Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110104

Proposed Models for Comprehensive Automobile Insurance Ratemaking in Egypt with Parametric and Semi-Parametric Regression: A case study

Eslam Seyam and Hussien Elsalmouny*

Statistics and Insurance Department, Faculty of Commerce, Zagazig University, 44519, Egypt

Received: 9 Oct. 2020, Revised: 13 Nov. 2020, Accepted: 30 Nov. 2020
Published online: 1 Jan. 2022

Abstract: The research proposes alternative tariff systems to estimate the pure premium for The Misr Insurance Company –the biggest insurance company in Egypt, which has 46.39% of the automobile insurance market share– for comprehensive automobile portfolio. The proposed tariff systems construct insurance rate for each risk class according to the risk factors that might affect the loss instead of a fixed rate that is applied by the company. Three different statistical models are used: Generalized Linear Model (GLM), Generalized Linear Mixed Model (GLMM) and Generalized Additive Model (GAM) using Gamma and Poisson distributions. The data consists of 576,381 cases during the years 2013, 2014, 2015 and 2016. Every case represents an insurance contract.The research found that GLMM is the most convenient model for ratemaking for Misr company because it has the lowest value of Akaike's Information Criterion (AIC) and takes into consideration the nature of the most insurance data that contains repeated measures.

Keywords: Ratemaking, Comprehensive Automobile Insurance, Generalized Linear Model (GLM), Generalized Linear Mixed Model (GLMM), Generalized Additive Model (GAM) and Semi-Parametric Regression.

1 Introduction

The Misr Insurance Company in its automobile line sets a fixed rate for all insureds regardless of the risk they represent. The rate can be changed for specific groups but depends on subjective decisions. This method has two drawbacks: The first is that it introduces adverse selection risk. For example, suppose company A sets constant rate £y for every risk unit and the competitive company B sets two different prices £$(y-t)$ for low risk units and pounds$(y+t)$, $(t>0)$, for high risk units. In this case, high risk units will move gradually from company B to company A, so this selection is against company A's interest. The second drawback is that the risk burden is not distributed fairly among the policyholders.

Equitable distribution of the risk burden must be structured using advanced statistical methods to divide the portfolio into homogenous classes, not just set up a fixed price for all the insured regardless of their amount of risk. Thus, each insured belonging to a particular class can pay the pure premium proportional to the risk degree of this class.

The research objective is to find proposed alternative statistical models that give Misr Insurance Company the opportunity to choose an equitable tariff system (the pure premium) and avoid the drawbacks related to the current, simple ratemaking plan used by the company. The research uses three statistical models: Generalized Linear Model (GLM), Generalized Linear Mixed Model (GLMM) and Generalized Additive Model (GAM). Every model has its merits and drawbacks. The company has to determine the model most compatible with it according to how much each model is successful in representing the sample data using Akaike's Information Criterion (AIC). As well as, considering the characteristics of the Egyptian market, and the competitors' reaction to these models when Misr Insurance Company apply one of these models.

There are many beneficiaries of our research. On the one hand, Misr Insurance Company most likely will have a bigger market share and consequently more profit. The main reason for that is twofold. First, the high risk insured units will find that the competitor's rate is cheaper, so it is expected for those unprofitable insureds to terminate their contract with the company and move to competitors, and this will decrease the adverse selection risk. Second, the company will be more

* Corresponding author e-mail: iaseam@commerce.zu.edu.eg

© 2022 NSP
Natural Sciences Publishing Cor.

attractive for the low risk insured units than the other competitors because they will get a cheaper rate to reflect the actual risk they have. On the other hand, for the insured units, the proposed tariff system will distribute the insurance cost among the insured units equitably. In addition, the proposed models might attract low-income people that have not considered insurance service before, and that will protect more people under the insurance umbrella from unexpected risks.

The remainder of this paper is organized as follows. In section 2, we illustrate a brief literature of ratemaking models. The research methodology is introduced in section 3, and we discuss many parametric and semi-parametric models such as GLM, GLMM and GAM. Then, section 4 encompasses data used and results. Lastly, section 5 contains the conclusion and recommendations.

2 Literature

Insurance researchers have used many approaches for ratemaking process. Some of them use credibility models using Log-normal/normal distribution assumption for loss data, see, e.g., [1] and [2]. Others used the Empirical Bayes credibility theory models, see, e.g.,[3] and [4].

Other methods allow researchers to classify the insureds into two classes or more according to specific risk factors such as discriminant analysis, see, e.g., [5], and fuzzy set theory, see, e.g., [6] and[7].

Simple regression models have been used also to set different rates for classes of insureds using many independent rating factors and variables, see, e.g., [8]. All the previous methods assume a linear relationship between the independent variable and dependent variables. However, this assumption is not appropriate to exhibit an actual relationship existing between the rating variables and the expected loss as a dependent variable. One of the methods that relaxes this assumption is Generalized Linear Models.

Over the last decade, Generalized Linear Models (GLM) has been a common statistical tool for a prior classification, see, e.g., [9], [10], [11], [12], [13], [14], [15] and [16]. The merits of this model are twofold. First, the random deviations from the mean may have a non-normal distribution, see, e.g., [17]. Second, the transformed mean of the dependent variable may be a linear function of the explanatory variables, which is determined by the link function.

However, this model has two drawbacks. First, it assumes that the random variables are independent, which may not be fulfilled in the case of longitudinal, repeated measurements and spatial data. Second, the model is not convenient for unknown nonlinear effects of independent variables because, in the GLM, the independent variables appear only linearly. Generalized Linear Mixed Model (GLMM)—a special case of hierarchical models—is used to recover the first shortage and superimpose credibility on a GLM setting, see, e.g., [18], [19], [20] and [16]. However, GLMM still models all the independent variables linearly. On the other hand, Generalized Additive Model (GAM) overcomes the linearity assumptions inherent to GLM, but GAM is not convenient to longitudinal data.

Another ratemaking statistical tool is the spatial model that gives us the chance to analyze the risk variation between different districts where data contain spatial dependence. Spatial dependence appears when variables observed in areas close to each other are related. Ignoring such spatial dependence patterns may cause overdispersion and erroneous conclusions. Unfortunately, the spatial models are not totally consistent because they depend on the Bayesian Approach to evaluate the risk associated with each district but use a frequentist approach to estimate the effect of the other risk factors, see, e.g., [18].

The Bayesian GAM model can tackle the previous inconsistency problem in spatial models by depending fully on a Bayesian approach to estimate simultaneously nonlinear effects of the continuous variables, factors, spatial effects, and interactions between factors. See, e.g., [21], [22] and [23] which considers a Bayesian implementation of the GAM.

Generalized Additive Model for location, scale and shape (GAMLSS) has recently been used in ratemaking, see, e.g., [24], where the exponential family assumption is relaxed and replaced by a very general distribution family assumption. Moreover, the model allows the mean, location and parameters of the conditional distribution of the dependent variable to be modeled as parametric or nonparametric functions for the fixed and random effects terms.

3 Research Methodology

In this section, we illustrate a brief background of the used models: Generalized Linear Models (GLM),Generalized Linear Mixed Models (GLMM) and Generalized Additive Models (GAM).

© 2022 NSP
Natural Sciences Publishing Cor.

