM. The first group of parameters are the initial probability of being in risk class i at time 1, which is denoted $\pi_i = P(X_1 = i)$. The time-independent transition probability from risk class i to j is defined as $a_{ij} = P(X_t = j|X_{t-1} = i)$, where $i, j \in \{1, 2, ..., N\}$. Furthermore, the time-independent conditional probability of Y being k at time t, given the risk class at time t and the driving behavior at time t-1, is denoted $b_{ijk} = P(Y_t = k|X_t = i, Y_{t-1} = j)$, where $k, j \in \{1, 2, ..., M\}$. Finally, we let $c_{ij}(l) = f(L_t = l|X_t = i, Y_t = j)$ be the conditional density of loss at time t given the risk class and the driving behavior at time t. We summarize the parameter space $\Phi = \{\Pi, \mathbb{A}, \mathbb{B}, \mathbb{C}\}$ as follows:

$$\Pi = \left\{\pi_i : P(X_1 = i)\right\}, \quad i \in \left\{1, 2, ..., N\right\}$$
$$\mathbb{A} = \left\{a_{ij} : P(X_t = j|X_{t-1} = i)\right\}, \quad i, j \in \left\{1, 2, ..., N\right\}$$
$$\mathbb{B} = \left\{b_{ijk} : P(Y_t = k|X_t = i, Y_{t-1} = j)\right\}, \quad i \in \left\{1, 2, ..., N\right\}, \quad j, k \in \left\{1, 2, ..., M\right\}$$
$$\mathbb{C} = \left\{c_{ij}(l) : f(L_t = l|X_t = i, Y_t = j)\right\}, \quad i \in \left\{1, 2, ..., N\right\}, \quad j \in \left\{1, 2, ..., M\right\}$$

Similar to the classic HMM, the number of risk classes needs to be specified prior to model estimation. There are multiple ways to determine the optimal number. First, Akaike's information criteria (AIC) and Bayesian information criteria (BIC) can be used with loss distribution to determine the optimal number of risk classes. Both AIC and BIC impose penalties on the number of parameters. Second, deviance statistics of loss distribution may also be used to select the number of risk classes. Lastly, to avoid overfitting due to a large number of risk classes, the cross-validation technique is widely used in machine learning area for model selection. Cross-validation divides a whole data set into training and test sets. A candidate model is trained or estimated using the training set, while the performance of the model is examined using the test set.

3.3 Auto Insurance Pricing

This section explores an asymptotic property of the 3-layer HMM and introduces a long-term pricing method for auto insurance pricing. Specifically, this section shows that the expected losses per unit of time may be associated with risk class only and the distribution of risk classes can be used as risk exposures for pricing. This leads to the second proposition, shown as follows:

Proposition 2

$$E[L_1 + ... + L_T | X_{1:T} = x_{1:T}, Y_{1:T} = y_{1:T}] \to \sum_{i=1}^{N} T_i E[L_t | X_t = i], \text{ when T is large,} \tag{3.3}$$

where $T = \sum_{i=1}^{N} T_i$ and T_i is the duration of the driver staying at risk class i.

The rest of this section gives the proof of this proposition. For a given time T, the expected total loss conditional on risk class and driving behavior is given as

$$E[L_1 + L_2 + ... + L_T | X_{1:T} = x_{1:T}, Y_{1:T} = y_{1:T}] \tag{3.4}$$

$$= \sum_{t=1}^{T} E[L_t | X_{1:T} = x_{1:T}, Y_{1:T} = y_{1:T}] \tag{3.5}$$

$$= \sum_{t=1}^{T} E[L_t | X_t = x_t, Y_t = y_t] \quad \text{by conditional independence} \tag{3.6}$$

$$= \sum_{t=1}^{T} \int_0^{\infty} L_t \, dP(L_t | X_t = x_t, Y_t = y_t). \tag{3.7}$$

To examine the asymptotic behavior, we first focus on the time points $k_1, ..., k_q$, where $X_t = i$, i is in the state space of X, and q is the number of these time points:

$$E[L_{k_1} + ... + L_{k_q} | X_{k_1:k_q} = i, Y_{k_1:k_q} = y_{k_1:k_q}] = \sum_{t=k_1}^{k_q} E[L_t | X_t = i, Y_t = y_t]. \tag{3.8}$$

Furthermore, among these q time points, it is assumed that there are $h_1, ..., h_j$ time points such that $Y_t = j$, where $j = 1, 2, ..., M$. It follows that $\sum_j h_j = q$ and

$$\sum_{t=k_1}^{k_q} E[L_t|X_t = i, Y_t = y_t] = \sum_{j=1}^{M} h_j E[L_t|X_t = i, Y_t = j]$$
$$= q(\sum_{j=1}^{M} \frac{h_j}{q} E[L_t|X_t = i, Y_t = j]). \tag{3.9}$$

When T becomes large, $\frac{h_j}{q}$ converges to $P(Y_t = j|X_t = i)$, and h_j covers all elements in the state space of Y. Therefore, we have the following asymptotic formula:

$$\sum_{t=k_1}^{k_q} E[L_t|X_t = i, Y_t = y_t] \to q(\sum_{j=1}^{M} P(Y_t = j|X_t = i) E[L_t|X_t = i, Y_t = j])$$
$$= qE[L_t|X_t = i]. \tag{3.10}$$

In other words, the expected loss of a driver given a certain risk class and driving behavior over time can be asymptotically estimated as the product of the expected loss of the driver in the given risk class and the time that the driver stays in the risk class. If we further carry i for the notation q, and let q_i be the number of total time points in state i for X_t, we can express the conditional expected loss (3.4) as:

$$E[L_1 + L_2 + ... + L_T|X_{1:T}, Y_{1:T}] \to \sum_{i=1}^{N} q_i E[L_t|X_t = i], \tag{3.11}$$

where $\sum q_i = T$. After replacing q_i by T_i, it proves Proposition 2.

This proposition shows that the pricing paradigm based on our 3-layer HMM helps mitigate the distortions from short-term individual driving behavior and enables insurers to determine insurance rates as the exposure-weighted average of expected loss per unit of time in different risk classes.

3.4 Evaluation and Decoding

This section introduces two important processes for the estimation of 3-layer HMM, namely evaluation and decoding. Evaluation is a process of calculating the output probability of the model, while decoding is a process of finding the hidden state sequence that fits the observations best. Both processes contain important algorithms (e.g., forward algorithm and Viterbi algorithm) and quantities (e.g. forward variable α) to be used later. Throughout this section, we assume that the parameters of the 3-layer HMM are given.

3.4.1 Evaluation

Output probability is the most vital quantity for model estimation and selection and the best hidden state path selection. In this section, we develop algorithms to calculate two different output probabilities in a recursive way.

The first output probability is **total output probability**, defined as $P(O|\Phi) = P(L_{1:T}, Y_{1:T}|\Phi)$, where $O = (L_{1:T}, Y_{1:T})$ denotes a sequence of observations up to a fixed time T. For any observation sequence and state sequence given Φ, we have

$$P(L_{1:T}, Y_{1:T}|\Phi) = \sum_{X_{1:T} \in X^T} P(L_{1:T}, Y_{1:T}, X_{1:T}|\Phi). \tag{3.12}$$

We can rewrite (3.12) as

$$\begin{aligned} &\sum_{X_{1:T} \in X^T} P(L_{1:T}, Y_{1:T}, X_{1:T}|\Phi) \\ &= \sum_{X_{1:T} \in X^T} P(X_T, Y_T, L_T|X_{T-1}, Y_{T-1}, L_{T-1})...P(X_2, Y_2, L_2|X_1, Y_1, L_1)P(X_1, Y_1, L_1) \\ &= \sum_{X_{1:T} \in X^T} \left(\prod_{t=2}^{T} P(X_t|X_{t-1})P(Y_t|X_t, X_{t-1})P(L_t|X_t, Y_t) \right) P(X_1, Y_1, L_1). \end{aligned} \tag{3.13}$$

Since $P(X_1, Y_1, L_1) = P(X_1)P(Y_1|X_1)P(L_1|X_1, Y_1)$, one finds that

$$P(O|\Phi) = \sum_{X_{1:T} \in X^T} \prod_{t=1}^{T} a_{X_{t-1}, X_t} b_{X_t, Y_{t-1}, Y_t} c_{X_t, Y_t}(l_t), \tag{3.14}$$

where $a_{X_0, X_1} = \pi_{X_1} = P(X_1)$ and $b_{X_1, Y_0, Y_1} = P(Y_1|X_1)$.

The algorithm is computational complex. Thus, the **forward algorithm** is introduced for faster computation (Rabiner, 1989) as it computes the total output probability in a recursive fashion. Similar to Rabiner (1989), we define the forward variable as follows:

$$\alpha_t(j,k) = P(L_{1:t}, Y_{1:t-1}, Y_t = k, X_t = j|\Phi). \tag{3.15}$$

The α_{t+1} can be expressed recursively in terms of the last period forward variable:

$$\begin{aligned}
\alpha_{t+1}(j,k) &= P(L_{1:t+1}, Y_{1:t}, Y_{t+1} = k, X_{t+1} = j|\Phi) \\
&= \sum_i P(L_{1:t+1}, Y_{1:t}, Y_{t+1} = k, X_{t+1} = j, X_t = i|\Phi) \\
&= \sum_i P(L_{t+1}, Y_{t+1} = k|L_{1:t}, Y_{1:t}, X_{t+1} = j, X_t = i, \Phi) \\
&\quad P(X_{t+1} = j|X_t = i, L_{1:t}, Y_{1:t}, \Phi)P(X_t = i, L_{1:t}, Y_{1:t-1}, Y_t|\Phi) \\
&= \sum_i P(L_{t+1}|Y_{t+1} = k, X_{t+1} = j, \Phi)P(Y_{t+1} = k|Y_t, X_{t+1}, \Phi) \\
&\quad P(X_{t+1} = j|X_t = i, \Phi)\alpha_t(i, Y_t) \\
&= \sum_i \alpha_t(i, Y_t)a_{ij}b_{j,Y_t,k}c_{j,k}(l_{t+1})
\end{aligned} \tag{3.16}$$

The forward algorithm can be implemented as follows:

1. Initialization

$$\alpha_1(i,h) = \pi_i b_{i,Y_0,h} c_{i,h}(l_1) \tag{3.17}$$

2. Recursion for $t = 2, ..., T-1$

$$\alpha_{t+1}(j,k) = \sum_i \alpha_t(i, Y_t)a_{ij}b_{j,Y_t,k}c_{j,k}(l_{t+1}) \tag{3.18}$$

3. Termination

$$P(L_{1:T}, Y_{1:T}|\Phi) = \sum_{i=1}^{N} \alpha_T(i, Y_T) \tag{3.19}$$

The forward variable defined here is essential to parameter estimation in the next section.

