Communication Papers of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 161–166
DOI: 10.15439/2025F5118
ISSN 2300-5963 ACSIS, Vol. 45

Claim Frequency Estimation in Motor Third-Party Liability (MTPL): Classical Statistical Models versus Machine Learning Methods

Ondřej Vít[†], Lubomír Seif[‡]
[†]ORCiD: 0009-0003-7317-5856
[‡]ORCiD: 0009-0003-7444-9425
Department of Statistics and Probability
Faculty of Informatics and Statistics
Prague University of Economics and Business
W. Churchill's square 4, 13067 Prague, Czech Republic
[‡]Email: ondrej.vit@vse.cz
[‡]Email: lubomir.seif@vse.cz

Lubomír Štěpánek[1, 2, 3]
ORCiD: 0000-0002-8308-4304
[1]Department of Statistics and Probability
[2]Department of Mathematics
Faculty of Informatics and Statistics
Prague University of Economics and Business
W. Churchill's square 4, 13067 Prague, Czech Republic
Email: lubomir.stepanek@vse.cz
&
[3]Institute of Biophysics and Informatics
First Faculty of Medicine
Charles University
Salmovská 1, 12000 Prague, Czech Republic
Email: lubomir.stepanek@lf1.cuni.cz

***Abstract*—This paper compares classical statistical models and machine learning techniques for claim frequency estimation in compulsory motor third-party liability insurance (MTPL). We evaluate Generalized Linear Models (GLMs), Hurdle models, and feedforward neural networks on real-world insurance data. Emphasis is placed on the trade-off between interpretability and predictive power, especially in segments with scarce data. Our findings show that expert-driven data preparation enables GLMs to perform competitively with complex neural networks. Hurdle models further improve performance in zero-inflated settings. While neural networks offer improved predictive performance in some segments, they struggle in underrepresented ones. Results highlight that careful preprocessing is as important as model complexity.**

***Index Terms*—claim frequency, motor third-party liability, neural network, generalized linear model, hurdle model**

I. Related Work

CLAIM frequency modeling in motor third-party liability (MTPL) insurance has traditionally been dominated by classical statistical techniques, especially Generalized Linear Models (GLMs) [1]. These models are widely used due to their interpretability and compatibility with insurance-specific assumptions, such as the use of exposure as an offset and count response distributions like Poisson or negative binomial.

To handle overdispersion and zero-inflated data, hurdle models and zero-inflated Poisson models have been proposed [2]. These models separate the claim occurrence process from the frequency process and are particularly useful when a large proportion of the policies report zero claims.

This research was supported by the grant no. F4/36/2025 which has been provided by the Internal Grant Agency of the Prague University of Economics and Business.

In recent years, machine learning (ML) techniques, including random forests, gradient boosting, and neural networks, have been introduced to actuarial problems [4]. Their flexibility allows them to capture non-linearities and interactions automatically, potentially leading to improved predictive performance. However, the trade-off between predictive accuracy and interpretability remains a critical consideration in insurance applications.

Classical models have shown limitations in highly heterogeneous MTPL segments or in portfolios with high zero inflation. This motivates exploration of machine learning methods which might overcome these shortcomings by capturing complex nonlinear interactions. This study investigates whether neural networks and hybrid models offer significant improvements over classical GLMs in MTPL frequency modeling, particularly in underrepresented risk segments.

II. Data and Methods

Modeling in actuarial science plays a key role in risk estimation, pricing, and reserving within the insurance industry. Traditionally, it relies on statistical methods using historical data to predict future outcomes. A widely used framework is the frequency-severity approach, where claim frequency and severity are modeled separately [5]. This modular approach supports flexible and interpretable analysis across various insurance products.

Claim frequency modeling focuses on counting the number of claims over a specific period or policy. Commonly used models include the Poisson and negative binomial distributions, which accommodate different levels of dispersion in the

data [1]. In the context of compulsory motor third-party liability (MTPL) insurance, claim frequency is typically influenced by observable risk factors such as driver age, accident history, or region [6].

