Risk Premium Prediction of Motor Hull Insurance Using Generalized Linear Models

Marek Strežo[1] | *University of Economics in Bratislava, Bratislava, Slovakia*
Vladimír Mucha[2] | *University of Economics in Bratislava, Bratislava, Slovakia*
Erik Šoltés[3] | *University of Economics in Bratislava, Bratislava, Slovakia*
Michal Páleš[4] | *University of Economics in Bratislava, Bratislava, Slovakia*

Abstract

Pricing is a quite complex endeavour, understood as a process with beginning and end where several different tasks have to be executed in a certain order. Set the price for some individual policy can be considered an art, taking into consideration various features of policyholder or the insured object. Actually, approach performed by insurance companies, is necessary to apply different premiums depending on the degree of risk because of presence of heterogeneity within insurance portfolio, which could lead to the appearance of asymmetric information.

The aim of this paper is to present the methodology of segmented pricing model with generalized linear models, known as GLMs, for setting the risk premium. Nowadays, the GLMs are widely recognized as the industry standard method for pricing motor, the other personal lines and the retail insurance in the European Union.

Keywords

GLMs, Poisson regression, Gamma regression, risk premium, motor hull insurance

JEL code

C21, G22

INTRODUCTION

Actuaries use many statistical methods to measure risk in process of setting the risk premium. Practically the most widely method used in practise is the regression analysis. Linear regression had been applied until the 1980s using various transformations of predicted variable. Nowadays, generalized linear models or GLMs for short are preferably applied. Restrictions in linear regression are discussed by (Anderson et al., 2007). The comprehensive reference for GLMs in actuarial field is (McCullagh and Nelder, 1989; Fahrmair and Tutz, 1996; Mildenhall, 1999; Kaas et al., 2001). Valecký (2017, p. 451) states that more

[1] Department of Mathematics and Actuarial Science, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1/b, 852 35 Bratislava 5, Slovakia. E-mail: marek.strezo@euba.sk.
[2] Department of Mathematics and Actuarial Science, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1/b, 852 35 Bratislava 5, Slovakia. E-mail: vladimir.mucha@euba.sk.
[3] Department of Statistics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1/b, 852 35 Bratislava 5, Slovakia. E-mail: erik.soltes@euba.sk.
[4] Department of Mathematics and Actuarial Science, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1/b, 852 35 Bratislava 5, Slovakia. E-mail: michal.pales@euba.sk.

applications of the GLMs occurred mostly after the 1990s when the insurance market was being deregulated in many countries and the models were used to perform tariff analysis. Even though the GLMs are used mainly in the non-life insurance practise, Haberman and Renshaw (1996) referred to their wide use in the actuarial practice, including life insurance (survival data analysis – SDA, health insurance modelling and mortality modelling). As David (2015) indicates, the GLMs allow modelling of non-linear behaviour and non-Gaussian distribution of residuals, which is very useful non-life insurance analysis. A random component (error term) in an ordinary linear model is assumed to be normally distributed. However, when the claim frequency (count of the claims per exposure) and the claim severity (average cost per claim) are modelled this condition is not fulfilled. For that reason, the GLMs are suitable for analysis with non-normal data, i.e. insurance data because the error term can follow the number of different distributions from the exponential dispersion family – EDF, which generalizes normal distribution used in the linear models. The Poisson distribution belongs to this family and represents the main tool for the claim frequency modelling meanwhile Gamma distribution allows econometric modelling of the claim costs (Ewald and Wang, 2015) and (Duan et al., 2018). It might be considered using a Tweedie model to analyze the risk premium directly (see Xacur and Garrido, 2015; Frees et al., 2016; Jörgensen and Souza, 1994).

In general, two approaches are commonly used to calculate the risk premium in the non-life insurance. In the first case, the risk premium is modelled directly. The second case describe the standard GLMs analysis with separated analysis for the claim frequency and severity. Goldburd et al. (2016) point out the reason for this separation where the claim frequency is more stable than the claim severity and much more predictive factors are associated with the claim frequency. Such a separate analysis represents greater accuracy and offers deeper insights to the risk w.r.t regression coefficients.

Here, both the claims count, and the claims amount are assumed to be independent in case of the separate claim frequency and claim severity analysis. When this fundamental assumption is not fulfilled, authors Shi et al. (2015) or Garrido et al. (2016) discuss about this problem. Charpentier and Denuit (2005) also prefer separate analyses for claim frequency and claim severity as the benefit of such approach is visible in fact that both models (frequency and severity) can be affected by different various factors. Mentioned facts give us the reason why to choose separate analysis in the GLMs for calculating the risk premium in motor hull insurance in Slovakia.

