Auto Insurance Actuarial Rating System and Related Risks under a Broadly Generalized Nonlinear Framework

Authors: Sun Huanyu, Xie Yuantao

Date: 2025-06-12T10:50:58+00:00

Abstract

Vehicle insurance pricing, as one of the core issues in actuarial science, impacts the profitability and capital liquidity of insurance companies. This literature review briefly elaborates on the development of motor insurance across different regions and discusses historical research related to motor insurance ratemaking systems, special motor insurance pricing, No-Claim Discount bonus-malus theory, and reserve evaluation methods. Taking statistical distributions and model structures as entry points, this review analyzes the specific operational logic of models and concentrates on discussing literature that establishes pricing frameworks and reserve evaluations using the Tweedie distribution family for index insurance pricing models, Generalized Additive Models and LocalGLMnet modeling methods, and most methods based on frequentist and Bayesian schools. To demonstrate the application effects of risk modeling concepts on real-world business, this article also includes relevant literature on Bonus-Malus Systems, GAM reserve evaluation models, etc., exploring a construction method for risk dynamic adjustment mechanisms in vehicle insurance based on these. Regarding reserve research, it includes literature on evaluation methods under homogeneity and heterogeneity assumptions, specifically divided into two categories: using traditional methods such as the chain-ladder method for predictive modeling of aggregate accident claim data, and using machine learning technology to improve models based on existing algorithmic models for predictive modeling of individual accident claim data. In addition to general motor insurance, weather index insurance is also studied and discussed, and by utilizing remote sensing data and ground observation data to establish multi-dimensional input methods, the model's ability to identify extreme weather events is enhanced. Empirical findings indicate that the model performs well in terms of claim payment stability and regional risk differences. Based on the data level, this literature review selects real vehicle insurance and weather insurance claim payment data as case studies, and sets up control groups for different modeling approaches to analyze and evaluate the similarities and differences in predictive performance. This article focuses on constructing a replicable and generalizable insurance modeling methodology in the property insurance field, comprehensively considering regulatory, data modeling, and other factors to enhance the quantitative pricing capabilities of property insurance companies, thereby promoting the property insurance industry to gain new practical experience and theoretical foundations in digital transformation, pricing actuarial science, and other aspects.

Full Text

Preamble

Dissertation Title: Research on Auto Insurance Actuarial Rating Systems and Related Risks with a Pan-Generalized Nonlinear Framework

Author: Sun Huanyu
Student ID: 202300930147
Institution: School of Insurance, University of International Business and Economics
Major: Actuarial Science
Advisor: Professor Xie Yuantao
Date: June 2025

Doctoral Dissertation Proposal Review Form
School of Insurance, Actuarial Science Program
Graduate Student: Sun Huanyu

After review by the advisor and school committee, the following decision is made regarding the doctoral dissertation topic and outline:

1. Approved to proceed with research and writing based on the proposed topic and outline (√)

2. Approved for topic, but requires modifications to outline and research methods ( )

3. Please select a different topic ( )

Signatures:
Advisor ( )
Department Head ( )

Advisor's Guidance on Proposal
This proposal demonstrates clear structure and addresses a topic of practical significance, covering both traditional methods and cutting-edge techniques. It reflects substantial research depth and application orientation. My recommendations are to further clarify research objectives, highlight model innovations, strengthen theoretical analysis of different models, and in subsequent research,

further specify comparison dimensions across models to emphasize the advantages of innovative methods such as LocalGAMnet. Additionally, combine business realities to reinforce analysis of interpretability and applicability, thereby enhancing research completeness.

Chapter 1: Dissertation Outline

Research on Auto Insurance Actuarial Rating Systems and Related Risks with a Pan-Generalized Nonlinear Framework

Chapter 1: Introduction

1.1 Research Background and Significance
1.2 Literature Review
1.3 Research Content, Methods, and Innovations

Chapter 2: Preliminary Knowledge

2.1 Auto Insurance Pricing Theory
2.2 Auto Insurance Data Characteristics
2.3 Traditional Actuarial Models
2.4 Machine Learning Methods
2.5 Credibility Theory

Chapter 3: Auto Insurance Data Analysis and Processing

3.1 Data Sources and Collection
3.2 Data Preprocessing and Feature Engineering
3.3 Data Quality Assessment
3.4 Summary

Chapter 4: Pan-Generalized Nonlinear Framework for Auto Insurance Ratemaking

4.1 Framework Design and Theoretical Foundation
4.2 Model Construction and Parameter Estimation
4.3 Model Validation and Comparison
4.4 Application in Special Vehicle Insurance
4.5 Summary

Chapter 5: Special Vehicle Insurance (ESG-Based)

5.1 ESG Risk Assessment and Integration
5.2 Weather Index Insurance Design
5.3 Case Analysis and Summary

Chapter 6: Theory of Auto Insurance Actuarial Rating Systems

6.1 System Architecture and Components
6.2 Dynamic Risk Adjustment Mechanisms
6.3 Regulatory Compliance and Fairness
6.4 Summary

Chapter 2: Research Background and Significance

2.1.1 Research Background on Auto Insurance Pricing

Auto insurance is specifically designed to cover a series of risks or predictable accidents that may occur during vehicle operation. It is a commercial insurance product that assumes compensation liability related to automotive mobile equipment, typically including vehicle damage insurance (property damage coverage) or third-party liability insurance.

In current practice, auto insurance primarily utilizes Generalized Linear Models (GLM) [**?**]. GLM consists of three major components: random component, systematic component, and link function. Specifically, the dependent variable Y is assumed to follow the exponential family of distributions, and explanatory variables X influence the response variable through a linear predictor. Since Y may not be normally distributed, directly modeling E(Y) might be inappropriate, hence a link function g(·) is introduced: g(E[Y]) = . By selecting an appropriate link function, the relationship between g(E[Y]) and X becomes linear. GLM assumes error terms follow the exponential family, such as Bernoulli, binomial, Poisson, gamma distributions, etc., greatly expanding the applicable data range [**?**].

However, since nonlinear relationships between certain features and claim amounts are more realistic and acceptable, theoretically Generalized Additive Models (GAM) [**?**, **?**] should replace GLM. GAM is an extension of GLM that allows nonlinear relationships between independent and dependent variables. The difference between GAM and GLM lies in that GAM relates the expected value E[Y] of the dependent variable to independent variables through smoothing functions rather than purely linear combinations. These smoothing functions can be splines or local weighted regression functions.

In reality, even when insurers consider most factors affecting claims, each policy and claim still implies different influencing conditions not captured by the model. These unconsidered conditions affect final claim amounts positively or negatively to varying degrees. To incorporate these effects into ratemaking, research has emerged on using Generalized Linear Mixed Models (GLMM) and Generalized

Additive Mixed Models (GAMM) for auto insurance pricing. In a sense, GLMM [?, ?] and GAMM [?] can be viewed as GLM and GAM models that incorporate random effects.

In the auto insurance field, both model selection and actuarial choice of loss distributions represent processes of gradual optimization forming industry consensus, then continuing to develop based on new theories. Two elements determine loss magnitude: loss frequency (number of claims) and loss severity (claim amount). Common theoretical distributions for loss frequency include Poisson, binomial, and negative binomial distributions. Common theoretical distributions for loss severity include log-normal, gamma distributions, and other right-skewed distributions such as log-gamma, Weibull distributions, as well as derived continuous distributions with more parameters like transformed gamma and transformed beta distributions [?]. In auto insurance, the Tweedie distribution is generally assumed as the practical distribution.

A potential problem exists: even when data fitting achieves ideal assumptions meeting actuarial standards, it remains insufficient because individual behavior consistently influences claims. This led actuaries to propose usage-based insurance (UBI), known as UBI auto insurance in this context. For example, Progressive Property Insurance Company provides UBI Snapshot service, requiring vehicle owners to install OBD devices so insurers can obtain driving data to calculate snapshot scores and determine premium discounts. Different states have different programs, but owners must provide at least 75 consecutive days of driving data. Features such as age, gender, driving experience, vehicle brand, vehicle usage period, and vehicle price are typically considered in UBI auto insurance, with telematics data being particularly important—daily average distance, for instance, is derived from widely used raw telematics data. To handle the dynamic data required by UBI auto insurance, neural networks, particularly Feed-Forward Networks (FFN), can automatically learn complex patterns in data and improve prediction accuracy [?].

Regarding model optimization methods in auto insurance, common algorithms include LASSO [?], XGBoost [?], and LightGBM [?]. XGBoost is an ensemble learning algorithm that can also perform feature selection, but it is less direct than LASSO for this purpose. Due to its use of decision trees, its computational complexity is much higher than LASSO, while LASSO is essentially a penalization algorithm typically built on linear models. From the perspective of auto insurance modeling, LightGBM's optimization efficiency is higher than XGBoost's.

2.1.2 Research Background on Weather Index Insurance

Weather index insurance, as the name suggests, is a non-traditional pricing model based on weather factors. Multiple studies have shown that precipitation and extreme temperatures affect driver behavior and vehicle mobility [?]. For example, adverse weather reduces free-flow speed, increases headway distance, and road surface friction coefficients affect vehicle handling. Basagaña X. et al. [?] used Poisson regression models controlling for rainfall, day of week, month, year, holidays, and other variables, finding that total vehicle accidents increased significantly by 2.9% during heatwave periods. Moreover, for every 1°C increase in maximum temperature, the risk of driver performance-related accidents increased by 1.1%. During vehicle operation, drivers primarily face fatigue driving and improper behavior (including speeding, road rage, sudden braking, etc.). Noelke et al. [?] studied 1.9 million Americans, finding that hot weather conditions more easily lead to irritability, nervousness, and other negative emotions and high-fatigue mental states compared to normal conditions. Makowiec-Dąbrowska T et al. [?] further used fatigue assessment questionnaires to find that "difficulty in decision-making" correlates with temperature. That is, higher temperatures make it more difficult for people to make correct judgments, which is dangerous for driving.

Zhai Xiaoqi et al. [?] integrated high-resolution weather and accident data through geographic information systems, finding that high temperature and rainfall significantly increase the risk of severe injury or death, and that improper behavior by pedestrians and drivers has greater impact under adverse weather conditions, demonstrating the importance of introducing real-time weather data. Commonly considered weather attributes include visibility and precipitation [?], while increased rainfall frequency and intensity correlate with poor visibility and low road friction, thereby increasing accident risk [?, ?, ?].

Like most traditional property insurance types, auto insurance is subject to information asymmetry, leading to moral hazard and adverse selection problems. For vehicles, insurers can only dispatch survey personnel to accident sites after claims investigation, increasing company costs and potentially enabling fraud. In contrast, weather index insurance uses "objective weather indicators" as the basis for claims settlement, which in a sense eliminates the need to consider moral hazard or adverse selection because compensation standards are based on clearly defined and pre-published weather data, making insurance contracts and business procedures simple.

2.1.3 Research Background on Insurance Risk Response Mechanisms

In addition to traditional models, auto insurance pricing can also employ Kappa-N models and Bonus-Malus System (BMS) to address adverse selection and moral hazard arising from renewal issues. The Kappa-N model is a generalization of count distributions that incorporates claim scores considering historical claims experience data, adding two covariates to the base formula [?]. However, since the Kappa-N model under penalty discounts does not consider actual penalty severity, it leads to excessive premium increases or decreases in the model. The BMS model is similar to the Kappa-N model but includes minimum and maximum value requirements. Here, BMS operates like a tiered system with limited levels, assigning relative proportions to each tier size. Policyholders generally move down tiers if no claims occur during the contract period; filing a claim moves them to a certain tier.

Future research will no longer simply interpret the parameter as a claim score but directly define it as BMS level. The primary risk response mechanism in the insurance industry is the reserve assessment mechanism [?]. Verrall [?] theoretically demonstrated that traditional reserve assessment methods (including the B-F method and GLM) are essentially equivalent in substance, providing a new interpretation of traditional methods. Mario V. W. [?] used regression trees to evaluate and predict individual reserves, showing that decision tree model structures can more accurately capture complex nonlinear relationships. Gabrielli et al. [?] cross-classified traditional chain ladder models to obtain feed-forward neural network structures with skip connections, demonstrating excellent fitting performance.

2.2.1 Impact of Pan-Generalized Nonlinear Framework on Loss Prediction

The high robustness and interpretability of Generalized Linear Models have led to their widespread acceptance, but they still have limitations when facing nonlinear problems and high-dimensional data issues. Using nonlinear structures becomes essential. Auto insurance loss prediction models under the pan-generalized nonlinear framework mainly include pricing models and reserve prediction models.

The rapid development of connected vehicles, ADAS, OBD, and remote information systems in recent years has enabled insurers to obtain large amounts of real-time driving behavior data (e.g., acceleration, mileage) in real time, enabling precise pricing that traditional pricing models cannot achieve. Ronald et al. [?] proposed LocalGLMnet, a model combining GLM interpretability with strong nonlinear modeling capabilities, enhancing real-world alignment by using neural networks for nonlinear modeling on top of a linear main architecture. However, pure neural networks cannot fully restore the nonlinear predictive capability after linear processing within the main architecture, necessitating GAM as a supplement—a model that can completely restore nonlinear processing on a nonlinear main architecture.

The pan-generalized nonlinear framework is a framework for modeling multi-stage, multi-source nonlinear information, possessing unified processing capabilities for data across all stages of the policy lifecycle (including pricing and reserve assessment stages). Compared with traditional linear models, the pan-generalized nonlinear framework breaks through linear assumptions by integrating static and dynamic data under a unified framework to analyze intrinsic relationships between input and output variables. It then uses a unified pattern to comprehensively represent multi-source information such as driver information, vehicle information, driving behavior characteristics, and historical claims information, improving risk characterization for auto insurance and thereby enhancing risk 刻画水平 for high-order insurance products. It can better handle low prediction accuracy issues caused by heterogeneity, providing feasible theoretical support for further improving the adaptability of auto insurance risk control systems.