3.1 Generalized Linear Models

The Generalized Linear Models (GLM) that will be used here to model the effect of p explanatory variables $x' := (x_1, x_2, \ldots, x_p)$ on the dependent variable Y given by the following system of equations

$$Y = E(Y) + \varepsilon, \quad Y = \beta_0 + x'\beta + \varepsilon, \quad Y = \eta + \varepsilon, \tag{1}$$

where $\beta := (\beta_1, \cdots, \beta_p)'$ is a vector of unknown real valued parameters.

In this section we assume the availability of a random sample of N units following the above model. Let $x_i := (x_{i1}, \cdots, x_{ip})'$ and Y_i denote the values of the ith unit explanatory variable vector and response, respectively. Let X denote the $N \times p$ matrix whose ith row consists of $(1, x_i')$, $1 \le i \le p$, $\mathbf{Y} := (Y_1, \cdots, Y_n)'$ and $\beta := (\beta_0, \beta_1, \cdots, \beta_p)$. Therefore, the model that describes the given data can be written as:

$$\mathbf{Y} = E[\mathbf{Y}] + \varepsilon, \qquad \mathbf{Y} = X\beta + \varepsilon, \quad \mathbf{Y} = \eta + \varepsilon, \qquad \eta = X\beta. \tag{2}$$

The Generalized Linear Models consists of the three components: random, systematic components and link function. The characteristics of each determine a unique model.

GLM components:

–**Random components:** The response variables $Y_i, 1 \le i \le N$ are independent r.v.'s and follow one of the exponential family distributions, e.g., normal, exponential, gamma, Bernoulli, Poisson, Tweedie. We write this as:

$$Y_i \sim \text{Exponential}(\mu_i, \phi), \qquad 1 \le i \le N. \tag{3}$$

$$\mu_i = E(Y_i) = b'(\theta_i), \quad V(Y_i) = b''(\theta_i).a(\phi), \tag{4}$$

where θ_i is natural or canonical parameter and ϕ is scale or dispersion parameter. Their values can be determined as illustrated by the following table:

Table 1: Canonical and dispersion parameters.

Distribution	Notation	Canonical parameter (θ)	Dispersion parameter (ϕ)
Normal	$N(\mu, \sigma^2)$	μ	σ^2
Poisson	P(μ)	ln μ	1
Gamma	G(μ,v)	$1/\mu$	v^{-1}
Binomial (m: trials)	B(μ, π)/m	$\ln(\mu/(1-\mu))$	$1/m$
Inverse Gaussian	IG(μ, σ^2)	$1/\mu^2$	σ^2

–**Systematic components:** The combination of $(p+1)$ β parameters to obtain the model prediction.

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \tag{5}$$

–**Link function:** Determines the relationship between the random components $E(Y_i) = \mu_i$ and the systematic components.

$$\text{g}(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \tag{6}$$

The link function g can take many forms. One of them is the ln link function. If g = ln, then the previous equation becomes:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \tag{7}$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \tag{8}$$

This link function transforms the additive effect of independent variables to multiplicative effect. This actually is more suitable for ratemaking in insurance companies because, unlike the additive models, it cannot produce negative prediction. Also, the multiplicative the effect of GLM makes effect of any independent variable nonlinear. Therefore, it is

© 2022 NSP
Natural Sciences Publishing Cor.

a more convenient and reasonable method for insurance ratemaking.
The exponential family probability function can be defined as:

$$f(Y_i;\theta,\phi) = exp\left(\frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i,\phi)\right) \tag{9}$$

Different choices of $a(\phi), b(\theta_i)$ and $c(Y_i,\phi)$ determine different types of distributions belonging to the exponential family as seen in the following table.

Table 2: Membership parameters of the exponential family.

Distribution	$a(\phi)$	$b(\theta_i)$	$c(Y_i,\phi)$
Normal	ϕ/w	$\theta_i^2/2$	$-\frac{1}{2}\left(\frac{wY_i^2}{\phi+ln\left(\frac{2\pi\phi}{w}\right)}\right)$
Poisson	ϕ/w	e^{θ_i}	$-lnY_i!$
Gamma	ϕ/w	$-ln(-\theta_i)$	$\frac{w}{\phi}ln\left(\frac{wY_i}{\phi}\right) - lny_i - ln(\Gamma(\frac{w}{\phi}))$
Binomial (m: trials)	ϕ/w	$m.ln(1+e^{\theta_i})$	$ln\binom{m}{Y_i}$
Inverse Gaussian	ϕ/w	$-\sqrt{-2\theta_i}$	$-\frac{1}{2}\left[ln(\frac{2\pi\phi Y_i^3}{w}) + \frac{w}{\phi Y_i}\right]$

In (Table 2), w represents constant prior weight. It equals one if we model claim count and equals number of claims if we model claim severity.
The log likelihood function of the exponential family distribution can be written as:

$$l = \sum_i \left(\frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i,\phi)\right). \tag{10}$$

Then, maximizing the log likelihood function by taking the derivative of l with respect to $\beta_j; j = 1,2,..,p$ and setting the result equals zero to get the estimated parameters.

3.2 Generalized Linear Mixed Models

Even though GLM is a common statistical tool having many merits and is widely used, it has two drawbacks. The first is that it assumes that variables are independent random variables. This assumption cannot be fulfilled in insurance ratemaking data because usually many years are used for rating (longitudinal data), so there is an autocorrelation over time in model variables. The two models can be used to tackle this issue. One of them is Generalized Estimating Equations GEE (marginal model). The other one is the Generalized Linear Mixed Model GLMM (conditional model).

The second drawback is that GLM gives full credibility to the data. That means the coefficient estimation of the model is trying to represent and describe the actual data perfectly in every level of independent factors regardless of how much this data is small and sparse. GLMM tackles this GLM drawback as it regards the coefficients of random effects as random variables, not fixed values. The more data is sparse and small, the less the model analyses will depend on that data in estimation for every level of random effect variables. Actually, GLMM is regarded as an alternative approach for introducing credibility theory in insurance.

In this section, GLMM will be introduced to eschew the drawbacks of GLM. Let Y be the response variable with conditional distribution given the exponencial random effects. Let $x_i; i = 1,\ldots,p$ be a set of p explanatory variables representing the fixed effects, and let $u_k; k = 1,\ldots,q$ be a set of q random effects, then The Generalized Linear Models of interest here written as:

$$\eta_{ij} = g(E(Y_i \mid u_1,\ldots u_q)) = \beta_0 + \sum_{i=1}^{q}\beta_i x_{ij} + \sum_{k=1}^{q} u_k z_{kj}, j = 1,\ldots,n, \tag{11}$$

© 2022 NSP
Natural Sciences Publishing Cor.

where β_0 is the intercept, β_i is the i^{th} fixed effect coefficents, x_{ij} is the i^{th} fixed effect independant variables for the j^{th} observation, u_k is the k^{th} random effect coefficents, z_{kj} is the k^{th} random effect variables for the j^{th} observation, η_{ij} is the expected value for the i^{th} variable and j^{th} observation and g is the link function. The model is written in matrix form as

$$\boldsymbol{\eta} = g(E(Y \mid U)) = X\boldsymbol{\beta} + ZU, \tag{12}$$

where Y is $n \times 1$ vector of the dependant variable, X is $n \times (p+1)$ design matrix of rank k representing the fixed effect variables, $\boldsymbol{\beta}$ is $(p+1) \times 1$ vector of fixed effect coefficients, Z is $n \times q$ design matrix representing the random effect variables, U is $q \times 1$ vector of random effect coefficients with multivariate normal distribution. The observations vector Y is obtained by adding residuals vector $\boldsymbol{\varepsilon}$:

$$Y = \boldsymbol{\eta} + \boldsymbol{\varepsilon} = X\boldsymbol{\beta} + ZU + \boldsymbol{\varepsilon}, \qquad Y \mid U \sim (g(E(Y \mid U), \boldsymbol{\Sigma}), \tag{13}$$

$$U \sim MVN(0, G) \tag{14}$$

$\boldsymbol{\Sigma}$ and G is the coveriance matrix of the conditional distribution of $Y \mid U$ and the random effect coefficients respectively.

