Journal of Information Systems Engineering and Management
2025, 10(29s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Price Forecasting Using Financial Technology by (GAMLSS) Theory of NIST AI Risk Management Framework (AI RMF) for Sustainable Development Goals

Randa Mostafa[1], Mostafa Abd-Elghany[2], Amr Soliman[3]
[1] *Assistant Lecturer in Insurance & Actuarial Sciences Department, Faculty of Commerce, Cairo University, Egypt.*
[2] *Professor in Insurance & Actuarial Sciences Department, Faculty of Commerce, Cairo University, Egypt.*
[3] *Associate Professor in Insurance & Actuarial Sciences Department, Faculty of Commerce, Cairo University, Egypt.*

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Introduction: This study explores the relationships between sustainability rating, price forecasting rating, financial technology, and the NIST AI risk management framework (AI RMF) in the digital age.

Objectives: The purpose is to crystallize an advanced price unified AI theory considering risk classifications to reinforce AI-RMF that facilitates the comprehensive transformation reshaping the competitive InsurTech industry from traditional Environmental, Social, and Governance (ESG) to Sustainable Development Goals (SDGs) Agenda 2030. However, "Risk Measurement" is the biggest challenge of AI-RMF application, including risks related to third-party software, hardware, methodology, and data, which are not uniform at the international or insurer levels.

Methods: the paper introduces GAMLSS machine learning for price forecasting in insurance companies, and compares it with traditional models, LMM, GLMM, and GAMM. GAMLSS is a flexible general framework for fitting semiparametric univariate regression-type models allowing adjustments through parametric/non-parametric additive smoothing functions or linear/nonlinear functions. Additionally, the empirical analysis examines a subset of actual heavy-tailed data observed from 2019 to 2023 from a major Egyptian "Non-life Insurance Company". Finally, the programming language "gamlss" packages software was used in data science.

Results: As a technological and economic instantiation, this article provides recommendations for ministries, regulatory bodies, and insurance companies to use GAMLSS algorithm as a price-reliable methodology enhancing risk assessment methodologies, standardizing sustainability metrics, including data accuracy, refining disclosure formats, and evaluating the influence of (SDGs) reporting on stakeholders, considering value-based AI principles according to the Economic Cooperation & Development (OECD).

Conclusions: In conclusion, this study examined the sustainable relationships of price forecasting, digital technology, the NIST AI Risk Management Framework (AI RMF) and Sustainable Development Goals (SDGs) in 2030. The traditional Environmental, Social, and Governance (ESG) dimensions are extended to include economic and technological considerations. However, "Risk Measurement" represents the most critical challenge to unify the methodologies at the international level. So, this paper reviewed the development of risk-based pricing strategies: LMM, GLMM, GAMM models. And this study suggested GAMLSS machine learning algorithm as a price-reliable AI methodology in insurance companies. Subsequently, this article reached several results, as follows: Firstly, GAMLSS achieved the lowest value AIC and GDEV test (568098.8 and 567569.2, respectively), compared to LMM (694230.8 and 694198.8, respectively), GLMM (600055.1 and 600091.6, respectively), and GAMM (597063.2 and 597031.2, respectively) for price forecasting. Secondly, GAMLSS as a semi-parametric model represents *Box-Cox t (BCT)* distribution as the most accurate distribution compared to about 100 distributions within "gamlss" software packages instead of the classic exponential distributions such as GAMMA within GLMMs, and GAMMs for claim forecasting. Thirdly, GAMLSS introduces the *Cubic-splines* as the most accurate compared to the *P-splines* smoother algorithms and other gamlss algorithms for auto-insurance price. Finally, GAMLSS developed *RS algorithm* reduces the computational burden of the Maximum Likelihood Estimation (MLE) of the model, and does not require chronological ordering, and can avoid overfitting problems. Additionally, the article recommends the importance of focusing future research on applying the NIST AI Risk Management Framework.

Keywords: **GAMLSS theory; Machine learning algorithms; Financial technology; Semiparametric models; NIST AI risk management framework (AI RMF); Sustainable development goals; Automated price forecasting; Predictive ratemaking, Economic development, Cybersecurity.**