The dual-model insurance system under this framework refers to a pricing mode that sequentially applies two different models meeting discrimination and fairness requirements (a basic model with few features and a precise model with many features). For example, a dual-model system composed of LocalGLMnet and LocalGAMnet models can also have its pricing model and reserve model serve as two subsystems to solve problems of traditional insurance models lacking interactivity, feedback, and inability to reflect the entire risk development process. By comprehensively updating traditional auto insurance pricing models through LocalGLMnet, BMS system, and GAM, among other methods, the model better captures more complex and realistic nonlinear features in auto insurance business while maintaining interpretability and stability, with noticeably improved regression prediction results. By adding optimization algorithms such as LightGBM to mitigate overfitting in claims prediction, the dual-model system can also improve overall modeling effectiveness and generalization capability to some extent.

Regarding reserve prediction, the chain ladder method, Bornhuetter-Ferguson method, and Mack model remain mainstream due to their stability and strong interpretability. However, because classical methods assume stable claims behavior, they cannot well capture changing characteristics during the claims process. Methods based on homogeneity assumptions also struggle to reflect individual risks in overall reserve estimates, and reserve estimates for certain high-risk niche markets may show deviations. For example, luxury sedans, high-performance sports cars, bulletproof vehicles, and imported new energy vehicle models have significantly different average claim amounts and claim types compared to family vehicles. Continuing to use aggregated models from the same broad category for reserve calculation would underestimate cumulative risk, leading to insufficient reserves or unidentified risk exposure, potentially causing solvency crises. The pan-generalized nonlinear dual-model system can accommodate such differences, embedding heterogeneity within the model structure to enable different dynamic trajectories for claims development processes across different vehicles and populations.

In terms of modeling technology, the pan-generalized nonlinear framework emphasizes model extensibility and composability, allowing selection of different model types based on business needs. For example, GAM can be used to model main effects at the pricing end, while local neural networks describe nonlinear interactions among certain variables. At the reserve end, mixed models can be introduced for dynamic modeling of payment timing and amounts. Using a dual-model system enables companies to more accurately predict risks and implement risk interventions, allowing insurers to timely grasp and defend against risk occurrence. Moreover, it can model pricing and reserves at the model level while connecting pricing and reserve modeling to establish a closed-loop mechanism for risk assessment and capital allocation, optimizing insurance product risk structure and enhancing insurers' risk control capabilities and resistance to tail risks.

2.2.2 Impact of Weather Index Insurance on Insurance Companies Under Pan-Generalized Nonlinear Framework

If insurance companies do not systematically integrate ESG factors into all aspects of claims management and operations, they will likely be excluded from major ESG investment standards (including but not limited to various funds and indices) in the future, losing capital market recognition. Moreover, in terms of climate change response, property and casualty insurers face more diverse climate-related risks and opportunities and should establish sound climate-resilient business structures and climate risk assessment models early on [**?**, **?**].

Currently, climate risks are mainly divided into two categories: physical risk and transition risk. In actuarial practice, the primary focus is on physical risk—the direct or indirect impact on insured objects and their related assets and liabilities caused by climate change itself or extreme events triggered by climate change, such as increased frequency and severity of catastrophic events like hurricanes, floods, and wildfires leading to significantly increased claims. Additionally, the impact of weather factors on auto insurance should be fully considered.

Precipitation events reaching weather index insurance thresholds have adverse effects on large vehicle populations, leading to high maintenance costs and traffic disruptions. [TABLE:A1]

Given that vehicle driving risks must consider not only direct impacts from extreme weather but also indirect risks such as large-area vehicle damage, high maintenance costs, and traffic disruptions, precipitation events exceeding threshold values in certain regions will cause accidents affecting large areas, bringing certain risks to insurers' loss calculations. Vehicle driving risk levels do not depend solely on direct damage from climate risks; related indirect impacts must also be emphasized. Specifically: drivers' inability to maintain good vehicle control due to high temperatures leads to various traffic violations; air quality problems triggered by meteorological changes reduce road visibility, increasing traffic mortality rates; frequent extreme weather events like heavy rain intensify risks of vehicle engine failure, potentially causing personal injury or unforeseen claim events. Without adequate ESG knowledge reserves and risk awareness, insurance companies may fail to identify high-risk exposure customers, negatively impacting overall competitiveness [**?**]. Auto insurance can also learn from agricultural insurance experience by introducing relevant weather indicators into the auto insurance system to enhance companies' ability to respond to ESG-related risks and ensure business resilience after disasters [**?**]. As an important component of property insurance, vehicle insurance must also integrate ESG strategic thinking to promote green transformation of business.

Chapter 3: Main Content and Basic Arguments

3.1 Auto Insurance Ratemaking Process Under Pan-Generalized Nonlinear Framework (Including Special Auto Insurance)

As mentioned above, this paper primarily studies the role of pan-generalized nonlinear methods in auto insurance ratemaking (including special auto insurance). With increasing data dimensions, complex data types, and external environmental impacts (mainly weather factors), single modeling approaches can hardly meet both accuracy and risk identification requirements simultaneously, making pan-generalized nonlinear modeling more suitable. Weather index auto insurance, as a new type of vehicle and insurance product, makes models more flexible.

In reality, vehicle insurance data naturally exhibit grouping characteristics, such as insurance agency affiliation, geographic scope of insured vehicles, and vehicle brand or model categories. In weather index auto insurance, hierarchical models can be constructed to combine meteorological exposure and accident risk across different geographic divisions, using weather indicators like rainfall and visibility as primary metrics with a city-level structure to better highlight how weather risk characteristics affect vehicle accident probability.

One typical fusion method is LocalGLMnet, which uses the interpretability of Generalized Linear Models and neural networks to fit more complex relationships between variables. Another fusion method combines Generalized Additive Models (GAM) with LocalGLMnet, specifically using GAM to model variable effects for each smoothed single-variable effect while possessing local regression and feed-forward neural network characteristics. This GAM approach enables independent modeling of main effect variables through smoothing functions, ensuring model interpretability of main effects while better fitting complex nonlinear relationships.

Traditional weather index insurance based on thresholds primarily covers situations where rainfall exceeds 50mm for compensation. Under pan-generalized nonlinear conditions, spline functions enable nonlinear representation of the impact of weather factors like temperature and precipitation on events, obtaining response functions corresponding to extreme weather accidents. Beyond requiring stability and fairness in index trigger mechanisms, weather index auto insurance pricing must address the greatest difficulty—the basis risk between compensation and actual loss—requiring models with uncertainty evaluation methods, using confidence interval estimation to assess uncertainty levels of compensation.

In insurance purchase decisions, consumers may exhibit clear risk aversion, and because risk and risk aversion are correlated (i.e., low-risk individuals are also more risk-averse), low-risk policyholders tend to purchase more insurance. Professor Shi Peng et al. [?] inferred that for the same reason, high-risk policyholders have lower risk aversion and may even be risk-seeking individuals, making them less likely to purchase insurance, potentially not buying at all. This mechanism may lead to a negative correlation between risk and insurance coverage, known as "advantageous selection." To avoid this outcome from a pricing perspective—ensuring high-risk policyholders purchase products with less coverage or higher prices while low-risk policyholders purchase products with more coverage or better prices—pricing model differentiation becomes necessary, forming part of the theoretical foundation for the dual-model system.

Most scholars point out that the main factors influencing consumer auto insurance purchasing behavior are consumers' personality characteristics and demographic features [?], which provides important value for insurance companies designing insurance products and marketing strategies. Current compulsory traffic accident liability insurance only provides basic protection for driver and passenger life safety, making it difficult to cover complex and diverse real-world situations or meet consumer protection needs. Therefore, in the transition from traditional to modern risk society, promoting the "dual-model system" to the auto insurance market is a necessary response to demand.

At this stage, the "TC pricing method" or "anchor pricing method" is adopted, setting three (or more) price point options for users (which should be ordered) to guide rational choices within a certain range. The middle price point is set as traditional premium for non-dual-model system users, determined by historical averages calculated using GLM pricing methods. The lowest price point takes the basic premium of the "dual-model system," i.e., the so-called "quasi-traditional premium" with smaller coverage scope. The highest price point is the comprehensive premium of the "dual-model system," i.e., the "advanced premium" with larger coverage scope. Other specific measures include providing insurance liability discounts, premium discounts, and more convenient and efficient claims services for new energy vehicles, low-carbon travel, and high-level autonomous driving vehicles. Conversely, for vehicle models with inherently high accident rates or high carbon emission intensity, premiums should be increased, using premium adjustments and dynamic risk assessment for feedback.

3.2 Post-Pricing Risk Response Mechanism for Auto Insurance Under Pan-Generalized Nonlinear Framework

Although insurance companies conduct auto insurance pricing based on rich historical data and actuarial models, future uncertainties mean insurers still face various risk challenges in subsequent operations.

Due to the potential high claim frequency characteristics of auto insurance in the short term, insurers need to consider increasing premiums for multiple-claim users. For such problems, Boucher et al. [?] adopted Kappa-N models and Bonus-Malus Scale (BMS) models. The commonality between these two models is directly embedding historical claim payment functions for this insurance type into the mean parameter of count distributions, with claim rates gradually decreasing when no claims occur and gradually increasing when claims appear.

Under pan-generalized nonlinear assumptions, letting the mean parameter have both linear and nonlinear logarithmic link function components with covariates yields expected premiums under three indicator variables: claim frequency, per-claim amount, and total claim amount (including intercept terms).

Formulas A3, A4, and A5 mainly differ in that function f(·) has different meanings, with only the specific meanings of independent and dependent variables being the same. As an auxiliary risk avoidance tool, the BMS system can be applied not only to auto insurance systems but also to other property insurance as one of the risk mitigation means to protect low-risk users' interests and maintain high-risk user thresholds.

Most existing auto insurance pricing models primarily consider general claim frequency and payment amounts but often lack methods to judge and resolve extreme risk points (such as major traffic accidents, chain collisions, or concentrated claims for certain models). Huge losses under extremely low probability, despite low occurrence frequency, have severe impacts. Failure to consider these factors in reserves may lead to decreased insurer solvency, affecting financial stability and industry confidence.

Key post-pricing risks for auto insurance include: First, claim payment uncertainty. Accidents are sudden and 偶然性, with probabilities potentially increasing sharply due to extreme weather conditions, major human-caused traffic accidents, etc., causing concentrated claim peaks in long-term business. For example, during year-end closing periods, numerous cases from such accidents may lead to insufficient survey and loss assessment capacity. Second, for some major accident auto insurance, tail risk exists where some large claims may require long-term pursuit for final determination, with compensation accumulating on the loss ratio from accident occurrence until claim settlement, resulting in higher payments after litigation concludes.

To prevent post-pricing risks, insurers generally establish reserves. Auto insurance actuarial work mainly studies outstanding claim reserves, requiring insurers to estimate future payment amounts based on available information for future disbursements. The two most important reserves are: Incurred But Not Reported (IBNR) reserves, which cover liabilities that have occurred but have not yet been noticed or reported by the insured; and catastrophe reserves, which address large claim phenomena caused by natural disasters, malignant accidents, large-scale traffic jams, or chain accidents. Such events have low occurrence probability but cause severe consequences, significantly impacting normal operations and solvency. Insurers generally extract a certain amount from collected premiums for catastrophe reserves and purchase reinsurance to prepare for catastrophes, making corresponding adjustments to catastrophe reserves based on historical catastrophe data.

Chapter 4: Research Methods, Key Difficulties, and Innovations

4.1.1 Literature Research Method

By reviewing domestic and foreign literature, this research systematically organizes knowledge of actuarial theory, risk pricing methods, and insurance reserves. By analyzing different theoretical perspectives, it summarizes applicable scenarios, advantages, and disadvantages of current methods, laying the foundation for subsequent research.

4.1.2 Data Analysis Method

In the pan-generalized nonlinear framework, models are highly data-dependent. This study will utilize real auto insurance data and appropriately reference simulated data at more complex levels to validate the effectiveness of different pricing models (such as Generalized Additive Models).

4.1.3 Model Construction and Optimization Method

Constructing the pan-generalized nonlinear framework requires parameter estimation and model optimization. When facing different module processing, different optimization methods (such as LightGBM) are used to improve models, with comparative analysis of different models.

4.1.4 Computer Simulation Method

Since insurance involves future risk prediction, future research needs to employ computer simulation.

4.1.5 Comparative Analysis Method

This research will compare practical experiences in auto insurance pricing and reserve assessment across different environments and explore impacts of different conditions on pricing models and reserve methods.

4.2 Key Difficulties

Property and Casualty (P&C) insurance faces numerous challenges in modeling claim frequency and claim severity: traditional GLM methods are too rigid; while machine learning methods are more flexible, they relatively lack interpretability; strong heterogeneity exists among insureds; nonlinear and interaction effects exist between claims and covariates; certain correlations exist between insurance types; and data exhibit characteristics such as multimodality and overdispersion [**?**].

4.2.1 Data Quality and Risk Heterogeneity

How to simultaneously address data quality and risk heterogeneity issues in actual actuarial modeling? This directly determines modeling effectiveness and the rationality of modeling approaches. Data quality affects modeling results—even with correct methods, poor input data quality leads to incorrect predictions. However, high-quality data is often difficult to obtain, typically requiring strict data validation and institutional guarantees. Except for some open datasets for reference, there are also artificially processed contents. For publicly published data, processing factors, methods, and whether processing occurred are unknown.

Such data facilitates compliance and transparency but significantly reduces model reliability and generalization. On one hand, scarcity of high-quality data becomes a major obstacle to modeling. Actuarial models heavily depend on data. In auto insurance pricing, numerous feature variables and high-dimensional data are involved. Data bias, missing values, or errors directly affect model stability and prediction accuracy, while high-quality, large-scale data is generally only held by insurance companies. Insurance data collection standards exceed those of other industries, with better completeness and update frequency than public data. However, due to user privacy concerns and various regulatory factors, future research cannot directly obtain these critical core data from insurers, only using artificially processed public datasets. Such datasets can only complete basic data preprocessing, meeting only fundamental functional requirements and cannot serve as complete basis for accuracy requirements, as some variables cannot present complete situations, ultimately leading to compressed or distorted feature spaces that affect conclusions from model training and validation, making them lack generalizability for actual business scenarios.