It is noteworthy that the total output probability might not select the model that performs the best for a certain sequence of hidden states. To address the problem, we can use **optimal output probability** as the criterion instead:

$$\max_{X_{1:T}} P(L_{1:T}, Y_{1:T}, X_{1:T}|\Phi) \tag{3.20}$$

This output probability can also be calculated in a recursive fashion. We define a variable:

$$\delta_t(i) = \max_{X_1,\ldots,X_{t-1}} P(L_{1:t}, Y_{1:t}, X_{1:t-1}, X_t = i|\Phi). \tag{3.21}$$

Then, the algorithm to compute the quantity fast is shown as follows:

1. Initialization

$$\delta_1(i) = \pi_i b_{i,Y_0,Y_1} c_{i,Y_1}(l_1) \tag{3.22}$$

2. Recursion for t = 2 ,..., T

$$\delta_t(j) = \max_i \delta_{t-1}(i) a_{ij} b_{j,Y_{t-1},Y_t} c_{j,Y_t}(l_t) \tag{3.23}$$

3. Termination

$$\max_{X_{1:T}} P(L_{1:T}, Y_{1:T}, X_{1:T}|\Phi) = \max_i \delta_T(i) \tag{3.24}$$

Thus, so far, we have the fast algorithms to compute two distinct output probabilities. These two quantities will be needed during parameter estimation in later section.

3.4.2 Decoding

As mentioned above, decoding is the process of finding the state sequence that best fits the observation. This process is important because in insurance, policyholders in the same risk class should share their risk and pay the same price. However, since a driver's risk class cannot be directly observed, we are unable to distinguish "bad" drivers from "good" drivers. This could lead to cross-subsidization between "bad" drivers and "good" drivers due to asymmetric information (Rothschild and Stiglitz, 1976). If such subsidization is overwhelmed, that is, "good" drivers are over charged much more than their fair premium, "good" drivers will leave the insurance industry, leaving "bad" drivers in the industry, which exacerbates the situation and creates a negative spiral.

Decoding provides a method to reveal the risk class of drivers based on driving behavior and resulting loss. Often the hidden states during decoding have no real life interpretation and thus decoding is not necessary. However, in our case, the hidden states can be meaningfully interpreted as risk classes.

First, we define the optimal state sequence as follows:

$$X^*_{1:T} = \underset{X_{1:T}}{\operatorname{argmax}} P(X_{1:T}|L_{1:T}, Y_{1:T}, \Phi) \tag{3.25}$$

This quantity can be calculated directly given model parameters, which is computationally expensive. Whereas, the **Viterbi algorithm** proposed by Viterbi (1967) is computationally efficient. Using the algorithm defined in Section 3.4.1 to calculate the optimal output probability, one can find the optimal path. The Viterbi algorithm is performed recursively and the optimal path is determined only when the algorithm reaches the end of the observation sequence. Let $\psi_1(i) := 0$ and $\psi_{t+1}(j) := \operatorname{argmax}_i \delta_t(i)a_{ij}$. The optimal hidden state, $X_{1:T}$ is picked up from the end of the algorithm. The probability of a sequence of hidden states is given below:

$$P(X_{1:T}|L_{1:T}, Y_{1:T}, \Phi) = \frac{P(X_{1:T}, L_{1:T}, Y_{1:T}|\Phi)}{P(L_{1:T}, Y_{1:T}|\Phi)}. \tag{3.26}$$

$P(L_{1:T}, Y_{1:T}|\Phi)$ is irrelevant since it is a constant for a fixed model and a given observation sequence. Thus, the optimal hidden state path can be the choice of the optimal output probability.

$$X^*_{1:T} = \underset{X_{1:T}}{\operatorname{argmax}} P(X_{1:T}|L_{1:T}, Y_{1:T}, \Phi) = \underset{X_{1:T}}{\operatorname{argmax}} P(X_{1:T}, L_{1:T}, Y_{1:T}|\Phi) \tag{3.27}$$

Then, using the Viterbi algorithm, we can identify the best hidden states one by one in a recursive fashion as shown below:[3]

1. Initialization

$$\delta_1(i) = \pi_i b_{i,Y_0,Y_1} c_{i,Y_1}(l_1) \tag{3.28}$$

2. Recursion for all times $t = 1, 2, ..., T-1$, and each hidden state j.

$$\delta_{t+1}(j) = \max_i \delta_t(i) a_{ij} b_{j,Y_t,Y_{t+1}} c_{j,Y_{t+1}}(l_{t+1}) \tag{3.29}$$

$$\psi_{t+1}(j) = \operatorname*{argmax}_i \delta_t(i) a_{ij} \tag{3.30}$$

3. Termination

$$\max_{X_{1:T}} P(X_{1:T}, L_{1:T}, Y_{1:T}|\Phi) = \max_i \delta_T(i) \tag{3.31}$$

$$X_T^* = \operatorname*{argmax}_j \delta_T(j) \tag{3.32}$$

4. Back tracking of the optimal path for $t = T-1, ..., 1$

$$X_t^* = \psi_{t+1}(X_{t+1}^*) \tag{3.33}$$

3.5 Estimation

With the variables, quantities, and algorithms introduced in Section 3.4, we can proceed to the estimation of the 3-layer HMM parameters. In this section, the EM algorithm, also known as the Baum-Welch algorithm (Baum et al., 1970) (focused on maximizing the total output probability), is used to estimate the parameters.

We first define several quantities that will be useful in the EM algorithm. In addition to the forward variable defined in Equation (3.15), we also define the backward variable (Rabiner, 1989) as follows:

$$\beta_t(j,k) = P(L_{t+1:T}, Y_{t+1:T}|Y_t = k, X_t = j, \Phi) \tag{3.34}$$

[3]For practical application, this reversal of the direction of the order of calculations implies a significant limitation. The optimal state sequence can be determined only after the end of the observation sequence is reached.

We claim in Proposition 3 that the backward variable can be written in a recursive fashion just like the forward variable, as shown below:

Proposition 3

$$\beta_t(j,k) = \sum_{X_{t+1}=i} \beta_{t+1}(i, Y_{t+1})\alpha_{ji} b_{i,k,Y_{t+1}} c_{i,Y_{t+1}}(l_{t+1}) \tag{3.35}$$

The proof of this proposition is shown in Appendix C.

Another important variable to define is the probability of being in state i at time t given an observation sequence:

$$\gamma_t(i) = P(X_t = i | L_{1:T}, Y_{1:T}, \Phi) \text{ for } i \in \{1, ..., N\}, \qquad t \geq 0. \tag{3.36}$$

This can be re-expressed using the forward and backward variables as

$$\begin{aligned}
\gamma_t(i) &= \frac{P(X_t = i, L_{1:T}, Y_{1:T} | \Phi)}{P(L_{1:T}, Y_{1:T} | \Phi)} \\
&= \frac{P(L_{t+1:T}, Y_{t+1:T} | X_t = i, L_{1:t}, Y_{1:t}, \Phi) P(X_t = i, L_{1:t}, Y_{1:t} | \Phi)}{P(L_{1:T}, Y_{1:T} | \Phi)} \\
&= \frac{P(L_{t+1:T}, Y_{t+1:T} | X_t = i, Y_t, \Phi) P(L_{1:t}, Y_{1:t}, X_t = i | \Phi)}{P(L_{1:T}, Y_{1:T} | \Phi)} \\
&= \frac{\alpha_t(i, Y_t) \beta_t(i, Y_t)}{P(L_{1:T}, Y_{1:T} | \Phi)}
\end{aligned} \tag{3.37}$$

Furthermore, $\gamma_t(i)$ can be decomposed into the sum of the probability of going from state i to j at time t, i.e., $\gamma_t(i) = \sum_j \gamma_t(i,j)$, where

$$\gamma_t(i,j) = P(X_t = i, X_{t+1} = j | L_{1:T}, Y_{1:T}, \Phi) = \frac{P(X_t = i, X_{t+1} = j, L_{1:T}, Y_{1:T} | \Phi)}{P(L_{1:T}, Y_{1:T} | \Phi)}, \tag{3.38}$$

for $i, j \in \{1, ..., N\}$. Note that the numerator of Equation (3.38) can be determined as

$$
\begin{aligned}
& P(X_t = i, X_{t+1} = j, L_{1:T}, Y_{1:T}|\Phi) \\
= \; & P(L_{1:t}, Y_{1:t}, X_t = i|\Phi) P(L_{t+1:T}, Y_{t+1:T}, X_{t+1} = j|L_{1:t}, Y_{1:t}, X_t = i, \Phi) \\
= \; & \alpha_t(i, Y_t) P(L_{t+1:T}, Y_{t+1:T}|X_{t+1} = j, X_t = i, Y_t, \Phi) P(X_{t+1}|X_t = i, \Phi) \\
= \; & \alpha_t(i, Y_t) a_{ij} P(L_{t+1}, Y_{t+1}|X_{t+1} = j, Y_t, |\Phi) \\
& P(L_{t+2:T}, Y_{t+2:T}|L_{t+1}, Y_{t+1}, X_{t+1} = j, Y_t, \Phi) \\
= \; & \alpha_t(i, Y_t) a_{ij} P(L_{t+1}, Y_{t+1}|X_{t+1} = j, Y_t, |\Phi) P(L_{t+2:T}, Y_{t+2:T}|Y_{t+1}, X_{t+1} = j, \Phi) \\
= \; & \alpha_t(i, Y_t) a_{ij} c_{j,Y_{t+1}}(l_{t+1}) b_{j,Y_t,Y_{t+1}} \beta_{t+1}(j, Y_{t+1}). \qquad (3.39)
\end{aligned}
$$