INTRODUCTION

Regarding 2030, The United Nations Sustainable Development Goals include 17 goals to address economic, environmental, social, and technological impacts for a sustainable future, the most important of which are: no poverty (SDG 1), good health and well-being (SDG 3), decent work and economic growth (SDG 8), and climate action (SDG 13) (United Nations, 2015). These are the effects that a comprehensive understanding of the multifaceted nature of sustainability requires (UNEP Finance Initiative, 2020). However, conventional international policy, Environmental, Social, and Governance (ESG) introduced according to the Millennium Development Goals (MDGs) policy as the first-ever global partnership strategy in 2000, adopted by the United Nations General Assembly by 189 countries with the support of 147 heads of state and the world's leading development institutions (such leaders of International Monetary Fund (IMF), and Organization for Economic Co-operation and Development (OECD)) (Woodbridge, 2015). Several challenges and shortcomings of ESG sustainability are represented as follows: Data quality & deficiency, scarcity of AI specializations and multidisciplinary expertise, lack of standardized metrics, regulatory uncertainty, climate liability risks, not aligning with market expectations, and absence of a more comprehensive strategy for economic and technological development leading the occurrence of the global economic crisis (Miao, X. 2024; Cruz, C. A., & Matos, F. 2023). Because *SDGs* and *AI* are two sides of the same coin (Misra et al., 2024), countries have competed to develop a framework to address the risks of generative *AI* under SDGs, including Canada, Japan, the United Kingdom, France, Germany, Italy, and the United States (OECD, 2023). The NIST-AI Risk Management Framework (AI RMF) framework was introduced by the National Institute of Standards and Technology and presents a structured approach to facilitate the imperative transformation toward a broader SDGs agenda framework. The AI RMF-risk-based approach is an Engineered or Machine-Based System composed of four functions: GOVERN, MAP, MEASURE, and MANAGE risks (Thomas, 2024; Adams, 2017; Anand et al., 2023).

CHALLENGES TO IMPLEMENTATION & OBJECTIVES

Risk Measurement is the most important challenge for AI RMF, some risk measurement challenges include risks related to third-party software, hardware, and data. AI risks or failures that are not well determined or adequately understood are difficult to measure qualitatively or quantitatively. Third-party data or systems can accelerate research and development and facilitate technological transition. However, this complicates the risk measurement. Risk can emerge from third-party data, software, or hardware. The risk metrics/methodologies used by the organization developing the AI system may not align with the risk methodologies or metrics used by the organization deploying or operating the system. Additionally, the organization developing the AI system may not be transparent to the risk methodology/metrics it uses. Risk measurement and management can be complicated by how customers use or integrate third-party data or systems into AI products or services, particularly without sufficient internal governance structures and technical safeguards (Bobby, 2024; Bruce, 2023). The OECD has developed a framework for classifying AI lifecycle activities according to five key sociotechnical dimensions, slightly modified by NIST: APPLICATION CONTEXT, DATA and INPUT, AI MODEL, TASK, and OUTPUT, each with properties regarding AI policy and governance, including risk management (OECD, 2022), highlighting the importance of evaluation, verification, and validation (TEVV) test processes throughout an AI lifecycle and generalizing the operational context of an AI System. AI RMF functions require diverse perspectives, disciplines, professions, and experience. Regarding Insurance Risk-Based Pricing Strategies, the accurate pricing of insurance contracts is a major challenge for insurance companies in a competitive international market environment and in the presence of price-sensitive customers. Predominant limitations regarding the framework and assumptions on which the model was built failed common traditional actuarial methods in loss rate prediction in insurance companies. *Subsequently, there is an* *urgent need for a practical, computationally efficient, and theoretically reliable methodology as a unified AI MODEL for price forecasting that supports the quantitative application of the (AI RMF) to keep pace with the advancement of insurance science to serve humanity in cutting-edge technology according to (SDGs) in 2030.*

LITERATURE REVIEW

- ***Linear Regression Models (LRMs).***

Linear Regression Models (LRMs) and their derivatives (Mason et al., 2006), such as Classical Multiple Regression (CMR), Classical Normal Linear Regression Model (CNLRM), Boosting Linear Regression Models (BLM), Bayesian Linear Regression Models (BLRM), and Linear Mixed Model (LMM), include Specification Error resulting from the observations follow a normal distribution, which does not allow for the appropriate modeling of frequencies or binary or skewed data. Additionally, the analysis of the common variables effect on the average value of the response variable only does not consider the degree of dispersion from the average value. The most important is the assumption that the independent variables affecting the model equation are non-random and fixed, in contrast to the practical reality that the explanatory variables are random Stochastic Explanatory Variables (Maddala and Lahiri1992; Mason et al., 2006, Bermúdez et al., 2020, Qazvini 2019, Davoudi et al. 2018).