Beyond these challenges, risk heterogeneity makes modeling more difficult. Traditional credibility theory assumes risk is homogeneous, i.e., risk parameter is constant, with errors completely determined by random terms. In reality, the auto insurance business environment is not so ideal, with significant differences in vehicle attributes, driving behavior, geographic location, and usage habits among insured objects. Due to differences between auto insurance objects, risk characteristics vary among different insureds, creating risk heterogeneity that increases modeling difficulty and demands higher flexibility and expressive power from models. For example, the LocalGLMnet model can theoretically model local nonlinearity and spatial structural differences, but in practice, due to complex modeling processes and high sensitivity to neural network structure design and hyperparameter tuning, over-modeling may occur if data quality and risk heterogeneity do not reach certain levels, leading to wasted computational resources and erroneous conclusions.

In this process, data quality and risk heterogeneity are not independent issues but rather interwoven factors affecting model construction success. Historical experience data shows that errors include both random disturbance effects and potential structural biases. When samples contain more high-information features, the theoretical predictive power of models improves, but due to increased interaction between features and risk differences, the "information enhancement-heterogeneity increase" phenomenon makes parallel enhancement of predictive power and robustness more difficult. Expanding credibility theory under Bayesian frameworks can reveal the nature of parameter modeling under risk heterogeneity, i.e., risk parameter should be modeled as a random variable following some prior distribution (structural function), creating statistical correlations between different risk levels.

For example, based on Bayesian axioms, the Bühlmann credibility model estimates risk levels using linear combinations of observations, providing credibility premium formulas as one of the important pricing tools for determining non-homogeneous risks. The model construction process must consider data quality and risk heterogeneity issues, requiring modelers to have strong data preprocessing and modeling capabilities to ensure models have predictive functions while demonstrating interpretability and practicality.

4.2.2 Feasibility Issues of Weather Index Auto Insurance

Although weather index insurance models have been applied to agricultural insurance and can theoretically be extended to auto insurance, some unforeseen problems remain when applied to auto insurance.

Spatiotemporal matching is a major challenge. Auto insurance accidents have strong individual attributes, but in reality, only regional average meteorological data or meteorological station data can be obtained, making it impossible to precisely locate weather conditions at the exact time of a specific accident. Even with high-resolution remote sensing data or ground meteorological stations, this "insufficient representativeness" problem cannot be completely solved, affecting compensation trigger accuracy and fairness [**?**]. Some scholars use windshield wiper status data, NCDC meteorological station data, and related weather accident data to better achieve this purpose.

There is no quantitative boundary for compensation triggers. Traditional insurance business loss determination is decided by surveyors with strong subjectivity. Index insurance triggers require strict adherence to objective standard procedures for loss determination. The problem lies in how to set a quantitative measurement standard that accurately reflects real risk while maintaining stability, and how this quantitative benchmark leads to "basis risk"—whether actual claims occur but cannot trigger compensation because the meteorological risk threshold is not met, or whether claims have actually occurred but the meteorological risk threshold is exceeded, leading to institutional-level compensation issues.

There are no precedents for index insurance application in the auto insurance field, with no mature successful examples for future research to draw upon. Only because agricultural product growth and maturity heavily depend on weather and geographic factors has agricultural index insurance development benefited from policy subsidies, reinsurance support, and strong government support. Auto insurance, as a high claim frequency insurance type, easily leads to extremely sensitive compensation standards. Moreover, weather index auto insurance is an innovative insurance type based on a new conceptual framework, and even if feasible, its market acceptance is unknown. Nevertheless, weather index auto insurance theoretically helps improve model prediction accuracy while avoiding moral hazard, benefiting pricing efficiency. Future research similarly requires in-depth study within a certain scope, such as matching meteorological and vehicle risk data, suitable model selection, reasonable compensation logic design, and maintaining product fairness. Solving these related problems requires deep research collaboration between meteorology, actuarial science, and big data, combined with market development environments and strong national-level promotion and support. Using experimental methods, such as operating demonstrations in simulated environments or meeting with potential insurance consumers, can help obtain evaluation information through testing, ultimately helping people understand and accept index insurance products [?]. Clark [?] criticized that economists should focus more on optimizing insurance product design to achieve optimal quality, avoiding situations where incorrectly designed insurance products cause harm to consumers or insureds due to insurance companies or government launches.

4.3.1 Constructing a Predictable Vehicle Insurance Actuarial Rating System

Currently, China's auto insurance system design still has many unreasonable aspects, requiring continuous enrichment and improvement in product structure, with problems in price formation mechanisms and risk response capabilities, especially evident for high-value vehicles, high-compensation accidents, and diversified usage scenarios [?]. Related to pricing, traditional insurance systems have become difficult to adapt to current requirements in risk identification, loss compensation, and cost allocation. Insurance companies only model static factors like people, vehicles, and roads through a few fixed variables (age, vehicle model, etc.), without considering consumers' dynamic factors (driving habits, road conditions, etc.), resulting in relatively average premium distribution and inability to achieve precise risk pricing. Meanwhile, traditional auto insurance products also suffer from high sales costs, overlapping channel charges, and other drawbacks, which not only hinder fair operation of the insurance industry but also fundamentally weaken the scientific nature of actuarial pricing and regulatory recognition.

Under current circumstances, building China's predictable vehicle insurance actuarial rating system is a critical step in adjusting unreasonable imbalances in the current insurance market structure and improving the entire society's risk governance system. "Predictable" refers to applying advanced data technology and modeling methods to discover potential risks in advance, proactively preparing before occurrence to achieve early warning, early prevention, and early disposal—a new working mechanism. “Vehicle-like” provides reference significance for other vehicle type insurance. Based on integrated modeling and complex data fusion modeling ideas from the pan-generalized nonlinear system, and leveraging the optimization capabilities of machine learning algorithms like LightGBM and neural networks’ complex data modeling capabilities, the model achieves unified high standards in interpretability, nonlinear expression capability, and high-dimensional variable applicability. In actuarial modeling, LocalGLMnet and LocalGAMnet models are used to fit nonlinear relationships between claim amounts and risk factors, ensuring predictive power while guaranteeing interpretability and considering issues like model overfitting and dimensionality disaster. Only by holistically utilizing institutional reconstruction, mechanism updates, and technology empowerment to form synergies can we truly achieve fundamental transformation of auto insurance from passive compensation to active risk control, from single products to comprehensive services, and from price orientation to risk orientation.

4.3.2 Pioneering ESG Pricing for Auto Insurance

Traditional auto insurance pricing is based on historical accident data combined with specific object data about vehicles and drivers, using empirical rate standards to determine underwriting costs. This empirical rate pricing based on past experience does not consider factors affecting accident occurrence rates and loss severity, ignoring the potential impact of climate conditions on vehicle accidents. Especially under climate warming background, climate conditions have shown strong systematic risk characteristics. Therefore, it is necessary to integrate climate factors into auto insurance pricing models, introducing climate factors in addition to human and vehicle factors.

To streamline climate types involved in weather indices, future research will only explore two typical regions: long-term hot regions and rainy regions. This approach helps reduce variable numbers, ensure model stability, and facilitates subsequent promotion and expansion to other climate region types.

During modeling, high-resolution meteorological data will be fused with historical auto insurance claims data to obtain weather-accident correlation regression models, generating “weather index auto insurance” combined with weather information and accident location information. When vehicles have accidents on high-temperature days with claim amounts below threshold values, the system can achieve automatic compensation based on matching relationships between weather information and vehicle accident information without manual survey and claim approval, greatly improving claim timeliness, reducing claim disputes, and providing good customer experience.

Practical operations attempt to introduce a second dual-model system combining traditional models with weather index models beyond the underlying modeling logic. One is the main model based on the pan-generalized nonlinear framework, using a pan-generalized nonlinear framework to model conventional characteristics (including age, mileage, driving score, and historical claims). The other is a sub-model specifically modeling relationships between weather factors and accidents (including accident frequency and accident compensation). Combining their outputs allows consideration of natural environment impacts during claim cost setting, making pricing more realistic.

The ESG auto insurance pricing and claims framework proposed in this paper can be easily extended or transplanted. As meteorological data acquisition accuracy continues to improve, technical means combining IoT (e.g., onboard meteorological monitoring) and 5G communication technology will continue to iterate, maintaining research advancement while enabling implementation in actual business. Most index insurance still concentrates on agricultural insurance and reinsurance levels. Extending such mechanisms to individual auto insurance dimensions remains an innovative attempt. Especially in the Chinese market, current auto insurance products suffer from serious homogenization and lack differentiated customer experience. The ESG pricing mechanism designed in this study can effectively promote auto insurance research development under dual drivers of policy and technology.

Chapter 5: Dissertation Writing Schedule

5.1 Dissertation Writing Schedule Arrangement

1. **Data Collection and Learning Phase (January 2025 - May 2025)**: Collect data, study relevant statistical knowledge, modeling methods, and related modeling techniques.

2. **Data Processing and Modeling Phase (June 2025 - September 2025)**: Conduct data preprocessing and feature engineering, perform modeling experiments.

3. **Draft Completion Phase (October 2025 - January 2026)**: Analyze experimental results, interpret models, conduct visualization and comparison experiments, and perform deep revisions after preliminary defense.

4. **Final Revision Phase (February 2026 - March 2026)**: Revise research reports, organize experimental code and data, conduct final summary and discussion, and finalize the manuscript.

Literature Review

"Research on Auto Insurance Actuarial Rating Systems and Related Risks with a Pan-Generalized Nonlinear Framework"

Auto insurance pricing, as one of the core issues in actuarial science, affects insurers' profitability and cash flow. This literature review briefly elaborates on different auto insurance developments across regions and discusses issues related to auto insurance ratemaking systems, special auto insurance pricing, no-claim discount reward-penalty theory, and reserve assessment methods. The review uses statistical distributions and model structures as entry points to analyze specific operational logic of models, focusing discussion on literature establishing pricing frameworks and reserve assessment based on Tweedie distribution families, Generalized Additive Models, LocalGLMnet modeling methods, and most methods based on frequentist and Bayesian schools. To demonstrate the application effects of risk modeling thinking on real business, this paper also includes literature on BMS systems, GAM reserve assessment models, etc., to explore a construction method for risk dynamic adjustment mechanisms in auto insurance. Regarding reserve research, literature includes assessment methods under both homogeneity and heterogeneity assumptions, specifically using traditional methods like chain ladder for aggregate accident claim data prediction modeling, and using machine learning technology to improve models for individual accident claim data prediction modeling on existing algorithmic foundations.

Beyond general auto insurance, the review also investigates weather index insurance, using remote sensing data and ground observation data to establish multi-dimensional input methods to improve model identification capabilities for extreme meteorological events. Empirical findings show this model performs well in compensation stability and regional risk differences.

Based on data levels, this literature review selects real auto insurance and weather insurance claim data as case studies, setting control groups for different modeling approaches to analyze and evaluate similarities and differences in prediction effects. This paper focuses on constructing a replicable and scalable insurance modeling method in the property insurance field, comprehensively considering regulatory and data modeling factors to improve property insurers' quantitative pricing capabilities, promoting new practical experience and theoretical foundations for property insurance industry digital transformation and actuarial pricing.

Keywords: Auto Insurance, LocalGLMnet, Generalized Additive Model, Feed-Forward Neural Network

I. Analysis of Auto Insurance Practice

Analysis is conducted on six representative countries and regions: Japan, UAE, Brazil, South Africa, China, and India (ranked by vehicle density at approximately 624, 540, 249, 200, 173, and 30 vehicles per thousand people respectively) and North America, Australia-New Zealand, UK-Ireland, Germany-France, Northern Europe, and Singapore-Malaysia regions. The total number of vehicles in these countries and regions is approximately 1.26 billion.

According to International Energy Agency (IEA) data, global vehicle ownership in 2020 was about 1.4 billion, with these countries and regions accounting for about 90% of the global total, covering six continents (all global continents except Antarctica), including North America (US-Canada AC), South America (Brazil B), Europe (UK-Ireland EI, Germany-France GF, Northern Europe NE), Africa (South Africa S), Oceania (Australia-New Zealand AN), and Asia (Japan J, UAE A, Singapore-Malaysia SM, China C, India I), with codes abbreviated using English initials.

(1) Analysis of American Auto Insurance Practice In the early 19th century, the United States established compulsory insurance laws providing legal support for subsequent auto insurance business development. However, without modern tools like computers, actuarial pricing could only rely on subjective judgment and experience, temporarily lacking scientific theoretical and analytical systems.

Entering the 1980s, Coutts' [**?**] auto insurance premium pricing method was highly complete and systematic. The research considered not only underwriting factors, claim frequency, and claim amounts but also had unique insights into reasonably incorporating external factors like inflation into premium calculations. To make premium pricing more scientific and accurate, Coutts used modern statistical methods such as Orthogonal Weighted Least Squares (OWLS) to study the impact of factors like vehicle age, policyholder age, and vehicle type on different vehicles' claim frequencies based on historical claims data, establishing a vehicle pricing model combining historical loss ratios and price indices. This model enables accurate premium assessment for different vehicles, achieving high coverage within reasonable price ranges and providing scientific pricing basis for insurers, making pricing work clearer and more straightforward. Brockman et al. [**?**] addressed Johnson & Hey's shortcomings in handling incomplete policy years through GLIM models and overdispersion parameters, proposing more accurate auto insurance rate pricing. More detailed parameterization improves pricing accuracy, enhances pricing precision, deepens theoretical research, and provides references for future rate pricing.