It follows that

$$\gamma_t(i, j) = \frac{\alpha_t(i, Y_t) a_{ij} c_{j,Y_{t+1}}(l_{t+1}) b_{j,Y_t,Y_{t+1}} \beta_{t+1}(j, Y_{t+1})}{P(L_{1:T}, Y_{1:T}|\Phi)}. \qquad (3.40)$$

The above three quantities, $\beta_t(j, k), \gamma_t(i)$, and $\gamma_t(i, j)$, are very important in the derivation of the EM algorithm. The EM algorithm has two steps: E step and M step. In the first step, we define the likelihood function for all variables, including values of X_t, Y_t, and L_t:

$$
\begin{aligned}
L_T^C(\Phi) &= P(L_{1:T} = l_{1:T}, Y_{1:T} = y_{1:T}, X_{1:T} = x_{1:T}) \\
&= \pi_{x_1} b_{x_1,y_0,y_1} c_{x_1,y_1}(l_1) \prod_{t=2}^{T} a_{x_{t-1},x_t} b_{x_t,y_{t-1},y_t} c_{x_t,y_t}(l_t)
\end{aligned} \qquad (3.41)
$$

Summing over $x_1, ..., x_T$, we obtain the likelihood for the data which includes only the observables, Y_t and L_t, i.e.,

$$
\begin{aligned}
L_T(\Phi) &= P(L_{1:T} = l_{1:T}, Y_{1:T} = y_{1:T}) \\
&= \sum_{X_1} ... \sum_{X_T} \pi_{X_1} b_{X_1,y_0,y_1} c_{X_1,y_1}(l_1) \prod_{t=2}^{T} a_{x_{t-1},x_t} b_{x_t,y_{t-1},y_t} c_{x_t,y_t}(l_t) \qquad (3.42)
\end{aligned}
$$

The goal is to maximize the expected log-likelihood function given complete data. Given a set of observed sequences, $O_{1:T} = \{L_{1:T}, Y_{1:T}\}$, the detailed steps of the EM algorithm are listed below:

1. Initialize the parameters in Φ^0, where the superscript stands for the iteration.

2. Compute the expected log-likelihood function given complete data over $X_{1:T}$ given observations and parameters from last iteration, that is,[4]

$$Q(\Phi; \Phi^k) = E_{X_{1:T}|O_{1:T},\Phi^k}[ln(L_T^C(\Phi; X_{1:T}, O_{1:T}))] \tag{3.43}$$

3. Find Φ^{k+1} that maximizes $Q(\Phi; \Phi^k)$

$$\Phi^{k+1} = \underset{\Phi}{\operatorname{argmax}}\, Q(\Phi; \Phi^k) \tag{3.44}$$

4. Repeat step 2 and 3 alternately until $ln(L_T(\Phi^{k+1})) - ln(L_T(\Phi^k))$ is sufficiently small.

Wu (1983) showed that the EM algorithm converges to at least a local maximum under certain conditions. In what follows, we provide more technical details for implementation of EM algorithm on the 3-layer HMM.

3.5.1 EM Algorithm on HMM: Hidden States and Driving Behavior

This subsection develops formulas for parameters of hidden states X_t and driving behavior Y_t. The first step is to initialize the parameters Φ^0. Then, in the E step, we need to calculate $E_{X_{1:T}|O_{1:T},\Phi^0}[ln(L_T^C(\Phi; X_{1:T}, O_{1:T}))]$, where $L_T^C(\Phi)$ is specified in Equation (3.41). After some manipulations, one finds that

[4]In this step, Φ and Φ^k have the same value. But the maximization step only optimizes Φ.

$$
\begin{aligned}
& E_{X_{1:T}|O_{1:T},\Phi^0}[ln(L_T^C(\Phi; X_{1:T}, O_{1:T}))] \\
= \; & E_{X_{1:T}|O_{1:T},\Phi^0}[ln(\pi_{X_1} b_{X_1,y_0,y_1} c_{X_1,y_1}(l_1) \prod_{t=2}^{T} a_{x_{t-1},x_t} b_{x_t,y_{t-1},y_t} c_{x_t,y_t}(l_t))|\Phi; X_{1:T}, O_{1:T}] \\
= \; & E_{X_{1:T}|O_{1:T},\Phi^0}[ln(\pi_{X_1})|\Phi; X_{1:T}, O_{1:T}] + \sum_{t=2}^{T} E_{X_{1:T}|O_{1:T},\Phi^0}[ln(a_{X_{t-1},X_t})|\Phi; X_{1:T}, O_{1:T}] \\
& + \sum_{t=1}^{T} E_{X_{1:T}|O_{1:T},\Phi^0}[ln(b_{X_t,y_{t-1},y_t}|\Phi; X_{1:T}, O_{1:T}] + \\
& \sum_{t=1}^{T} E_{X_{1:T}|O_{1:T},\Phi^0}[ln(c_{X_t,y_t}(l_t))|\Phi; X_{1:T}, O_{1:T}] \\
= \; & \sum_{i=1}^{N} P(X_1 = i|O_{1:T}, \Phi^0) ln(\pi_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T} ln(a_{ij}) P(X_{t-1} = i, X_t = j|O_{1:T}, \Phi^0) \\
& + \sum_{i=1}^{N}\sum_{t=1}^{T} ln(b_{i,y_{t-1},y_t}) P(X_t = i|O_{1:T}, \Phi^0) + \sum_{i=1}^{N}\sum_{t=1}^{T} ln(c_{i,y_t}(l_t)) P(X_t = i|O_{1:T}, \Phi^0) \\
= \; & \sum_{i=1}^{N} \gamma_1(i) ln(\pi_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T} ln(a_{ij}) \gamma_{t-1}(i,j) \\
& + \sum_{i=1}^{N}\sum_{t=1}^{T} ln(b_{i,y_{t-1},y_t}) \gamma_t(i) + \sum_{i=1}^{N}\sum_{t=1}^{T} ln(c_{i,y_t}(l_t)) \gamma_t(i). \qquad (3.45)
\end{aligned}
$$

Equation (3.45) can be calculated With an initial set of parameters Φ^0, and this completes the E step. The M step maximizes (3.45) with respect to Φ:

$$
\max_{\text{w.r.t } \Phi=\{\Pi,\mathbb{A},\mathbb{B},\mathbb{C}\}} Q(\Phi; \Phi^0) = E_{X_{1:T}|O_{1:T},\Phi^0}[ln(L_T^C(\Phi; X_{1:T}, O_{1:T}))] \qquad (3.46)
$$

subject to the following constraints:

1. $g_1(\Pi) : \sum_{i=1}^{N} \pi_i = 1$
2. $g_{2i}(\mathbb{A}) : \sum_{j=1}^{N} a_{ij} = 1$ for all $i \in \{1, 2, ..., N\}$. (N constraints)
3. $g_{3ij}(\mathbb{B}) : \sum_{k=1}^{M} b_{ijk} = 1$ for all $i \in \{1, 2, ..., N\}$ and $j \in \{1, 2, ..., M\}$. ($N \times M$ constraints)
4. $g_{4ij}(\mathbb{C}) : \int_l c_{ij}(l) = 1$ for all $i \in \{1, 2, ..., N\}$ and $j \in \{1, 2, ..., M\}$. ($N \times M$ constraints)

To do the maximization, we introduce a new function F using Lagrange multiplier theory for constrained optimization.

$$\begin{aligned} F(\Phi^0, \kappa) &= Q(\Phi;\Phi^0) + \kappa_1(1 - \sum_{i=1}^{N} \pi_i) + \sum_{i=1}^{N} \kappa_{2i}(1 - \sum_{j=1}^{N} a_{ij}) \\ &+ \sum_{i=1}^{N}\sum_{j=1}^{M} \kappa_{3ij}(1 - \sum_{k=1}^{M} b_{ijk}) + \sum_{i=1}^{N}\sum_{j=1}^{M} \kappa_{4ij}(1 - \int_l c_{i,j}(l)) \end{aligned} \tag{3.47}$$

where $\kappa = \{\kappa_1, \kappa_{2i}, \kappa_{3ij}, \kappa_{4ij}\}$ are the Lagrange multipliers.