The GLMs are an efficient and reliable tool used in various fields of predictive modelling. According to (Xie and Lawniczak, 2018, p. 2) the main reason for the prevalence of GLMs is that it enables a simultaneous modelling of all possible risk factors as well as the determination of the retention of risk factors in the model.

The main effort of this paper is not only to estimate the claim frequency and claim severity and then set price of transfer risk from the insured to an insurer, but also to identify relevant risk factors as well as to quantify their impact in the claim frequency, claim severity and also on the expected loss per exposure.

Data on which the research was based are real and comes from an unnamed insurance company operating in the Slovak insurance market. All calculations in this paper have been realized in R environment (R Core Team, 2019) using *glm()* function and packages *data.table* (Dowle et al., 2015) and *MASS* (Venables et. al., 2002).

1 METHODS OF ANALYSIS

The expected loss (also known as a risk premium) consists of the claim frequency and claim severity that are in the multiplicative relation:

$$Risk\ Premium = Frequency \cdot Severity \tag{1}$$

The frequency refers to the number of claims that an insurer anticipates will occur for a specific risk over a given time period. The severity represents the average cost of claims for specific risk. This article focuses on separate modelling of the claim frequency and claim severity using generalized linear models and determining the risk premium. This part of the paper provides a brief description of the methodology of sophisticated mathematical and statistical methods associated with the GLMs.

1.1 Theoretical framework of Generalized linear models

Generalized linear models include a wide set of statistical models consisting of three keystone elements – random component, linear predictor and the link function.

A *random component* refers to the conditional distribution of the response variable Y given the values of the explanatory variables in the model. Nelder and Wedderburn (1972) present the basics of the GLMs theory and declare that distribution of Y with independent observations y_i ($i = 1, 2, \ldots, n$) is a member of an exponential dispersion family. Exponential dispersion family, shortly EDF, has the probability density function in the following form:

$$f(y_i) = c(y_i, \phi) \exp\left\{\frac{y_i\theta_i - a(\theta_i)}{\phi}\right\}, \tag{2}$$

where θ_i and ϕ are the parameters, θ_i is called canonical or natural parameter and ϕ is a dispersion parameter (Agresti, 2015; Kafková and Křivánková, 2014). So called cumulant function $a(\theta_i)$ is assumed twice differentiable, where the first derivative is invertible. EDF includes the univariate Bernoulli, binomial, Poisson, geometric, Gamma, normal, inverse Gaussian, lognormal, Rayleigh, and von Mises distributions (Forbes et al., 2011).

The claim severity is modelled by two commonly used distributions the Gamma and inverse Gaussian distribution. Both these distributions are right-skewed with a lower bound at zero. According to Goldburd et al. (2016) inverse Gaussian compared to the Gamma distribution has a sharper peak and a wider tail and is therefore appropriate for the situations where the skewness of the severity curve is expected to be more extreme.

The claim frequency is modelled by the GLMs with Poisson noise. Some members of EDF such as Poisson and Bernoulli distribution have the distribution determined by the mean. When fitting models to data with binary or count dependent variables, it is common to observe that variance exceeds and anticipated by the fit of the mean parameters. This phenomenon is known as overdispersion (Edward, 2010). One way to check for and deal with it is to run negative binomial distribution or overdispersed Poisson distribution (Valecký, 2016; Ohlsson and Johansson, 2010). There are also several probabilistic models available to explain this phenomenon, depending on the application on hand. For a more detailed inventory see McCullagh and Nelder (1989).

A *linear predictor* is a linear function of the regressors:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \text{ or } \eta_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}, \tag{3}$$

where:

$\boldsymbol{\beta}$ is $p \times 1$ vector of model parameters ($p = k + 1$) including intercept β_0 and the regression coefficients $\beta_j (j = 1, 2, \ldots, k)$,

$\mathbf{X}$ is $n \times p$ matrix of the regressors (known from the classical regression) and x_{ij} is i-th observation of j-th regressor X_j.

The regressor can be expressed as quantitative explanatory variable, transformation of quantitative explanatory variable, e.g. polynomial regressor, dummy variable (coding the particular categorical variable), interaction, etc. (see Wooldridge, 2013).

