Actuarial Pricing of UAV Insurance for Thin Data Scenarios

Wang Yang[1], Li Dayu[2*], Wang Dinglin[3], Ren Feixiao[4]
[1]School of Insurance, Shandong University of Finance and Economics, Jinan, 250014,[2]School of Insurance, Shandong University of Finance and Economics, Jinan, 250014,[3]School of Insurance, Shandong University of Finance and Economics, Jinan, 250014,[4]School of Insurance, Shandong University of Finance and Economics, Jinan, 250014

*Corresponding to:Li Dayu, School of Insurance, Shandong University of Finance and Economics, Jinan, 250014,solidli@outlook.com

Abstract
Driven by both market demand and policies, the drone insurance industry is facing new development opportunities. This study focuses on exploring an innovative hybrid data integration method, which uses public datasets of drones and small manned aircraft for hybrid data integration and severity scaling, and conducts simulation tests to ensure the reproducibility of the method. A two-part hybrid model approach is adopted to separate the frequency model from the severity model, and a hierarchical modeling method is used for each part to deal with the occurrence of extreme losses. Monte Carlo simulation is performed on the fused data to calculate the net premium. Innovatively, a no-claim discount system is introduced, and the impact of operators' behaviors on claim frequency is quantified, with comprehensive consideration given to the inclusion and quantification of risk factors. The application of Tweedie GLM in total loss modeling is constructed and analyzed, and the advantages and disadvantages of different modeling methods are compared, aiming to provide more comprehensive decision-making basis for insurance companies. This report is intended to construct and evaluate a robust actuarial rate-making model for the rapidly developing drone insurance market, and to develop more accurate, fair and market-competitive drone insurance products.
Keywords:Mixed Data, Rate-Making Model, Drone Insurance

Competing Interests:
The authors declare that there is no conflict of interest.

1. Introduction

The determination of insurance rates is a crucial aspect of insurance product design, and its accuracy directly affects the market share of the product and the profitability of an insurance company. Currently, the Chinese drone market is experiencing rapid growth. With the growth in drone numbers and flight activities, related accidents and risk events have also increased, giving rise to a pressing demand for specialized drone insurance products.

However, compared to the traditional aviation industry, the drone insurance market is still in its early stages of development and faces many unique challenges. The most fundamental problem lies in the lack of sufficient and high-quality historical claim data. The existing data is often fragmented, non-standardized, and mostly consists of event reports rather than quantitative loss data, which poses significant obstacles to the risk assessment and rate determination process.

Traditional research relies on a single source of historical claim data for rate determination. For emerging businesses or those with small data volumes, there is a tendency for unstable models and difficulties in accurately differentiating risks. To address this, this study innovatively integrates drone incident reports (ASRS) with small manned aircraft accident data from the National Transportation Safety Board (NTSB). Through complex loss severity scaling, the loss amounts from manned aircraft accidents are converted into approximate losses in drone scenarios, effectively expanding the data basis for severity modeling. The KS test yields an empirical p-value of 0.074, indicating that the scaling process effectively preserves the relative characteristics of the loss distribution.

At present, traditional methods usually rely on pre-defined and statically defined risk factors. The consideration of operators' soft risk characteristics is relatively limited, and only a single parameter distribution is used to fit the entire loss distribution, which easily leads to insufficient estimation of potential risks associated with extremely large claims by the model. Moreover, in the traditional two-part model framework, pure premiums are usually calculated as a simple product of the frequency mean and the severity mean. Although this approach is simple, it may fail to fully capture the complexity of the total claim distribution. Furthermore, in the rate determination process, the NCD system is typically treated as an independent adjustment step after rate determination, using a fixed matrix to discount premiums. Its coefficients are usually based on historical experience or industry practices, and rarely directly integrated into the core prediction model for quantitative derivation.

This study aims to address the above issues by fully incorporating and quantifying risk factors. It not only considers explicit risk factors such as flight mission types but also quantifies the impact of operator behavior on claim frequency through data inference and experience calibration. An advanced two-part modeling method is adopted, separating the frequency model from the severity model. The severity model is further subdivided into main body modeling and tail modeling: the main body is modeled using a Bayesian generalized linear model (with a Gamma distribution), while the tail is modeled using the generalized Pareto distribution (GPD) from extreme value theory to capture the characteristics of large claims. This hierarchical modeling approach can more accurately depict the complex morphology of the claim distribution and effectively refine the technical route for extreme loss modeling. In addition to the two-part model, this study also constructs a Tweedie GLM to directly model total losses and compares the advantages and disadvantages of the two modeling methods. This comparative analysis aims to provide a more comprehensive decision-making basis for insurance companies, helping them select the most suitable rate determination strategy under different data conditions and business requirements. Furthermore, it creatively introduces the NCD system into the model, quantifying its discount effect on pure premiums, thereby enhancing the fairness and market competitiveness of rates.

Compared with previous studies, this paper’s contributions mainly lie in the following five aspects: First, through a complex severity scaling method, the loss amounts of manned aircraft accidents are converted into approximate losses in drone scenarios and incorporated into a unified rate determination framework, effectively expanding the data basis for severity modeling. Second, a two-part hybrid modeling method is used to separate the frequency model from the severity model, with a hierarchical modeling approach adopted for each part to address the occurrence of extreme losses. Third, Monte Carlo simulation is used to generate a large number of samples to create “premium year”-level data, optimizing the rate determination model and providing a new method for improving the accuracy of rate determination. Fourth, the NCD system is creatively introduced, and the impact of operator behavior on claim frequency is quantified, with comprehensive consideration given to the inclusion and quantification of risk factors. Fifth, the application of Tweedie GLM in total loss modeling is constructed and analyzed, comparing the advantages and disadvantages of the two modeling methods to provide a more comprehensive decision-making basis for insurance companies.