The estimation of the maximum likelihood in GLMM is difficult due to the fact that it requires high dimentional integration. Therefore, there is no closed form of the maximum likelihood function, and the approximation methods are required. In the Bayesian framework, there are Markov Chain Monte Carlo (MCMC) methods and the hybrid approach. In the frequentist framework, there are three approximation methods: Laplace, Penalized Quasi Likelihood (PQL) and Adaptive Guassian Quadrature (AGQ). [25] finds that PQL gives biased results and recommends using Laplace and AGQ. The Laplace and AGQ methods approximation procedures use intensive computations and sometimes become abstruse and arduous to solve, such as our case. The R software got stuck in calculations with these two methods. In this research, we shall use an approximation of the Laplace method using the penalized iteratively reweighted least squares (PIRLS) algorithm, see, e.g., [26].

3.3 Generalized Additive Models

The main advantage of Generalized Additive Models (GAM) is its resilience to depict the nonlinear relationship between dependent and independent variables. However, the link function or some transformation of the variables can represent nonlinear relationship, like log link function, but this method introduces a specific relationship form and is not as flexible as GAM. The previous models (GLM, GLMM) assume that the relationship between dependent variable and independent variables is linear or semi linear (with log link function), as it represents the relationship in fixed coefficients. That means one unit increase in any independent variable leads to a specific fixed amount of increase in the dependent variable along with the whole domain of every independent variable, but it is not the realistic case in our data, as we will see later. GAM is GLM but in addition, we replace the parametric coefficients in GLM with a non-parametric one or add new nonparametric covariates.

Let Y be the response variable, $x_i; i = 1, \ldots, p$ be a set of p explanatory variables corresponding to the parametric effect and $u_k; k = 1, \ldots, q$ be a set of q variables representing non-parametric effects. The Generalized Additive Model is written as

$$\eta_j = g(E(Y_i)) = \beta_0 + \sum_{i=1}^{q} \beta_i x_{ij} + \sum_{k=1}^{q} f_k(u_{kj}), j = 1, \ldots, n, \tag{15}$$

where β_0 is the intercept, β_i is the i^{th} parametric effect coefficients, x_{ij} is the i^{th} parametric effect independent variables for j^{th} observation, $f_k(u_{kj})$ is k^{th} non-parametric effect functions for j^{th} observation, η_j is the expected value for the j^{th} observation and g is the link function. Every non-parametric function of $f_k(u_k)$ is a summation of many weighted base or smooth spline functions, for that the term additive came from.

There are two main estimating smoothing methods, REstricted Maximum Likelihood (REML) and Generalized Cross Validation (GCV). The REML smoothing method has less optimization problems, lower volatility of smoothing parameter, and it penalizes ovefit more efficiently than GCV, see, e.g., [27] and [28]. Therefore, we select the REML method for modeling. GAM gives us the chance to understand how the covariates affect both number of accidents and compensation variables without a restriction of linear relationship

© 2022 NSP
Natural Sciences Publishing Cor.

4 Numerical Application

4.1 DATA USED

The research will depend on the data set related to an Egyptian motor insurance portfolio observed during 2013, 2014, 2015 and 2016. The portfolio relates to Misr Insurance Company, the biggest insurance company in Egypt, which has 46.39% of the market share.

The data consists of 576,381 cases, and every case represents an insurance contract. Variables of the research can be illustrated by the following table:

Table 3: Research variables.

Categorical and Ordinal Variables	Number of Levels	Notes
District	9	The place of insurance contract
Car_Type	3588	
Car_Type_group	18	Car_Type variable divided in to 18 interval according to popularity of car in the Egyptian insurance market
Use	5	• Bank • Companies • Employees of Company • Exhibition • Private
Coverage	4	• Comprehensive • Fire and theft • Liability • liability,Fire and theft
Model	56	manufacture year
Year	4	year of insurance contract 2013 ,2014 ,2015 ,2016
Continuous Variables		**Notes**
ID		ID for insurance contract
Agec		Age of car
Insurance_Amount		
Premium		
Num_acc		Number of accidents
Compensation		

4.2 RESULTS

4.2.1 Generalized Linear Model (GLM)

Finding suitable regression models for insurance to predict pure premiums is challenging because of the special characteristics of the data. Insurance data has a lot of zeros, and the positive values are highly right skewed. The common regression models cannot deal with this situation.

One alternative models to predict pure premium is estimating two separate models. One is for the severity of accidents using the cases that achieve losses only using Gamma GLM (GLM 2). One could also use the inverse Gaussian GLM model here, but this is not justified because the AIC of this model is approximately 8.3 times larger than the Gamma GLM in our case. The other model estimates the probability of accident occurrence using Poisson GLM (GLM 1) or negative binomial GLM, which has a greater AIC by 2226 units. Therefore, the proposed GLM model for severity will be Gamma GLM with AIC 2676123.1 and for number of accidents Poisson GLM with AIC 771277.9 . We can write the GLM equation as:

© 2022 NSP
Natural Sciences Publishing Cor.

$$(\mu_{\text{Compensation}}) = \beta_0 + \text{factor}(\text{District}) + \text{factor}(\text{Car\_age\_group}) + \text{factor}(\text{Use}) + \text{factor}(\text{Coverage}) + \beta_{33}\text{Agec} \quad (16)$$

$$(\mu_{\text{Num\_acc}}) = \beta_0 + \text{factor}(\text{District}) + \text{factor}(\text{Car\_age\_group}) + \text{factor}(\text{Use}) + \text{factor}(\text{Coverage}) + \beta_{33}\text{Agec} \quad (17)$$

Table 4: Model 1 parameters.

factor	levels	Beta
District	9	from$\beta_1$toβ_8
Car_type_group	18	from$\beta_9$toβ_{25}
Use	5	fromβ_{26}toβ_{29}
Coverage	4	fromβ_{30}toβ_{32}

The output analysis is shown in tables 5 and 6. These Tables represent severity and probability of loss, respectively. The first GLM equation 16 models the positive loss values of compensation variables because it is not possible to model zeros with Gamma GLM. It can be modeled also by inverse Gaussian GLM, but Gamma distribution fits the research data better. The second GLM equation 17 represents the probability of loss occurrence modeled using the whole data with Poisson distribution. Negative binomial distribution can be used also, but Poisson GLM fit the data better with lower standard error and confidence interval for estimated parameters. The pure premium can be calculated by multiplying the expected compensation by the expected corresponding probability. If we multiply GLM 1 relatives by GLM 2 relatives, it produces the loadings that might be used in the ratemaking process.

The car type variable has 3589 different car types, so it is not possible to include it as a categorical variable even if it is reduced to the brand name only. On the other hand, it does not make any sense to model it as a continuous variable because it is not a numerical variable. Therefore, the variable is sorted in descending order according to frequencies (popularity) and then divided into 18 intervals with 200 car types for each interval, introduced by Car_type_group variable.