The *link function* $g(\cdot)$ is strictly monotone and twice differentiable. This fundamental object links the mean of the response variable to the linear predictor through:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} \text{ or } g(\mu_i) = \eta_i\,, \tag{4}$$

where:
$\boldsymbol{\mu} = E(\mathbf{y})$ or $\mu_i = E(y_i)$,
$\mathbf{y}$ is $n \times 1$ vector of observations of target variable Y (called also response variable, explained variable or dependent variable),
$\boldsymbol{\mu}$ is $n \times 1$ vector of expected values of the elements of $\mathbf{y}$.

The link function that transforms μ_i to the natural parameter θ_i of distribution from exponential family is called canonical (or natural) link function (Agresti, 2015; Fox, 2015; Littell et al., 2010).

A maximum likelihood method is used to estimate the regression parameters $\boldsymbol{\beta}$ in Formula (4) (De Jong and Heller, 2008; Littell et al., 2010). As a result of this method is system of equations:

$$(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}^{*}, \tag{5}$$

where:
$\mathbf{W} = \mathbf{D}\mathbf{V}^{-1}\mathbf{D}$, whereby $\mathbf{V} = diag[\phi \cdot Var(\boldsymbol{\mu})]$ and $\mathbf{D} = diag\left[\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}}\right]$ is $n \times n$ diagonal matrix whose elements are derivates of the elements of $\boldsymbol{\eta}$ with respect to $\boldsymbol{\mu}$ and $Var(\boldsymbol{\mu})$ is a covariance matrix of $\boldsymbol{\mu}$.
$\mathbf{y}^{*} = \hat{\boldsymbol{\eta}} + \mathbf{D}^{-1}\,(\mathrm{y} - \boldsymbol{\mu})$ [more detailed in (Littell et al., 2010; McCullagh and Nelder, 1989)].

We note that for the normal error model is $\mathbf{V} = \sigma_{\varepsilon}^{2}\mathbf{I}$ where $\mathbf{W}$ is the unit matrix and system (5) is reduced to the well-known system of normal equations, that we can estimate parameters of classical linear regression model (Agresti, 2015; Littell et al., 2010). In general, the system of equations from (5) is nonlinear in $\hat{\boldsymbol{\beta}}$, therefore the iterative methods are used for solving nonlinear equations such as Newton-Raphson method using a Hessian matrix itself and Fisher scoring method which uses expected values of Hessian matrix (Allison, 2012; Agresti, 2015).

1.2 Assessment of impact of explanatory variables on target variable and model selection

After estimating the generalized linear model, it is important to verify its statistical significance and verify if influence of the individual explanatory variables on probability target variable is significant. The significance of model is revealed by zero-hypothesis test $\boldsymbol{\beta} = (\beta_1\ \beta_2\ \ldots\ \beta_k) = \mathbf{0}^{\mathrm{T}}$ against an alternative hypothesis – at least one regression coefficient should not be zero, while three different chi-square statistics are prevalently used (Likelihood ratio, Score statistics, Wald statistics). Allison (2012) discusses differences between mentioned statistical methods and notes that in the large samples, there is no reason to prefer any of these statistics and they will be quite close in value.

In order to validate the significance of the explanatory variable influence, a Wald test is used. It tests the zero-hypothesis showing that the respective explanatory variable does not affect the probability of occurrence of explored event. To verify hypothesis, Wald statistic

$$Wald = \hat{\boldsymbol{\beta}}^{\mathrm{T}} \cdot \hat{\mathbf{S}}_{\mathrm{b}}^{-1} \cdot \hat{\boldsymbol{\beta}} \tag{6}$$

is used, where $\hat{\boldsymbol{\beta}}$ is the vector of regression coefficients estimates that stand at dummy variables for the respective factor (categorical explanatory variable) and $\hat{\mathbf{S}}_{\mathbf{b}}$ is the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$. Wald statistic has asymptotically χ^2 distribution with degrees of freedom equal to the number of parameters estimated for a given effect. A special case of the test above is the Wald test, which verifies the statistical significance of one regression coefficient. In this case Wald statistics is asymptotically distributed as χ^2 with 1 degree of freedom. The test statistic has an equation:

$$Wald = \left(\frac{\hat{\beta}_j}{s_{\hat{\beta}_j}}\right)^2, \tag{7}$$

where $s_{\hat{\beta}_j}$ is an estimated standard error of the j-th estimated coefficient.

When the process of the model building starts, there is a wide set of potential regressors and not all of them have significant impact on the dependent variable. It is obvious to use methods for variable selection, namely, the stepwise regression (see Draper and Smith, 1981; Hebák et al., 2005). In the stepwise regression, the selection procedure is automatically performed by statistical packages. In the practical part, it is used one of the main approaches of the stepwise selection known as backward elimination (see Agresti, 2015).