A prominent framework for frequency modeling is the generalized linear model (GLM), which provides a flexible yet interpretable approach to capturing linear effects on the log scale [2]. However, insurance data often exhibit structural properties such as excess zeros and overdispersion, which GLMs may not sufficiently handle. To address these issues, hurdle models [3] have been introduced as a semi-parametric extension of GLMs, where zero and positive counts are modeled separately. This two-part structure allows for more robust modeling of claim occurrence and intensity, particularly in MTPL datasets.

With the increasing availability of computational power and large-scale data, modern machine learning techniques such as neural networks have become attractive alternatives for predictive modeling [7]. Although these models typically lack the interpretability of classical approaches, they are capable of capturing nonlinear interactions and complex relationships between predictors. Recently, their application in actuarial science has gained attention, and comparative studies of traditional and machine learning-based frequency models have started to emerge [4]. This paper contributes to this line of research by evaluating the performance of GLMs, hurdle models, and feedforward neural networks in the task of frequency modeling for MTPL insurance.

Since their introduction by Nelder and Wedderburn [8], generalized linear models (GLMs) have become a standard tool in actuarial modeling. They link the expected value of a dependent variable to a linear combination of covariates through a specified link function. GLMs support a variety of distributions, including Poisson, Gamma, and Tweedie, making them suitable for different types of insurance data.

Despite their strengths, GLMs may struggle with highly sparse or zero-inflated data structures. In such contexts, hurdle models offer a valuable alternative by explicitly modeling the excess zeros separately from the positive outcomes. This is particularly relevant in claim frequency modeling, where the majority of policyholders may not report any claim, while a smaller subset reports one or more claims.

With the growing availability of detailed policyholder and behavioural data, new approaches such as neural networks are increasingly considered. These models are capable of capturing complex nonlinear relationships and interactions that traditional methods may miss, thereby enhancing predictive accuracy [9].

A. Claim Frequency modeling in Actuarial Science

In actuarial science, modeling the frequency of insurance claims is traditionally addressed through count data models. The fundamental statistical framework for this task begins with the *Poisson regression model*, which assumes that the number of claims Y_i for observation i follows a Poisson distribution,

$$Y_i \sim \text{Poisson}(\lambda_i), \quad \text{with} \quad \log(\lambda_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{1}$$

where λ_i represents the expected number of claims for i-th observation, $\mathbf{x}_i$ is a vector of covariates (such as age, region, or vehicle type) of i-th observation, and $\boldsymbol{\beta}$ is a vector of coefficients to be estimated [8]. This model is embedded in the framework of *generalized linear models (GLMs)*, which relate to the conditional mean of the response variable to linear predictors via a link function and specify a distribution from the exponential family [2].

However, a common issue with real insurance data is *overdispersion*, i.e., the variance of Y_i exceeds the mean, violating the commonly known Poisson assumption, i.e., $\mathbb{E}(Y_i) = \text{var}(Y_i)$. A standard solution is the *negative binomial model* (NB), which introduces an additional parameter θ to model the dispersion,

$$\begin{aligned} Y_i &\sim \text{NB}(\mu_i, \theta), \\ \log(\mu_i) &= \mathbf{x}_i^\top \boldsymbol{\beta}, \\ \text{var}(Y_i) &= \mu_i + \frac{\mu_i^2}{\theta}, \end{aligned} \tag{2}$$

keeping the mathematical notation the same as before. The NB model maintains the GLM structure and is estimated using quasi-likelihood or maximum likelihood methods [1], [10].

In many practical applications, especially in motor third-party liability (MTPL) insurance, datasets are *zero-inflated*: a large proportion of policyholders report no claims. To account for this, *hurdle models* [3] have become a useful extension. A hurdle model separates the modeling of zeros and positive counts. Formally, it combines

- a binary model for the probability of at least one claim,

$$P(Y_i > 0) = \pi_i, \quad \text{with} \quad \text{logit}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\gamma}, \tag{3}$$

- a truncated count model (typically truncated Poisson or NB) for $Y_i \mid Y_i > 0$,

$$Y_i \mid Y_i > 0 \sim f_{\text{trunc}}(\mu_i), \tag{4}$$

where π_i is the probability that policyholder i reports at least one claim, $\mathbf{x}_i$ is the vector of explanatory variables for policyholder i, $\boldsymbol{\gamma}$ is the parameter vector of the binary component, μ_i is the conditional expected number of claims given $Y_i > 0$, and $f_{\text{trunc}}(\mu_i)$ denotes the probability distribution of the count component truncated at zero (e.g., zero-truncated Poisson or negative binomial) with mean μ_i.