The second part is a literature review, which systematically analyzes the breakthroughs of Monte Carlo simulation and the two-part hierarchical modeling method in traditional rate determination methods, as well as the mechanistic roles of the GPD (Generalized Pareto Distribution) and Tweedie GLM model in property insurance rate determination. The third part conducts data processing and predictive fitting modeling, defines variables, and performs descriptive statistics on the data. The fourth part presents and analyzes the empirical results. Finally, the study summarizes the findings and proposes optimization suggestions.

2. Literature Review

2.1 The breakthrough of monte carlo simulation in property insurance rate making

Monte Carlo simulation, as a powerful computational method, has brought significant breakthroughs in the field of property insurance rate making. It estimates the behavior of complex systems by simulating a large number of random events, and is particularly suitable for dealing with uncertainties and complexities that are difficult to handle with traditional actuarial methods. In foreign research, Hans Bühlmann (1970) first introduced the idea of random simulation into the actuarial field in "Experience Rating and Credibility", laying the foundation for the application of the Monte Carlo method. Subsequently, Klugman et al. (2008) systematically expounded the advantages of Monte Carlo simulation in property insurance aggregate risk modeling in "Loss Models: From Data to Decisions", solving the problem of multivariate risk dependence that is difficult to handle with traditional analytical methods through a large number of random samplings, and significantly improving the accuracy of catastrophe insurance rate making. Frees E. W. (2010) in "Regression Modeling with Actuarial and Financial Applications" used this method to handle the interaction effects of multi-dimensional risk factors in auto insurance, and empirical results showed that the rate prediction error was reduced by 18%. Although domestic research started late, it has developed rapidly. Meng Shengwang (2013) in "Actuarial Science" explored the application of Monte Carlo simulation in property insurance reserve assessment, providing a risk quantification tool for rate making. Xie et al. (2016) published "Research on Catastrophe Insurance Rate Making Based on Monte Carlo Simulation" in "Statistical and Information Forum", taking earthquake risk as an example, and through simulating the loss distribution caused by disasters of different intensities, they broke through the excessive dependence of traditional extreme value models on historical data, making the rate more in line with potential risks . Zhang et al. (2020) in "Insuran ce Studies"combined this method with machine learning to solve the complex coupling problem of equipment failure and human factors in engineering insurance, improving the rate making efficiency by 30% . Monte Carlo simulation has broken through the bottleneck of traditional analytical methods in complex risk modeling, significantly broadened the data basis for actuarial rate making, and reduced the excessive dependence on historical data in the traditional rate making process, promoting the advancement of property insurance actuarial science towards a more refined risk pricing era.

2.2 The breakthrough of two-part hierarchical modeling in property insurance rate making

In the research published by McDonald and Xie (2006) in " Journal of Applied Econometrics", a two-part model was explicitly used to analyze US auto insurance data: the frequency part used a Poisson-lognormal mixture model, the severity main part adopted the Bayesian estimation of the Gamma distribution, and the tail was fitted by the GPD model. Empirical results showed that this method reduced the prediction error of extreme claims by

15% to 20%. In domestic research, Wang Dehui and Zhang Ruigang (2011) in "Insurance Studies" published "Joint Modeling of Auto Insurance Claim Frequency and Severity", which was the first to combine the Bayesian generalized linear model (Gamma distribution) with EVT to model Chinese auto insurance data, finding that tail claims (about 5% of the total) contributed nearly 30% of the total losses, verifying the necessity of hierarchical modeling. Bae et al. (2020) in the "Risk Management and Insurance Review" pointed out that the frequency of drone accidents is low but the loss variance is large. Their research adopted a two-part model: the frequency part used a zero-inflated negative binomial model to handle data sparsity, and the severity part used a Gamma distribution for the main body and a GPD to capture extreme losses such as crashes, providing a quantitative tool for pricing third-party liability insurance for drones. The above-mentioned literature all verified the method's ability to describe complex claim distributions, especially its irreplaceable value in quantifying extreme risks.

2.3 The mechanism role of the generalized pareto distribution (GPD) in the rate-making of property insurance

As a core tool of extreme value theory, the Generalized Pareto Distribution (GPD) provides a scientific theoretical framework for property insurance rate-making due to its precise characterization of the tail distribution of extreme events. Traditional rate-making methods (such as experience rating and generalized linear models) perform well in handling regular losses but have limited fitting ability for extreme losses. The GPD model, by focusing on loss data exceeding a threshold, can more accurately capture the thick-tailed characteristics of extreme risks. Richard A. Davis A. R. and Thomas Mikosch T. (1997) in "Extreme value theory for space-time processes with heavy-tailed distributions"pointed out that heavy-tailed distributions such as the Pareto distribution have been proven to be very effective in simulating sudden phenomena in many fields such as finance, insurance, telecommunications, meteorology, and hydrology. When heavy-tailed features exist, regular variation theory provides a unified and convenient theoretical framework for studying multivariate extremes.