We first consider the initial state probability π_i. Taking the derivative of F with respect to π_i and setting to zero, we have

$$\frac{\partial F}{\partial \pi_i} = \frac{\gamma_1(i)}{\pi_i} - \kappa_1 = 0 \tag{3.48}$$

for all $i \in \{1, 2, ..., N\}$. It follows that

$$\pi_i = \frac{\gamma_1(i)}{\kappa_1}. \tag{3.49}$$

Substituting Equation (3.49) into constraint (1), we obtain

$$g_1(\Pi) = \sum_{i=1}^{N} \pi_i = \sum_{i=1}^{N} \frac{\gamma_1(i)}{\kappa_1} = 1,$$

and $\kappa_1 = \sum_{i=1}^{N} \gamma_1(i) = 1$. Therefore, the optimal initial state probability is given by

$$\pi_i^* = \gamma_1(i), \text{ for all } i \in \{1, 2, ..., N\}. \tag{3.50}$$

Next, we optimize the transition probability in a similar fashion. Take the derivative with respect to a_{ij} and setting it to zero, we have

$$\frac{\partial F}{\partial a_{ij}} = \frac{\sum_{t=2}^{T} \gamma_{t-1}(i,j)}{a_{ij}} - \kappa_{2i} = 0 \quad \text{for all } i, j \in \{1, 2, ..., N\}. \tag{3.51}$$

Utilizing constraint (2), one finds that

$$a_{ij}^* = \frac{\sum_{t=2}^{T} \gamma_{t-1}(i,j)}{\sum_{t=2}^{T} \gamma_{t-1}(i)} \quad \text{for all } i, j \in \{1, 2, ..., N\}. \tag{3.52}$$

Third, applying the same procedure, the emission probabilities satisfy

$$\frac{\partial F}{\partial b_{ikh}} = \frac{\sum_{t:y_{t-1}=k,y_t=h} \gamma_t(i)}{b_{ikh}} - \kappa_{3ik} = 0. \tag{3.53}$$

Under constraint (3), we obtain the following:

$$\sum_{h=1}^{M} b_{ikh} = 1 = \sum_{h=1}^{M} \frac{\sum_{t:y_{t-1}=k,y_t=h} \gamma_t(i)}{\kappa_{3ik}} \tag{3.54}$$

and

$$\kappa_{3ik} = \sum_{h=1}^{M} \sum_{t:y_{t-1}=k,y_t=h} \gamma_t(i) = \sum_{t:y_{t-1}=k} \gamma_t(i). \tag{3.55}$$

Thus,

$$b_{ikh} = \frac{\sum_{t:y_{t-1}=k,y_t=h} \gamma_t(i)}{\sum_{t:y_{t-1}=k} \gamma_t(i)}. \tag{3.56}$$

Lastly, loss parameters can be obtained by taking derivatives with respect to $c_{ij}(L_t = l_t)$ and set them to zero. However, loss related parameters largely depend on the form of $c_{ij}(L_t)$ and will be demonstrated in the next subsection after specifying the model assumptions for loss frequency and severity.

It is worth noting that setting first derivatives to zero only enables us to find the critical point. To further check whether the obtained critical point is a local maximum, we examine whether the associated Hessian matrix is negative definite. The Hessian matrix can be written as

$$H = \begin{bmatrix} \frac{\partial^2 F}{\partial \pi_1^2} & 0 & \dots & 0 \\ 0 & \frac{\partial^2 F}{\partial \pi_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \frac{\partial^2 F}{\partial c_{ij}(L_t=l_t)^2} \end{bmatrix}, \tag{3.57}$$

or alternatively

$$H = \begin{bmatrix} \frac{-r_1(1)}{\pi_1^2} & 0 & \dots & 0 \\ 0 & \frac{-r_1(2)}{\pi_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & . \end{bmatrix}. \tag{3.58}$$

When only the parameters for hidden states and driving behavior are considered, the corresponding part of the Hessian matrix is negative definite, and the point that the parameters are approaching using the EM algorithm is a local maximum. However, to show that the whole Hessian matrix is negative definite, the assumptions of loss distribution are necessary. The next section adds the detailed implementation of the EM algorithm on loss related parameters.

3.5.2 EM Algorithm on HMM: Loss Frequency and Loss Severity

The previous derivation gives a general way to estimate parameters of hidden states and driving behavior without incorporating loss distribution. This section, we derive a more detailed solution within an auto insurance context. Specifically, we assume the probabilistic distributions for both loss frequency and loss severity. It leads to the last proposition:

Proposition 4 *The Hessian matrix, H, of the 3-layer HMM is negative definite, when loss frequency at any single time point follows Bernoulli distribution and loss severity follows Gamma distribution.*

Let L be the loss of a policyholder in auto insurance. In actuarial science, we can use a collective risk model to write L. For a period from 0 to T:

$$L(0,T) = \sum_{n=1}^{N(0,T)} S_n, \tag{3.59}$$

where $N(0,T)$ is the number of claims during $(0,T)$ and S_n is the claim severity for the n^{th} claim. For illustrative purpose, We assume that S_n follows a gamma distribution with parameters k_n and θ_n. In our setting, the observation frequency is every second, and at most one claim can be observed at each time point. Therefore, we assume that the number of claims at each time unit, N_t, follows a Bernoulli distribution, where t is the unit of time (i.e., one second) for the observation.[5]

[5]We have the following relation: $N(0,T) = \sum_{i=1}^{T} N_i$.

As in traditional setting, we can assume that N_t and S_t are independent of each other. S_t now becomes the claim severity at time t given a claim occurrence. Also, both N_t and S_t depend on the hidden state, X and the driving behavior, Y. Thus, we have the following assumptions about loss at time t:[6]

1. $L_t = N_t S_t$.

2. N_t follows Bernoulli distribution with parameter $p(X_t, Y_t)$.

3. S_t follows gamma distribution with parameters $k(X_t, Y_t)$ and $\theta(X_t, Y_t)$.

Then, the parameter space of the model becomes

$$\Phi = \{\mathbb{A}, \Pi, \mathbb{B}, \mathbb{P}, \mathbb{K}, \Theta\}, \tag{3.60}$$

where $\mathbb{P}$ is a matrix of the parameters for the Bernoulli distribution:

$$\mathbb{P} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & M \end{array} \\ \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & \cdots & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & \cdots & \cdots & p_{NM} \end{pmatrix} \end{array} \tag{3.61}$$

Similarly, $\mathbb{K}$ and Θ are matrices of the parameters for gamma distributions.

Since L_t is fully characterized by the random vector of observations $L_t = (N_t, S_t)$, we proceed the E step of EM algorithm as follows:

$$\begin{aligned} Q(\Phi; \Phi_0) &= E_{X_{1:T}|O_{1:T},\Phi^0}[ln(L_T^C(\Phi; X_{1:T}, O_{1:T}))] \\ &= \sum_{i=1}^{N} \gamma_1(i) ln(\pi_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T} ln(a_{ij})\gamma_{t-1}(i,j) \\ &\quad + \sum_{i=1}^{N}\sum_{t=1}^{T} ln(b_{i,y_{t-1},y_t})\gamma_t(i) + \sum_{i=1}^{N}\sum_{t=1}^{T} ln(c_{i,y_t}(l_t))\gamma_t(i). \end{aligned} \tag{3.62}$$

[6]For different objectives, N_t and S_t can have different distributions. For example, S_t can follow Pareto distribution if the objective is to estimate extreme loss. The derivation demonstrated in this section can be easily modified to accommodate different loss distributions.

Note that $c_{i,y_t}(l_t)$ satisfies

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T} ln(c_{i,y_t}(l_t))\gamma_t(i) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} ln(Pr(N_t S_T = n_t s_t | X_t = i, Y_t = y_t))\gamma_t(i) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} ln(Pr(N_t = n_t | X_t = i, Y_t = y_t) Pr(S_t = s_t | X_t = i, Y_t = y_t))\gamma_t(i) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} \left[ln(Pr(N_t = n_t | X_t = i, Y_t = y_t)) + ln(Pr(S_t = s_t | X_t = i, Y_t = y_t))\right]\gamma_t(i).
\end{aligned} \tag{3.63}$$

Substituting (3.63) into (3.62), the Q function can be rewritten as

$$\begin{aligned}
Q(\Phi; \Phi_0) &= E_{X_{1:T}|O_{1:T},\Phi^0}[ln(L_T^C(\Phi; X_{1:T}, O_{1:T}))] \\
&= \sum_{i=1}^{N} \gamma_1(i) ln(\pi_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T} ln(a_{ij})\gamma_{t-1}(i,j) \\
&+ \sum_{i=1}^{N}\sum_{t=1}^{T} ln(b_{i,y_{t-1},y_t})\gamma_t(i) + \sum_{i=1}^{N}\sum_{t=1}^{T} ln(Pr(N_t = n_t | X_t = i, Y_t = y_t))\gamma_t(i) \\
&+ \sum_{i=1}^{N}\sum_{t=1}^{T} ln(Pr(S_t = s_t | X_t = i, Y_t = y_t))\gamma_t(i).
\end{aligned} \tag{3.64}$$

In the M step, the estimation of the parameters of hidden states and driving behavior remain the same as in Section 5.1, and we only need to further obtain the optimal loss parameters. Taking partial derivatives of Q with respect to p_{ij} and focusing only on the terms involving p_{ij}, we have

$$\begin{aligned}
\frac{\partial Q}{\partial p_{i,y_t}} &= \sum_{t:Y_t=y_t} \left(\frac{n_t}{p_{i,y_t}} + \frac{1-n_t}{p_{i,y_t}-1}\right)\gamma_t(i) = 0 \\
&\Rightarrow \sum_{t:Y_t=y_t} \left[(p_{i,y_t}-1)n_t + p_{i,y_t}(1-n_t)\right]\gamma_t(i) = 0 \\
&\Rightarrow \sum_{t:Y_t=y_t} (p_{i,y_t} - n_t)\gamma_t(i) = 0 \\
&\Rightarrow p_{i,y_t} = \frac{\sum_{t:Y_t=y_t} n_t\gamma_t(i)}{\sum_{t:Y_t=y_t} \gamma_t(i)}.
\end{aligned} \tag{3.65}$$