This two-part model allows separate covariate effects for claim occurrence and claim frequency conditional on having a claim, offering greater flexibility.

More recently, *neural networks* and other *machine learning models* have been explored in actuarial applications. Neural networks estimate nonlinear functions of covariates without requiring a prespecified parametric form,

$$\hat{y}_i = f(\mathbf{x}_i, \boldsymbol{\theta}) \tag{5}$$

where f is a composite function defined by layers of transformations [7]. These models are particularly powerful in large datasets with complex interactions, although they often lack

TABLE I
SUMMARY OF KEY VARIABLES USED IN THE GLM

Variable	Description
Vehicle power	Engine power grouped into categories (e.g., <50, 50–74, 75–89, 90–109, 110+)
Vehicle weight	Weight category of the vehicle in kilograms
Driver age	Age group of the main driver (e.g., 18–22, 23–29, ..., 65+)
Driver–owner	Whether the driver differs from the policyholder
Vehicle status	Vehicle usage classification (e.g., personal, company)
Bonus–Malus	Claim-free discount class (e.g., -1, 0–10, M1–M6)
Region	Region category (e.g., Prague, Town, Rural)
Payment frequency	Frequency of premium payments
Fuel type	Type of fuel (e.g., petrol, diesel, other)

the interpretability of GLMs and often require regularization to prevent overfitting.

While GLMs remain the backbone of actuarial modeling due to their interpretability and regulatory acceptability, the flexibility of hurdle models and the predictive power of neural networks provide valuable complements, especially in large-scale portfolios with heterogeneous policyholder characteristics.

B. Data preprocessing

The data created for this paper are derived from real data of the Czech compulsory liability insurance market and include only passenger cars up to 3.5 tonnes. In total, 130,585 contracts have an aggregate insurance period equal to 115,492 years. Each contract specifies an exposure period expressed as a fraction of a full year. These individual exposures are aggregated across contracts to obtain the total insurance exposure, also referred to as the aggregate insurance period. Since compulsory liability insurance policies are typically written for one year, the exposure values range from zero to one. In the case of policyholder retention when the insured renews coverage with the same insurer, the subsequent period is treated as a new contract and includes information on the policyholder's prior behavior within the portfolio.

Each contract includes an exposure period, expressed as a proportion of a full year, which is aggregated across policies to form the total insurance exposure (referred to as the aggregate insurance period).

An important aspect of data preparation step was the transformation of selected continuous predictors into categorical variables. This expert-driven segmentation allowed Generalized Linear Models (GLMs) to better capture non-linear effects and improve interpretability. We refined them based on observed claim frequency trends. While segmentation increases the number of levels and requires a sufficient sample size [11], the size of our dataset allowed for stable estimates. Final category definitions were chosen to balance homogeneity within groups and predictive performance.

C. Used models

In this section, we revisit principles and usages of the models applicable for claim frequency estimation. Models' overview is in Table II. The models are compared against a homogeneous benchmark that assigns each policyholder the average annual claim frequency observed in the training data.

D. Homogeneous model

In this study, a homogeneous model is used as a benchmark to evaluate the performance of more complex predictive models. This baseline approach assigns the same predicted annual claim frequency to every policyholder, specifically the average claim frequency observed in the training data. The homogeneous model does not incorporate any individual-level features or covariates, effectively treating all policyholders as identical with respect to risk.

E. Generalized Linear Model (GLM)

A generalized linear model was developed in a Poisson regression framework to model claim frequencies, incorporating an offset to adjust for different exposure periods between policies. Predictor variables representing policyholder and risk characteristics were standardised prior to modeling to improve numerical stability and ensure comparability of coefficient estimates. An intercept term was explicitly included to capture the baseline level of risk. The model links the expected number of claims to the linear predictor via a logarithmic link function, which is consistent with the canonical specification for Poisson results.