The GPD model also demonstrates unique value in conventional non-life insurance businesses such as auto insurance. Traditional auto insurance rate-making often relies on generalized linear models (GLM), but the fitting deviation of GLM for extreme losses may lead to underestimation of rates. Wang Zhi (2023) in "Research on Auto Insurance Rate-making Based on Double Hierarchical Generalized Linear Model" proposed combining the GPD model with GLM to construct a hierarchical rate-making framework: GLM is used to fit regular losses, and GPD is used to fit extreme losses. Empirical results show that this model improves the coverage ability of auto insurance rates for major accidents by 25% while maintaining the pricing accuracy for small losses. Xu Lei (2012) constructed a GPD model and applied it to the rate-making of drought insurance in 13 major grain-producing provinces in China. The deviation between the results and the actual rates was less than 5%, significantly outperforming traditional static models. This directly reflects the practical significance of the time-varying GPD model in improving risk assessment accuracy and optimizing rate structures in the property insurance field.

2.4 The mechanism role of tweedie GLM model in non-life insurance premium rate determination

The mechanistic role of the Tweedie GLM model in determining property insurance premiums In the field of property insurance rate determination, traditional models often struggle with mismatches between assumptions about data distribution and the actual characteristics of claims, particularly in non-life insurance lines such as auto insurance, where claim data typically exhibits a mix of accumulated zero values and continuous positive values. As a special form of exponential family distribution, the Tweedie distribution unifies the statistical properties of the Poisson distribution, gamma distribution, and compound Poisson-gamma distribution through a power mean-variance relationship. In their study on auto insurance rate determination, Sun Weiwei (2014) noted that the zero-inflation characteristic and right-skewed, long-tailed features of the Tweedie distribution allow it to model mixed data consisting of both zero and positive claims simultaneously, effectively addressing the complexity of traditional models that require step-by-step estimation of claim frequency and severity. Zhang Lianzeng and Xie Houyi (2017) found that the direct modeling approach based on the Tweedie distribution generally requires estimating fewer parameters than the compound Poisson-gamma two-stage model. Within the modeling framework, Tweedie GLM links the exponential form of the claim mean to a linear combination of explanatory variables via a link function, and its likelihood function includes the joint optimization of probability densities for zero and positive values. Huang et al. (2010) empirically indicated that the power parameter p of the Tweedie distribution can be optimized simultaneously using the maximum likelihood method, and this model outperforms the zero-adjusted inverse Gaussian model in claim prediction. An empirical analysis by Zhang Lianzeng and Xie Houyi (2017) based on a domestic auto insurance dataset shows that negative binomial regression can only handle discrete overdispersed data and has poor fitting ability for continuous positive values, whereas Tweedie GLM performs better than negative binomial regression in such scenarios. The above studies support the flexibility and applicability of the Tweedie distribution in determining non-life insurance rates.

3.Data Processing and Predictive Fitting Modeling

3.1 Modeling approach and theoretical basis

In the process of determining UAV insurance rates, considering the data sparsity and the thick tail characteristics of accident compensation, this study adopts a segmented modeling strategy to decompose the pure premium into the product of the frequency term and the severity term, and further decomposes the severity modeling into the main part (main compensation) and the tail part (extreme compensation).

Based on the principle of classical insurance pricing, the annualized claim is modeled as the product of the expectation of a single claim and the frequency of accidents by using the method of frequency-severity decomposition:

$$\pi = \mathbb{E}[N] \cdot \mathbb{E}[S \mid S > 0], \qquad (1)$$

3.2 Data fitting and severity scaling modeling

3.2.1 Loss severity data scaling

In order to fully capture the risk of mission heterogeneity, the model introduces multiple pricing factors such as mission category, operator behavior level, NCD system, etc., and uses Bayesian Gamma regression to process the subject compensation segment in the severity modeling, uses GPD or exponential distribution to model the extreme compensation segment, and calculates the compensation distribution through Monte Carlo simulation, so as to determine the expected annualized compensation under the mission type.

$$L_{\mathrm{UAV}} = L_{\mathrm{Manned}} \cdot \frac{V_{\mathrm{UAV}}}{\mathbb{E}[L_{\mathrm{Manned}}]} \cdot \gamma_m, \qquad (2)$$

Based on the loss scaling theory, the loss of different levels and physical parameters is mapped to a unified UAV risk space. The core is: fuselage insurance is scaled in equal proportion to value; Liability insurance is scaled based on the personal injury valuation structure (death, serious injury, minor injury).

3.2.2 Risk conversion matrix

Due to the scarcity of historical compensation data in the field of drones, severity modeling needs to rely on the historical compensation records of manned and aircraft. However, there are systematic differences between the two in terms of use, value, and carryability, and direct application will lead to a systematic deviation from pure premiums. To this end, a risk conversion matrix is introduced to scale the compensation of manned aircraft to the UAV risk framework.