To evaluate how well the model fits the experience criteria AIC (Akaike information criterion) and BIC (Bayesian information criterion) are used. These measures are based on logarithmic transformation of the likelihood function (see Kim and Timm, 2006; Agresti, 2015). Preferred model is considered have with the lowest AIC and BIC, respectively. As state (De Jong and Heller, 2008, p. 63) BIC applies a greater penalty for the number of the parameters. When number of observations is large, as it is in most of cases of insurance data sets, the BIC tends to select the model which most of analysts consider too simple. In this case the AIC is preferable.

2 DATA PROCESSING AND MODEL BUILDING

In this part, we demonstrate practical usage of GLMs in actuarial practice which have been described in previous sections of this paper. We will try to set price of a non-life insurance policy, taking into consideration various properties of the insured object and policyholder as well. In this empirical study, we will go through models for short-term insurance schemes based on the Slovak market´s conditions. The study in this paper works with a very basic feature of the portfolio of risks – heterogeneity, which means that risks generate different values of claims. Consequently, charging each policy with the same premium (flat rate) is both unjust and uncompetitive. Therefore, we will try to classify each risk into the homogeneous risk groups where the *i*th risk has the same risk premium. Basic assumption that will give foundation to our statistical models is policy independence. This means that independence between random variables $Y_1, \ldots, Y_n$ is made in modelling the value of single claims and in the number of claims as well. Presented frequency-severity models will decompose the aggregate claim amount for a single risk into two parts, where the frequency part examines the number of claims by Poisson regression, the severity part by the GLMs Gamma regression. The R software will be used to calculate and analyse the results of these different multiplicative models.

2.1 Motor hull insurance data and descriptive analysis

Before the modelling it is useful to provide certain preliminary analyses, such as data checks, identification of observations with negative claim counts, zero or negative exposures, etc. The portfolio D consists of $n = 91\,685$ car insurance policies for which we have features information $\mathbf{x}_i \in \mathbf{X}$ and exposure - years at risk information, denotes as $v_i \in [0; 1]$, for $i = 1, 2, \ldots, n$. Nature of the data comprises a Slovak motor

hull insurance with corresponding claim sizes and counts for calendar year 2018. Now, we briefly describe the list of variables in our dataset D:

- *ID profile*: represents unique identifier; policy number;
- *Claim.No*: number of claims which occurred on each policy;
- *Claims*: total claim cost per every policy;
- *Policyholder_Age*: the owners age in years, between 0 and 91, non-linear continuous feature portioned as nominal categorical variables;
- *Vehicle_Age*: age of cars in years, narrowly defined categorical factor;
- *Policy_Exposure*: the exposure is widely applied in non-life pricing. In order to illustrate this concept, we take GLM for the frequency claims. Policies that begin in a given calendar last year until the end of the coverage period. This period is longer for annual contracts than for short-term policies, which results in a higher number of expected claims for longer contracts. Therefore, it is necessary to include this effect in the model as exposure with the use of weights;
- *Region*: regional divisions of Slovakia according to the company´s internal policy, categorical feature with 11 labels;
- *B-M Class*: bonus class, taking values for bonus from 0 to 7 and for malus from 1 to 2, with the reference level 0;
- *Vehicle_Engine_Volume*: represents engine volume of car, continuous feature;
- *Total Sum_Insured (TSI)*: specified car value which represents the upper limit of what would be pay out for the claim;
- *Power*: power of car, non-linear continuous feature split as categorical variables;
- *Payment_Frequency*: expresses the frequency of premium payments (payment option is 1,2,4 and 12);
- *Vehicle_Weight*: weight of car, non-linear continuous feature portioned as nominal categorical variable;
- *Policyholder_entity*: categorical variable which can obtain 2 values;
- *Mileage_per_Year*: total length in miles per given period (calendar year);
- *Deductible_group*: policyholders can choose the excess at level that exploits reduction in premium, categorical variable.

In the next step, we provide a short summary of the data D. Since the policy number is not considered to be an explanatory variable, we drop this feature from all our next considerations.