As the auto insurance industry matured in pricing, Generalized Linear Models gradually became the foundation of modern auto insurance pricing, though some theories hold that Generalized Additive Models are more suitable for real-world auto insurance pricing. Meanwhile, Usage-Based Insurance (UBI) products began to emerge. Since 2004, companies have launched behavior-based insurance products such as Progressive's Snapshot, Allstate's Drivewise, GEICO's DriveEasy, Liberty Mutual's RightTrack, and Tesla's Autopilot. Boucher J.P. et al. [**?**] noted that with the explosion of telematics data, UBI insurance is the trend the auto insurance market will follow.

Compared to North America, South American auto insurance markets are much smaller, with Brazil as the largest South American market accounting for only about half of the entire South American market. Although Brazil's auto insurance market accounts for 60% of its automotive industry, only 27% of vehicles have adequate insurance coverage. For Brazil, many factors influence auto insurance rates, with the most important being vehicle and driver personal information. For example, sports cars and vehicles with premium parts (alloy wheels, fog lights) are more susceptible to theft or robbery, resulting in higher premiums. Although most Brazilian vehicles use gasoline, diesel, and other energy sources, insurance amounts do not differ based on fuel type. Brazilians fear their cars being stolen more than being in accidents, so insurance rates are not determined solely by vehicle price and brand, nor are luxury vehicles necessarily subject to high premiums while economy vehicles enjoy preferential policies. For instance, Volkswagen Golf is one of Brazil's lowest-priced models, but due to expensive parts, its insurance price is slightly higher relative to its vehicle price.

Beyond vehicle type, insurance rates are also determined by vehicle parking location. For Brazil, vehicle location is an important factor in determining premiums, whether vehicles are in garages, on streets, or parked at workplaces or schools, affecting insurers' risk assessments. Connected vehicle technology can enhance vehicle safety and reduce premiums. Using vehicle registration postal codes can determine premium prices. Although some areas have higher crime rates leading to premium increases, more importantly, different regions have different driver habits and accident probabilities.

(2) Analysis of European Auto Insurance Practice The UK-Ireland region has been at the forefront of global maritime insurance due to "the empire on which the sun never sets" maritime hegemony, global economic trade networks, and unique maritime and geographical structural and policy system advantages.

Before 1930, the UK passed the Road Traffic Act, becoming the world's first country to implement compulsory auto insurance, requiring owners or drivers to provide protection for property damage or personal injury caused to others on roads. Socioeconomic development and non-life insurance prosperity also provided impetus for new insurance types like auto insurance, presenting diversified development trends in the insurance industry.

Ireland's auto insurance development history demonstrates the development of social transportation methods and changes in transportation legislation. At the end of World War I in the early 20th century, people acquired their first automobiles, bringing risks of traffic accidents, injuries, and property damage. In the UK, the 1933 Road Traffic Act brought all motor vehicles into compulsory insurance scope, providing help for risks endangering people's lives, health, and property damage, greatly protecting victims' rights. Over time, insurers launched various insurance products for increasingly numerous cars, including fire and theft insurance and special insurance types for cars or electric vehicles in recent years. Some insurance is legally mandated, such as Ireland's Road Traffic Act, which clearly states that no motor vehicle can operate on roads without legally required insurance, with violators facing various negative consequences including high fines, license revocation, and even vehicle confiscation, strengthening effective management of public safety and road risks [?].

In 1939, Germany implemented compulsory motor vehicle third-party liability insurance, requiring all motor vehicle owners to purchase third-party liability insurance, marking the transition of auto insurance from voluntary to state-mandated management. In France, all vehicle owners must purchase auto insurance, with 70% of vehicles insured beyond compulsory traffic accident liability insurance, including third-party liability, vehicle damage, and other commercial insurance. Different insurers propose different deductibles, though third-party liability does not allow deductibles. France's reward-penalty system is legally mandated: the initial coefficient is 1, increasing by 25% if an accident occurs within a year, and decreasing by 5% if no accident occurs. The coefficient lower limit is between 0.5 and 3.5 (with some exceptions). Final premiums are calculated by multiplying differentiated premiums by reward-penalty coefficients [?].

Currently, besides the US, Germany and France have the world's most powerful property insurance companies and insurers themselves. In the decades after World War II, they have been one of Europe's major industrialized regions, with the largest global market share in property and insurance companies, covering almost any accessible region worldwide. However, the global pandemic's impact spread to Germany and France's auto insurance markets. Although initial data showed decreased travel kilometers and accident rates, they soon rose, with both incidence and damage severity increasing. Additionally, supply chain issues caused parts price increases, requiring insurers to spend more on auto accident claims processing, so auto losses will exceed pre-pandemic prices. These factors will continue to affect insurers' need to adjust rates upward. Considering recent high accident incidence and continuously increasing claim amounts, introducing an efficient and fast auto insurance claims process is necessary, highlighting the importance and necessity of auto insurance pricing research [?].

Northern Europe's auto insurance dates back to the early 20th century. With automobile popularization and frequent traffic accidents, auto insurance gradually improved. In the early 20th century, Sweden already had its first company, the Swedish National Insurance Company, providing auto insurance for first-time car buyers, mainly covering physical damage and owner liability compensation. Later development continued through the late 1910s when Norway already had auto insurance, though still following the form of owners' own selection with certain industry practices. Before and after World War II, Sweden adopted compulsory third-party liability insurance requiring owners to purchase insurance to compensate third-party property, influencing not only domestic auto insurance market development but also improving insurance industry standardization. Besides France, Denmark and Finland also adopted compulsory insurance systems in the 1940s-1950s, greatly promoting insurance industry development. Sweden's auto insurance research is also very important. Northern Europe's actuarial development is relatively advanced, including Jung J. et al.'s [?] insurance rate estimation improvement models, and Delong Ł. et al. [?] analyzing claim influencing factors based on Swedish motorcycle insurance, providing more experimental support for industry preference in Poisson-gamma parameterization.

From the perspective of auto insurance ESG compliance, the EU can be considered the most active region in legislation globally, having successively introduced strong laws and regulations including the Sustainable Finance Disclosure Regulation (SFDR) and EU Sustainable Taxonomy (EUT). These provisions mainly aim to gradually meet environmental, social, and governance requirements in business operations for enterprises or financial institutions. Insurers must comply with the EU Sustainable Taxonomy, especially regarding product transparency and environmental responsibility, and can incorporate climate adaptation measures (such as weather index insurance) into non-life insurance underwriting processes to address emerging risks. For example, weather index auto insurance based on high-temperature conditions can provide product responses to indirect greenhouse gas emissions mentioned in PCAF initiatives.

(3) Analysis of African and Oceanian Auto Insurance Practice In global economic development, the African continent can be considered the least developed region, with its insurance industry overall development also relatively backward. Limited by low motor vehicle ownership, weak insurance awareness, and imperfect systems, auto insurance struggles to form scaled market mechanisms.

To date, only a very small number of African countries have compulsory motor insurance laws, largely depending on motor vehicle owners' voluntary participation in commercial insurance and public risk awareness. South Africa is a relatively mature insurance market on the African continent, with premium income accounting for about 70% of total African premium income. However, there is no compulsory motor insurance system, with only 1/3 of vehicles on roads having insurance [?]. Therefore, the South African government formulated corresponding laws and regulations and established a "Road Accident Fund" as an alternative social security mechanism, forcibly collecting part of the fees or fuel surcharges from vehicles purchasing fuel locally as fund sources to provide compensation for traffic accident victims.

Similarly, Namibia is one of five sub-Saharan African countries (including Eswatini, Botswana, and Lesotho) that levy fuel taxes to support road injury victims [?]. Since MVAF implementation in 1990, there has been little research on its impact, particularly from the perspective of users or health workers providing services. Any person injured in a car accident or family members of those killed in car accidents can apply for assistance from MVAF. Some surveyed individuals reported that MVAF covered all their medical expenses [?]. Like some compulsory insurance, MVAF operation also has many problems, such as its system design relying on accident reporting and police filing, leading to large numbers of accidents in rural areas lacking law enforcement that fail to obtain claim qualifications.

In the Australia-New Zealand region, legislation progressed unevenly across states, with all states achieving compulsory auto insurance only after Western Australia passed relevant acts in 1943. Previously, Victoria attempted multiple times but failed, with Tasmania taking the lead in introducing it in 1935 [?]. Victoria, Australia also implemented a plan similar to MVAF called the Traffic Accident Commission (TAC), funded through statewide vehicle registration taxes. TAC helps promote communication between health professionals, provides consistent and up-to-date information, and coordinates service delivery. Case managers' care coordination not only facilitates timely access to assistance but also achieves cost reduction and efficiency improvement [?].

(4) Analysis of Asian Auto Insurance Practice At the end of the 20th century, the UAE's gradual opening of its insurance market led to rapid development, but due to religious reasons, many insurance products were restricted from listing, leading to the sale of many religious or Sharia-compliant products not found in other countries, creating a special situation in the UAE's insurance industry—coexistence of traditional insurers and Islamic insurers (Takaful).

Among Asian countries, Singapore and Malaysia's special history places their actuarial development at the world forefront, such as early introduction of BMS systems [?]. Malaysia is a religious country that, like the UAE, has religious insurance products, but differs in considering this type of auto insurance as an important component of the national Islamic financial strategy core [?]. Takaful auto insurance has developed for nearly 50 years since 1985, representing a relatively mature special type of vehicle insurance market globally.

Singapore's rapid increase in automobile numbers in the 1970s, combined with limited land space and high population density, exacerbated traffic congestion [?], leading the government to formulate strict traffic rules such as area licensing schemes, which indirectly facilitated future auto insurance development through electronic device access identification. Nowadays, in the Singapore-Malaysia region, auto insurance is the largest proportion business for comprehensive insurers, accounting for about 36% of the entire market.

Japan, as an Asian country that established modern insurance systems earlier, has a relatively complete actuarial system and mature insurance regulatory system. As early as 1914, Tokyo Marine Insurance Company, which pioneered motor vehicle insurance business in Japan, first operated motor vehicle insurance. In reality, Japanese car owners attach great importance to risk distribution management. According to statistics, since 2011, "limited driver" auto insurance contracts have exceeded "unlimited driver" contracts in market proportion, with "policyholder and spouse only" insurance contracts reaching 70% [?].

India, as one of the world's most populous countries, has huge differences in traffic management status between states. Therefore, the Indian government requires all motor vehicles operating in public places to be insured, with driving uninsured vehicles considered illegal behavior. Third-party liability insurance belongs to the compulsory insurance category, intended to provide corresponding compensation for life and property damage caused by motor vehicles operating in public places. The reason is that motor vehicles operating in public places inherently possess certain danger, and vehicle users may have limited financial capacity to provide sufficient compensation to victims. Regarding insurance duration, Indian law originally stipulated mostly one-year motor vehicle insurance, but according to Supreme Court guidance, since September 1, 2018, newly purchased four-wheeled vehicles must have three-year third-party insurance, and two-wheeled vehicles five-year insurance. The purpose is to improve auto insurance continuity, reduce uninsured driving, and strengthen insurance system enforcement and social security functions.

Since 1980, China has gradually recognized insurance's role in risk management and resource allocation in socio-economic development, shifting from exclusive operation by People's Insurance Company of China to the current multi-operator business model. China officially implemented the "Regulations on Compulsory Traffic Accident Liability Insurance for Motor Vehicles" in 2006 and launched three sets of industry-unified clauses (A, B, C) in 2009. However, due to short development time and difficulty adapting to economic development and traffic condition changes, problems gradually emerged including high legal risks, narrow coverage scope, and non-transparent clauses. For a long time, the auto insurance industry also faced consumer complaints about high premiums, difficult claims, and over-insurance with under-compensation, with service levels far from public expectations. To break these bottlenecks, promote market-oriented premium mechanisms, and improve risk pricing capabilities, the China Banking and Insurance Regulatory Commission launched commercial auto insurance reform pilots in 2015, marking China's commercial auto insurance entering a new stage of refinement, specialization, and differentiation, with auto insurance pricing systems evolving from unified rates to limited autonomous pricing [**?**]. In 2017, commercial auto insurance reform was implemented nationwide, introducing the no-claim reward-penalty mechanism implemented in Taiwan region since 1996, providing premium discounts for customers with consecutive claim-free years and increasing premiums for those with claims, providing more reliable institutional tools for professional auto insurance actuarial pricing in China.

II. Auto Insurance Data Analysis and Processing

(1) Nature of Auto Insurance Data The accuracy of auto insurance rate determination is an important manifestation of insurance companies' risk management capabilities. For reasonable pricing, insurers must consider various uncertain factors in risk analysis that can predict partial parameters of loss distributions. Loss distributions describe the probability distribution of loss amounts (i.e., insurance claim amounts) during specific periods and are necessary tools for insurers to assess risks.

In practice, loss distributions play crucial roles in insurance pricing, especially for more complex insurance contracts where loss distribution selection and parameter estimation are important. For insurance, loss distributions are constrained not only by accident incidence but also by accident severity, claim amounts, and economic data conditions. In auto insurance pricing, loss distributions are generally divided into two parts: loss frequency distribution and loss severity distribution. The former represents the number of accidents, while the latter represents claim amounts per accident, with their combination forming total loss distribution.

Frequency distributions record accident occurrence counts within certain time periods, generally following discrete distributions such as Poisson Distribution or Negative Binomial Distribution, helping insurers predict how many accidents occur during certain periods. Severity distributions use continuous distributions such as Gamma Distribution or Lognormal Distribution, enabling insurers to predict approximate loss amounts per insurance event.

Common loss distributions in auto insurance pricing include:

Poisson Distribution: Models insurance accident frequency, assuming accident occurrence probability is independent per time unit with average occurrence λ. The probability density function is:

$$P(k;\lambda)=\frac{\lambda^k e^{-\lambda}}{k!}$$

where λ is average accident frequency and k is actual accident count. In auto insurance pricing, Poisson distribution estimates accident occurrence counts during certain periods. For example, assuming a region's vehicles have average annual traffic accidents $\lambda = 0.15$, Poisson distribution helps insurers calculate expected accidents per policy in the next year.