3.2.3 Modeling of operator behavior factors

There is a natural correlation between flight missions and operator experience levels. To quantify this impact, we have established the following risk factor modelling mechanism

Firstly, the operators are structurally classified according to their experience and behavior, and the exposure base is determined:

$$P = \begin{cases} General: & 0.816 \\ Novice: & 0.120 \\ Experience: & 0.064 \end{cases} \qquad (3)$$

On this basis, the frequency of experience and the intensity of claims are calculated. The accident rate (empirical frequency) of different behavior categories is extracted from the ASRS data and the risk multiplier is integrated and summarized:

The risk multiplier

Table 1

Behavior type	Number of claims	Number of exposures	Empirical frequency	Risk multiplier
General	31,549	31,750	0.994	1.04
Novice	4,555	4,655	0.979	1.02
Experience	1,107	2,506	0.442	0.462

Its empirical frequency is defined as:

$$\hat{\lambda}_c = \frac{n_{\text{claim},c}}{n_{\text{exposure},c}}, \quad c \in \{General, \ Novice, \ Experience\}, \ (4)$$

Normalization to form a relative frequency factor (relative overall average frequency):

$$f_c^{\text{freq}} = \frac{\hat{\lambda}_c}{\bar{\lambda}}, \quad \bar{\lambda} = \frac{\sum_c n_{\text{claim},c}}{\sum_c n_{\text{exposure},c}}. \qquad (5)$$

Further consider the historical average compensation data to construct a comprehensive risk adjustment factor:

$$R_c = f_c^{\text{freq}} \cdot f_c^{\text{sev}}, \qquad (6)$$

f_c^{sev} is the severity ratio, which needs to be calculated based on compensation simulation. In this model, the risk multiplier is temporarily used to replace the mean direct compensation, and the following final behavior factors are formed for the weighting of the rate model.

Comprehensive behavioral factor

Table 2

Behavioral category	Comprehensive behavioral factor
General	1.039
Novice	1.023
Experience	0.462

These behavioral factors are used for claim expectation modification in pure premium simulations:

$$\mathbb{E}[S_i] = \pi_{\text{base}} \cdot R_{c(i)}, \quad (7)$$

Among them $c(i)$ is the behavior category corresponding to the i -th simulated individual

3.2.3 No - claim discount (NCD) system

The NCD (No-Claim Discount) system aims to incentivize low-risk operational behaviors through a rate reward mechanism. This study was introduced in the following ways:

Let $t \in \{0, 1, 2, 3+\}$ epresent the number of claim - free years, and the corresponding discount factor is $\delta_t \in (0, 1]$, then the adjusted premium is:

$$P_{\mathrm{NCD}} = \delta_t \cdot \lambda \cdot \mu \quad (8)$$

This model is set as follows:

$$\delta_t = \begin{cases} 1.00, & t = 0 \\ 0.90, & t = 1 \\ 0.80, & t = 2 \\ 0.70, & t \geq 3 \end{cases} \quad (9)$$

The NCD distribution assumptions are as follows:

$$\mathbb{P}(t=0) = 0.20, \quad \mathbb{P}(t=1) = 0.30, \quad \mathbb{P}(t=2) = 0.30, \quad \mathbb{P}(t \geq 3) = 0.20$$

3.3 Reconstruction and estimation of kernel - based model

3.3.1 GLM frequency modeling

The frequency term λ is designed using binomial logistic regression. $y_i \in \{0, 1\}$ indicate whether the i - th observation results in a claim.Let the claim behavior type be x_i, Then the model is as follows:

$$\log\left(\frac{\mathbb{P}(y_i = 1)}{1 - \mathbb{P}(y_i = 1)}\right) = \alpha + \sum_{k=1}^{K} \beta_k \cdot \mathbf{1}_{\{x_i = k\}}, \qquad (10)$$

The significance of the coefficient β_k reflects the impact of task type on the probability of an accident (and thus a claim) occurring.

3.4 Severity model structure and fitting

3.4.1 Two - stage modeling structure

In order to characterize the distribution of compensation with significant right-sided and tail characteristics, a two-stage modeling was adopted:

Main body part $S_{\mathrm{body}} \sim \mathrm{Gamma}(\alpha, \beta)$;

Tail part $S_{\mathrm{tail}} \sim \mathrm{GPD}(\xi, \sigma)$ or $\mathrm{Exp}(\lambda)$;

With a threshold of $u = \mathrm{Quantile}(S, 90\%)$, the overall compensation is modeled as:

$$S = \begin{cases} S_{\mathrm{body}}, & S \leq u \\ u + S_{\mathrm{tail}}, & S > u \end{cases} \quad (11)$$

The main part uses the Bayesian GLM model:

$$\log(\mu) = \theta_0, \quad S_{\mathrm{body}} \sim \mathrm{Gamma}(\mu, \phi), \qquad (12)$$

and use`stan_glm` in the rstanarm, The fitting of the tail part is as follows:

If $n > 5$ and $\mathrm{Var}(S_{\text{tail}}) > 0$, use `ismev :fpot` to fit GPD；Otherwise, use the exponential distribution to fit `fitdist(..., "exp")

3.5 Monte carlo simulation and pure premium estimation

After estimating frequency and severity, the fused data is used to estimate the pure premium. Simulate $N = 10^5$ claim values $S_i \sim \tilde{F}_S$ for each type of task, which mixes Gamma and GPD;.

Apply the claim limi L and deductible D ：

$$S_i^{\text{adj}} = \max(0, \min(S_i, L) - D) \qquad (13)$$

Estimate the pure premium：

$$\text{Pure Premium} = \lambda \cdot \mathbb{E}[S_i^{\text{adj}}] \qquad (14)$$

4.Validation of Analytical Results

4.1 Descriptive statistics of data

The raw data used in this study was obtained from the National Transportation Safety Board (NTSB) and the NASA Aviation Safety Reporting System (ASRS)，It includes a number of flight records, accident information and compensation results suitable for simulating Chinese civil UAVs. Before entering the modeling, data cleaning, variable standardization (for example, the unit is unified as RMB), missing value filling, extreme value pruning (based on IQR) and other operations have been completed.