The fitting was performed using a maximum likelihood method [1] assuming a Poisson distribution, where the logarithm of exposure was treated as an offset to normalize the number of claims by exposure duration. The resulting model estimates were then used to generate predictions on both the training and validation datasets, resulting in exposure-adjusted expected claim frequencies. This approach supports transparent derivation of risk factors and is consistent with established actuarial methodologies for modeling frequencies, providing a sound basis for pricing and reserving tasks.

F. Neural Networks

Feedforward neural networks were selected for their ability to model smooth nonlinear relationships and their previous application in insurance frequency modeling. Two neural network models have been developed to improve the prediction of claim frequencies by capturing complex non-linear relationships in the data that traditional GLMs may not adequately model. Both models use a feedforward architecture with three hidden layers that progressively reduce dimensionality from 20 to 10 neurons, each using ReLU activation to introduce nonlinearity. Importantly, the logarithm of the exposure was incorporated as an additional input via concatenation before the final output layer, ensuring that different policy exposure times were accounted for analogously to offsets in the GLM. The output layer applies an exponential transformation to ensure a strictly positive prediction of the number of claims, which is consistent with the Poisson modeling framework.

Both neural networks were trained using feedforward architectures with exponential activation in the output layer to

TABLE II
OVERVIEW OF FREQUENCY MODELS

Model	Formula	Description	Setting
Homogeneous model	$\mathbb{E}[Y_i] = \bar{y}$	Baseline model assigning each policyholder the same average claim frequency from training data.	No covariates. No offset. Serves as a naive benchmark for comparison.
Generalized linear model (GLM)	$\log(\mathbb{E}[Y_i]) = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \log(e_i)$	Standard Poisson regression with log link and offset for exposure. Captures additive effects of standardized covariates on log-scale.	Fitted via MLE. Offset: $\log(e_i)$. Exposure-adjusted prediction. Standardized inputs. Intercept included.
Neural Network (NN)	$\hat{y}_i = \exp(f(\mathbf{x}_i, \log(e_i)))$	Feedforward network with 3 hidden layers and ReLU activation. Exposure included as input; exponential output ensures positivity.	Two variants: one with repeated 5x10 CV and Nadam optimization over 200 epochs, the other trained once with early stopping at 150 epochs. Both use Poisson loss. The number of epochs was determined based on convergence diagnostics and validation loss.
Hurdle model with GLM Hurdle_GLM	$\mathbb{P}(Y_i > 0 \mid \mathbf{x}_i) \cdot \mathbb{E}[Y_i \mid Y_i > 0, \mathbf{x}_i]$	Two-part model: logistic regression for probability of claims; Poisson GLM (on $Y_i > 0$) for positive counts. Decouples occurrence and frequency.	Binary part: calibrated logistic regression (Platt). Count part: Poisson GLM with offset. Standardized inputs. Applied to positive-claim subsample.
Hurdle model with XGBoost Hurdle_XG	$\mathbb{P}(Y_i > 0 \mid \mathbf{x}_i) \cdot \mathbb{E}[Y_i \mid Y_i > 0, \mathbf{x}_i]$	As above, but second part is a gradient-boosted zero-truncated Poisson model via XGBoost. Captures nonlinearity and interactions.	Binary: logistic with Platt scaling. Count: XGBoost with custom zero-truncated Poisson loss. Hyperparameter tuning, early stopping, offset included.

ensure non-negative predictions, appropriate for count data. The Poisson loss function was applied throughout, reflecting the underlying assumption that claim counts follow a Poisson distribution. Optimization was performed using the Nadam algorithm [12].

The first model employed 5-fold cross-validation, repeated 10 times, to robustly estimate predictive performance. The model was trained for 200 epochs with a batch size of 256. The best configuration was selected based on minimum Poisson deviance on validation folds.

The second model used the full training set and internal validation split for early monitoring, trained over 150 epochs. Both models integrated exposure as an input feature rather than as an offset, ensuring compatibility across model classes.