- ***Generalized Linear Models (GLMs).***

The predominant limitation is the failure of common traditional actuarial methods in price forecasting in insurance companies. Especially; Generalized Linear models (GLMs) (Nelder and Wedderburn 1972), and their derivatives such as Generalized Linear Mixed Model (GLMM), Vector Generalized Linear Model (VGLM), Double Generalized Linear Model (DGLM), Hierarchical Generalized Linear Model (HGLM), Boosting Generalized Linear Model (BGLM), and Bayesian Generalized Linear Mixed Model (BGLMM) (Goldburd et al., 2016; Denuit and Trufin 2019): firstly, GLMs parametric models do not include how to handle the non-linear additive effects of non-parametric smoothing functions or machine learning algorithms such as Loess Curves, Decision Trees, and Neural Networks. Second, the shape of the classical exponential family is fixed on the response variable. Third, GLM models do not fit a relatively large set of data, the inadequacy of longitudinal data, data with heavy tails, right-skewed data, or overdispersion, which require a more flexible skewness or kurtosis model. Finally, the difficulty of mathematically estimating and inferring spatial data because the sample size is large with the development of remote sensing technology and automated sensor networks (Bermúdez et al., 2020; Qazvini 2019; Davoudi et al., 2018; Hürlimann 2007; Guillén et al., 2018; Zhang and Miljkovic 2019; Omerašević and Selimović 2020; Jeong and Valdez 2020; Alemany et al., 2020).

- ***Generalized Additive Models (GAMs).***

Several limitations directed at Generalized Additive models (GAMs) (Hastie and Tibshirani 1990), and their derivatives such as Generalized Additive Mixed Model (GAMM), Vector Generalized Additive Model (VGAM), Generalized Geoadditive Model (GGAM), Dynamic Generalized Additive Model (DGAM), Generalized Geoadditive Mixed Model (GGAMM), Bayesian Generalized Additive Model (BGAM), Hierarchical Generalized Additive Model (HGAM), Double Generalized Additive Model (DGAM), Quantile Generalized Additive Model (QGAM), Boosting Generalized Additive Model (BGAM), and Bayesian Generalized Additive Mixed Model (BGAMM), Generalised Additive Index Models (GAIMs), and Generalized linear additive smooth structures (GLASS): First, the GAM framework only accommodates the linear exponential family of distributions, without considering heavy-tailed data, right-skewed data, and overdispersion suitable for modeling claims in insurance companies. Second, they are more likely to be overprocessed. Third, less interpretability and predictive accuracy compared to GLMs, such as performing estimation and inference on spatial data, providing measures of the non-linear effect of covariates, and assessing the impact of uncertainty, is challenging. Finally, the GAM package may include processing the nonlinear additive effects of non-parametric smoothing functions but lacks other more advanced machine learning algorithms, such as Loess Curves and Neural Networks, which are used to predict the independent variables in the grading scale (Wuthrich and Buser 2021; Boucher and Turcotte 2020; Joao 2019; Tingting 2018; Staudt and Wagner 2019; Lee 2020; Czado et al., 2015).

Therefore, the following hypothesis is proposed:

***H1**. There is a direct correlation between the high sustainability rating in risk management and the pricing forecast rating using AI, where the AI RMF can be applicable in this area.*

METHODOLOGY

GAMLSS machine-learning strategic framework:

The GAMLSS theory is a flexible general framework for fitting semi-parametric univariate regression-type models, where "parametric" means that the distribution of the response variable does not have to belong to the exponential family and includes highly skewed and kurtotic continuous and discrete distribution, and "semi" means that the systematic part is expanded to allow modeling not only the mean but also all the location, scale, and shape parameters of the response variable distribution as functions of explanatory variables or random effects, and allows adjustments through (Stasinopoulos and Rigby 2007; Rigby et al., 2005; Stasinopoulos and Rigby 2017).

- Parametric or non-parametric additive-smoothing functions.
- or linear/nonlinear functions.
- Supervised machine-learning algorithms.

GAMLSS Assumptions:

GAMLSS assumes that independent observations for $i = 1,\ldots n$ are conditionally independent given a set of covariates. The response variable *is* a general distribution with conditional density function $f(\mathcal{Y}_i|\boldsymbol{\theta^i})$, where the vector $\theta^i = (\theta_{i1}, \theta_{i2}, .., \theta_{iP})^T = (\mu_i, \emptyset_i, v_i, \tau_i)^T$ represents a vector of the four distribution parameters that represent the location ($\theta_{i1} = \mu_i$), scale ($\theta_{i2} = \emptyset_i$), skewness ($\theta_{i3} = v_i$), and kurtosis ($\theta_{i4} = \tau_i$). Each distribution parameter θ^i was modeled its additive predictor η_k for $k = 1,..,4$ and depends additively on the covariates, including possible smooth predictor effects. Let $g_k(\boldsymbol{\theta}_k)$ be the known monotonic link function relating the distribution parameters to the set of explanatory variables x_{kj} that links the regular complex $x_k\beta_k$ to the distribution parameter θ_{ki}, the parameter vector β_k, and the covariate design matrix x_k (Rigby et al., 2005). The configuration of distribution D $(\mu_i, \emptyset_i, v_i, \tau_i)$ is general and only implies that the distribution must be in a parametric configuration (Stasinopoulos et al., 2024).