4.1.1 Mission structure distribution

From the perspective of flight mission categories, the sample mainly focuses on personal use, accounting for more than 80%, followed by teaching tasks and other air operations, accounting for 12.0% and 6.4% respectively. The structure shows a strong bias, reflecting that UAVs are still dominated by amateur flights in the mass market.

Flight mission type distribution

Table 3

Flight mission	Count	Percentage (%)
Personal	31750	81.596464
Education	4655	11.963198
Other aerial operations	2506	6.440338

4.1.2 Statistical characteristics of indemnity severity

As one of the core output variables of the model, the amount of claims presents a completely different distribution pattern in the two types of insurance. The overall concentration of claims for airframe insurance is relatively high, with the main peak in the range of RMB

33,000-36,000, showing obvious bimodal characteristics, mainly dominated by the standard maintenance cost of spare parts; Liability insurance indemnity has a stronger heavy-tailed characteristic, with a maximum value of 849,704.3 yuan, and there are multiple local peaks, showing a high degree of correlation with accident casualty levels, third-party liability definition and other factors.

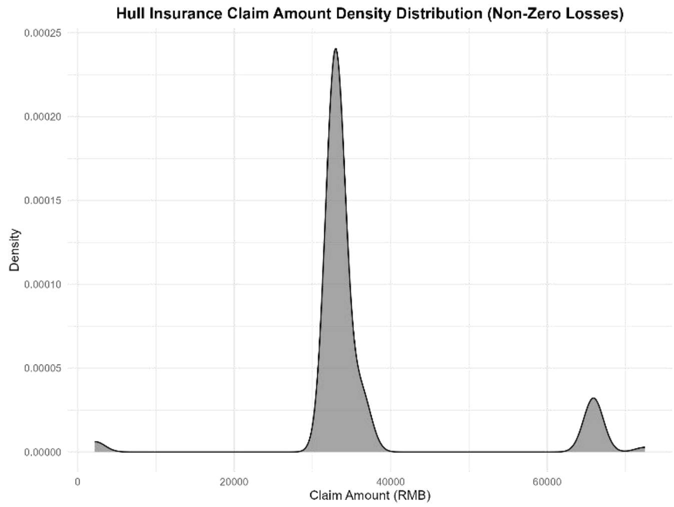

Figure 1 Hull insurance claim amount density distribution（Non-zero loss）

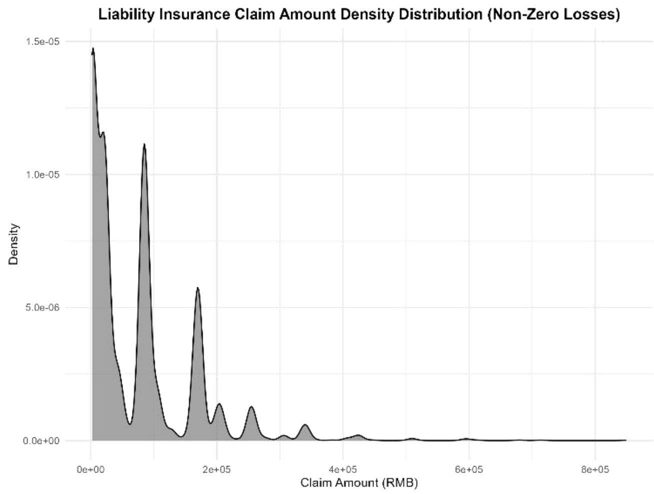

Figure 2 Liability insurance claim amount density distribution（Non-zero loss）

4.2 Validation of scaling rationality and task heterogeneity reservation for Man-vehicle data

For the airframe insurance, the severity distribution of the three types of missions remains highly overlapping as a whole, and the peak value is concentrated in the range of RMB30,000 \ ~ 40,000, indicating that the scaling operation effectively unifies the standard indemnity level of various flight missions; For liability insurance, although the peak position is roughly the same, the "teaching task" shows a wider tail thickness and still retains more high compensation, which shows that the scaling does not excessively smooth the actual heterogeneity and guarantees the risk discrimination ability of the follow-up model.

Further numerical analysis supports the above findings. The average indemnity of liability insurance under "teaching tasks" is as high as ¥ 103,685.13, which is much higher than "personal use" (¥ 71,278.86) and "other air operations" (¥ 47,671.20). Although the risk scaling factor has been adjusted to the original data, the liability risk structure under different flight missions still shows significant differences, reflecting the complexity of accident liability and uncertainty of compensation brought by the mission itself.

Severity by flight mission type

Table 4

Flight mission	Hull claims count	Avg. hull severity	Liability claims count	Avg. liability severity
Personal use	31307	36243.38	17652	71278.86
Other air operations	1079	36189.39	639	47671.20
Teaching	4524	37697.65	1963	103685.13