G. Hurdle models

Hurdle modeling was implemented as a two-part approach to effectively address the zero-inflation commonly observed in claim frequency data. The first part consisted of a binary classification model predicting the probability of a positive number of claims versus zero claims. This classification was primarily performed using logistic regression, enhanced with calibration techniques such as Platt scaling [13] to improve probabilistic accuracy. The model was trained on normalized features and used to estimate the hurdle probability, i.e., the probability that the policyholder reports at least one claim.

The second part modeled positive numbers of claims conditional on non-zero occurrences. Initially, a classical Poisson GLM was applied to a subset of the data with positive counts, including exposure as an offset, to account for different risk durations. Subsequently, more advanced Poisson models with truncated zeros were explored using gradient boosting machines (GBMs) implemented via XGBoost, which allow for flexible nonlinear effects and complex interactions between predictors. Extensive tuning of hyperparameters (gradient boosting) - adjusting learning rate, tree depth, penalization, undersampling frequency, and early stopping – was performed to optimize predictive performance and avoid overfitting. GBM models used truncated Poisson likelihood to correctly handle the absence of zeros in the frequency component.

Predictions from the hurdle model combined the probability of a positive count from the calibrated binary classifier with the expected number of claims conditional on positivity from the Poisson or zero truncation Poisson models. This product provided an overall frequency estimate that explicitly accounted for zero inflation and heterogeneity in the incidence and severity of claims. Performance metrics, such as Poisson deviance (defined in the following section), were calculated on both training and test samples to evaluate fit and generalization.

The hurdle modeling approach employed combines a binary component modeling the probability of positive claim counts with a zero-truncated count component predicting the frequency given a claim occurs. The binary component used logistic regression with Platt's scaling calibration to accurately estimate the probability of non-zero insurance claims, effectively addressing zero inflation in the data. For positive counts, both classical Poisson regressions and gradient boosted Poisson models with zero truncation (via XGBoost) were used, incorporating exposure as offset and using careful tuning of hyperparameters to balance bias and variance. The final hurdle prediction is the result of multiplying the predicted probability of a claim occurrence by the expected claim frequency conditional on positivity, allowing flexible and interpretable modeling of claim frequency that decouples the occurrence and severity processes while accounting for complex nonlinear relationships and regularization to avoid overfitting.

TABLE III
COMPARISON OF MODEL PERFORMANCE METRICS

Model	In-sample Poisson Deviance	Out-sample Poisson Deviance	Average frequency
Homogeneous model	22.1	22.2	0.0354
NN_cat	19.6	20.2	0.0362
NN_nocat	19.5	19.9	0.0364
GLM	19.9	19.6	0.0362
Hurdle_cat_GLM	20.0	19.7	0.0329
Hurdle_cat_XG	20.3	20.1	0.0358
Hurdle_nocat_GLM	20.9	20.9	0.0338
Hurdle_nocat_XG	20.9	20.9	0.0333

H. Model validation

Model validation was conducted on an independent dataset comprising approximately 39,176 contracts that were not used during training, ensuring an unbiased evaluation of predictive performance. Model accuracy was assessed using the Poisson deviance metric, defined as

$$\text{Poisson Deviance} = \frac{200}{n} \sum_{i=1}^{n} \left(\hat{y}_i - y_i + y_i \log \left(\frac{y_i}{\hat{y}_i} \right) \right), \quad (6)$$

where y_i are observed claim counts and $\hat{y}_i$ are predicted values. This measure is commonly used in actuarial science and generalized linear modeling to quantify goodness-of-fit for count data models under the Poisson assumption [2]. The lower Poisson deviance is, the better predictive performance a model does reach.

III. RESULTS

Based on the Poisson deviation, the homogeneous benchmark model performed worst, as expected – see Table III. It predicted the same expected claim count for all policies, equal to individual exposure times the average annual frequency in training data (0.0354). Its out-of-sample Poisson deviation was 22.2, exceeding all other models.

In contrast, all other models — including GLM, neural networks, and hurdle models — achieved lower deviance values ($\langle 19, 21 \rangle$) and better in-sample fit, with consistent average frequency on validation data (0.0355), indicating robust training.

GLM and *NN_cat*, both using discretized inputs, produced similar results and effectively modeled typical risk thresholds. However, GLM proved more stable in low-frequency segments, benefiting from regularization and pooling.