Similarly, we can obtain the parameter matrices $\mathbb{K}$ and Θ for Gamma distributions, where the probability density function (pdf) is given by

$$f(S_t = s_t | X_t = i, Y_t = y_t) = \frac{s_t^{k_{i,y_t}-1} e^{-\frac{s_t}{\theta_{i,y_t}}}}{\Gamma(k_{i,y_t})\theta_{i,y_t}^{k_{i,y_t}}}. \tag{3.66}$$

The relevant part of Q for determining $\mathbb{K}$ and Θ is

$$\begin{aligned} &\sum_{i=1}^{N}\sum_{t=1}^{T} ln(Pr(S_t = s_t | X_t = i, Y_t = y_t))\gamma_t(i) \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} \big[(k_{i,y_t} - 1)ln(s_t) - \frac{s_t}{\theta_{i,y_t}} - ln(\Gamma(k_{i,y_t})) - k_{i,y_t} ln(\theta_{i,y_t})\big]\gamma_t(i). \end{aligned} \tag{3.67}$$

By setting the partial derivative of Q with respect to Θ to zero, we obtain

$$\begin{aligned} \frac{\partial Q}{\partial \theta_{i,y_t}} &= \sum_{t:Y_t=y_t} (\frac{s_t}{\theta_{i,y_t}^2} - \frac{k_{i,y_t}}{\theta_{i,y_t}})\gamma_t(i) = 0 \\ &\Rightarrow \sum_{t:Y_t=y_t} (s_t\gamma_t(i) - k_{i,y_t}\theta_{i,y_t}\gamma_t(i)) = 0 \\ &\Rightarrow \theta_{i,y_t} = \frac{\sum_{t:Y_t=y_t} s_t\gamma_t(i)}{\sum_{t:Y_t=y_t} k_{i,y_t}\gamma_t(i)} \\ &\Rightarrow \theta_{i,y_t} = \frac{\bar{s}_t}{k_{i,y_t}} \text{ where } \bar{s}_t = \frac{\sum_{t:Y_t=y_t} s_t\gamma_t(i)}{\sum_{t:Y_t=y_t} \gamma_t(i)}. \end{aligned} \tag{3.68}$$

Furthermore, taking partial derivatives with respect to k_{i,y_t}, one arrives at

$$\sum_{t:Y_t=y_t} \big[ln(s_t) - \frac{s_t}{\bar{s}_t}\psi_0(k_{i,y_t}) - ln(\frac{\bar{s}_t}{k_{i,y_t}}) + 1\big]\gamma_t(i) = 0, \tag{3.69}$$

where $\psi_0(k_{i,y_t})$ is the digamma function. Equation (3.69) can be evaluated numerically.

Finally, we examine whether the Hessian matrix is negative definite under this model assumption. The Hessian matrix is again diangonal one, i.e.,

$$H = \begin{bmatrix} \frac{-r_1(1)}{\pi_1^2} & 0 & \dots & 0 \\ 0 & \frac{-r_1(2)}{\pi_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sum_{t:Y_t=y_t}(-\psi_1(k_{NM}) + \frac{1}{k_{NM}})\gamma_t(N) \end{bmatrix}, \tag{3.70}$$

as the cross second derivatives with respect to two different parameters θ_{ij} and k_{ij} are zero as before. Besides, the second derivative of Q with respect to p_{i,y_t} is smaller than zero:

$$\frac{\partial^2 Q}{\partial p_{i,y_t}^2} = \sum_{t:Y_t=y_t} (-\frac{n_t}{p_{i,y_t}^2} - \frac{(1-n_t)}{(p_{i,y_t}-1)})\gamma_t(i) < 0. \tag{3.71}$$

The second derivative of Q with respect to k_{i,y_t} can be expressed as

$$\frac{\partial^2 Q}{\partial k_{i,y_t}^2} = \sum_{t:Y_t=y_t} (-\psi_1(k_{i,y_t}) + \frac{1}{k_{i,y_t}})\gamma_t(i), \tag{3.72}$$

where $\psi_1(k_{i,y_t}) = \psi_0^{'}(k_{i,y_t})$ is the trigamma function. Since $\gamma_t(i) > 0$, $k_{i,y_t} > 0$, and $\psi_1(k_{i,y_t}) > 0$, we need to show that $\psi_1(k_{i,y_t}) > \frac{1}{k_{i,y_t}}$ so that $\frac{\partial^2 Q}{\partial k_{i,y_t}^2} < 0$. We discuss in two different cases: $(0, 1]$ and $(1, \infty)$.

When $k_{i,y_t} \in (0, 1)$, the trigamma function can be re-expressed as:

$$\psi_1(k_{i,y_t}) = \sum_{p=0}^{\infty} \frac{1}{(p+k_{i,y_t})^2} = \frac{1}{k_{i,y_t}^2} + \sum_{p=1}^{\infty} \frac{1}{(p+k_{i,y_t})^2} > \frac{1}{k_{i,y_t}} \tag{3.73}$$

(Sebah and Gourdon, 2002). Clearly, the inequality also holds for $k_{i,y_t} = 1$.

When $k_{i,y_t} \in (1, \infty)$, let

$$\rho_1(k_{1,y_t}) = (\psi_1)^{-1}(\frac{1}{k_{i,y_t}}) - k_{i,y_t}. \tag{3.74}$$

Batir (2007) shows that $\rho_1(k_{i,y_t}) > 0$ for all $k_{i,y_t} > 0$. Since $\psi_1(k_{i,y_t})$ is a decreasing function of k_{i,y_t}, we have

$$\frac{1}{k_{i,y_t}} = \psi_1(k_{i,y_t} + \rho_1(k_{i,y_t})) < \psi_1(k_{i,y_t}), \tag{3.75}$$

which is equivalent to $\frac{\partial^2 Q}{\partial k_{i,y_t}^2} < 0$.

In summary, the second derivatives of all loss parameters are negative, and the Hessian matrix is negative definite, which completes the proof of Proposition 4.

3.6 Conclusion

In this study, we propose a 3-layer HMM which links risk classes, driving behavior, and losses in an integrated framework. The proposed model has several advantages.

First, the 3-layer HMM mitigates the number of intermediate steps typically needed in existing predictive models, such as feature selections, for estimating loss frequency and loss severity. Second, the model significantly preserves the time series structure of telematics data and directly predicts losses with minimal telematics feature aggregation. Utilizing the Markov property of the model, the study also shows that the expected loss per unit of time can be asymptotically estimated by exposure-weighted rates of different risk classes. Finally, we show that the proposed model can be estimated in an efficient way. In particular, a set of equations based on the EM algorithm are provided for parameter estimation, when Gamma loss severity is assumed. Other distributional models may be analyzed in a similar pattern.

REFERENCES CITED

Agrawal, A. and G. N. Mandelker (1987). Managerial incentives and corporate investment and financing decisions. *The Journal of Finance 42*(4), 823–837.

Ahmed, M., R. Seraj, and S. M. S. Islam (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics 9*(8), 1295.

Al-Sultan, S., A. H. Al-Bayatti, and H. Zedan (2013). Context-aware driver behavior detection system in intelligent transportation systems. *IEEE Transactions on Vehicular Technology 62*(9), 4264–4275.

Allegaert, T. (1996). Derivative actions by policyholders on behalf of mutual insurance companies. *The University of Chicago Law Review 63*(3), 1063–1097.

Amsalu, S. B. and A. Homaifar (2016). Driver behavior modeling near intersections using hidden markov model based on genetic algorithm. In *2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, pp. 193–200.

Aoude, G. S., V. R. Desaraju, L. H. Stephens, and J. P. How (2012). Driver behavior classification at intersections and validation on large naturalistic data set. *IEEE Transactions on Intelligent Transportation Systems 13*(2), 724–736.

Appel, I. (2019). Governance by litigation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2532278.

Ayuso, M., M. Guillén, and A. M. Pérez-Marín (2014). Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis & Prevention 73*, 125–131.

Ayuso, M., M. Guillen, and A. M. Pérez-Marín (2016). Telematics and gender discrimination: Some usage-based evidence on whether men's risk of accidents differs from women's. *Risks 4*(2), 10.

Baranoff, E. G. and T. W. Sager (2002). The relations among asset risk, product risk, and capital in the life insurance industry. *Journal of Banking Finance 26*(6), 1181–1197.

Baranoff, E. G. and T. W. Sager (2003). The relations among organizational and distribution forms and capital and asset risk structures in the life insurance industry. *Journal of Risk and Insurance 70*(3), 375–400.

Batir, N. (2007). On some properties of digamma and polygamma functions. *Journal of Mathematical Analysis and Applications 328*(1), 452–465.

Baum, L. E. and T. Petrie (1966). Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics 37*(6), 1554–1563.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics 41*(1), 164–171.

Berndt, H., J. Emmert, and K. Dietmayer (2008). Continuous driver intention recognition with hidden markov models. In *2008 11th International IEEE Conference on Intelligent Transportation Systems*, pp. 1189–1194.

Bertrand, M. and S. Mullainathan (2003). Enjoying the quiet life? corporate governance and managerial preferences. *Journal of Political Economy 111*(5), 1043–1075.

Bholowalia, P. and A. Kumar (2014). Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications 105*(9), 17–24.

Blier-Wong, C., H. Cossette, L. Lamontagne, and E. Marceau (2021). Machine learning in p&c insurance: A review for pricing and reserving. *Risks 9*(1), 4.

Block, D. J., S. A. Radin, and M. J. Maimone (1993). Derivative litigation: Current law versus the american law institute. *The Business Lawyer 48*(4), 1443–1483.

Boucher, J.-P., S. Côté, and M. Guillen (2017). Exposure as duration and distance in telematics motor insurance using generalized additive models. *Risks 5*(4), 54.

Bourveau, T., Y. Lou, and R. Wang (2018). Shareholder litigation and corporate disclosure: Evidence from derivative lawsuits. *Journal of Accounting Research 56*(3), 797–842.