District factor has 9 levels with 8 parameters from β_1 to β_8 and 8 variables from x_1 to x_8. Every variable takes values zero or one. For instance, if the insurance contract belongs to Alexandria, x_1 will take value one and the other variables for the same factor will take value zero and so on. Otherwise, if the insurance contract belongs to the central area of Cairo (the base for district factor), all variables in the same level will take value zero, and the central area effect will be included into the intercept. The same concept is applied to the other factors.

The omnibus test sig. in the following tables indicates a significant effect of the models. This means the explained variance of the dependent variable is significantly greater than the unexplained variance. The test of model effect is also significant. And accordingly, all independent variables have a significant effect on the dependent variable. Exp(B) column shows factor relatives or loadings. In property insurance ratemaking, log link function is the most commonly used function for two reasons. First, it makes the model coefficients easy to be interpreted. For example, if B_{33} equals -.006 as in GLM 1, that means for every increase in Agec variable by 1, the probability of accident will be decreased by $1 - Exp(-.006) = .006$. Second, the insurance companies tend to select the ratemaking model that will be easy for insureds to understand. For These reasons, we used log link function for all models in this research. The standard error is an indicator of how fast log likelihood function will fall from its maximization value as parameters move away from the point of maximization.

© 2022 NSP
Natural Sciences Publishing Cor.

Table 5: GLM 1 estimated parameters. Estimates of probability of loss occurrence using Number of Accidents as a dependent variable.

Parameter	B	Std. Error	Hypothesis Test sig.	Exp(B)	%95Wald Confidence Interval for Exp(B) Lower	Upper	Tests of Model Effects sig.
Probability Distribution = Poisson				Link Function = Log			
Dependent Variable = Num_acc				Omnibus Test sig. = 0.00			
(Intercept)	-3.700	0.7244	0	0.025	0.006	0.102	
District=Alexandria	-.422	0.1549	0.006	0.656	0.484	0.888	
District=Canal	-1.304	0.1562	0	0.271	0.2	0.369	
District=Central West Delta	-.559	0.1549	0	0.572	0.422	0.775	
District=East Delta	-1.278	0.1561	0	0.279	0.205	0.378	
District=North and Central South	-.461	0.1557	0.003	0.631	0.465	0.856	0.00
District=North of Cairo	-.304	0.1546	0.049	0.738	0.545	0.999	
District=South cairo	-.330	0.1546	0.033	0.719	0.531	0.973	
District=South South	-.930	0.156	0	0.394	0.29	0.535	
District=The Central Area	0	.	.	1	.	.	
Car_Type_group=(0,200]	0.121	0.0296	0	1.128	1.065	1.196	
Car_Type_group=(1.2e+03,1.4e+03]	0.07	0.0545	0.199	1.073	0.964	1.193	
Car_Type_group=(1.4e+03,1.6e+03]	0.072	0.0607	0.239	1.074	0.954	1.21	
Car_Type_group=(1.6e+03,1.8e+03]	0.204	0.0643	0.002	1.226	1.081	1.39	
Car_Type_group=(1.8e+03,2e+03]	-.242-	0.0925	0.009	0.785	0.655	0.941	
Car_Type_group=(1e+03,1.2e+03]	0.06	0.0451	0.185	1.062	0.972	1.16	
Car_Type_group=(2.2e+03,2.4e+03]	0.076	0.1124	0.499	1.079	0.866	1.345	
Car_Type_group=(2.4e+03,2.6e+03]	-.001	0.1257	0.996	0.999	0.781	1.279	
Car_Type_group=(2.6e+03,2.8e+0]	0.164	0.1106	0.137	1.179	0.949	1.464	
Car_Type_group=(2.8e+03,3e+03]	-.121	0.1504	0.423	0.886	0.66	1.19	0.00
Car_Type_group=(200,400]	0.013	0.031	0.676	1.013	0.953	1.076	
Car_Type_group=(2e+03,2.2e+03]	-.178	0.0928	0.056	0.837	0.698	1.004	
Car_Type_group=(3.2e+03,3.4e+03]	0.042	0.2153	0.846	1.043	0.684	1.59	
Car_Type_group=(3.4e+03,3.6e+03]	-.558	0.2689	0.038	0.573	0.338	0.97	
Car_Type_group=(3e+03,3.2e+03]	-.073	0.1649	0.657	0.929	0.673	1.284	
Car_Type_group=(400,600]	-.014	0.0336	0.67	0.986	0.923	1.053	
Car_Type_group=(600,800]	0.066	0.0364	0.071	1.068	0.994	1.147	
Car_Type_group=(800,1e+03]	0	.	.	1	.	.	
Use=Bank	-.046	0.0091	0	0.955	0.938	0.972	
Use=Companies	0.1	0.0061	0	1.105	1.092	1.118	
Use=Employees of Company	0.061	0.0182	0.001	1.063	1.025	1.101	0.00
Use=Exhibition	-.200	0.0077	0	0.818	0.806	0.831	
Use=Private	0	.	.	1	.	.	
Coverage=comprehensive	2.755	0.7072	0	15.715	3.93	62.844	
Coverage=Fire and theft	-.019	0.9129	0.983	0.981	0.164	5.87	
Coverage=liability	-2.388	1	0.017	0.092	0.013	0.652	0.00
Coverage=liability,Fire and theft	0	.	.	1	.	.	
Agec	-.006	.0005	.000	0.994	.9930	.995	0.00
Goodness of Fit	Deviance (value/df)= 0.84		Pearson Chi-Square (value/df)= 1.1		Akaike's Information Criterion (AIC) = 771277.9		

NOTE: In this table, Sig. is the p-vlaue, EXP(B) returns the natural exponential of B, Std. error is the standard error, and Df is the degree of freedom.

© 2022 NSP
Natural Sciences Publishing Cor.

Table 6: GLM 2 estimated parameters. The dependent variable represents the expected mean of compensation.