NN_nocat, trained on raw inputs, better captured smooth trends but missed local discontinuities, such as a frequency spike for drivers aged 40–50. This highlights the trade-off between flexibility and the ability to model structural effects.

The hurdle models (*hurdle_no_cat_GLM*, *hurdle_no_cat_XG*) used categorised data and combined a logistic part (claim occurrence) with a count part (conditional frequency). This dual structure is standard for handling excess zeros in insurance data, as mentioned before.

Both hurdle models reduced average predicted frequency to match the validation set more closely, addressing the tendency of other models to overestimate. Using domain-informed segmentation, they responded well to risk changes.

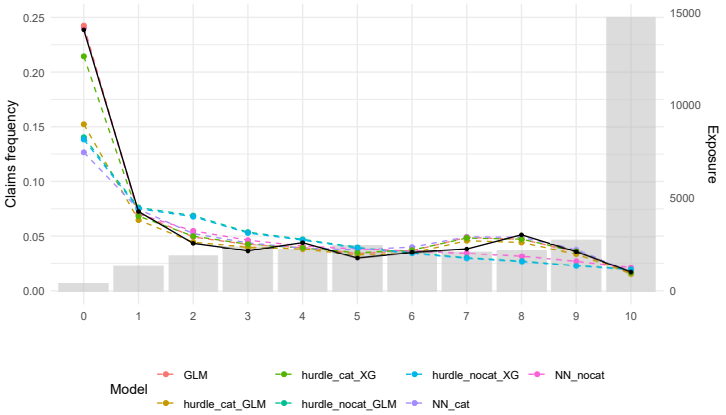

Fig. 1. Bonus segmentation: Models' claim frequency predictions (coloured lines) and segment exposures (grey bars).

The GLM-based hurdle model gave lower estimates than XGBoost, suggesting a more conservative bias.

The Fig. 1 shows model predictions (lines) and relative exposure sizes (bars) for each bonus-malus segment. Y-axis represents claim frequency, and X-axis bonus level from 0 to 10. GLM closely tracked the observed trend, especially in sparse segments like bonus 0 and 10. Neural networks slightly underestimated in extreme segments and overestimated in mid-range, reflecting sensitivity to data distribution.

Neural networks (models denoted as *NN_cat* and *NN_nocat*) tend to slightly overestimate in segments with medium bonus and slightly underestimate in very risky segments (especially segment 0). For example, *NN_cat* predicts only 0.127 in segment 0, which is clearly an underestimate relative to reality. In contrast, in the middle segments 2-6, the predictions of these models approximate the observed frequencies very well, sometimes more accurately than the GLM. The accuracy in the most overlapping segment of the bonus 10, for which it predicts 0.170, is important. The model *NN_nocat*, working with bonuses as numbers rather than as categorical categories, was better able to estimate bonus zero and on average predicted 0.240, but worse for bonus 10 (0.207). It also failed to capture non-monotonic fluctuations for bonus 8, for example.

Similarly, other models using numerical variables without categorization, namely *hurdle_nocat_XG* and *hurdle_nocat_GLM*, failed to capture the bias in bonus 8. Models using boosting stay relatively close to the other models in

most segments, but systematically underestimate frequencies in higher bonuses (lower risk). This was not the case for *hurdle_nocat_GLM* and both models significantly underestimated the 0 bonus. In particular, the poor performance for bonus 0 highlighted a key weakness of hurdle models in sparsely represented segments, where the two-stage structure increases the risk of error.

Models based on the hurdle approach with categorical numerical variables show higher variability. For the *hurdle_cat_GLM* or *hurdle_cat_XG* variants, there is a more pronounced overestimation of atypical bonuses. The better predictor of bonus damage frequency 0 was the *hurdle_cat_XG* model, which estimated 0.214, while *hurdle_cat_GLM* also underestimated with a prediction of 0.152. As a result, *hurdle_cat_XG* most closely resembled GLM in its predictions, while *hurdle_cat_GLM*, often underestimated.