Negative Binomial Distribution: Represents the probability of exactly k successes (or n failures) on the (n+k)th attempt, typically modeling overdispersed frequency data (where actual data variance exceeds the mean). The probability density function is:

$$P(n;k,p)=\binom{n+k-1}{k-1}p^k(1-p)^n$$

where n+k is trial count, p is success probability, n is accident count, and k is non-accident count. Negative binomial distribution better captures accident occurrence volatility and is more suitable than Poisson distribution for some overdispersed data situations.

Gamma Distribution: Models severity distributions, especially suitable for describing skewed and asymmetric claim amounts. The probability density function is:

$$f(x;\alpha,\lambda) = \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$$

where α and λ are shape and scale parameters. Gamma distribution can describe situations with small claim amounts but occasional large compensations, suitable for describing right-skewed insurance claims.

Lognormal Distribution: Suitable for modeling claim amounts composed of products of multiple independent random variables. The probability density function is:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$$

where and σ are the mean and variance of data logarithms, used to describe extremely asymmetric loss amounts, especially suitable for modeling extreme claim events.

After determining reasonable loss distributions and parameters, insurers can estimate future compensation situations based on certain confidence levels for policy pricing, risk management, and reserve extraction. Starting from historical experience data, using loss frequency and severity calculations, pricing is determined according to insurance contracts. Through loss distribution models, insurers can better predict potential compensation losses and price reasonably. For example, gamma distribution calculates large compensation risks; Poisson distribution represents accident occurrence probability distributions. Through loss distribution analysis, insurers can further test and adjust capital adequacy. If large compensation losses occur, reserves need to be increased for future claim needs. Loss distribution analysis also enables insurers to anticipate future huge compensation situations and configure appropriate mechanisms for catastrophe transfer.

Based on loss distributions, frequency and severity distributions jointly calculate auto insurance compound losses, serving as important tools for reasonable pricing, reserve extraction, and risk evaluation. The Tweedie distribution is a compound Poisson distribution widely used in insurance fields, mostly used to characterize data with zero values and continuous positive value distributions, such as insurance claim data.

In statistics, the Tweedie distribution family is a very important distribution family, widely used for modeling overdispersed data, continuous non-negative data, and discrete data. The Tweedie distribution family is valued for its ability to adjust distributions according to different exponential parameters (typically denoted ϕ) to describe different loss data. For example, in insurance pricing, Tweedie distribution families describe situations with both continuous losses and a few large losses.

The Tweedie distribution family is an exponential family distribution with probability density function (PDF):

$$f(y;\mu,\phi) = a(y,\phi)\exp\left(\frac{y\theta-\kappa(\theta)}{\phi}\right)$$

where is the mean parameter, ϕ is the distribution parameter generally related to distribution variance and scale; is a function related to mean ; a(y, ϕ) is a normalization constant ensuring PDF integration equals 1.

The Tweedie distribution family's notable characteristic includes logarithmic link functions in exponential functions, enabling determination of specific distribution types based on actual conditions and suitability for describing various data types with different characteristics. For different ϕ values, the Tweedie distribution family can produce different loss characteristics, including common probability distributions such as normal, Poisson, and gamma distributions. When $\phi = 0$, Tweedie distribution degenerates to normal distribution with PDF:

$$f(y) = \frac{1}{\sqrt{2\pi\phi}}e^{-\frac{(y-\mu)^2}{2\phi}}$$

When $\phi = 1$, Tweedie distribution becomes Poisson distribution for describing event occurrence counts. When $\phi = 2$, Tweedie distribution becomes gamma distribution, particularly important in insurance risk modeling for describing continuous positive losses. Typically, auto insurance uses Tweedie distribution with ϕ in the (1,2) interval, around 1.65, belonging to Poisson-gamma compound distribution. This distribution generally assumes claim event counts follow Poisson distribution while each claim amount follows gamma distribution.

When $\phi = 3$, Tweedie distribution becomes inverse Gaussian distribution, commonly used to model work completion within fixed time periods with PDF:

$$f(y;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi y^3}}\exp\left(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right)$$

The Tweedie distribution family has important properties making it suitable for handling insurance data, financial risks, and other random phenomena. Important properties include: second-order moment existence (except for special cases with parameters less than or equal to two, most Tweedie distribution families have second-order moments with variance typically dependent on distribution parameters); exponential family property (belonging to exponential family distributions, enabling easy separation and parameter estimation using statistical methods to improve accuracy); representation as sums of simple distributions (Tweedie distribution is also a collection class function of simple distributions); flexibility in skewness and kurtosis adjustable through parameter ϕ changes (larger ϕ leads to larger skewness and kurtosis).

In auto insurance pricing models, Tweedie distribution is often used to model compensation amounts. Under normal conditions, most compensation amounts are relatively small, but serious accidents or high-value vehicle damage may require large payments. Tweedie distribution with ϕ close to 2 better represents this characteristic. Tweedie distribution family parameters are typically obtained using Maximum Likelihood Estimation (MLE), finding parameters that maximize sample data likelihood functions. Since Tweedie distribution likelihood functions are difficult to solve, numerical optimization methods are required. Assuming n independent samples (y$_1$, ..., y), the Tweedie distribution likelihood function is:

$$L(\mu, \phi; y_1, ..., y_n) = \prod_{i=1}^{n} f(y_i; \mu, \phi)$$

Maximum likelihood estimates are obtained through log transformation and derivative calculation of the likelihood function.

(2) Discrimination and Fairness in Pricing In auto insurance, insurers typically consider driver characteristics like age and gender for parametric pricing, but some factors may raise compliance and regulatory issues for discrimination and fairness reasons.

For example, based on age factors, young drivers without driving experience or with short driving time have higher rates than experienced drivers. In terms of gender, males have relatively higher accident rates, so from an actuarial perspective, male rates should be higher. In 2011, the EU ruled to prohibit gender discrimination in auto insurance pricing [**?**]. Similar regulations exist that prohibit insurers from using certain characteristics within pricing frameworks, constituting illegal discrimination. Simply ignoring protected policyholder attributes is not an appropriate solution, as this still allows inference of protected attributes from unprotected covariates, leading to proxy discrimination phenomena. Under pressure, insurers must find new variables that are acceptable and do not reduce accident prediction accuracy. For example, policyholder personal information (age, gender, claim history) and vehicle characteristics (price, seat count, operational purpose) have only indirect correlations with insurance losses in subsequent modeling. New technological developments should prompt insurers to explore UBI premium modeling methods [**?**].

Chen et al. [**?**] studied non-discriminatory pricing formulas for weighted mixed male-female mortality rates, avoiding direct and indirect discrimination. Although the NDP formula only uses non-discriminatory features as rating factors, Lindholm M. [**?**] introduced an adjustment requiring knowledge of policyholders' discriminatory characteristics, providing an explicit mathematical method to eliminate indirect discrimination, forming part of the theoretical foundation for feature selection in pan-generalized nonlinear framework modeling. Meanwhile, Lindholm M. [**?**] argues that avoiding proxy discrimination does not guarantee group fairness. Simply ignoring protected information cannot ensure absence of discrimination in pricing. When statistical associations exist between covariates used in pricing, unprotected covariates may serve as proxies for undesirable variables like gender or race. From a regulatory perspective, closely linking insurance policies to actual risks compared to overly broad actuarial categories can improve actuarial fairness and reduce cross-subsidization [?].

III. Auto Insurance Actuarial Rating Framework

(1) Known Pricing Models and Methods Auto insurance actuarial rate pricing methods—assuming claim counts follow Poisson distribution and individual claim amounts follow gamma distribution, with total claim costs conforming to Tweedie compound Poisson distribution.

Smyth and Jørgensen [?] addressed the problem of formulating fair and accurate insurance rates based on aggregated insurance data, proposing the use of Tweedie compound Poisson models for dispersed modeling of insurance claim data, establishing linked linear models for mean cost and dispersion separately to simultaneously model mean and dispersion in generalized linear models. When only total claim costs are observed, approximate maximum likelihood estimates for mean and dispersion coefficients are obtained by alternating between two generalized linear models; approximate REML methods are also introduced to adjust dispersion estimates. When both claim costs and frequencies are observed, joint likelihood functions are used for parameter estimation, also using approximate REML methods. Using 1977 Swedish third-party auto insurance data as an example, log-linear models were fitted, finding all three explanatory factors (vehicle brand, annual kilometers, no-claim reward level) had significant main effects on both mean and dispersion, with dispersion effects being more significant. When claim counts are unavailable, Gao G. [?] fits marginal Tweedie compound Poisson models and proposes a new model fitting method using Expectation-Maximization (EM) algorithm, equivalent to iteratively reweighted Poisson-gamma regression on expanded datasets to improve pure premium model fitting effects. Jørgensen and de Souza [?] applied Tweedie compound Poisson models to insurance claim data, using iterative least squares methods from generalized linear models, employing stable Newton-Raphson algorithms, and utilizing parameter orthogonality to calculate standard errors for parameter estimates.

Traditional auto insurance pricing models are typically GLM models dependent on historical data, with premiums depending on self-reported rating variables (e.g., age, gender, driving history, postal code) that capture policyholder and insured vehicle characteristics and are typically only indirectly related to accident risk, with key factors being age, license age, postal code, engine power, vehicle usage, and claim history. However, the relationship between accident count and driving distance is not necessarily linear; in other words, vehicle driving distance and accident risk are not necessarily proportional [?, ?].

Regarding modeling random effects and nonlinear relationships, Breslow N.E. [?] proposed approximate inference methods for generalized linear mixed models, comparing PQL, MQL, Bayesian methods (such as Gibbs sampling), and other methods (such as Generalized Estimating Equations GEE) through simulation studies, elaborating on their advantages and disadvantages. Verbelen R. [?] used generalized additive models and component predictors to quantify and explain the impact of telematics variables on expected claim frequency, using Wood S N.'s [?] approximate tests to conclude that random effects were unnecessary. Boucher J.P. [?] used GAM models to model 71,489 PAYD auto insurance policies in the Spanish market, covering mileage, risk exposure time, policyholder age, vehicle age, gender, and parking type information, analyzing the combined impact of mileage and risk exposure time on accident risk. Research shows accident risk changes slowly after mileage reaches certain values, meaning long-time drivers' skill improvement may be one reason for reduced accident risk. Beyond GAM models, Lee et al. [?] proposed a flexible generalized varying coefficient regression model that relaxes restrictions on covariate type differentiation and interaction terms in traditional varying coefficient models, providing penalized least squares estimation, sieve estimation, and kernel estimation methods, deriving their convergence rates and asymptotic properties. The GAM framework only applies to loss distribution modeling under exponential families; however, many distributions do not belong to this category, such as negative binomial, multivariate negative binomial, and beta-negative binomial distributions. To handle loss assessment problems under broader frameworks, generalized additive models for location, scale, and shape (GAMLSS) can be considered. In property insurance, GAMLSS can be used for spatial data analysis, such as for auto theft data [?]. De Bastiani et al. [?] considered spatial components of Gaussian Markov random fields in GAMLSS models, while Ramires et al. [?] proposed a GAMLSS clustering method considering latent variables to minimize or correct anomalies, such as addressing difficult-to-explain bimodal phenomena.

GAM, GAMM, and GAMLSS models all involve spline functions such as cubic regression splines, B-splines, and P-splines. They are all forms of basis functions, typically using linear combinations of given functions to represent nonlinear relationships. Basis functions construct relationships between response and explanatory variables, while link functions establish connections between response variables and linear predictors.

Beyond regression methods like GLM, Bayesian methods are also convenient and efficient choices, especially when dealing with random or complex scenarios. Unlike GLM's fixed nature that may lead to unreasonable results due to incomplete or uncertain data, Bayesian methods can continuously adjust models based on known situations, with each new change improving upon previous changes. Unlike traditional pricing methods, Bayesian methods can not only increase accuracy by adding more features but also accommodate complex interactions between features.

Dimakos X. & Frigessi A. [?] proposed a full Bayesian method based on hierarchical models with latent variables for non-life risk premium assessment, using Bayesian estimation-based interactive regional latent variables to correct traditional GLM estimates when calculating claim counts, while considering uncertainty factors in claim counts. A simulated investment portfolio containing 5,000 policies was used to compare different models. Results showed that if missing covariates had relatively smooth inter-regional distributions, models with interactive regional latent variables showed stronger predictive power than ordinary models. Using the above model to analyze actual Norwegian auto theft insurance data found that under data sparsity and lack of spatial smoothing, latent variable models did not perform much better than classic GLM for making good predictions. Due to hierarchical Bayesian models' characteristics that fully reflect various uncertainty factors, they can theoretically improve performance, but when applied to real data, they still cannot provide effective data support for real auto theft situations.

Chung Y. [?] used nonparametric Bayesian methods for conditional distribution modeling, specifically examining variable selection and hypothesis testing under conditional distributions with multiple predictors. Few studies have addressed this area, especially the inability to simultaneously handle discrete and continuous multi-predictor situations. The paper proposes a general Bayesian nonparametric model applicable to most situations, aiming to construct flexible sparse models without pre-fixing mixture weights, establishing Probit Stick-Breaking Process (PSBP) to fully facilitate calculation of marginal likelihood functions and posterior probabilities. PSBP mixture models (PSBPM) also allow selection mechanisms in both regression coefficients and mixture weights, enabling selection of any number of parameters under PSBP to examine local and overall contributions of various predictors to conditional distributions. Besides applicability to continuous predictors, this method also naturally incorporates multiple categorical variables within the framework. PSBPM methods show high test power and low Type I error rates in most simulation experiments, though improvements are still needed for modeling and good interaction hypothesis testing under high-dimensional predictors with small sample sizes.