Probability Distribution = Gamma				Link Function = Log			
Dependent Variable = Compensation				Omnibus Test sig. = 0.00			
Parameter	B	Std. Error	Hypothesis Test sig.	Exp(B)	%95Wald Confidence Interval for Exp(B)		Tests of Model Effects sig.
					Lower	Upper	
(Intercept)	9.37	0.9874	0	11732	1693	81252	
District=Alexandria	0.017	0.1702	0.921	1.017	0.729	1.42	0.00
District=Canal	0.211	0.1713	0.217	1.236	0.883	1.729	
District=Central West Delta	-0.121	0.1702	0.477	0.886	0.635	1.237	
District=East Delta	0.043	0.1713	0.803	1.044	0.746	1.46	
District=North and Central South	-0.22	0.171	0.198	0.803	0.574	1.122	
District=North of Cairo	-0.064	0.1699	0.707	0.938	0.672	1.309	
District=South cairo	-0.034	0.1699	0.841	0.966	0.693	1.348	
District=South South	-0.205	0.1713	0.232	0.815	0.582	1.14	
District=The Central Area	0	.	.	1	.	.	
Car_Type_group=(0,200]	-0.187	0.0312	0	0.829	0.78	0.881	0.00
Car_Type_group=(1.2e+03,1.4e+03]	0.062	0.0579	0.283	1.064	0.95	1.192	
Car_Type_group=(1.4e+03,1.6e+03]	-0.179	0.064	0.005	0.836	0.737	0.948	
Car_Type_group=(1.6e+03,1.8e+03]	0.156	0.0692	0.024	1.169	1.02	1.338	
Car_Type_group=(1.8e+03,2e+03]	0.531	0.0976	0	1.7	1.404	2.058	
Car_Type_group=(1e+03,1.2e+03]	0.123	0.0479	0.01	1.131	1.03	1.243	
Car_Type_group=(2.2e+03,2.4e+03]	-0.094	0.1165	0.422	0.911	0.725	1.144	
Car_Type_group=(2.4e+03,2.6e+03]	0.275	0.136	0.043	1.316	1.008	1.718	
Car_Type_group=(2.6e+03,2.8e+0]	0.236	0.1229	0.055	1.266	0.995	1.611	
Car_Type_group=(2.8e+03,3e+03]	0.571	0.1499	0	1.77	1.319	2.374	
Car_Type_group=(200,400]	0.021	0.0327	0.53	1.021	0.957	1.088	
Car_Type_group=(2e+03,2.2e+03]	0.59	0.1007	0	1.803	1.48	2.197	
Car_Type_group=(3.2e+03,3.4e+03]	-0.14	0.2196	0.524	0.869	0.565	1.337	
Car_Type_group=(3.4e+03,3.6e+03]	-0.052	0.2618	0.841	0.949	0.568	1.585	
Car_Type_group=(3e+03,3.2e+03]	0.659	0.1747	0	1.933	1.372	2.722	
Car_Type_group=(400,600]	0.138	0.0355	0	1.148	1.071	1.231	
Car_Type_group=(600,800]	-0.001	0.0385	0.978	0.999	0.926	1.077	
Car_Type_group=(800,1e+03]	0	.	.	1	.	.	
Use=Bank	-0.069	0.0096	0	0.933	0.915	0.951	0.00
Use=Companies	-0.056	0.0065	0	0.946	0.934	0.958	
Use=Employees of Company	-0.074	0.0193	0	0.929	0.894	0.965	
Use=Exhibition	0.068	0.0081	0	1.07	1.053	1.087	
Use=Private	0	.	.	1	.	.	
Coverage=comprehensive	-0.102	0.9723	0.917	0.903	0.134	6.072	0.26
Coverage=Fire and theft	1.785	1.3752	0.194	5.962	0.403	88.287	
Coverage=liability	-0.442	1.1908	0.711	0.643	0.062	6.635	
Coverage=liability,Fire and theft	0	.	.	1	.	.	
Agec	-.052	.0006	.000	0.949	.9480	.950	0.00
Goodness of Fit	Deviance (value/df)= 1.08		Pearson Chi-Square (value/df)= 2.12		Akaike's Information Criterion (AIC) = 2676123.08		

NOTE: In this table, Sig. is the p-vlaue, EXP(B) returns the natural exponential of B, Std. error is the standard error, and Df is the degree of freedom.

4.2.2 Generalized Linear Mixed Model (GLMM)

As mentioned at the beginning of this section, the main disadvantages of GLM are twofold: first, the autocorrelation over time in model variables and second, it gives full credibility for every class in the data, and that might create a problem for classes which have small and sparse data. Therefore, we resort to GLMM to tackle these flaws and suggest (GLMM 1, GLMM 2) models. The unique serial number of an insured car is labeled as ID. The structures and the output of 2 models are delineated in the following tables:

© 2022 NSP
Natural Sciences Publishing Cor.

Table 7: GLMM models

Model	Dependent variable	Distribution	Independent variables	
			Fixed effect	Random effect
GLMM1	Number of accidents	Poisson	Agec, Coverage, Car_Type_group, Use	ID
GLMM2	Compensation	Gamma	Agec, Coverage, Car_Type_group, Use	ID

Table 8: GLMM 1 model output

Model (GLMM 1)				Link Function = Log		
Dependent Variable = Number of accidents				Probability Distribution = Gamma		
Parameter	B	Std. Error	Hypothesis Test sig.	Exp(B)	%95Wald Confidence Interval for Exp(B)	
					Lower	Upper
(Intercept)	-1.317	0.015	< 2e-16	0.268	0.260	0.276
District_Canal	-0.857	0.026	< 2e-16	0.424	0.403	0.447
District_Central West Delta	-0.16	0.017	< 2e-16	0.852	0.824	0.881
District_East Delta	-0.843	0.026	< 2e-16	0.43	0.409	0.453
District_North and Central South	-0.036	0.024	0.139	0.965	0.919	1.012
District_North of Cairo	0.115	0.012	< 2e-16	1.122	1.095	1.149
District_South cairo	0.087	0.013	0	1.091	1.065	1.119
District_South South	-0.502	0.026	< 2e-16	0.606	0.576	0.637
District_The Central Area	0.37	0.164	0.024	1.448	1.050	1.998
Car_Type_group(200,400]	-0.100	0.011	< 2e-16	0.905	0.885	0.924
Car_Type_group(400,600]	-0.128	0.018	0.000	0.880	0.850	0.912
Car_Type_group(600,800]	-0.049	0.024	0.040	0.952	0.908	0.998
Car_Type_group(800,1e+03]	-0.126	0.033	0.000	0.882	0.828	0.940
Car_Type_group(1e+03,1.2e+03]	-0.058	0.038	0.128	0.944	0.876	1.017
Car_Type_group(1.2e+03,1.4e+03]	-0.033	0.051	0.511	0.967	0.875	1.068
Car_Type_group(1.4e+03,1.6e+03]	-0.040	0.058	0.497	0.961	0.857	1.078
Car_Type_group(1.6e+03,1.8e+03]	0.081	0.064	0.208	1.084	0.956	1.230
Car_Type_group(1.8e+03,2e+03]	-0.343	0.094	0.000	0.710	0.590	0.854
Car_Type_group(2e+03,2.2e+03]	-0.305	0.096	0.002	0.737	0.610	0.890
Car_Type_group(2.2e+03,2.4e+03]	-0.030	0.120	0.804	0.971	0.766	1.229
Car_Type_group(2.4e+03,2.6e+03]	-0.135	0.135	0.315	0.874	0.671	1.137
Car_Type_group(2.6e+03,2.8e+03]	0.079	0.117	0.500	1.082	0.861	1.359
Car_Type_group(2.8e+03,3e+03]	-0.201	0.156	0.196	0.818	0.602	1.110
Car_Type_group(3e+03,3.2e+03]	-0.166	0.170	0.327	0.847	0.607	1.181
Car_Type_group(3.2e+03,3.4e+03]	-0.055	0.220	0.803	0.947	0.614	1.458
Car_Type_group(3.4e+03,3.6e+03]	-0.594	0.273	0.030	0.552	0.324	0.943
Use_Companies	0.125	0.010	< 2e-16	1.133	1.110	1.156
Use_Employees of Company	0.093	0.021	0.000	1.097	1.053	1.143
Use_Exhibition	-0.138	0.011	< 2e-16	0.871	0.852	0.890
Use_Private	0.044	0.010	0.000	1.045	1.025	1.065
Coverage_Fire and theft	-2.710	0.581	0.000	0.067	0.021	0.208
Coverage_liability	-5.119	0.708	0.000	0.006	0.001	0.024
Coverage_liability,Fire and theft	-2.704	0.713	0.000	0.067	0.017	0.271
Agec	-0.008	0.001	< 2e-16	0.992	0.991	0.993
Goodness of Fit	Deviance (value/df) = 762606		Bayesian information criterion (BIC) = 763070		Akaike's Information Criterion (AIC) = 762676	

NOTE: In this table, Sig. is the p-vlaue, EXP(B) returns the natural exponential of B, Std. error is the standard error, and Df is the degree of freedom.