IV. Discussion

The results show that traditional GLMs remain competitive with more complex machine learning models when expert knowledge is embedded in data preprocessing. In particular, GLMs demonstrated strong performance in low-frequency segments, where neural networks (NNs) often struggled due to insufficient training data. GLMs benefit from coefficient regularization and data pooling, which enhance extrapolation in underrepresented segments and provide robustness against overfitting.

In contrast, neural networks captured nonlinear trends in better-populated regions of the feature space, but their predictions were unstable in sparse areas.

A key observation is the trade-off between modeling smooth relationships and preserving discontinuities. Models using raw numerical inputs, such as *NN_nocat* or hurdle models with continuous features, offered smooth approximations but failed to capture local structural effects—like the spike in claim frequency for policyholders aged 40–50 or bonus-specific discontinuities.

Hurdle models, particularly those with gradient boosting components (*Hurdle_cat_XG*), effectively addressed excess zeros by separating claim occurrence from frequency. The hurdle models generally provided lower average predicted frequencies, aligning more closely with the validation set and mitigating the tendency of other models to overestimate.

These findings reinforce the importance of domain-informed feature engineering. Categorical transformations allowed models to capture nonlinearities more effectively and to respond to behavioral thresholds commonly observed in actuarial data. At the same time, the gap in performance between classical and modern models suggests that complexity alone does not guarantee better results. Interpretability, especially in regulated environments, remains a crucial advantage of traditional models.

Future work may investigate ensemble approaches, interpretability tools for neural networks, or applications of explainable AI techniques to bridge the gap between predictive power and transparency in complex models.

V. Conclusion

This study compared classical and modern approaches for claim frequency estimation in MTPL insurance using real-world Czech data. The results show that Generalized Linear Models, supported by domain-informed preprocessing, remain strong contenders in predictive tasks, particularly in sparse or regulated segments.

While neural networks and hurdle models offer greater flexibility and potential in modeling complex patterns, they are more sensitive to data sparsity and less transparent. The experiments demonstrate that modeling success depends not only on algorithmic complexity but also on careful feature engineering and understanding of the domain.

Future research should explore hybrid or interpretable machine learning models that can combine the predictive power of modern methods with the robustness and clarity required in actuarial practice.

References

[1] M. Denuit, X. Marechal, S. Pitrebois, a J.-F. Walhin, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Chichester, West Sussex, England; Hoboken, NJ: Wiley-Interscience, 2007.

[2] P. McCullagh, *Generalized Linear Models*, 2nd ed. New York: Routledge, 2019.

[3] J. Mullahy, "Specification and testing of some modified count data models," *Journal of Econometrics*, vol. 33, no. 3, pp. 341–365, Dec. 1986.

[4] M. V. Wuthrich, "From Generalized Linear Models to Neural Networks, and Back," Social Science Research Network, Rochester, NY, Dec. 2019. [Online]. Available: https://papers.ssrn.com/abstract=3491790

[5] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2012.

[6] Dong-Young Lim, "A Neural Frequency-Severity Model and Its Application to Insurance Claims," [Online]. Available: https://paperswithcode.com/paper/a-neural-frequency-severity-model-and-its

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.

[8] J. A. Nelder and R. W. M. Wedderburn, "Generalized Linear Models," *J. Roy. Statist. Soc. A*, vol. 135, no. 3, pp. 370–384, 1972.

[9] M. V. Wuthrich and M. Merz, "Statistical Foundations of Actuarial Learning and its Applications," Social Science Research Network, Rochester, NY, Jun. 2022. [Online]. Available: https://papers.ssrn.com/abstract=3822407

[10] L. Štěpánek, P. Martinková. "Feasibility of computerized adaptive testing evaluated by Monte-Carlo and post-hoc simulations", Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, vol. 21, pp. 359–367, FedCSIS, Sep. 2020. Available: http://dx.doi.org/10.15439/2020F197.

[11] A. Agresti, *Categorical Data Analysis*, 3rd ed. Hoboken, NJ: Wiley, 2013.

[12] T. Dozat, "Incorporating Nesterov Momentum into Adam," in *Proc. 4th Int. Conf. Learn. Representations (ICLR)*, 2016

[13] J. C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," in *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA: MIT Press, 1999