GAMLSS Equations.

$$g_1(\mu_i) = \eta_1 = x_1\beta_1 + \sum_{j=1}^{J_1} Z_{j1}\gamma_{j1}$$

$$g_2(\emptyset_i) = \eta_2 = x_2\beta_2 + \sum_{j=1}^{J_2} Z_{j2}\gamma_{j2}$$

$$g_3(v_i) = \eta_3 = x_3\beta_3 + \sum_{j=1}^{J_3} Z_{j3}\gamma_{j3}$$

$$g_4(\tau_i) = \eta_4 = x_4\beta_4 + \sum_{j=1}^{J_4} Z_{j4}\gamma_{j4} \tag{1}$$

RESULTS

In the beginning, the GAMLSS Methodology can be applied and managed practically for Insurance Pricing Forecast Machine Learning Projects. it was applied to a subset of heavy-tailed and over-dispersed actual data from automobile insurance policies –the most influence branch on the portfolio and the highest loss rates– observed over 4 years from 2019 till 2023 from a major Egyptian "Non-life Insurance Company". There were (n = 112398) observations that met our criteria. Ratemaking entails classifying policyholders into homogeneous categories. For this purpose, the basic hypothesis that there is a strong relationship for price forecasting between claim severities Y_i as the response variable and the priori rating variables with the highest predictive power for response variable at fault registered for each insured vehicle in the dataset. Furthermore, an exploratory analysis was carried out to accurately select the subset of explanatory variables with the highest predictive power for the response variable Y_i.

The available rating factors for price forecasting are summarized in Table 1.

x_n	Variables	Description
		Categorical Variables:
***VB* (x_1)**	**Vehicle Brand (Origin country)**	Four categories: "Germany" (C1), "Korea" (C2), "Japan" (C3), and "France" (C4).
***VT* (x_2)**	**Vehicle Type**	Three categories: Bus (C1), Private (C2), and Tanker (C3).
***VU* (x_3)**	**Vehicle Usage (Driver Activity Type)**	Two categories: Individual Sector (C1), and Corporate Sector (C2).
***VF* (x_4)**	**Vehicle Fuel**	Three categories: Electric (C1), Petrol (C2), and Diesel (C3).
***GA* (x_5)**	**Geographic Area**	Seven categories: Alex (C1), Channel (C2), Tribal (C3), Delta (C4), External Branches (C5), and Cairo (C6).
		Continuous Variables:
***VA* (x_6)**	**Vehicle Age**	Three categories: age "between 0 to 7 years" (C1), "between 7 to 14 years" (C2) and "greater than 14 years" (C3).
***HP* (x_7)**	**Vehicle Horsepower**	Three categories: HP of "0-1400 cc" (C1), "1400–1800 cc" (C2), and "greater than 1800 cc" (C3).
***VV* (x_8)**	**Vehicle Value**	Three categories: VV of "0-200000 EG" (C1), "200000–400000 EG" (C2), and "greater than 400000 EG" (C3).
***CD* (x_9)**	**Contract Duration (Exposure)**	Three categories: CD of "0-3 months" (C1), "3–7 months" (C2), and "7–12 months" (C3).
***BM* (x_{10})**	**Bonus Malus (Posteriori Experience)**	Three categories: BM of "40%-60%" (C1), "60%-80%" (C2), and "greater than 80%" (C3) applied on the gross premium (Staudt, Y., & Wagner, J., 2019).

Table 1: Description of the explaining variables for price forecasting.

The subject of this paper is purely economic: if explanatory variables affect the probability of being involved in an accident or the severity of a vehicle claim (and thus the company's economic losses), insurance companies may require different rates according to the size of different categories of explanatory variables. The descriptive statistics of the effects of the covariates for the rating factors are summarized in Table 2.

	y_1	*VB* (x_1)	*VT* (x_2)	*VU* (x_3)	*VF* (x_4)	*GA* (x_5)	*VA* (x_6)	*HP* (x_7)	*VV* (x_8)	*CD* (x_9)	*BM* (x_{10})
Min.:	**0**	**C1: 12768**	**C1: 7411**	**C1: 56583**	**C1: 20135**	**C1: 10795**	**2.10**	**1000**	**300**	**0.002732**	**50.00**
1st Qu.:	**0**	**C2: 21329**	**C2: 78346**	**C2: 55815**	**C2: 60122**	**C2: 8526**	**5.30**	**1500**	**120000**	**0.280000**	**50.00**
Median:	**0**	**C3: 66267**	**C3: 26641**		**C3: 32141**	**C3: 13421**	**7.60**	**1600**	**197000**	**0.670000**	**50.00**
Mean:	**3854**	**C4: 12034**				**C4: 9555**	**10.01**	**2372**	**250782**	**0.624504**	**59.41**
3rd Qu:	**798**					**C5: 5057**	**12.50**	**1800**	**285000**	**1.000000**	**64.00**
Max:	**987500**					**C6: 65044**	**63.90**	**15800**	**7200000**	**2.010000**	**230.00**

Table 2: Descriptive Statistics of Claim Severities y_1 - the Number or Size of Categories of the Categorical Explanatory Variables for the Database and Continuous Explanatory Variables in Database.