Table 9: GLMM 2 model output

Model (GLMM 2)			Link Function = Log				
Dependent Variable = Compensation			Probability Distribution = Poisson				
Parameter	B	Std. Error	Hypothesis Test sig.	Exp(B)	%95Wald Confidence Interval for Exp(B)	
					Lower	Upper
(Intercept)	8.529	0.016	< 2e-16	5060.7	4904.95	5223.9
District_Canal	0.234	0.028	< 2e-16	1.264	1.197	1.335
District_Central West Delta	-0.128	0.019	0	0.88	0.848	0.913
District_East Delta	-0.085	0.027	0.002	0.919	0.871	0.969
District_North and Central South	-0.277	0.026	< 2e-16	0.758	0.720	0.798
District_North of Cairo	0.019	0.014	0.178	1.019	0.992	1.047
District_South cairo	0.048	0.014	0.001	1.049	1.020	1.078
District_South South	-0.308	0.028	< 2e-16	0.735	0.696	0.776
District_The Central Area	-0.017	0.154	0.912	0.983	0.727	1.330
Car_Type_group(200,400]	0.130	0.012	< 2e-16	1.139	1.113	1.165
Car_Type_group(400,600]	0.191	0.019	< 2e-16	1.210	1.166	1.256
Car_Type_group(600,800]	0.123	0.026	0.000	1.131	1.076	1.190
Car_Type_group(800,1e+03]	0.159	0.035	0.000	1.172	1.095	1.255
Car_Type_group(1e+03,1.2e+03]	0.186	0.040	0.000	1.205	1.113	1.305
Car_Type_group(1.2e+03,1.4e+03]	0.160	0.054	0.003	1.174	1.057	1.305
Car_Type_group(1.4e+03,1.6e+03]	-0.006	0.059	0.918	0.994	0.884	1.116
Car_Type_group(1.6e+03,1.8e+03]	0.159	0.069	0.021	1.173	1.024	1.342
Car_Type_group(1.8e+03,2e+03]	0.313	0.098	0.001	1.367	1.129	1.655
Car_Type_group(2e+03,2.2e+03]	0.392	0.102	0.000	1.480	1.210	1.808
Car_Type_group(2.2e+03,2.4e+03]	0.084	0.130	0.522	1.087	0.842	1.404
Car_Type_group(2.4e+03,2.6e+03]	0.421	0.148	0.004	1.524	1.140	2.036
Car_Type_group(2.6e+03,2.8e+03]	0.517	0.124	0.000	1.678	1.314	2.140
Car_Type_group(2.8e+03,3e+03]	0.129	0.154	0.404	1.137	0.841	1.537
Car_Type_group(3e+03,3.2e+03]	0.511	0.167	0.002	1.667	1.202	2.312
Car_Type_group(3.2e+03,3.4e+03]	0.084	0.212	0.692	1.088	0.717	1.649
Car_Type_group(3.4e+03,3.6e+03]	-0.164	0.264	0.535	0.849	0.507	1.423
Use_Companies	0.006	0.010	0.528	1.006	0.987	1.026
Use_Employees of Company	-0.003	0.020	0.864	0.997	0.959	1.036
Use_Exhibition	0.093	0.011	< 2e-16	1.098	1.074	1.122
Use_Private	0.039	0.009	0.000	1.040	1.021	1.059
Coverage_Fire and theft	2.342	1.061	0.027	10.39	1.298	83.263
Coverage_liability	-0.115	0.602	0.849	0.892	0.274	2.901
Coverage_liability,Fire and theft	-0.698	0.825	0.398	0.498	0.099	2.507
Agec	-0.037	0.001	< 2e-16	0.964	0.963	0.966
Goodness of Fit	**Deviance (value/df) =** 2627541		**Bayesian information criterion (BIC) =** 2627966		**Akaike's Information Criterion (AIC) = 2627613**	

NOTE: In this table, Sig. is the p-vlaue, EXP(B) returns the natural exponential of B, Std. error is the standard error, and Df is the degree of freedom.

The two models show acceptable CI except for coverage factor that have a wide CI. The coverage factor has four levels. One of them is a base level in the model–Comprehensive level–and the others appeared in the model with a wide CI. This can be a result of the fact that the majority of the data are comprehensive coverage and there is a little information available about the rest of the levels.

4.2.3 Generalized Additive Model (GAM)

The estimation model of number of accidents (claim frequency GAM 1) and compensation (claim cost GAM 2) for our case study can be written respectively as:

$$(\mu_{\text{Compensation}}) = \beta_0 + \text{factor}(\text{District}) + \text{factor}(\text{Use}) + \text{factor}(\text{Coverage}) + \text{S}(\text{Agec}) + \text{S}(\text{Car\_Type}) \tag{18}$$

© 2022 NSP
Natural Sciences Publishing Cor.

$$(\mu_{\mathrm{Num\_acc}}) = \beta_0 + \mathrm{factor(District)} + \mathrm{factor(Use)} + \mathrm{factor(Coverage)} + \mathrm{S(Agec)} + \mathrm{S(Car\_Type)} \quad (19)$$

The first equation is modeled with Poisson distribution and the second one with GAM distribution. The link function is log for both of them. Car type variable is modeled as a continuous variable since we coded it according to its popularity in the Egyptian market. Therefore, it can be regarded as car value or popularity index.

The outcome analysis can be illustrated in the following table:

Table 10: Generalized Additive Models parameters estimation.

Link Function = Log	Probability Distribution = GAMMA				Probability Distribution = Poisson			
Dependent Variable	Compensation (GAM 2)				Num_acc (GAM 1)			
Parametric coefficients	B	Std. Error	p-value	Exp(B)	B	Std. Error	p-value	Exp(B)
(Intercept)	8.824	0.021	2.00E-16	6792.063	-1.334	0.014	2.00E-16	0.263
District=Alexandria	0.130	0.038	0.0006	1.139	-0.848	0.026	2.00E-16	0.428
District=Canal	-0.121	0.024	6.62E-07	0.886	-0.135	0.016	2.00E-16	0.874
District=Central West Delta	0.020	0.038	0.6	1.020	-0.845	0.026	2.00E-16	0.429
District=East Delta	-0.290	0.035	2.50E-16	0.748	-0.006	0.023	0.808053	0.994
District=North and Central South	-0.051	0.018	0.00459	0.951	0.109	0.012	2.00E-16	1.115
District=North of Cairo	-0.027	0.018	0.14	0.973	0.087	0.012	1.66E-13	1.091
District=South cairo	-0.240	0.038	2.28E-10	0.787	-0.495	0.025	2.00E-16	0.610
District=South South	0.029	0.253	0.91	1.030	0.375	0.163	0.02131	1.455
Use=Bank	0.037	0.015	0.01	1.038	0.131	0.010	2.00E-16	1.140
Use=Companies	0.029	0.031	0.35	1.029	0.081	0.021	7.91E-05	1.085
Use=Employees of Company	0.083	0.017	9.53E-07	1.087	-0.120	0.011	2.00E-16	0.887
Use=Exhibition	0.070	0.014	1.22E-06	1.072	0.045	0.010	2.60E-06	1.046
Coverage=comprehensive	1.832	1.432	0.2	6.247	-2.575	0.599	1.69E-05	0.076
Coverage=Fire and theft	-0.296	1.026	0.77	0.743	-5.126	0.724	1.45E-12	0.006
Coverage=liability	-0.053	1.446	0.97	0.948	-2.663	0.733	0.000279	0.070
Nonparametric variables	effective degrees of freedom (EDF)			p-value	effective degrees of freedom (EDF)			p-value
s(Agec)	6.405			<2e-16	7.499			<2e-16
s(Car_Type)	8.872			<2e-16	7.330			<2e-16
Akaike's Information Criterion (AIC)	2676506				995180.7			

NOTE: In this table, EXP(B) returns the natural exponential of B, Std. error is the standard error, and Df is the degree of freedom.