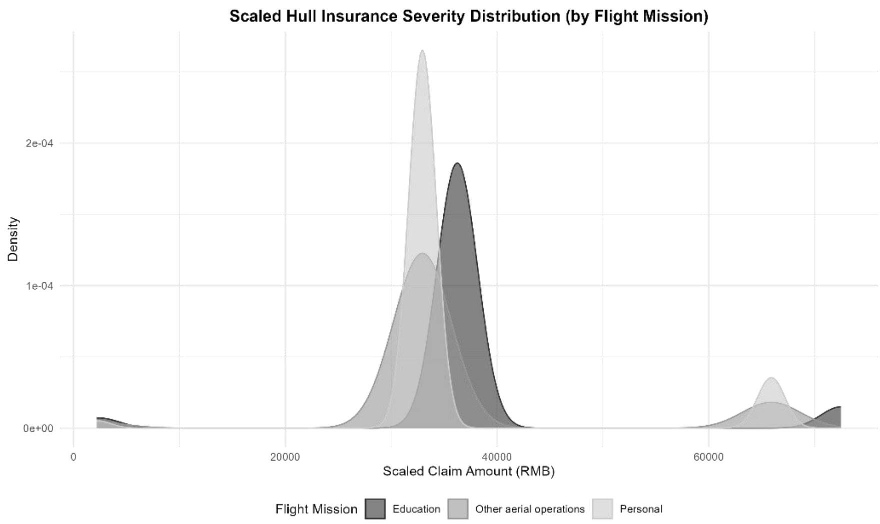

Figure 3 Scaled hull insurance severity distribution (by flight mission)

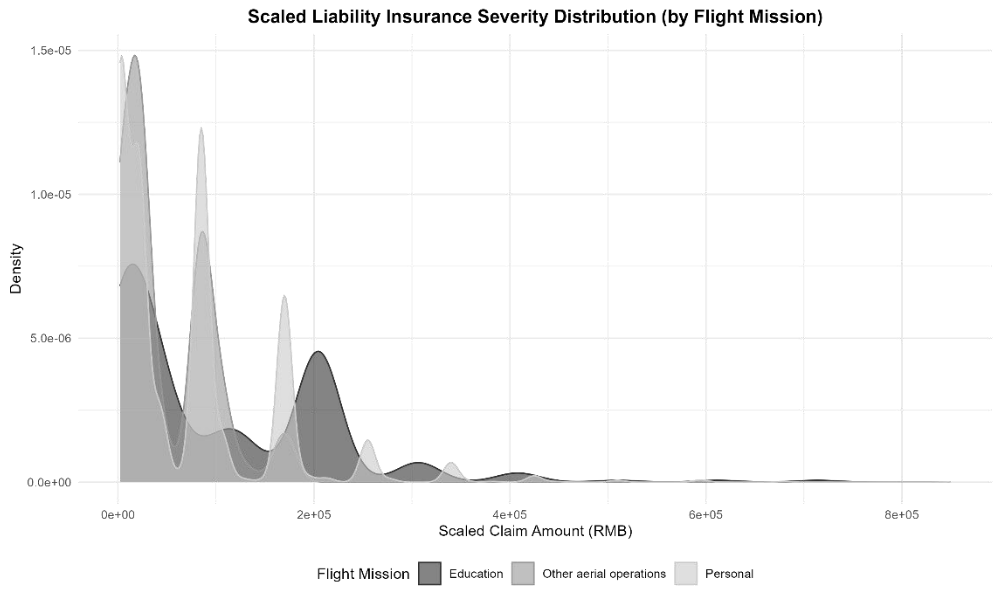

Figure 4 Scaled liability insurance severity distribution (by flight mission)

According to the analysis, we can find that the risk scaling matrix has a significant effect in reducing the systematic bias, while still retaining the risk structure differences brought by the flight mission category, indicating that the use of calibrated manned data for modeling has good interpretability and generalization. This result verifies the rationality of cross-platform data integration and lays a solid empirical foundation for the next stage of frequency-severity decomposition modeling.

4.3 Pure premium determination: SPG method and NCD factor adjustment

In this study, the traditional frequency-severity decomposition (SPG) method was used to determine the net premium:

$$\pi_i = f_i \cdot s_i \qquad (15)$$

Among, $\pi_i$is i Is the net premium for the task of type I, f_i as frequency (loss probability) , s_i as the mean severity. The loss rates of each task category are shown in the table below:

The loss rates of each task category

Table 5

Flight mission	Hull loss rate	Liability loss rate
Personal use	0.0700	0.0050
Teaching	0.1155	0.0120
Other air operations	0.2100	0.0300

Based on the above modeling results of frequency and severity, the pure premium structures of task categories are constructed respectively, and the simulation values are used to assist the tail correction. As an important adjustment mechanism for liability insurance, the NCD system defines the following discount factors:

$$\text{Adjusted Premium}_i = \pi_i \cdot \gamma_i, \quad \gamma_i \in [0.3, 1] \qquad (16)$$

γ_i

Among It indicates the discount multiples of liability insurance corresponding to different compensation records, which can be designed as a five-level system: 0 compensation for three consecutive years to enjoy the maximum discount, 1 compensation to restore the original price, 2 compensation to rise, etc.

4.4 Conversion Factor Setting

Define the risk adjustment factors for each type of mission γ_m， to reflect the degree of deviation from the standard compensation Denote：

$$m \in \{\text{PERS}, \text{INST}, \text{OTHER\_MANNED\_PURPOSE}\} \quad (17)$$

represents the mission purpose; L_{orig} is the original compensation; L_{adj} is the converted compensation.