Insurance companies' classic actuarial risk factors classify insured populations by region, age, gender, etc., generally involving about fifty different covariate factors including vehicle condition, owner information, policy content, and geography. Delong et al. [?] research results show that age, category, and other factors affect claim frequency and amounts. Gao [?] used waiting time models to establish relationships between claim frequency and severity, finding vehicle value also significantly affects claim frequency through AIC solutions for suitable covariates.

With telematics technology development, insurers have more driving data including GPS, dashboard readings, and three-axis acceleration, and can record driving behavior and style, generating high-density time series data about GPS location, speed, and acceleration. Therefore, statistical analysis of telematics auto driving data has become a rapidly developing field in actuarial science. Since traditional auto insurance risk factors and causal relationships caused by actual situations fail to fully reflect, leading to unstable rates, market demand prompted UBI product emergence. This product's price is based on insured driving behavior habits, greatly improving insurance rate determination accuracy. Moreover, UBI can monitor driver behavior in real time, closely linking insurance premiums with driver habits, achieving finer customer segmentation and higher driving behavior supervision that breaks information asymmetry-induced adverse selection and moral risk between insurers and policyholders [?].

Due to the short development history of UBI insurance pricing models, many recent studies have attempted new systems. For example: Toledo T. et al. [?] introduced a new system called IVDR and conducted verification tests, showing in short-term tests that this system had certain testing value for driving behavior and reduced car collision numbers and risk indices. Based on the above research, Paefgen J. et al. [?] used more data materials (such as including data from 1,600 motor vehicles, PAYD measurement platforms with onboard data recorders, and actual accident risk multivariate exposure models). They used collected data (i.e., auto insurance data from 1,600 vehicles from European PAYD insurance service providers from 2009-2011, selecting 600 vehicles with traffic accidents and 1,000 vehicles without accidents), fusing these into total exposure matrices and adding different variable factors (such as time and road type) based on actual conditions to infer different driving exposure impacts on vehicles. They ultimately found that driving distance and accident risk are not linearly related. Guo [?], as one of the main contributors in such database research, used natural driving data from 100 vehicles (including 102 drivers with about 2 million kilometers of driving mileage and nearly 43,000 hours of driving time) as data sources, calculating an indicator for final consideration. He classified safety-related events into three categories: Collision, Near collision, and Critical incident event (CIE), evaluating individual driver risk levels based on three event types and predicting high-risk drivers. In more recent research, Gao Y. [?] identified major driving risk factors: average speed, daily trip count, nighttime driving proportion, sudden braking times, and intersection turning proportions. Research proves these factors are typically nonlinear [?], with complex interactions between features, providing new perspectives and data support for UBI insurance pricing models.

Huang Y. [?] worked on auto insurance classification ratemaking based on telematics driving data, proposing a new UBI product ratemaking framework and validating it from risk classification and claim frequency prediction perspectives. To address model interpretability issues, Huang used regression tree algorithms for continuous variable binning, building prediction models using five techniques: logistic regression, support vector machines, random forests, XGBoost, and artificial neural networks, using annual mileage logarithm as offset terms and Poisson regression models for model testing. AUC and RMSE were used to test risk classification models and claim frequency prediction models respectively. Research found that adding driving behavior variables to risk classification greatly improves framework performance, with better results after binning variable processing than raw variables. Binning methods effectively mitigate overfitting problems, and combining traditional variables with driving behavior variables can provide more accurate claim frequency prediction results in theory and practice.

Huang Y. [?] further focused on insurance loss prediction problems based on BNP regression frameworks, proposing a Gaussian mixture model using Dirichlet processes as priors, incorporating covariate effects into mixture component weights through probit stick-breaking methods (PSBM) to improve regression coefficient hierarchical structures. Slice sampling methods are used for parameter estimation. BNP regression shows better performance in goodness-of-fit and prediction accuracy compared to gamma regression, inverse Gaussian regression, GB2 regression, and Dirichlet Process ANOVA (DDP ANOVA) models. Compared to previous BNP regression studies, this model breaks through "single-p" model limitations, better characterizing loss distributions in insurance data, can incorporate prior knowledge to avoid biases in classic sampling methods, and achieves good results across different auto insurance dataset types.

Current UBI pricing research mostly approaches from risk classification and traditional actuarial dimensions. While having certain calculation accuracy in practical applications, driving behavior variable applications are relatively single, lacking focus on specific influencing factors, leading to poor model interpretability. Gao [?] proposed two claim frequency prediction methods based on telematics data—using speed heatmaps and single-trip scores to improve prediction effects. Scholars represented by Shengwang Meng [?] focused analysis on covariate extraction and descriptive statistical calculations, introducing confidence-based average risk scores, combining single-trip 1D CNN and GLM modeling methods to directly predict risk values for data preprocessing. After removing null values, calculating speed and acceleration changes, selecting stable factors, and time series characterization, experimental results show this method effectively improves out-of-sample prediction capability and reveals deeper differences in insurance customer groups, helping drive safe driving habits formation. Various driving methods mentioned above can better complete claim frequency prediction work theoretically and practically.

Additionally, weather risks cannot be ignored [?], with current safety field papers mostly focusing on visibility impacts on driving behavior [?].

(2) Post-Pricing Risk Response Mechanisms For post-pricing risk response mechanisms, whether regulatory and technical balance can be maintained is debatable. The BMS discount system, as an important means of auto insurance premium adjustment, mostly sets up tiered systems based on driver conditions, vehicle equipment, safety devices, or insurance behavior, lacking mechanisms for post-pricing premium adjustments based on dynamic data.

Although rigidly embedded structured data maps with effective explanatory rules exist, achieving true compatibility between explanatory power and flexibility remains difficult, especially in auto insurance markets with large-scale multi-dimensional interactions. For example, Japan's Insurance Business Law Implementation Rules impose rigid regulations on pure rate difference multiples caused by different risk factors used in insurance pricing: if age is used as a factor, pure rate differences between age groups cannot exceed 3 times; if gender is used, differences between genders can only be limited to 1.5 times; if regional attributes are input, pure rate differences between all regional factors cannot exceed 1.5 times. This means even if models are well-built and can detect high-risk and low-risk groups, results must be "artificially compressed" during output to comply with regulations. Artificially narrowing variable gaps affects real explanatory power of explanatory variables, weakening model predictive power and economic rationality. Some auto insurance pricing systems have adjustment mechanisms like "no claims but tier upgrade," and some stipulate that if insured objects do not appear in insurance contract terms for a certain year, they risk not enjoying corresponding protection. For example, very strict age regulations for drivers (e.g., only those 21 or 26 years and older can drive) are actually policy precautions for high-risk youth, but these rules are rigid prevention ideas without detailed risk group characterization based on risk types.

Many countries adopt multi-tier rate systems, dividing different policyholders into 1-20 levels based on claim history, age, gender, vehicle conditions, and insurance factors, grouping different vehicles into 1-9 applicable rate groups based on historical claim payment situations. An ordinary hatchback rising to tier 3 for one personal injury claim, with other accident types causing classifications: property damage rises to tier 3, driver/passenger injury rises to tier 4, and vehicle damage rises to tier 3, while high-end sports cars under the same accident circumstances may rise to tiers 6, 5, 4, or 9. Different vehicle models also have 1-9 applicable rate levels set according to historical claim risk. While this mechanism appears to strengthen risk identification and differentiation, it still relies on preset rules and personal experience judgments, failing to fully reflect data-based individualized risk characteristics. Consequently, although luxury models have advantages in active and passive safety equipment and theoretically should have lower personal injury claim rates, higher repair costs and stronger theft attraction objectively cause more severe claim losses after accidents, leading to more significant rate level adjustments for such high claim intensity risks, representing trade-offs insurers make based on claim frequency and severity.

However, note that new vehicles' first insurance basically have no historical claim records, and the initial stage is based on observable factors (e.g., vehicle displacement, price) without historical data support, unable to accurately reflect each individual's real risk.

The main reason for insurance company crisis events is insufficient reserve extraction. China's "Insurance Company Non-life Business Reserve Management Measures" stipulate that non-life business reserves consist of unearned premium reserves and outstanding claim reserves. In practice, some insurers also establish catastrophe risk reserves and indirect claim expense reserves. Outstanding claim reserves are subdivided into Reported But Not Settled (RBNS) and Incurred But Not Reported (IBNR).

Mainly used reserve assessment methods include: loss development methods (chain ladder, average claim payment method, reserve development method), loss ratio method, and combined methods (Bornhuetter-Ferguson method) [?]. For RBNS reserve assessment methods, regulations advocate case-by-case estimation and average assignment methods; for IBNR portions, chain ladder, average claim payment, reserve development, loss ratio methods, or other methods approved by regulators are advocated. Notably, the chain ladder method is most commonly used in practice due to its convenience and easy implementation. However, deterministic models represented by chain ladder can only produce point estimates of reserves, not providing estimated uncertainty ranges. Using stochastic models with probability distributions can estimate both expected values and confidence intervals, increasing scientific risk identification.

In specific modeling paths, stochastic chain ladder models are considered natural extensions of chain ladder methods. Kremer [?] first proposed lognormal distribution-based stochastic chain ladder models using claim amount logarithms as linear regression response variables, introducing accident year and development year dimensions to construct regression models. During this period, Pollard [?] proposed PPCI (average claim payment for incurred cases) and Reid [?] proposed PPCF (average claim payment for finalized cases) models. PPCI models the stage from occurrence to payment, while PPCF represents the process from occurrence to final payment completion.

Beyond parametric models, Mack's [?] nonparametric stochastic chain ladder model is currently one of the most commonly used models. This model only has mean and variance assumptions without specifying random term distribution forms, making it more general than previous models. However, some believe the Mack model fails to consider real factors in historical data, such as calendar year effects. Mario V. W. [?] proposed a more generalized nonparametric model including the Mack model as a special case, modifying the Mack model to better fit actual conditions.

In addition to frequentist methods, Bayesian methods and their extensions are introduced. In non-life actuarial science, Bühlmann [?] first attempted Bayesian methods by combining prior knowledge with data observations to obtain posterior parameter distributions for reserve inference. Commonly used Bayesian inference techniques today include conjugate distribution methods, MCMC sampling methods, and credibility theory approximation methods. Variational inference in recent years has improved Bayesian method operability and computational speed for large-scale data [?]. In Bayesian modeling, De Alba [?] constructed three-parameter lognormal distribution models to address potential negative values in claim data; Mario V. W. [?, ?] detailed Bayesian estimation accuracy research based on Mean Squared Error (MSE); Gisler and Mario V.

W. [?] studied chain ladder methods from a Bayesian framework, proving that under exponential family and conjugate distribution assumptions, credibility distributions are exact Bayesian estimates.

The pan-generalized nonlinear framework focuses on individual reserve calculation, conducting theoretical research on model selection, optimization algorithms, and post-pricing risks based on Ticconi [?] and Gao Guangyuan et al. [?] research on individual reserves. When GLM structural assumptions are overly restrictive, neural networks are one solution to help GLM address nonlinear relationships and higher data complexity [?]. In reality, most stochastic claim reserve methods ignore the huge impact of outliers, with some extreme observations potentially appearing in upper triangular areas, negatively affecting existing reserve models [?]. Robust GAM models use stratified sampling bootstrap methods [?], producing results similar to traditional models when outliers are absent, and showing significant advantages in estimation accuracy and efficiency when outliers exist through accident period effect and development period effect spline smoothing.

(3) Weather Index Insurance Related Research According to National Highway Traffic Safety Administration (NHTSA) data, about 22% of accidents annually are caused by weather factors, with over 6,000 deaths and more than 445,000 injuries [?], making weather an unavoidable topic in future auto insurance. In the pan-generalized nonlinear framework, weather index auto insurance mainly considers weather factors' biological or physical impacts on drivers, such as fatigue levels and driving visibility.

Driver fatigue driving has become the most prominent hidden danger in current traffic accidents, with fatigue driver accident incidence accounting for about 10%-15% of accident cases, and fatigue driving accident probability being 8 times that of sober driving [?]. In the long term, some anti-drowsiness methods (such as opening windows or radios) have no positive effects on driving, with the only countermeasure to drowsiness being sleep [?]. Once in a drowsy state for long periods, only rapid sleep supplementation can provide relief.

Increased environmental temperature can lead to shortened sleep duration, most noticeably shortened deep sleep periods [?], revealing human sensitivity to weather changes to some extent. Ahn S et al. [?] used multimodal EEG, ECG, eye movement, and near-infrared spectroscopy signals to more deeply explore sleep's impact on drivers, pointing out that sleep deprivation causes severe mental fatigue. When drivers are sleep-deprived, relative power levels of alpha waves in right central parietal regions significantly increase; relative power levels of beta waves in frontocentral regions decrease; average heart rate significantly decreases; driving condition levels and relative driving condition levels show significant increases in driver fatigue. These conclusions provide evidence that "weather affects drivers' mental states and physiological indicators, thereby increasing driving risk."

[TABLE:A2]
[TABLE:A3]

Sun Xianglong et al. [?] used international standard thermal comfort level standards for experimental grouping, using data from bus driver sleep quality scales, fatigue scales, and driving behavior scales under high-temperature weather to analyze and find that bus drivers' physical and mental states under different high temperatures affect safe driving conditions.

High-temperature weather mainly affects fatigue levels, while rainy weather mainly affects driving visibility. Some studies concluded that accidents increase by 100% or more during rainfall due to blocked vision [?], while other studies found milder but still statistically significant increases [?]. Druta et al. [?] used SHRP 2 NDS data to analyze driver adaptive behavior under adverse weather (rain, low visibility). By matching driving scenarios of the same drivers under normal and adverse weather conditions, they compared changes in speed, situational awareness, and visibility responses. Results found most accidents and critical situations occurred in rainy conditions, with critical situation numbers being twice accident numbers, showing drivers often avoided collisions through sudden braking and steering. Most critical situations occurred in rear-end collisions and side slips during merging or lane changes, mainly due to failure to adjust speed in time. Ahmed et al. [?] developed a method based on windshield wiper status variables to extract rain-related trips from the SHRP2 database, finding drivers were more likely to reduce speed by over 5 kph on rainy days, with probabilities of speed reduction 23% and 29% higher in light and heavy rain than in sunny conditions [?].