Now, it is better to visualize the smoothing function of the non-parametric portion of the model. The following graphs clarify the marginal effect of Agec and Car_type variable on the dependent variable in GAM 1 and GAM 2 models. To show the marginal effect, the other factors are fixed on the base level and other continuous variable are fixed on its mean.

The first chart on the left side of (Figure 1) illustrates the nature of the relationship between car age and number of accidents. It reveals a positive relationship with approximately +1 slope to age 10 (a ten years old vehicale), which has the highest number of accidents among all ages, and then the curve turns to a negative relationship with nearly -1 slope. After that, the probability of accidents has remained constant from age 32 to 35, followed by a downward trend with the same previous trend until age 45. From age 45 to 65 the slope is declined to -.5 approximately. The lowest number of accidents belongs to the oldest car in our data. The gray area represents the confidence interval for the expected values of number of accidents, which becomes very large from age 50 due to the small points available in that area.

The second chart on the left side shows a fluctuating positive and negative relationship with a declining pattern. The turning points on the graph are for car types 500, 1450, 2200 and 2750.

The first chart on the right side illustrates a negative relationship between car age and its expected risk cost with slope less than -1 until age 5. Then the slope rises gradually to age 35. After that, there is no effect of the car age on the risk cost.

The last chart depicts a positive effect of the car type on compensation until car type 400, then an approximately fixed effect to car type 1500. After that, the curve pulsates with two peaks on 2100 and 3100 car types.

© 2022 NSP
Natural Sciences Publishing Cor.

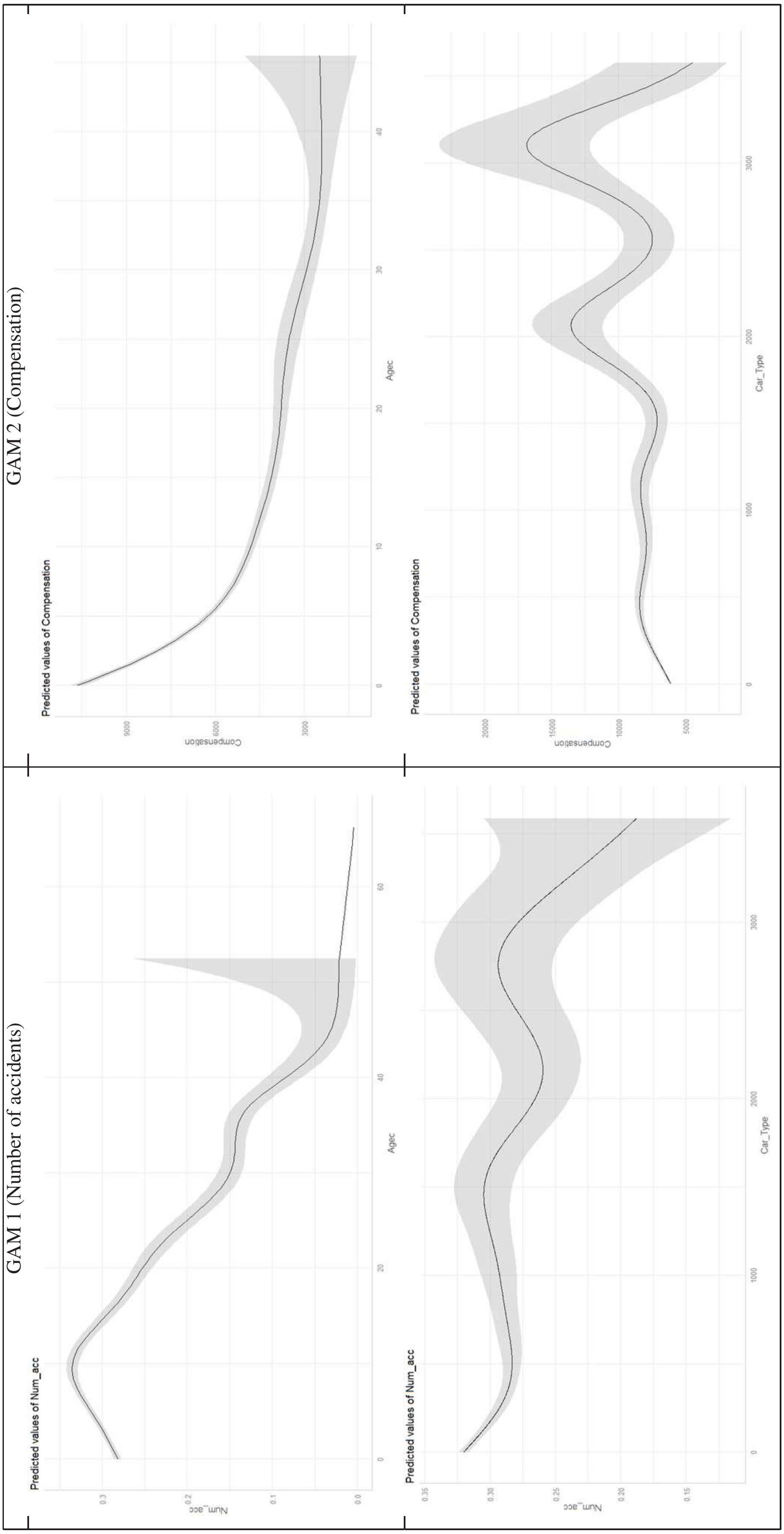

Fig. 1: Partial effects of the non-parametric variables in GAM.

© 2022 NSP
Natural Sciences Publishing Cor.

So the question at this point is what is the ideal model to represent our data? One of the criteria that be used to compare GLM, GLMM and GAM is Akaike's Information Criterion (AIC). As illustrated in the following table, GAM has the worst AIC both for number of accidents models and for compensation models. The best model statistically regarding total AIC value is alternative number 2. GLM models come in the second place and they have two shortages as we illustrated before in 3.2 . Hence, The recommended models for the company are GLMM 1 and GLMM 2 models (alternative number 2).

Table 11: Akaike's Information Criterion (AIC) of all regression models in the research.