Boyraz, P., M. Acar, and D. Kerr (2007). Signal modelling and hidden markov models for driving manoeuvre recognition and driver fault diagnosis in an urban road scenario. In *2007 IEEE Intelligent Vehicles Symposium*, pp. 987–992.

Che, X., A. Liebenberg, and J. Xu (2021). Usage-based insurance - impact on insurers and potential implications for insurtech. *North American Actuarial Journal 0*(0), 1–28.

Cheng, J., E. Elyasiani, and J. J. Jia (2011). Institutional ownership stability and risk taking: Evidence from the life—health insurance industry. *The Journal of Risk and Insurance 78*(3), 609–641.

Chu, Y. and Y. E. Zhao (2021). The dark side of shareholder litigation: Evidence from corporate takeovers. *Financial Management 50*(3), 845–873.

Cox, J. D. (1999). The social meaning of shareholder suits. *Brooklyn Law Review 65*, 3–45.

Craye, C., A. Rashwan, M. S. Kamel, and F. Karray (2016). A multi-modal driver fatigue and distraction assessment system. *International Journal of Intelligent Transportation Systems Research 14*, 173–194.

Cummins, J. A. and N. A. Doherty (2002). Capitalization of the property-liability insurance industry: Overview. *Journal of Financial Services Research 21*, 5–14.

Cummins, J. D. and D. W. Sommer (1996). Capital and risk in property-liability insurance markets. *Journal of Banking Finance 20*(6), 1069–1092.

Cummins, J. D., M. A. Weiss, and H. Zi (1999). Organizational form and efficiency: The coexistence of stock and mutual property-liability insurers. *Management Science 45*(9), 1254–1269.

Dalla Pellegrina, L. and M. Saraceno (2016). Can shareholder litigation discipline ceo bonuses in the financial sector? the role of securities class actions. *Economic Notes 45*(1), 3–36.

Deng, Q. and D. Söffker (2019). Classifying human behaviors: Improving training of conventional algorithms. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 1060–1065.

Deng, Q. and D. Söffker (2022). A review of hmm-based approaches of driving behaviors recognition and prediction. *IEEE Transactions on Intelligent Vehicles 7*(1), 21–31.

Deng, Q., J. Wang, K. Hillebrand, C. R. Benjamin, and D. Söffker (2020). Prediction performance of lane changing behaviors: A study of combining environmental and eye-tracking data in a driving simulator. *IEEE Transactions on Intelligent Transportation Systems 21*(8), 3561–3570.

Deng, Q., J. Wang, and D. Soffker (2018). Prediction of human driver behaviors based on an improved hmm approach. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 2066–2071.

Denuit, M., M. Guillen, and J. Trufin (2019). Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Annals of Actuarial Science 13*(2), 378–399.

Derbel, O. and R. J. Landry (2015). Driving style assessment based on the gps data and fuzzy inference systems. In *2015 IEEE 12th International Multi-Conference on Systems, Signals & Devices (SSD15)*, pp. 1–8.

Desai, J., H. Li, J. K. Mathew, Y.-T. Cheng, A. Habib, and D. M. Bullock (2021). Correlating hard-braking activity with crash occurrences on interstate construction projects in indiana. *Journal of Big Data Analytics in Transportation 3*, 27–41.

Dionne, G., R. Gagné, A. Nouira, and J. D. Cummins (2007). Determinants of insurers' performance in risk pooling, risk management, and financial intermediation activities. https://ssrn.com/abstract=988991.

Donelson, D. C. D., L. Kettell, J. McInnis, and S. Toynbee (2022). The need to validate exogenous shocks: Shareholder derivative litigation, universal demand laws and firm behavior. *Journal of Accounting and Economics 73*(1), 101427.

Duval, F., J.-P. Boucher, and M. Pigeon (2022). How much telematics information do insurers need for claim classification? *North American Actuarial Journal 0*(0), 1–21.

Easterbrook, F. H. and D. R. Fischel (1991). *The Economic Structure of Corporate Law*. Harvard University Press.

Erickson, J. (2010). Corporate governance in the courtroom: An empirical analysis. *William and Mary Law Review 51*, 1749 – 1831.

Erickson, J. (2017). The gatekeepers of shareholder litigation. *Oklahoma Law Review 70*, 237 – 279.

Fama, E. F. and M. C. Jensen (1983a). Agency problems and residual claims. *The Journal of Law Economics 26*(2), 327–349.

Fama, E. F. and M. C. Jensen (1983b). Separation of ownership and control. *The Journal of Law Economics 26*(2), 301–325.

Ferris, S. P., T. Jandik, R. M. Lawless, and A. Makhija (2007). Derivative lawsuits as a corporate governance mechanism: Empirical evidence on board changes surrounding filings. *The Journal of Financial and Quantitative Analysis 42*(1), 143–165.

Filipova-Neumann, L. and P. Welzel (2010). Reducing asymmetric information in insurance markets: Cars with black boxes. *Telematics and Informatics 27*(4), 394–403.

Gadepally, V., A. Kurt, A. Krishnamurthy, and U. Özgüner (2011). Driver/vehicle state estimation and detection. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 582–587.

Galai, D. and R. W. Masulis (1976). The option pricing model and the risk factor of stock. *Journal of Financial Economics 3*(1), 53–81.

Gao, G., S. Meng, and M. V. Wüthrich (2019). Claims frequency modeling using telematics car driving data. *Scandinavian Actuarial Journal 2019*(2), 143–162.

Gao, G., S. Meng, and M. V. Wüthrich (2022). What can we learn from telematics car driving data: A survey. *Insurance: Mathematics and Economics 104*, 185–199.

Gao, G., H. Wang, and M. V. Wüthrich (2022). Boosting poisson regression models with telematics car driving data. *Machine Learning 111*, 243–272.

Gao, G. and M. V. Wüthrich (2018). Feature extraction from telematics car driving heatmaps. *European Actuarial Journal 8*, 383–406.

Gao, G. and M. V. Wüthrich (2019). Convolutional neural network classification of telematics car driving data. *Risks 7*(1), 6.

Ghahramani, Z. (2001). An introduction to hidden markov models and bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence 15*(1), 9–42.

Guillen, M., J. P. Nielsen, and A. M. Pŕez-Marín (2021). Near-miss telematics in motor insurance. *Journal of Risk and Insurance 88*(3), 569–589.

Guillen, M., J. P. Nielsen, A. M. Pŕez-Marín, and V. Elpidorou (2020). Can automobile insurance telematics predict the risk of near-miss events? *North American Actuarial Journal 24*(1), 141–152.

Hamada, R., T. Kubo, K. Ikeda, Z. Zhang, T. Shibata, T. Bando, K. Hitomi, and M. Egawa (2016). Modeling and prediction of driving behaviors using a nonparametric bayesian method with ar models. *IEEE Transactions on Intelligent Vehicles 1*(2), 131–138.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2 ed.). Springer.

He, E. and D. W. Sommer (2010). Separation of ownership and control: Implications for board composition. *The Journal of Risk and Insurance 77*(2), 265–295.

Helland, E. and J. Klick (2009). *The Relation between Regulation and Class Actions: Evidence from the Insurance Industry*. RAND Corporation.

Henckaerts, R. and K. Antonio (2022). The added value of dynamically updating motor insurance prices with telematics collected driving behavior data. *Insurance: Mathematics and Economics 105*, 79–95.

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

Houston, J. F., C. Lin, and W. Xie (2018). Shareholder protection and the cost of capital. *The Journal of Law and Economics 61*(4), 677–710.

Huang, Y., N. Li, Y. Yu, and X. Zhou (2020). The effect of managerial litigation risk on earnings warnings: Evidence from a natural experiment. *Journal of Accounting Research 58*(5), 1161–1202.

Huang, Y. and S. Meng (2019). Automobile insurance classification ratemaking based on telematics driving data. *Decision Support Systems 127*, 113–156.

Jensen, M. C. and W. H. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics 3*(4), 305–360.

Jiang, Q. and T. Shi (2022). Auto insurance pricing using telematics data: Application of a hidden markov model. Working paper.

John, K., L. Litov, and B. Yeung (2008). Corporate governance and risk-taking. *The Journal of Finance 63*(4), 1679–1728.

Karpoff, J. M., D. Scott Lee, and G. S. Martin (2008). The consequences to managers for financial misrepresentation. *Journal of Financial Economics 88*(2), 193–215.

Ketchen, D. J. and C. L. Shook (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal 17*(6), 441.

Kinney, T. P. (1994). Stockholder derivative suits: Demand and futility where the board fails to stop wrongdoers. *Marquette Law Review 78*, 172–189.

Kleffner, A. E. and N. A. Doherty (1996). Costly risk bearing and the supply of catastrophic insurance. *The Journal of Risk and Insurance 63*(4), 657–671.

Lamm-Tennant, J. and L. T. Starks (1993). Stock versus mutual ownership structures: The risk implications. *The Journal of Business 66*(1), 29–46.

Lee, S.-J., D. Mayers, and C. W. Smith (1997). Guaranty funds and risk-taking evidence from the insurance industry. *Journal of Financial Economics 44*(1), 3–24.

Lefèvre, S., A. Carvalho, Y. Gao, H. E. Tseng, and F. Borrelli (2015). Driver models for personalised driving assistance. *Vehicle System Dynamics 53*(12), 1705–1720.

Lin, C., S. Liu, and G. Manso (2021). Shareholder litigation and corporate innovation. *Management Science 67*(6), 3346–3367.

Lin, W.-C., Y.-H. Lai, and M. R. Powers (2014). The relationship between regulatory pressure and insurer risk taking. *The Journal of Risk and Insurance 81*(2), 271–301.

Liu, S., K. Zheng, L. Zhao, and P. Fan (2020). A driving intention prediction method based on hidden markov model for autonomous driving. *Computer Communications 157*, 143–149.