Current index insurance pricing models include compound Poisson process models as earlier methods, modeling catastrophe indices as compound Poisson processes with nonnegative jumps to capture disaster event suddenness and intensity, or using exponential Lévy processes to model loss dynamics at various stages.

Agricultural weather index insurance mainly considers precipitation factors. Conradt S [?] used cumulative precipitation indices to explore special insurance pricing methods. Biagini et al. [?] proposed whole-process models encompassing entire loss accumulation to subsequent assessment result calculations, using time-nonhomogeneous compound Poisson processes and Lévy processes respectively for modeling. Chen et al. [?] used deep learning methods like neural networks to reduce basis risk in traditional contracts, providing theoretical support for pan-generalized nonlinear frameworks. Ken Seng Tan et al. [?] introduced penalized spline methods to design more flexible compensation functions, giving this method advantages in depicting relationships between meteorological variables and agricultural yield losses. Applied to empirical data from Illinois corn producers with rainfall and evapotranspiration as meteorological index variables to formulate corresponding index insurance.

Weather insurance mostly serves as an additional contract to traditional agricultural insurance or as a supplementary product providing more protection for policyholders. However, research on whether optimal contracts exist remains in early stages. Zhang et al. [?] theoretically proved their existence and noted their structure depends on policyholders' utility functions, premium levels, and compensation limits. To partially overcome weather insurance basis risk, Zhang [?] created a hybrid insurance product combining traditional compensation insurance and index insurance, using multi-output neural networks to design trigger mechanisms and determine index compensation levels, organically combining these two insurance methods. Comparative results from its application to Iowa soybean insurance show that when using traditional compensation insurance or index insurance alone, policyholders' average utility is lower; only hybrid insurance introducing more production-investment factors in trigger links achieves policyholder utility maximization. Because this method retains model interpretability advantages while maintaining strong flexibility, it can be expanded to more property and casualty insurance applications, such as weather index auto insurance.

References

(1) Chinese Monographs

[1] Li Xiaolin, et al. Research on Japan's Compulsory Insurance System[M]. Beijing: China Financial & Economic Publishing House, 2017.
[2] Han Tianxiong, et al. Non-life Insurance Actuarial Science[M]. Beijing: China Financial & Economic Publishing House, 2010.

(2) Chinese Journal Articles

[3] Sun Xianglong, Guan Huanhuan. Impact of High Temperature Weather on Risky Driving Behavior of Bus Drivers[J]. Traffic Science and Technology and Economy, 2022, 24(6): 16-22.
[4] Wei Li, Yang Feiyan. Analysis of China's Commercial Auto Insurance Reform[J]. Insurance Research, 2018 (5): 16-32.

(3) Dissertations

[5] Fung T C S. A Class of Mixture of Experts Models for General Insurance Ratemaking and Reserving[D]. University of Toronto (Canada), 2020.
[6] Zhu W. Actuarial Ratemaking in Agricultural Insurance[D]. University of Waterloo, 2015.

(4) English Monographs

[7] Wood S N. Generalized Additive Models: An Introduction with R[M]. Chapman and Hall/CRC, 2017.

(5) English Literature

[8] Abreu M. and Fernandes F.T. The insurance industry in Brazil: a long-term view[J]. 2010.
[9] Almer B. Risk analysis in theory and practical statistics[C]//15th International Congress of Actuaries. 1957, 2: 314.
[10] Ahmed M M. and Abdel-Aty M A. The viability of using automatic vehicle identification data for real-time crash prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 13(2): 459-468.
[11] Ahmed M M, Ghasemzadeh A. The impacts of heavy rain on speed and headway Behaviors: An investigation using the SHRP2 naturalistic driving study data[J]. Transportation research part C: emerging technologies, 2018, 91: 371-384.
[12] Ahn S, Nguyen T, Jang H, et al. Exploring neuro-physiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data[J]. Frontiers in human neuroscience, 2016, 10: 219.
[13] Andreescu M P, Frost D B. Weather and traffic accidents in Montreal, Canada[J]. Climate research, 1998, 9(3): 225-230.
[14] Basagaña X., Escalera-Antezana J P., Dadvand P. et al. High ambient temperatures and risk of motor vehicle crashes in Catalonia, Spain (2000–2011): a time-series analysis[J]. Environmental health perspectives, 2015, 123(12): 1309-1316.
[15] Bennett M C. Models in motor insurance[J]. Journal of the Staple Inn Actuarial Society, 1978, 22: 134-160.
[16] Biagini F, Bregman Y, Meyer-Brandis T. Pricing of catastrophe insurance options written on a loss index with reestimation[J]. Insurance: Mathematics and Economics, 2008, 43(2): 214-222.
[17] Bolker B.M., Brooks M.E. and Clark C.J. et al. Generalized linear mixed models: a practical guide for ecology and evolution[J]. Trends in ecology & evolution, 2009, 24(3): 127-135.
[18] Bornhuetter R L. and Ferguson R E. The actuary and IBNR[C]//Proceedings of the casualty actuarial society, 1972: 181-195.
[19] Boucher J P. Bonus-malus scale models: Creating artificial past claims history[J]. Annals of Actuarial Science, 2023, 17(1): 36-62.
[20] Boucher J P., Côté S., Guillen M. Exposure as duration and distance in telematics motor insurance using generalized additive models[J]. Risks, 2017, 5(4): 54.
[21] Boucher J P, Coulibaly R. Bonus-Malus Scale premiums for Tweedie's compound Poisson models[J]. Annals of Actuarial Science, 2024, 18(2): 509-533.
[22] Boucher J P., Pérez-Marín A M. and Santolino M. Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident[C]//Anales del Instituto de Actuarios Españoles. Madrid: Instituto de Actuarios Españoles, 2013, 19(3): 135-154.
[23] Braaf S., Ameratunga S. and Nunn A. et al. Patient-identified information and communication needs in the context of major trauma[J]. BMC health services research, 2018, 18: 1-13.

[24] Breslow N.E. and Clayton D.G. Approximate inference in generalized linear mixed models[J]. Journal of the American statistical Association, 1993, 88(421): 9-25.
[25] Bressan S. Effects from ESG scores on P&C insurance companies[J]. Sustainability, 2023: 12644.
[26] Brockman M J. and Wright T S. Statistical motor rating: making effective use of your data[J]. Journal of the Institute of Actuaries, 1992, 119(3): 457-543.
[27] Brodsky H, Hakkert A S. Risk of a road accident in rainy weather[J]. Accident Analysis & Prevention, 1988, 20(3): 161-176.
[28] Bühlmann H. Experience rating and credibility[J]. ASTIN Bulletin: The Journal of the IAA, 1967, 4(3): 199-207.
[29] Castles A C. Compulsory automobile liability insurance in Australasia[J]. The American Journal of Comparative Law, 1957: 257-283.
[30] Chang L, Gao G, Shi Y. Claims reserving with a robust generalized additive model[J]. North American Actuarial Journal, 2024, 28(4): 840-860.
[31] Chantarat S, Mude A G, Barrett C B, et al. Designing index-based livestock insurance for managing asset risk in northern Kenya[J]. Journal of Risk and Insurance, 2013, 80(1): 205-237.
[32] Chatukuta M., Groce N. and Mindell J.S. et al. Road traffic injuries in Namibia: health services, public health and the motor vehicle accident fund[J]. International journal of injury control and safety promotion, 2021: 167-178.
[33] Chen A. and Vigna E. A unisex stochastic mortality model to comply with EU Gender Directive[J]. Insurance: Mathematics and Economics, 2017, 73: 124-136.
[34] Chen T. and Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.
[35] Chen Z, Lu Y, Zhang J, et al. Managing weather risk with a neural network-based index insurance[J]. Management Science, 2024, 70(7): 4306-4327.
[36] Chung Y. and Dunson D B. Nonparametric Bayes conditional distribution modeling with variable selection[J]. Journal of the American Statistical Association, 2009, 104(488): 1646-1660.
[37] Clarke D J. A theory of rational demand for index insurance[J]. American Economic Journal: Microeconomics, 2016, 8(1): 283-306.
[38] Conradt S, Finger R, Spörri M. Flexible weather index-based insurance design[J]. Climate Risk Management, 2015, 10: 106-117.
[39] Coutts S M. Motor insurance rating, an actuarial approach[J]. Journal of the Institute of Actuaries, 1984, 111(1): 87-148.
[40] de Alba E. Claims reserving when there are negative values in the runoff triangle: Bayesian analysis using the three-parameter log-normal distribution[J]. North American Actuarial Journal, 2006, 10(3): 45-59.
[41] De Bastiani F., Rigby R A. and Stasinopoulous D M. et al. Gaussian Markov random field spatial models in GAMLSS[J]. Journal of Applied Statistics, 2018, 45(1):
[42] Delong Ł., Lindholm M. and Wüthrich M V. Making Tweedie's compound Poisson model more accessible[J]. European Actuarial Journal, 2021, 11(1):

185-226.
[43] Desyllas P. and Sako M. Profiting from business model innovation: Evidence from Pay-As-You-Drive auto insurance[J]. Research Policy, 2013, 42(1): 101-116.
[44] Druta C, Kassing A, Gibbons R et al. Assessing driver behavior using shrp2 adverse weather data[J]. Journal of safety research, 2020, 73: 283-295.
[45] Filipova-Neumann L. and Welzel P. Reducing asymmetric information in insurance markets: Cars with black boxes[J]. Telematics and Informatics, 2010, 27(4):
[46] Gabrielli A., Richman R. and Wüthrich M V. Neural network embedding of the over-dispersed Poisson reserving model[J]. Scandinavian Actuarial Journal, 2020, 2020(1): 1-29.
[47] Gatzert N. and Reichel P. Awareness of climate risks and opportunities: empirical evidence on determinants and value from the US and European insurance industry[J]. The Geneva Papers on Risk and Insurance-Issues and Practice, 2022: 1-22.
[48] Gisler A. and Wüthrich M V. Credibility for the chain ladder reserving method[J]. ASTIN Bulletin: The Journal of the IAA, 2008, 38(2): 565-600.
[49] Gao G. Fitting Tweedie's compound Poisson model to pure premium with the EM algorithm[J]. Insurance: Mathematics and Economics, 2024, 114: 29-42.
[50] Gao G. and Li J. Dependence modeling of frequency-severity of insurance claims using waiting time[J]. Insurance: Mathematics and Economics, 2023, 109: 29-
[51] Gao Y., Huang Y. and Meng S. Evaluation and interpretation of driving risks: Automobile claim frequency modeling with telematics data[J]. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2023, 16(2): 97-119.
[52] Ghasemzadeh A, Hammit B E, Ahmed M M, et al. Complementary methodologies to identify weather conditions in naturalistic driving study trips: Lessons learned from the SHRP2 naturalistic driving study & roadway information database[J]. Safety Science, 2019, 119: 21-28.
[53] Ghazali P L, Ghani P P S A, Mamat M, et al. Integration model in auto takaful insurance[J]. Far East J. Math. Sci, 2015, 98(5): 599-611.
[54] Guangyuan G., Shengwang M. and Wüthrich M V. What can we learn from telematics car driving data: A survey[J]. Insurance: Mathematics and Economics, 2022, 104: 185-199.
[55] Guo F. and Fang Y. Individual driver risk assessment using naturalistic driving data[J]. Accident Analysis & Prevention, 2013, 61: 3-9.
[56] Hastie T. J. Generalized additive models[J]. Statistical models in S, 2017: 249
[57] Hamdar S H, Qin L, Talebpour A. Weather and road geometry impact on longitudinal driving behavior: Exploratory analysis using an empirically supported acceleration modeling framework[J]. Transportation research part C: emerging technologies, 2016, 67: 193-213.
[58] Hoerl A E. and Kennard R W. Ridge regression: Biased estimation for

nonorthogonal problems[J]. Technometrics, 1970, 12(1): 55-67.

[59] Huang Y. and Meng S. Automobile insurance classification ratemaking based on telematics driving data[J]. Decision Support Systems, 2019, 127: 113156.

[60] Huang Y. and Meng S. A Bayesian nonparametric model and its application in insurance loss prediction[J]. Insurance: Mathematics and Economics, 2020, 93: 84-

[61] Husnjak S., Peraković D and Forenbacher I. et al. Telematics system in usage based motor insurance[J]. Procedia Engineering, 2015, 100: 816-825.

[62] Jiang X., Zhuang D. and Zhang X. et al. Uncertainty quantification via spatial-temporal tweedie model for zero-inflated and long-tail travel demand prediction[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023: 3983-3987.

[63] Johnson P D. and Hey G B. Statistical studies in motor insurance[J]. Journal of the Institute of Actuaries, 1971, 97(2/3): 199-249.

[64] Jørgensen B. and Paes De Souza M C. Fitting Tweedie's compound Poisson model to insurance claims data[J]. Scandinavian Actuarial Journal, 1994, 1994(1): 69-

[65] Jung J. On automobile insurance ratemaking[J]. ASTIN Bulletin: The Journal of the IAA, 1968, 5(1): 41-48.

[66] Ke G., Meng Q. and Finley T. et al. Lightgbm: A highly efficient gradient boosting decision tree[J]. Advances in neural information processing systems, 2017, 30.

[67] Kremer E. IBNR-claims and the two-way model of ANOVA[J]. Scandinavian Actuarial Journal, 1982, 1982(1): 47-55.

[68] Lee Y K., Mammen E. and Park B U. Flexible generalized varying coefficient regression models[J]. 2012.

[69] Lemaire J., Park S C. and Wang K C. The use of annual mileage as a rating variable[J]. ASTIN Bulletin: The Journal of the IAA, 2016, 46(1): 39-69.

[70] Li A., Luo H., Zhu Y. et al. Climate warming may undermine sleep duration and quality in repeated-measure study of 23 million records[J]. Nature Communications, 2025, 16(1): 2609.