Algorithm	LMM		GLMM		GAMM		GAMLSS					
Parametric Part	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept	12594.7	6200.24	8.5626	0.5413***	8.56843	0.2202***	5.7e+03	(5.3e+01)***	-8.2e-01	(3.2e-02)***	-8.3e-01	(4.1e-02)***
VB C2	-3637.4	525.62***	-0.1214	0.0140***	-0.1058	0.0170***	-1.2e+02	(2.4e+01)***	5.6e-02	(1.5e-02)***	-2.5e-01	(2.1e-02)***
VB C3	-4149.2	462.35***	-0.1333	0.0124***	-0.1211	0.0150***	-1.0e+02	(2.1e+01)***	5.4e-02	(1.3e-02)***	-2.5e-01	(1.8e-02)***
VB C4	-1985.8	611.68**	-0.0607	0.0163***	-0.0448	0.0197*	-5.1e+01	(2.9e+01)·			-1.1e-01	(2.5e-02)***
VT C2	2459.7	712.25***	0.1405	0.0186***	0.12845	0.0335***			5.6e-02	(2.1e-02)**	2.6e-01	(2.6e-02)***
VT C3	3818.4	604.70***	0.0680	0.0156***	0.02112	0.0289	-3.1e+02	(3.1e+01)***	9.4e-02	(1.7e-02)***	-1.1e-01	(2.3e-02)***
VU C2	-2456.3	342.50***	-0.0781	0.0090***	-0.0690	0.0110***	-7.4e+01	(1.4e+01)***	3.9e-02	(1.0e-02)***	-1.2e-01	(2.0e-02)***
VF C2	1059.1	425.50*	0.0480	0.0113***	0.03788	0.0138**	1.2e+02	(2.4e+01)***			1.8e-01	(1.3e-02)***
VF C3	1450.0	516.09**	-0.0013	0.0138	-0.0093	0.0168	-1.1e+02	(3.1e+01)***			-2.4e-01	(1.4e-02)***
VA	-3706.9	288.03***	-0.0761	0.0075***	-0.0762	0.0098***						
HP	401.6	247.546	0.0272	0.0066***	0.02723	0.0084**						
VV	1533.0	146.06***	0.0508	0.0044***	0.05088	0.0049***						
CD	459.4	281.33	0.0281	0.0074***	0.02806	0.0095**						
BM	335.8	523.76	0.0520	0.0139***	0.05207	0.0178**						
Random Effects												
σ^2_{GA}	156794982		1.046									
$\sigma_{Residual}$	23853		0.8128									
Smooth Terms					*edf*	*Ref.df*						
s(GA)					3.31	4***						
P-Splines												
pb (VA)							-6.7e+00	(1.0e+00)***	8.2e-03	(6.8e-04)***	-1.0e-02	(8.6e-04)***
pb (HP)									3.7e-06	(2.1e-06)·		
pb (VV)							2.2e-04	(2.4e-05)***	-7.2e-08	(1.1e-08)***	2.6e-07	(1.6e-08)***
pb (CD)							4.2e+01	(1.8e+01)*			1.3e-01	(1.5e-02)***
pb (BM)							8.2e-01	(4.0e-01)*	6.5e-04	(2.5e-04)**	3.8e-03	(3.4e-04)***
Cubic-Splines												
Cs (VA)							-5.6e+00	(9.8e-01)***	8.6e-03	(6.8e-04)***	-1.1e-02	(1.9e-02)***
Cs (HP)							-2.0e-02	(3.6e-03)***	9.1e-06	(2.1e-06)***	9.6e-05	(8.2e-04)***
Cs (VV)							2.9e-04	(2.4e-05)***	-7.5e-08	(1.1e-08)***	1.6e-07	(1.6e-08)***
Cs (CD)							6.4e+01	(1.7e+01)***			6.9e-02	(1.4e-02)***
Cs (BM)									7.9e-04	(2.5e-04)**	7.4e-03	(3.0e-04)***
GDEV	694198.8		600091.6		597031.2		567569.2					
AIC	694230.8		600055.1		597063.2		568098.8					

Table 3: Comparison Between Coefficients Estimation of Four Machine Learning Algorithms for Price Forecasting: ***LMM, GLMM, GAMM***, and ***GAMLSS.*** In GLMM and GAMM "GAMMA Distribution" was fitted. In GAMM without smooth terms, a random effect smoother is applied to the predictor *GA*. In GAMLSS (BCT, μ_i, $\emptyset_i$, v_i, τ_i) modeled with two algorithms: *Cubic splines* compared to penalized *P-splines* smooth functions for nonparametric predictors, penalized random effect smoother is applied to the categorical variable *GA* when necessary. Ineffective estimates have been removed. All variables Using (RS) Algorithm. the corresponding link function is displayed. Additionally, *C1* is a categorical dummy variable, Goodness-of-Fit Statistics (*GDEV, AIC*), and *Stands for 5% significance for fitting claim severities.