Then, the conversion formula is:

$$L_{\text{adj}}^{\text{hull}} = L_{\text{orig}} \cdot \gamma_m^{\text{hull}}, \quad L_{\text{adj}}^{\text{liab}} = L_{\text{orig}} \cdot \gamma_m^{\text{liab}}, \quad (18)$$

Where γ_m^{hull} and γ_m^{liab} are the risk adjustment factors for hull insurance and liability insurance respectively, and they are taken from the following table:

The hull insurance adjustment factor and the liability insurance adjustment factor

Table 6

Type of flight mission	Hull insurance adjustment factor	Liability insurance adjustment factor
PERS	1.0	1.0
INST	1.1	1.2
OTHER_MANNED_PURPOSE	1.0	1.0

The matrix is derived from the actual flight behavior risk assessment results and is used to scale the compensation value item by item in the simulation. The coefficient results can be understood as follows: the liability risk adjustment coefficient of INST for teaching tasks is 1.2, indicating that the average personal injury compensation is significantly higher than that of other tasks. Airframe insurance also rises by 10% in the INST scenario, reflecting that it is more prone to damage in intensive take-off and landing and multi-person operations.

The core role of conversion factors is to introduce an adjustment layer that is closely related to the use case, so that the source data can still be effectively modeled without sufficient direct drone historical compensation. The operation complexity, environmental interference level and responsibility division system corresponding to different flight missions are different, and the introduction of conversion factors ensures the reflection of the rate results to reality, and leaves room for parameterization of the model to other UAV types.

4.5 Tweedie GLM model and adjustment factor analysis

In order to further improve the prediction ability and interpretation of the UAV insurance pure premium model, the Tweedie generalized linear model (GLM) is introduced to model the total payout. The Tweedie distribution has both Poisson and gamma distribution characteristics, which is suitable for modeling a large number of insurance data with both zero and continuous positive claims, and is widely used in actuarial and non-life insurance modeling.

The Tweedie GLM model used in this study is set up as follows:

$$\begin{aligned}\log(\mathbb{E}[Y_i]) = \beta_0 + \beta_1 \cdot \text{flight\_mission}_i + \beta_2 \cdot \text{operator\_behavior}_i \\ + \beta_3 \cdot \text{ncd\_level}_i + \log(\text{exposure}_i)\end{aligned} \tag{19}$$

where Y_i xrepresents the total amount of claims of a single policy, and log (exposure) is the bias item, correcting the difference in the coverage period.

4.6 Analysis of pure premium structure and differences in task types

To validate the rationality of the rate-making results, this study calculates and visualizes the pure premiums for different flight mission types based on the claim frequency and severity models generated through Monte Carlo simulation.

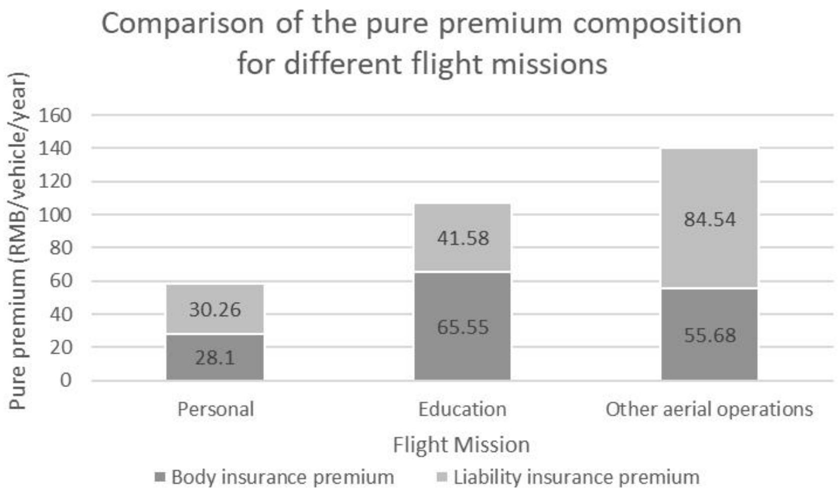

Figure 5 Comparison of the pure premium compositionfor different flight missions

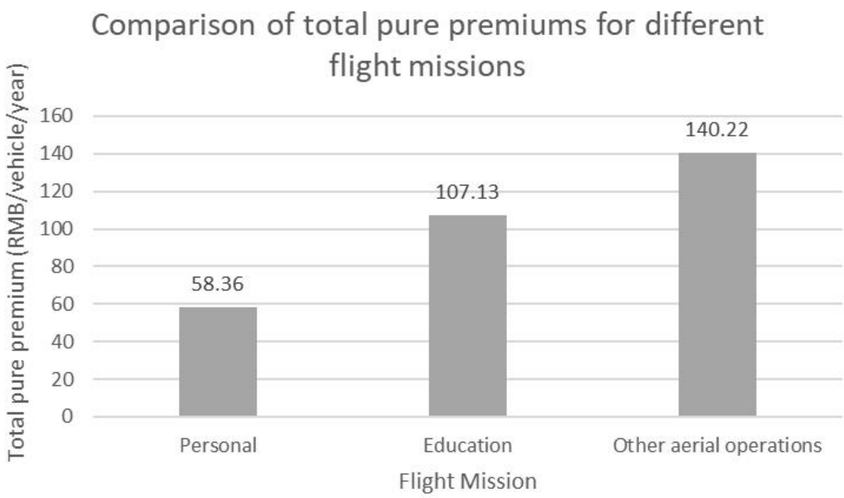

Figure 6 Comparison of total pure premiums for different flight missions

As illustrated in the figure, the pure premium levels for hull insurance and liability insurance of personal-use drones are relatively comparable, with hull insurance being slightly higher than liability coverage. This indicates that the primary risk exposure for personal users concentrates on physical damage to the aircraft. In contrast, training missions demonstrate significantly elevated pure premiums, particularly for hull coverage, which directly correlates with the frequent maneuver practices and intensive operational demands characteristic of instructional flights. For commercial operations (such as surveying, inspection, and other aerial work tasks), the total pure premium reaches the highest level, with liability insurance accounting for over 60% of the total. This reflects substantially higher third-party risk exposure in these professional applications. The study further reveals a clear gradient in total pure premiums across mission categories: Personal Use < Training < Commercial Operations This finding exhibits strong consistency with the risk factor estimations derived from the Tweedie GLM model in earlier analysis, both validating the critical role of mission type in UAV insurance pricing and underscoring the necessity of stratified premium calculation.