Alternatives	Number of accidents models	Akaike's Information Criterion (AIC)	Compensation models	Akaike's Information Criterion (AIC)	Total
1	GLM 1	771278	GLM 2	2676123	3447401
2	GLMM 1	762676	GLMM 2	2627613	3390289
3	GAM 1	995181	GAM 2	2676506	3671687

5 Conclusion

Among all models in this research, we suggest GLMM 1 and GLMM 2 models as a recommended tariff system for loss probability and loss severity respectively for Misr insurance company because they are the only models that have the following merits. First, they have the lowest AIC value. Second, they take into consideration the repeated measure nature in insurance data that regards as a violation of the premises of the GLM model. Third, they can apply credibility theory in one step with the risk classification process. Hence, the prediction power of the ratemaking model will be increased . Hence, GLMM can be regarded as the most convenient tariff system for our case study. GLM is the second best statistical model to be used in the ratemaking process for Misr insurance company. In spite of GAM is having the worst AIC both for number of accidents models and for compensation models, it gives us the chance to understand how the covariates affect both number of accidents and compensation variables without a restriction of linear relationship.

To sum up, the optimal model selection depends on many factors like the reliability of the used statistical models, to what extent that model was tested before in the market, statistical tests, insurance company financial position, and the expected reaction of the rivals in the market upon the application of this tariff system.

Acknowledgement

A major portion of the first author's research for this paper was conducted under the guidance of Professor Hira L. Koul at the Department of Statistics and Probability of the Michigan State University (MSU), during his visit there as a Research Scholar from May 1, 2018 to September 1, 2020. He would like to express his sincere thanks to Professor Koul for numerous brilliant comments and suggestions which contributed greatly to the improvement of this research. His friendly guidance, patience and expert advice have been invaluable throughout all stages of this work.

The first author would also like to thank the people in the Department of Statistics and Probability, MSU, for providing excellent working conditions and to the Egyptian Ministry of Higher Education for providing the financial support for his visit to MSU. He would also like to extend his sincere gratitude and thanks to the Statistics and Insurance Department, Zagazig University for giving him the opportunity to study at the MSU. He also extends his sincere gratitude and thanks to Professor Mohamed Abdelmawla of the Insurance Department, Tanta University and Professor Mohamed Adel Elshahat of the Statistics and Insurance Department, Zagazig University, for providing interesting and valuable feedback for this research. Finally, both authors are grateful to the anonymous reviewers for a careful checking of the details and for helpful comments that improved the presentation.

Conflict of interest: The authors declare that they have no conflict of interest.

References

[1] M. T. AlBalkiny and R. A. Ibrahim, The use of credibility theory in pricing insurance from compulsory civil liability for car accidents in egypt. Egyptian Journal of Commercial Studies **23**, 563 – 580(1999).

© 2022 NSP
Natural Sciences Publishing Cor.

[2] Q. C.Chukwudum, Credibility Premium Estimation of Insurance Claims in Nigeria (2018). URL https://hal. archives-ouvertes.fr/hal-01855972. Working paper or preprint.

[3] I. M. Mahdi, I. A.ElHusseini and M. M. ElMadawy, Alternative models for supplementary car insurance pricing (an applied study). The Egyptian Journal of Commercial Studies **34**, 573-594 (2010).

[4] B. Linda and J. Kubanová, Credibility premium calculation in motor third-party liability insurance. In *WSEAS* conference Advances in Mathematical and Computational Methods, Malta, 259-263 (2012).

[5] E. M.ElFeky, Using differentiation analysis to classify risks in car insurance: An applied study on the kuwaiti insurance market. The Egyptian Journal of Commercial Studies **26**, $145-166(2002)$.

[6] M. T. AlBalkiny, The use of fazian categories theory in estimating the volume of claims in supplementary insurance on cars. The Egyptian Journal of Commercial Studies **25**, $1-25(2001)$.

[7] M. S. Ahmed, On rationalizing underwriting decisions in property and liability insurance in the egyptian market by using vasian neural networks: applying to the supplementary car insurance branch. PhD Thesiss (2013).

[8] W. A. Mustafa, quantitative method for estimating the number of traffic accidents and the rates of injury and death due to those accidents in the kingdom of saudi arabia. The Egyptian Journal of Commercial Studies **28**, $449\ 473(2004)$

[9] G. Dionne and C. Vanasse, A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. ASTIN Bulletin: The Journal of the IAA **19**, 199-212 (1989).

[10] S. Haberman and A. E. Renshaw, Generalized linear models and actuarial science. Journal of the Royal Statistical Society: Series *D* (The Statistician) **45**, 407-436 (1996).

[11] Pinquet, J. Allowance for hidden information by heterogeneous models and applications to insurance rating. In Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation, 47-78 (Springer, 1999).

[12] G. A. Wasef, A quantitative method for predicting the dangerousness of the car and the driver as a basis for pricing in car insurance. The Egyptian Journal of Commercial Studies **24**, $119-143(2000)$.

[13] L. Bermúdez and D. Karlis, Bayesian multivariate poisson models for insurance ratemaking. Insurance: Mathematics and Economics **48**, $226-236(2011)$.

[14] J. P. Bouche and M. Denuit, Fixed versus random effects in poisson regression models for claim counts: A case study with motor insurance. ASTIN Bulletin: The Journal of the IAA **36**,$285-301(2006)$.

[15] M. David, Auto insurance premium calculation using generalized linear models. Procedia Economics and Finance **20** ,$147-156(2015)$

[16] M. Oeben, Generalized Linear Mixed Models in the competitive non-life insurance market. Master's thesis, MSc thesis, School of Mathematical Sciences (2015).

[17] R. Kaas, M. Goovaerts, J. Dhaene and M. Denuit, Modern actuarial risk theory: using *R*, vol. **128** (Springer Science & Business Media, 2008).

[18] X. K. Dimakos and A. F. Di Rattalma, Bayesian premium rating with latent structure. Scandinavian Actuarial Journal ,$162-184(2002)$

[19] F. Klinker, Generalized linear mixed models for ratemaking: a means of introducing credibility into a generalized linear model setting. In Casualty Actuarial Society E-Forum, vol. **2** ,$1-25(2011)$.

[20] K. Antonio and J. Beirlant, Actuarial statistics with generalized linear mixed models. Insurance: Mathematics and Economics **40**, $58-76(2007)$.

[21] M. Denuit and S. Lang, Non-life rate-making with bayesian gams. Insurance: Mathematics and Economics **35**, $627-647(2004)$

[22] L. Fahrmeir, F. Sagerer and G. Sussmann, Geoadditive regression for analyzing small-scale geographical variability in car insurance. Blätter der $DGVFM$**28**, $47-65$ (2007).

[23] L. E. Echavarria, Semiparametric Bayesian count data models. Ph.D. thesis, PhD thesis, Ludwig-MaximiliansUniversität München (2004).

[24] N. Klein, M. Denuit, S. Lang and T. Kneib, Nonlife ratemaking and risk management with bayesian generalized additive models for location, scale, and shape. Insurance: Mathematics and Economics **55**, $225-249(2014)$.

[25] J. C. Pinheiro and E. C. Chao, Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. Journal of Computational and Graphical Statistics **15**, $58-81(2006)$.

[26] D. Handayani, K. A. Notodiputro, K. Sadik and A. Kurnia A comparative study of approximation methods for maximum likelihood estimation in generalized linear mixed models (glmm). In AIP Conference Proceedings, vol. **1827**,020033 (AIP Publishing LLC, 2017).

[27] P. T. Reiss and R. T. Ogden, Smoothing parameter selection for a class of semiparametric linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **71** ,$505-523(2009)$.

[28] S. N. Wood, Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series *B* (Statistical Methodology) **73** ,$3-36$ (2011).

© 2022 NSP
Natural Sciences Publishing Cor.