Source: Author`s contributions.

DISCUSSION

Table 3 summarizes a comparison between Classic LMM, GLMM, and GAMM models and advanced GAMLSS machine learning theory for price forecasting using the Programming language “gamlss” R packages software. Firstly, there is a growing use of LMM in data analysis where random effects are considered in addition to fixed effects to better estimate the response variable distribution. Secondly, in GAMM (with no smoother), it could be used a random effect smoother for GA which appears significantly larger than in the GLMM model. This illustrates the sensitivity of the results to the algorithm. One can detect variations in the estimates, standard errors, effective degrees of freedom (edf), and p-values; in particular, the estimate for GA is significantly larger than in the GLMM model. This illustrates the sensitivity of the results to the algorithm. Thirdly, in terms of AIC, i.e., taking into account the number of degrees of freedom used by the models, the (BCT, μ_i, $\emptyset_i$, and v_i) and (BCT, $\mu_i, \emptyset_i, v_i$, and τ_i) models were the best (AIC = 631981.7), with only a slight difference between them, but are followed very closely by the Generalized Beta Type 2 (GB2) distribution (AIC = 631701.6) via the maximum likelihood (ML) approach. Finally, in GAMLSS degrees of freedom *df* equals (141.5796 and 36.40257 for *Cubic splines* and *P-splines* algorithms, respectively), and the higher the degrees of freedom, the better the smooth algorithms will improve the semi-parametric model.

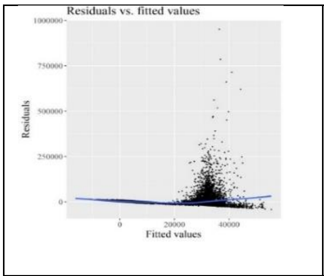

FIGURE 1 | Diagnostics for ***LMM Model***: deviance residuals vs. fitted values and Q-Q plot of these residuals.

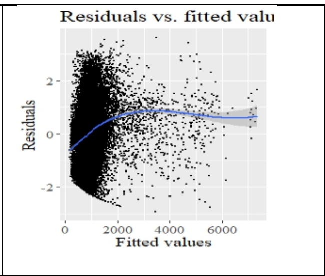

FIGURE 2 | Diagnostics for ***GLMM Model***: deviance residuals vs. fitted values and Q-Q plot of these residuals.

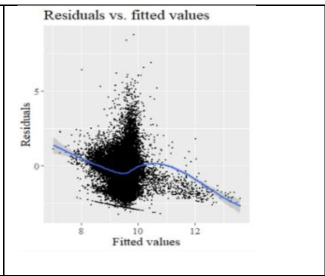

FIGURE 3 | Diagnostics for ***GAMM Model***: deviance residuals vs. fitted values and Q-Q plot of these residuals.

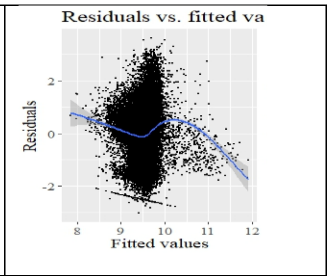

FIGURE 4 | Diagnostics for ***BCT Cubic-splines GAMLSS Algorithm***: deviance residuals vs. fitted values and Q-Q plot.

Source: Author`s contributions.