Manchiraju, H., V. Pandey, and K. R. Subramanyam (2020, 07). Shareholder Litigation and Conservative Accounting: Evidence from Universal Demand Laws. *The Accounting Review 96*(2), 391–412.

Mankaï, S. and A. Belgacem (2016). Interactions between risk taking, capital, and reinsurance for property–liability insurance firms. *The Journal of Risk and Insurance 83*(4), 1007–1043.

Mayers, D. and C. W. Smith (1981). Contractual provisions, organizational structure, and conflict control in insurance markets. *The Journal of Business 54*(3), 407–434.

Mayers, D. and C. W. Smith (1988). Ownership structure across lines of property-casualty insurance. *The Journal of Law Economics 31*(2), 351–378.

Mayers, D. and J. Smith, Clifford W. (2002, Feb). Ownership structure and control: Property-casualty insurer conversion to stock charter. *Journal of Financial Services Research 21*(1), 117–144.

McDonald, A. D., C. Schwarz, J. D. Lee, and T. L. Brown (2012). Real-time detection of drowsiness related lane departures using steering wheel angle. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 56*(1), 2201–2205.

McTier, B. C. and J. K. Wald (2011). The causes and consequences of securities class action litigation. *Journal of Corporate Finance 17*(3), 649–665. Financial Flexibility and Corporate Liquidity.

Meng, S., H. Wang, Y. Shi, and G. Gao (2022). Improving automobile insurance claims frequency prediction with telematics car driving data. *ASTIN Bulletin 52*(2), 363–391.

Myers, S. C. and N. S. Majluf (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics 13*(2), 187–221.

NAIC (2021). Telematics/usage-based insurance. https://content.naic.org/cipr-topics/telematicsusage-based-insurance.

Nguyen, H. T., H. V. Phan, and L. S. Sun (2018). Shareholder litigation rights and corporate cash holdings: Evidence from universal demand laws. *Journal of Corporate Finance 52*, 192–213.

Ni, X. and S. Yin (2018). Shareholder litigation rights and the cost of debt: Evidence from derivative lawsuits. *Journal of Corporate Finance 48*, 169–186.

Paefgen, J., T. Staake, and E. Fleisch (2014). Multivariate exposure modeling of accident risk: Insights from pay-as-you-drive insurance data. *Transportation Research Part A: Policy and Practice 61*, 27–40.

Parry, I. W. H. (2005). Is pay-as-you-drive insurance a better way to reduce gasoline than gasoline taxes? *American Economic Review 95*(2), 288–293.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Powell, L. S. (2017). What it means to be mutual.

Pukthuanthong, K., H. Turtle, T. Walker, and J. Wang (2017). Litigation risk and institutional monitoring. *Journal of Corporate Finance 45*, 342–359.

Qin, H., J. Liu, and T. Hong (2012). An eye state identification method based on the embedded hidden markov model. In *2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012)*, pp. 255–260.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*(2), 257–286.

Rabiner, L. and B. Juang (1986). An introduction to hidden markov models. *IEEE ASSP Magazine 3*(1), 4–16.

Reimers, I. and B. R. Shiller (2019). The impacts of telematics on competition and consumer behavior in insurance. *The Journal of Law and Economics 62*(4), 613–632.

Romano, R. (1991). The shareholder suit: Litigation without foundation? *Journal of Law Economics & Organization 7*, 55–87.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics 90*(4), 629–649.

Sathyanarayana, A., P. Boyraz, and J. H. Hansen (2008). Driver behavior analysis and route recognition by hidden markov models. In *2008 IEEE International Conference on Vehicular Electronics and Safety*, pp. 276–281.

Saunders, A., E. Strock, and N. G. Travlos (1990). Ownership structure, deregulation, and bank risk taking. *The Journal of Finance 45*(2), 643–654.

Scarlett, A. M. (2013). Shareholder derivative litigation's historical and normative foundations. *Buffalo Law Review 61*, 837 – 908.

Sebah, P. and X. Gourdon (2002). Introduction to the gamma function. https://www.csie.ntu.edu.tw/~b89089/link/gammaFunction.pdf.

Shaner, M. W. (2014). The (un)enforcement of corporate officers' duties. *UC Davis Law Review 48*(1), 271–336.

Smyth, P. (1996). Clustering sequences with hidden markov models. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS'96, Cambridge, MA, USA, pp. 648–654. MIT Press.

So, B., J.-P. Boucher, and E. A. Valdez (2021). Cost sensitive multi-class adaboost for understanding driving behavior based on telematics. *ASTIN Bulletin 51*(3), 719–751.

Soleymanian, M., C. B. Weinberg, and T. Zhu (2019). Sensor data and behavioral tracking: Does usage-based auto insurance benefit drivers? *Marketing Science 38*(1), 21–43.

Stein, J. C. (1988). Takeover threats and managerial myopia. *Journal of Political Economy 96*(1), 61–80.

Stein, J. C. (1989). Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The Quarterly Journal of Economics 104*(4), 655–669.

Tadesse, E., W. Sheng, and M. Liu (2014). Driver drowsiness detection through hmm based dynamic modeling. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4003–4008.

Vantage Market Research (2022). Usage-based insurance (ubi) market size, share & trends analysis report - global industry assessment (2016-2021) & forecast (2022-2028). https://www.vantagemarketresearch.com/industry-report/usagebased-insurance-ubi-market-1658.

Verbelen, R., K. Antonio, and G. Claeskens (2018). Unravelling the predictive power of telematics data in car insurance pricing. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 67*(5), 1275–1304.

Verisk Insurance Solutions (2016). The challenge of auto insurance premium leakage. https://www.verisk.com/insurance/campaigns/challenge-of-auto-insurance-premium-leakage/.

Virgilis, M. D., D. Lupton, L. McGrath, M. Qazvini, S. Roby, and L. Vernon (2022). Machine learning in insurance. https://www.casact.org/sites/default/files/2022-03/01_Winter-Eforum-2022-ML_in_Insurance.pdf.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory 13*(2), 260–269.

Wang, W., J. Xi, and D. Zhao (2018). Learning and inferring a driver's braking action in car-following scenarios. *IEEE Transactions on Vehicular Technology 67*(5), 3887–3899.

Wang, W., J. Xi, and D. Zhao (2019). Driving style analysis using primitive driving patterns with bayesian nonparametric approaches. *IEEE Transactions on Intelligent Transportation Systems 20*(8), 2986–2998.

Weidner, W., F. W. Transchel, and R. Weidner (2016). Classification of scale-sensitive telematic observables for risk individual pricing. *European Actuarial Journal 6*, 3–24.

Wood, F. S. (1944). *Survey and Report Regarding Stockholders' Derivative Suits*. New York.

Wu, C. F. J. (1983). On the convergence properties of the em algorithm. *The Annals of Statistics 11*(1), 95–103.

Wüthrich, M. V. (2017). Covariate selection from telematics car driving data. *European Actuarial Journal 7*, 89–108.

Xie, X., Y. Wang, G. Zhao, and W. Lu (2017). Cash holdings between public and private insurers – a partial adjustment approach. *Journal of Banking Finance 82*, 80–97.

Xiong, X., L. Chen, and J. Liang (2018). A new framework of vehicle collision prediction by combining SVM and HMM. *IEEE Transactions on Intelligent Transportation Systems 19*(3), 699–710.

Zhang, C., J. Zhu, W. Wang, and J. Xi (2022, jul). Spatiotemporal learning of multivehicle interaction patterns in lane-change scenarios. *IEEE Transactions on Intelligent Transportation Systems 23*(7), 6446–6459.

Zhang, L., X. Wu, and D. Luo (2015). Human activity recognition with HMM-DNN model. In *2015 IEEE 14th International Conference on Cognitive Informatics Cognitive Computing (ICCI*CC)*, pp. 192–197.

Zhu, T., A. Khajepour, C. Zong, L. Mai, and H. Deng (2015). Driving condition identification based on hhmm model – applicationto the simulation of heavy duty vehicle. *The Dynamics of Vehicles on Roads and Tracks*, 663–668.

Zong, C., C. Wang, D. Yang, and H. Yang (2009). Driving intention identification and maneuvering behavior prediction of drivers on cornering. In *2009 International Conference on Mechatronics and Automation*, pp. 4055–4060.

APPENDIX A
BAYESIAN NETWORKS

This section introduces Bayesian networks. The graph from a Bayesian network can reveal the dependency and conditional dependency clearly, which can facilitate the decomposition of joint probabilities. HMM is a special case of Bayesian networks. Thus, any change in conditional dependency in a HMM can be revealed in the graph.

As introduced in Ghahramani (2001), "Bayesian network is a graphical model for representing conditional independence between a set of random variables." Specifically, a Bayesian network includes directed acyclic path consisting of nodes, representing random variables, and edges, representing relationships between nodes with direction. For a Bayesian network, the relationships between random variables are completed defined, such as correlation and independence.