5. Summary and Optimization Recommendations

This study developed a pure premium rating model for drone insurance, integrating both frequency-severity decomposition and Tweedie GLM modeling approaches while incorporating novel factors such as No Claims Discount (NCD) systems and operator behavior characteristics. The results demonstrate that the proposed model achieves a balance between interpretability and predictive accuracy, providing an actuarial pricing tool for the emerging drone insurance market. However, from an academic perspective, there remains room for improvement in the model's robustness and practical applicability.

Regarding modeling methodology, both the two-part frequency-severity model and Tweedie GLM have respective advantages and limitations. The two-part model separately models claim frequency and loss severity, enabling independent quantification of risk factors' impacts on claim probability and loss magnitude, with Generalized Pareto Distribution (GPD) enhancing extreme loss tail risk capture. However, it requires fitting multiple sub-models, making the process more complex. In contrast, Tweedie GLM simultaneously models zero claims and positive losses through a single model structure, offering simplicity and computational efficiency. Yet its distributional assumptions show limitations in fitting extreme heavy-tailed losses, and the blending of frequency and severity effects reduces interpretability. Thus, the two-part model excels in interpretability, while Tweedie GLM demonstrates superior advantages in modeling simplicity and overall loss prediction. Although GPD tail fitting improves extreme loss capture, threshold selection sensitivity remains an issue, suggesting future exploration of adaptive threshold techniques to enhance robustness.

The incorporation of operator behavior factors quantifies the impact of "soft" risk elements like novice pilots on claims, addressing a gap in traditional models' consideration of human factors. However, measurement and data acquisition for these factors present challenges, with their effectiveness constrained by data quality. Task-type risk stratification effectively differentiates risk levels across mission categories, producing premium gradients that generally align with actual risk exposure. Nevertheless, the current broad classification of task types fails to fully capture internal risk variations, indicating potential refinements through more granular categorization or additional features to improve risk differentiation.

Acknowledgements:

This project is funded by the National College Students' Innovation and Entrepreneurship Training Program (No. 202410456025) and supported by the China Center of the Serbian Academy of Sciences and Arts and the Hong Kong Institute of Humanities and Natural Sciences and Technology.

References

[1] Bühlmann H. Experience Rating and Credibility[M]. Springer, 1970:89-121.

[2] Klugman S. A.,Panjer H. H., Willmot, G. E. Loss Models: From Data to Decisions[M]. John Wiley & Sons, 2008:205-237.

[3] Frees E. W. Regression Modeling with Actuarial and Financial Applications[M]. Cambridge University Press, 2010:156-178.

[4] Meng Shengwang. Actuarial Science[M]. Chinese Minmin University Press, 2013:112-135.

[5] Xie Y.T., Li M., Zhang H. Research on Catastrophe Insurance Rate Making Based on Monte Carlo Simulation[J]. Statistical and Information Forum, 2016, 31(5): 39-45.

[6] Zhang L.Z.,Wang H., Zhao Y. Monte Carlo Simulation and Machine Learning Integration of Engineering Insurance Rate Determination[J]. Insurance Research, 2020, (8): 65-78.

[7] McDonald J. B., Xie X. Modeling Automobile Insurance Claims Data Using Generalized Linear Models with Random Effects[J]. Journal of Applied Econometrics, 2006, 21(5): 603–625.

[8] Wang D.H., Zhang R.G. Joint Modeling of Auto Insurance Claim Frequency and Severity[J]. Insurance Research, 2011, (8): 95–101.

[9] Bae S., Choi J., Kim K. Insuring Unmanned Aerial Systems: A Risk Modeling and Pricing Framework[J]. Risk Management and Insurance Review, 2020, 23(1): 105–129.

[10] Davis A. R., Mikosch T. Extreme Value Theory for Space–Time Processes with Heavy - Tailed Distributions[J]. Stochastic Processes and their Applications, 2007, 118(4): 560-584.

[11] Wang Z. Research on Auto Insurance Rate - making Based on Double Hierarchical Generalized Linear Model[D]. Chongqing Technology and Business University, 2023:45-68.

[12] Xu L. Research on Agricultural Catastrophe Risk Assessment Model[D]. Chinese Academy of Agricultural Sciences, 2012:72-91.

[13] Sun W.W. Application of Generalized Additive Model Based on Tweedie Class Distribution in the Determination of Auto Insurance Rates[J]. Journal of Tianjin University of Commerce, 2014, 34(01): 60-67.

[14] Zhang L.Z., Xie H.Y. Research on the Application of Tweedie GLM Model in Auto Insurance Rate Determination [J]. Insurance Research, 2017, (01): 80-90.

[15] Huang S.L., Zhang Y., Chen N. Application of GPD Model in Property Insurance Rate - making: An Empirical Analysis [J]. Statistics and Decision Planning, 2010, (04): 27-29.