[71] Lindholm M., Richman R. and Tsanakas A. et al. Discrimination-free insurance pricing[J]. ASTIN Bulletin: The Journal of the IAA, 2022, 52(1): 55-89.

[72] Lindholm M., Richman R. and Tsanakas A. et al. What is fair? Proxy discrimination vs. demographic disparities in insurance pricing[J]. Scandinavian Actuarial Journal, 2024, 2024(9): 935-970.

[73] Litman T. Pay-As-You-Drive Pricing and Insurance Regulatory Objectives[J]. Journal of insurance regulation, 2005, 23(3).

[74] Mack T. Distribution-free calculation of the standard error of chain ladder reserve estimates[J]. ASTIN Bulletin: The Journal of the IAA, 1993, 23(2): 213-225.

[75] Makowiec-Dąbrowska T., Gadzicka E., Siedlecka J. et al. Climate conditions and work-related fatigue among professional drivers[J]. International journal of biometeorology, 2019, 63: 121-128.

[76] Maze T H, Agarwal M, Burchett G. Whether weather matters to traffic demand, traffic safety, and traffic operations and flow[J]. Transportation research record, 2006, 1948(1): 170-176.

[77] Mohamed M G, Saunier N, Miranda-Moreno L F, et al. A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada[J]. Safety science, 2013, 54: 27-37.

[78] Müller S, Welsh A H. Outlier robust model selection in linear regression[J]. Journal of the American Statistical Association, 2005, 100(472): 1297-1310.

[79] Naik B, Tung L W, Zhao S, et al. Weather impacts on single-vehicle truck crash injury severity[J]. Journal of safety research, 2016, 58: 57-65.

[80] Nelder J. A. and Wedderburn R. W. M. Generalized Linear Models[J]. Journal of the Royal Statistical Society, 1972:370-384.

[81] N van Huyssteen N. and Rudansky-Kloppers S. Factors influencing consumers' purchase decisions regarding personal motor vehicle insurance in South Africa[J]. Cogent Business & Management, 2024.

[82] Noelke C., McGovern M., Corsi D J. et al. Increasing ambient temperature reduces emotional well-being[J]. Environmental research, 2016, 151: 124-129.

[83] Paefgen J., Staake T. and Fleisch E. Multivariate exposure modeling of accident risk: Insights from Pay-as-you-drive insurance data[J]. Transportation Research Part A: Policy and Practice, 2014, 61: 27-40.

[84] Park S.C. and Han S. Externalities from driving luxury cars[J]. Risk Management and Insurance Review, 2017: 391-427.

[85] Pollard J H. Outstanding claims provisions: a distribution-free statistical approach[J]. Journal of the Institute of Actuaries, 1982, 109(3): 417-433.

[86] Ramires T G., Nakamura L R. and Righetto A J. et al. Incorporating clustering techniques into GAMLSS[J]. Stats, 2021, 4(4): 916-930.

[87] Reid D H. Claim reserves in general insurance[J]. Journal of the Institute of Actuaries, 1978, 105(3): 211-315.

[88] Ronald R. and Mario V. W. LocalGLMnet: interpretable deep learning for tabular data[J], Scandinavian Actuarial Journal, 2023a.

[89] Richaudeau D. Automobile insurance contracts and risk of accident: An empirical test using French individual data[J]. The Geneva Papers on Risk and Insurance Theory, 1999, 24: 97-114.

[90] Richman R. and Wüthrich M V. LASSO regularization within the LocalGLMnet architecture[J]. Advances in Data Analysis and Classification, 2023b, 17(4): 951-981.

[91] Shengwang M. An Application of Generalized Linear Model to Auto Insurance Pricing [J], Application of Statistics and Management, 2007:24.

[92] Shengwang M., He W., Yanlin S. and Guangyuan G. Improving Automobile Insurance Claims Frequency Prediction with Telematics Car Driving Data[J]. ASTIN Bulletin:The Journal of the IAA, Dec.27th.2021:363 - 391.

[93] Shi P., Zhang W. and Valdez E A. Testing adverse selection with two-dimensional information: evidence from the Singapore auto insurance market[J]. Journal of Risk and Insurance, 2012, 79(4): 1077-1114.

[94] Schiffman L G., Kanuk L. and Hansen H. Consumer behaviour: a European

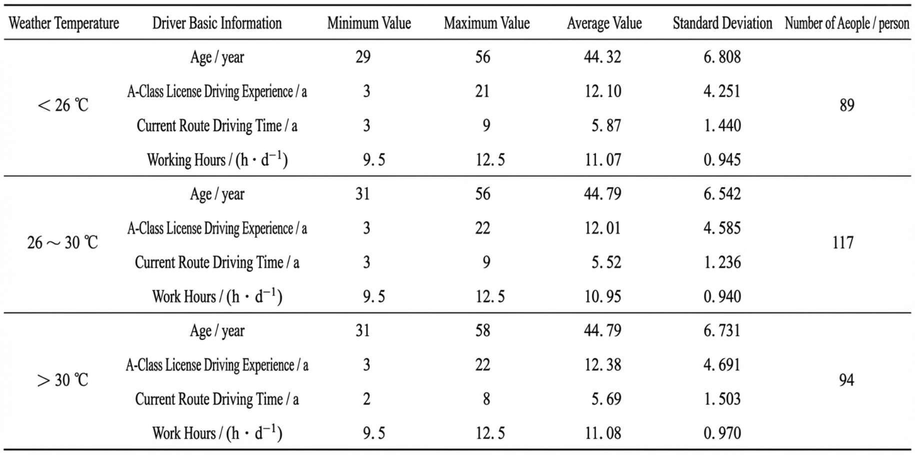
outlook.[J]. Pearson Education, 2012.
[95] Smyth G K. and Jørgensen B. Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling[J]. ASTIN Bulletin: The Journal of the IAA, 2002, 32(1): 143-157.
[96] Stricker L., Pugnetti C. and Wagner J. et al. Green insurance: a roadmap for executive management[J]. Journal of Risk and Financial Management, 2022, 15(5):
[97] Tan K S, Zhang J. Flexible weather index insurance design with penalized splines[J]. North American Actuarial Journal, 2024, 28(1): 1-26.
[98] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 1996, 58(1): 267-288.
[99] Ticconi D. Individual claims reserving in Credit insurance using GLM and Machine Learning[J]. Dipartimento di Scienze Statistiche, Sapienza Università di Roma: Rome, Italy, 2018.
[100] Theofilatos A, Yannis G. A review of the effect of traffic and weather characteristics on road safety[J]. Accident Analysis & Prevention, 2014, 72: 244-256.
[101] Toledo T., Musicant O. and Lotan T. In-vehicle data recorders for monitoring and feedback on drivers' behavior[J]. Transportation Research Part C: Emerging Technologies, 2008, 16(3): 320-331.
[102] Turcotte R. and Boucher J P. Gamlss for longitudinal multivariate claim count models[J]. North American Actuarial Journal, 2024, 28(2): 337-360.
[103] Verbelen R., Antonio K. and Claeskens G. Unravelling the predictive power of telematics data in car insurance pricing[J]. Journal of the Royal Statistical Society Series C: Applied Statistics, 2018, 67(5): 1275-1304.
[104] Verrall R J. A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving[J]. North American Actuarial Journal, 2004, 8(3):
[105] Wood S N. A simple test for random effects in regression models[J]. Biometrika, 2013, 100(4): 1005-1010.
[106] Wörle J, Metz B, Steinborn M B, et al. Differential effects of driver sleepiness and sleep inertia on driving behavior[J]. Transportation research part F: traffic psychology and behaviour, 2021, 82: 111-120.
[107] Wüthrich M V. Prediction error in the chain ladder method[J]. Insurance: Mathematics and Economics, 2008, 42(1): 378-388.
[108] Wüthrich M V. Challenges with non-informative gamma priors in the Bayesian over-dispersed Poisson reserving model[J]. Insurance: Mathematics and Economics, 2013, 52(2): 352-358.
[109] Wüthrich M V. Neural networks applied to chain–ladder reserving[J]. European Actuarial Journal, 2018, 8: 407-436.
[110] Zhang J, Tan K S, Weng C. Index insurance design[J]. ASTIN Bulletin: The Journal of the IAA, 2019, 49(2): 491-523.
[111] Zhang J. Blended insurance scheme: A synergistic conventional-index insurance mixture[J]. Insurance: Mathematics and Economics, 2024, 119: 93-105.

[112] Zhai X, Huang H, Sze N N, et al. Diagnostic analysis of the effects of weather condition on pedestrian crash severity[J]. Accident Analysis & Prevention, 2019, 122: 318-324.

[113] Zhou H., Qian W. and Yang Y. Tweedie gradient boosting for extremely unbalanced zero-inflated data[J]. Communications in Statistics-Simulation and Computation, 2022, 51(9): 5507-5523.

Figures

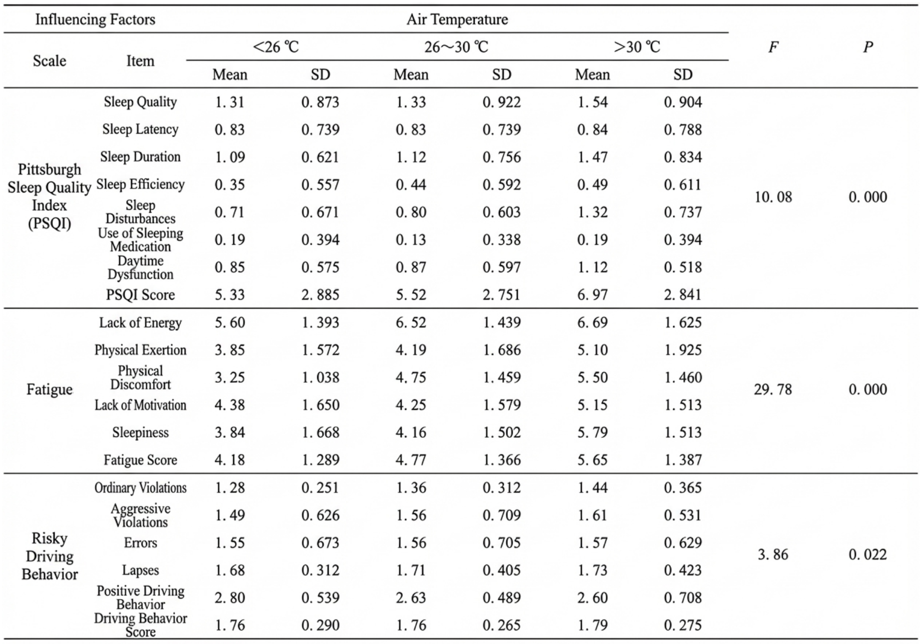
Figure 1: Figure 1



Weather Temperature	Driver Basic Information	Minimum Value	Maximum Value	Average Value	Standard Deviation	Number of Aeople / person
< 26 ℃	Age / year	29	56	44. 32	6. 808	89
	A-Class License Driving Experience / a	3	21	12. 10	4. 251	
	Current Route Driving Time / a	3	9	5. 87	1. 440	
	Working Hours / (h · d^{-1})	9. 5	12. 5	11. 07	0. 945	
26 ～ 30 ℃	Age / year	31	56	44. 79	6. 542	117
	A-Class License Driving Experience / a	3	22	12. 01	4. 585	
	Current Route Driving Time / a	3	9	5. 52	1. 236	
	Work Hours / (h · d^{-1})	9. 5	12. 5	10. 95	0. 940	
> 30 ℃	Age / year	31	58	44. 79	6. 731	94
	A-Class License Driving Experience / a	3	22	12. 38	4. 691	
	Current Route Driving Time / a	2	8	5. 69	1. 503	
	Work Hours / (h · d^{-1})	9. 5	12. 5	11. 08	0. 970	

Figure 2: Figure 3

Source: ChinaXiv — Machine translation. Verify with original.



Influencing Factors		Air Temperature							
Scale	Item	<26 ℃		26～30 ℃		>30 ℃		*F*	*P*
		Mean	SD	Mean	SD	Mean	SD		
Pittsburgh Sleep Quality Index (PSQI)	Sleep Quality	1.31	0.873	1.33	0.922	1.54	0.904	10.08	0.000
	Sleep Latency	0.83	0.739	0.83	0.739	0.84	0.788		
	Sleep Duration	1.09	0.621	1.12	0.756	1.47	0.834		
	Sleep Efficiency	0.35	0.557	0.44	0.592	0.49	0.611		
	Sleep Disturbances	0.71	0.671	0.80	0.603	1.32	0.737		
	Use of Sleeping Medication	0.19	0.394	0.13	0.338	0.19	0.394		
	Daytime Dysfunction	0.85	0.575	0.87	0.597	1.12	0.518		
	PSQI Score	5.33	2.885	5.52	2.751	6.97	2.841		
Fatigue	Lack of Energy	5.60	1.393	6.52	1.439	6.69	1.625	29.78	0.000
	Physical Exertion	3.85	1.572	4.19	1.686	5.10	1.925		
	Physical Discomfort	3.25	1.038	4.75	1.459	5.50	1.460		
	Lack of Motivation	4.38	1.650	4.25	1.579	5.15	1.513		
	Sleepiness	3.84	1.668	4.16	1.502	5.79	1.513		
	Fatigue Score	4.18	1.289	4.77	1.366	5.65	1.387		
Risky Driving Behavior	Ordinary Violations	1.28	0.251	1.36	0.312	1.44	0.365	3.86	0.022
	Aggressive Violations	1.49	0.626	1.56	0.709	1.61	0.531		
	Errors	1.55	0.673	1.56	0.705	1.57	0.629		
	Lapses	1.68	0.312	1.71	0.405	1.73	0.423		
	Positive Driving Behavior	2.80	0.539	2.63	0.489	2.60	0.708		
	Driving Behavior Score	1.76	0.290	1.76	0.265	1.79	0.275		

Figure 3: Figure 4