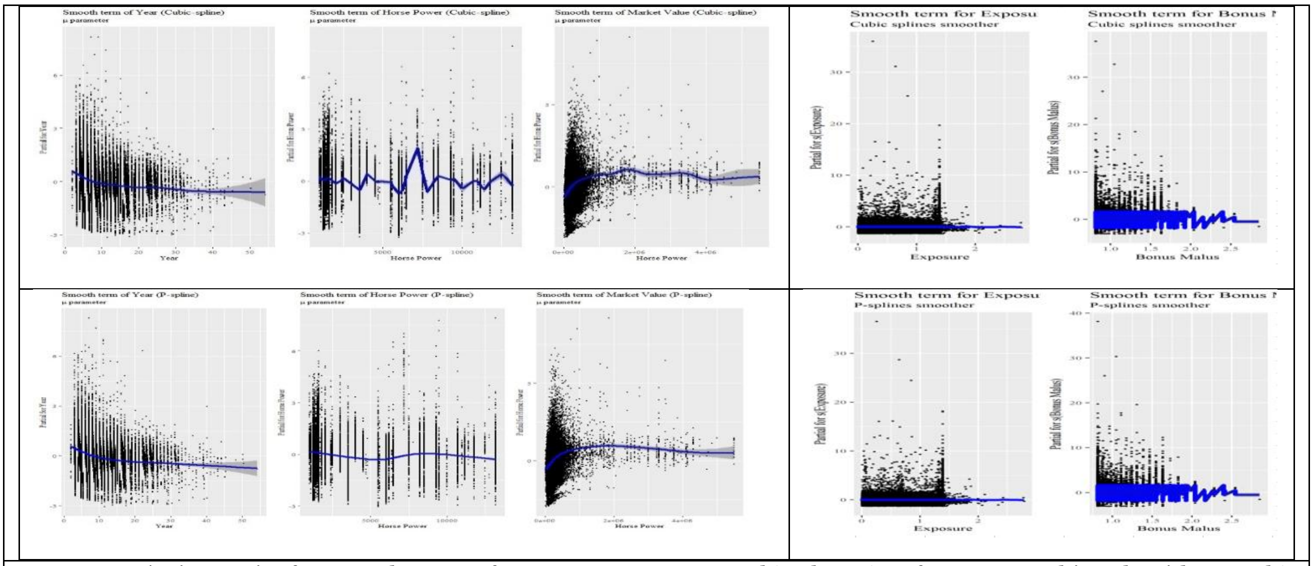

FIGURE 5 | Diagnostics for smooth terms of *Box-Cox t GAMLSS machine learning*: for two smoothing algorithms: *cubic splines, and P-splines*. Additive terms for *VA, HP, CD, VV, and BM* for the μ_i components of the (BCT, $\mu_i, \emptyset_i, v_i$, and τ_i) model reported in Table 7. A *cubic spline and P-splines* smooth function is constant to *CD and BM* for μ_i.

Source: Author`s contributions.

The various graphs illustrate that heteroscedasticity in GAMLSS is more moderate and more efficiently captures the stylized characteristics of data for modeling heavy-tailed claim severities compared to LMM, GLMM, and GAMM. Additionally, the *BCT (Cubic-splines GAMLSS)* algorithm achieved more efficient assumptions of the actual claims amount distribution compared to the *BCT (P-splines GAMLSS)* Algorithm. *P-splines* smooth algorithm results in relationships displays little “wiggliness” compared to *Cubic-splines*. *Cubic splines* lead to high degrees of nonlinearity, especially for (HP, and VV), which is reflected by the larger values of *df*, lower standard errors, higher p-values, and hence lower values in (AIC = 568098.8) comparison to *P-splines* smooth algorithm. The effect of smooth terms on two variables CD and BM for μ_i remains constant in both cases, *Cubic-splines* and *P-splines*.

ACKNOWLEDGMENT

My profound gratitude goes to my supervisors, from whom the study was extracted from my Ph.D. thesis registered in 2020/2021 under the research program academic Ph.D in the Insurance & Actuarial Sciences Department, Faculty of Commerce at Cairo University, Egypt.

REFRENCES

[1] Adams, C. A. (2017). The Sustainable Development Goals, integrated thinking, and the integrated report. Integrated Reporting.

[2] Alemany, R., Bolancé, C., Rodrigo, R., & Vernic, R. (2020). Bivariate Mixed Poisson and Normal Generalised Linear Models with Sarmanov Dependence-An Application to Model Claim Frequency and Optimal Transformed Average Severity. *Mathematics*, *9*(1), p.73.

[3] Anand et al., (2023). Sustainable Development Goals in Northeast India: Challenges and Achievements. *Springer*.

[4] Bermúdez, L., Karlis, D., & Morillo, I. (2020). Modelling unobserved heterogeneity in claim counts using finite mixture models. *Risks*, *8*(1).

[5] Bobby, K. J., (2024), AI-RMF A Practical Guide for NIST AI Risk Management Framework, *Packt Publishing; 1st edition*, *New York*.

[6] Bruce, B. (2023). RMF Security Control Assessor: NIST 800-53A Security Control Assessment Guide (NIST 800 Cybersecurity), *Packt Publishing; 1st edition*, *New York*.

[7] Boucher, J. P., & Turcotte, R. (2020). A longitudinal analysis of the impact of distance driven on the probability of car accidents. *Risks*, *8*(3), p.91.

[8] Cruz, C. A., & Matos, F. (2023). ESG Maturity: A Software Framework for the Challenges of ESG Data in Investment. Sustainability, 15(3), 2610,

[9] Czado, C., Pfettner, J., Gschlößl, S., & Schiller, F. (2015). Nonnested model comparison of GLM and GAM count regression models for life insurance data. *ASTIN Bulletin*.