Some basic definitions in Bayesian networks are introduced below. As defined in Ghahramani (2001), a node A is a parent of another node B if there is a directed edge from A to B. In this case, B is called a child of A. An undirected path from A to B is a sequence of nodes starting from A and ending in B such that each node in the sequence is a parent or child of the following node. For an arbitrary Bayesian network with nodes $x_1, x_2, ..., x_n \in X$, we can derive the joint distribution $P(X)$ as the product of each node x_i given its parents $e(x_i)$:

$$P(X) = P(x_1, x_2, ..., x_n) = \prod_{x_i} P(x_i|e(x_i)) \tag{3.76}$$

For example, assuming there are four random variables, A, B, C, D, without considering the structure of a Bayesian network and only applying Bayes' theorem, we obtain the following decomposition:

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|A, B, C) \tag{3.77}$$

However, when we impose a Bayesian network structure as in Figure 3.2, following the rule above, we obtain the another decomposition:

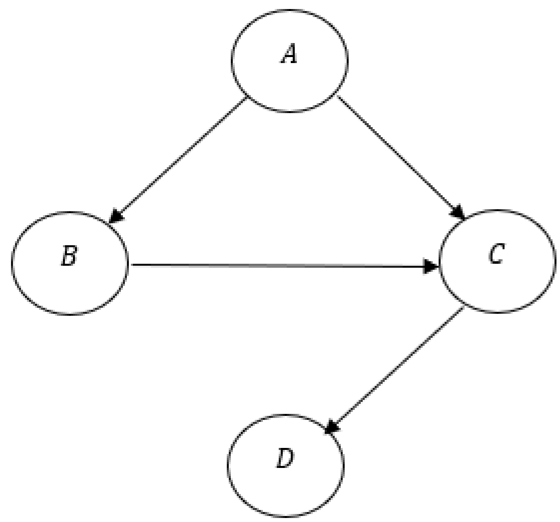

Figure 3.2.: Bayesian Network Example 1

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|C) \tag{3.78}$$

The differences between those two decomposition is the conditional independence between (A, B) and D given C, which is implied by the Bayesian network structure.This factorization strategy only works if there are no cycles in the graph, and that Bayesian network is acyclic by definition. In other words, there is no direct circle in the graph.

There are two ways to determine dependency, that is, blockage and d-separate. **Blockage** is to follow all undirected paths between two variables and check if the path is blocked. Two variables are independent if all available paths between them are "blocked". Blockages are determined by visiting each node on a path and checking the structure of surrounding nodes and edges against 3 rules explained below.

Rule 1 is called Markov Chain, shown in Figure 3.3. C is independent of A if B is given. The Markov Chain is blocked given B.

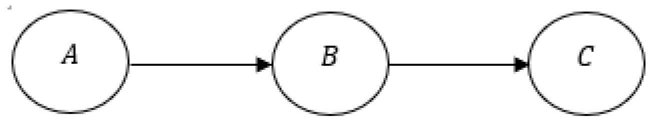

Figure 3.3.: Rule 1: Markov Chain

Rule 2 is called Two Parents, One Child, shown in Figure 3.4. A and B are independent if C is not known. A and B are dependent if C is known.

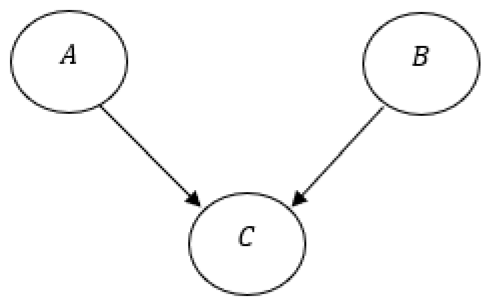

Figure 3.4.: Rule 2: Two Parents, One Child

Rule 2 Extension is an extension of Rule 2, where C causes D, shown in Figure 3.5. If a descendant of the node into which the arrows converge is know, the path is unblocked. The path is only blocked if C and any descendant of C are unknown.

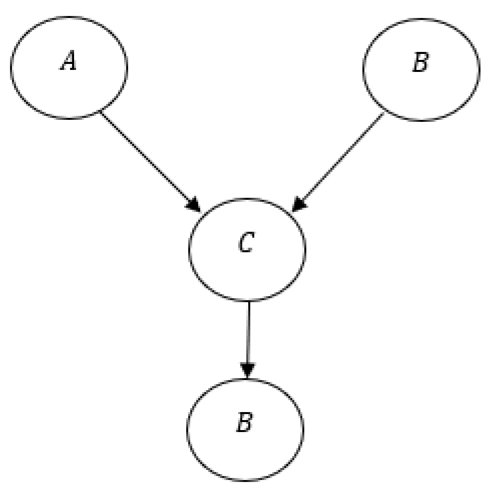

Figure 3.5.: Rule 2 Extension

Rule 3 is called One Parent, Two Child, shown in Figure 3.6. A and B are dependent with C unknown. A and B are independent given C. So One Parent, Two Children case is blocked given C.

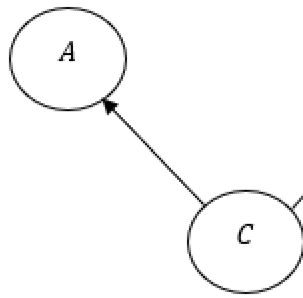

Figure 3.6.: Rule 3: One Parent, Two Child

The second method **d-separate** is another way to determine conditional independence. According to Ghahramani (2001), two disjoint sets of nodes A and B are conditionally independent given C, if C d-separates A and B, that is, if along every undirected path between a node in A and a node in B, there is a node D such that: (1) D has converging arrows (D is a child of both the previous and following nodes in the path) and neither D or its descendants are in C, or (2) D does not have converging arrow and D is in C.

APPENDIX B
PROPOSITION 1 PROOF

$$
\begin{aligned}
&P(X_t,Y_t,L_t|X_{t-1},Y_{t-1},L_{t-1})\\
&=\frac{P(X_{t-1},Y_{t-1},L_{t-1},X_t,Y_t,L_t)}{\int P(X_{t-1},Y_{t-1},L_{t-1},X_t,Y_t,L_t)\,dX_t\,dY_t\,dL_t}\\
&=\frac{P(X_{t-1})P(Y_{t-1}|X_{t-1})P(L_{t-1}|X_{t-1},Y_{t-1})P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)}{\int P(X_{t-1})P(Y_{t-1}|X_{t-1})P(L_{t-1}|X_{t-1},Y_{t-1})P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)\,dX_t\,dY_t\,dL_t}\\
&=\frac{P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)}{\int P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)\,dX_t\,dY_t\,dL_t}
\end{aligned}
\tag{3.79}
$$

The numerator above can be rewritten as follows:

$$
\begin{aligned}
&P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)\\
&=\frac{P(X_t,X_{t-1},Y_{t-1})}{P(X_{t-1},Y_{t-1})}\frac{P(X_t,Y_t,X_{t-1},Y_{t-1})}{P(X_t,X_{t-1},Y_{t-1})}\frac{P(X_t,Y_t,L_t,X_{t-1},Y_{t-1})}{P(Y_t,X_t,X_{t-1},Y_{t-1})}\\
&=\frac{P(X_t,Y_t,L_t,X_{t-1},Y_{t-1})}{P(X_{t-1},Y_{t-1})}
\end{aligned}
\tag{3.80}
$$

We plug the ratio above back into the Equation (3.79) and obtain:

$$
\begin{aligned}
P(X_t,Y_t,L_t|X_{t-1},Y_{t-1},L_{t-1}) &=\frac{P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)P(X_{t-1},Y_{t-1})}{\int P(X_t,Y_t,L_t,X_{t-1},Y_{t-1})\,dX_t\,dY_t\,dL_t}\\
&=\frac{P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)P(X_{t-1},Y_{t-1})}{P(X_{t-1},Y_{t-1})}\\
&=P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t)
\end{aligned}
\tag{3.81}
$$

Moreover, after applying conditional independence, we obtain Proposition 1:

$$
\begin{aligned}
P(X_t,Y_t,L_t|X_{1:t-1},Y_{1:t-1},L_{1:t-1}) &=P(X_t,Y_t,L_t|X_{t-1},Y_{t-1},L_{t-1}) \text{ by Markov Property}\\
&=P(X_t|X_{t-1})P(Y_t|X_t,Y_{t-1})P(L_t|X_t,Y_t) \text{ for } t>1
\end{aligned}
\tag{3.82}
$$

APPENDIX C
PROPOSITION 3 PROOF

First, we have the initial condition as follows:

$$\beta_T(j,k) = 1, \text{for all } j \in \{1,...,N\}, k \in \{1,...,M\} \tag{3.83}$$

For $1 \leq t-1 \leq T-1$, conditioning on the value of X_t, the backward variable $\beta_t(j,k)$ can be written as

$$\begin{aligned}\beta_t(j,k) &= P(L_{t+1:T}, Y_{t+1:T} | Y_t = k, X_t = j) \\ &= \sum_{X_{t+1}} P(L_{t+1:T}, Y_{t+1:T}, X_{t+1} | Y_t = k, X_t = j) \\ &= \sum_{X_{t+1}} P(L_{t+1:T}, Y_{t+1:T}, | X_{t+1}, Y_t = k, X_t = j) P(X_{t+1} | X_t = j, Y_t = k).\end{aligned} \tag{3.84}$$

Note that

$$\begin{aligned}&P(L_{t+1:T}, Y_{t+1:T}, | X_{t+1}, Y_t = k, X_t = j) \\ &= P(L_{t+1:T}, Y_{t+1:T}, | X_{t+1}, Y_t = k) \\ &= P(L_{t+1:T}, Y_{t+2:T}, | X_{t+1}, Y_{t+1}, Y_t = k) P(Y_{t+1} | X_{t+1}, Y_t = k) \\ &= P(L_{t+1} | X_{t+1}, Y_{t+1}) P(L_{t+2:T}, Y_{t+2:T}, | X_{t+1}, Y_{t+1}) P(Y_{t+1} | X_{t+1}, Y_t = k) \\ &= \beta_{t+1}(X_{t+1}, Y_{t+1}) b_{X_{t+1},k,Y_{t+1}} c_{X_{t+1},Y_{t+1}}(l_{t+1})\end{aligned} \tag{3.85}$$

Substituting (3.85) into (3.84), one concludes that

$$\beta_t(j,k) = \sum_{X_{t+1}=i} \beta_{t+1}(i, Y_{t+1}) \alpha_{ji} b_{i,k,Y_{t+1}} c_{i,Y_{t+1}}(l_{t+1}). \tag{3.86}$$