[10] Davoudi Kakhki, F., Freeman, S. A., & Mosher, G. A. (2018). Analyzing large workers' compensation claims using generalized linear models and Monte Carlo simulation. *Safety, 4(4)*.

[11] Denuit, M., & Trufin, J. (2019). Effective statistical learning methods for actuaries.

[12] Goldburd, M., Khare, A., Tevet, D., & Guller, D. (2016). Generalized linear models for insurance rating. *Casualty Actuarial Society, CAS Monographs Series*, *5*, 77.

[13] Guillén, M., Nielsen, J. P., Ayuso, M., & Pérez Marín, A. M. (2018). Exposure to risk increases the excess of zero accident claims frequency in automobile insurance. *IREA–Working Papers, IR18/10*.

[14] Hastie, T., & Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, pp.1005-1016.

[15] Hürlimann, W. (2007). Benchmark rates for XL reinsurance revisited: model comparison for the Swiss MTPL market. *Belgian Actuarial Bulletin, 7(1)*.

[16] Jeong, H., & Valdez, E. A. (2020). Predictive compound risk models with dependence. *Insurance: Mathematics and Economics*, 94.

[17] Joao Pedro. (2019). Bayesian Generalized Additive Models for Car Insurance Data. *Master Thesis of Science Degree in Mathematics and Applications*.

[18] Lee, S. C. (2020). Delta boosting implementation of negative binomial regression in actuarial pricing. *Risks, 8(1), 19*.

[19] Maddala, G. S., & Lahiri, K. (1992). Introduction to econometrics (Vol. 2, p. 525). *New York: Macmillan*.

[20] Mason, R.L. et al, (2006). Regression Analysis and Problems of Multicollinearity, *Commun, Statist*.

[21] Miao, X. (2024). Challenges and Responses to ESG Risk Management. International Conference on Finance and Economics (Vol. 6, No. 1).

[22] Misra, S., Siakas, K., & Lampropoulos, G. (Eds.). (2024). Artificial Intelligence of Things for Achieving Sustainable Development Goals (No. 192). Springer.

[23] OECD. (2022). OECD Framework for the Classification of AI systems. *OECD Digital Economy Papers, No. 323*.

[24] OECD (2023), G7 HIROSHIMA PROCESS ON GENERATIVE ARTIFICIAL INTELLIGENCE (AI**):** TOWARDS A G7 COMMON UNDERSTANDING ON GENERATIVE AI, *OECD Publishing*.

[25] Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3),* pp.370-384.

[26] Omerašević, A., & Selimović, J. (2020). Risk factors selection with data mining methods for insurance premium ratemaking·. *Zbornik Radova Ekonomski Fakultet u Rijeka, 38(2).*

[27] Rigby et al., (2005). Generalized Additive Models for Location, Scale, and Shape. *Applied Statistics*, 54, pp.507–554.

[28] Rigby, R.A., Stasinopoulos, D.M. and Voudouris, V., (2013). Discussion: A comparison of GAMLSS with quantile regression. *Statistical Modelling*, *13*(4), pp.335-348.

[29] Stasinopoulos et al., (2024). Generalized Additive Models for Location, Scale and Shape: A Distributional Regression Approach, with Applications (Vol. 56). *Cambridge University Press*.

[30] Stasinopoulos et al., (2017). Flexible regression and smoothing: using GAMLSS in R. *CRC Press*.

[31] Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*.

[32] Staudt, Y., & Wagner, J. (2019). Comparison of Machine Learning and Traditional Severity-Frequency Regression Models for Car Insurance Pricing. Working Paper. *Lausanne: University of Lausanne*.

[33] Thomas, M. (2024). Unveiling the NIST Risk Management Framework (RMF): A practical guide to implementing RMF and managing risks in your organization, *Packt Publishing; 1st edition, New York*.

[34] Tingting, C. (2018). Generalized Additive Models for Dependent Frequency and Severity of Insurance Claims. *Doctoral dissertation, University of Guelph*.

[35] UNEP Finance Initiative. (2020). Managing environmental. social and governance risks in non-life insurance business. Technical Report. Principles for Sustainable Insurance.

[36] United Nations (2015), "Transforming our world: the 2030 agenda for sustainable development".

[37] Woodbridge, M. (2015). From MDGs to SDGs: What Are the Sustainable Development Goals. Urban Issues, 1(1), 4.

[38] Wuthrich, M. V., & Buser, C. (2021). Data analytics for non-life insurance pricing. *Swiss Finance Institute Research Paper, (68).*

[39] Zhang, J., & Miljkovic, T. (2019). Ratemaking for a New Territory: Enhancing GLM Pricing Model with a Bayesian Analysis.