Table 1 Split of portfolio w.r.t. number of claims and the severity claims

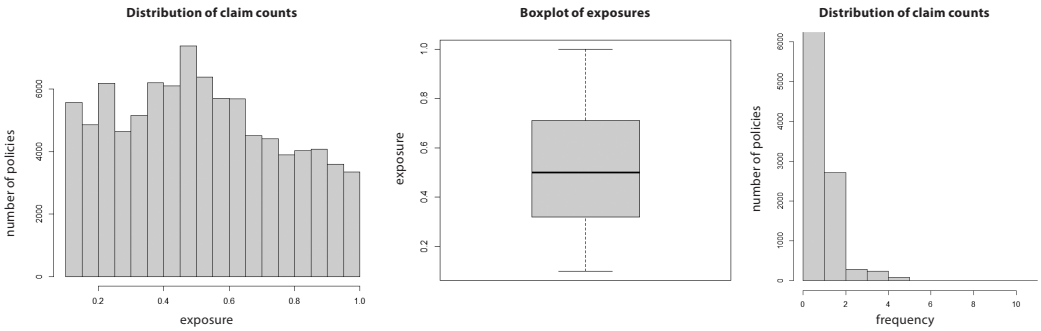
# of claims	0	1	2	3	4	5	6	7	8	9	10	11
# of policies	79 667	8 667	2 715	285	244	93	7	2	2	1	1	1
# of policies in %	86.89	9.45	2.96	0.31	0.27	0.10	0.01	0	0	0	0	0
Total exposures	40 243	5 256	1 741	182	171	65	4	1.60	1.63	0.96	0.98	0.75

Source: Own construction

In Table 1 you can be see the distribution of the observed claims $(N_i)_{1\leq i\leq n}$ across the whole portfolio of our dataset D with the attributable policy exposure. We note that 86.89% of the policies don't have a claim. In practice, this claim imbalance can often causes difficulties in the model calibration. Next, we provide helpful preliminary analysis to determine distribution of the key data items to investigate any problems or unfamiliar features prior the modelling. This concerns the distributions for claim counts and for the claim severity. Typical claim distribution is shown in Figure 1 (lhs) and in Figure 2.

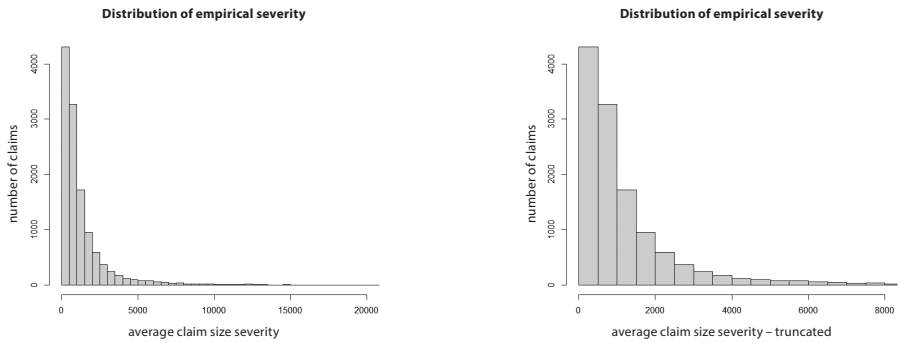
The years at risk *Policy_Exposure* is illustrated in Figure 1 (lhs and middle). For this feature, we have following properties $\min_i v_i = 0.1$ and $\max_i v_i = 1$, that is, minimal time insured in our portfolio is 36.5 days and the maximal insurance time is 1 year. The average insured time is represented as $\sum_i v_i/n = 0.3461$, which corresponds to 126 days. Median time insured is 183 days and only 35.33% of policies are in force the whole year.

Figure 1 Histogram (lhs) and (middle) boxplot of the years at risk, (rhs) histogram of frequencies of whole portfolio dataset *D*



Source: Own construction, customized in R

Figure 2 Histogram of empirical severity (lhs) and histogram of truncated empirical severity over the interval (0; 8 000]



Source: Own construction, customized in R

The heavy tail of the severity distribution is obvious. The average claim size of whole portfolio is 1 300.87 EUR. In practise, when the severity is modelled, it is often useful to provide a large loss threshold to certain claims. This helps to assess the possible thresholds. Presented study does not work with large claims in the dataset *D*.

Before modelling, it is necessary to investigate if and how the explanatory variables should be categorized, and if some of variables should be modelled as the continuous component. Some features used in our models are (highly) non-linear which does not support the log-linear assumption. This is certainly true for the components like policyholder age, vehicle age, vehicle power and volume, etc. Our approach, for these continuous feature components is to group values into intervals, where treated values in the same interval are identical. This approach is based purely on the expert judgement. Next Table 2 shows final predictors with chosen categorization used in presented risk frequency-severity model. In GLMs, it is advised to select the level with maximum exposure as reference for each predictor, because

it minimizes the standard errors of parameter estimates. The sign ® in Table 2 refers to the reference level of the particular predictor.

Table 2 The predictors used in the final step of the frequency and the severity modelling

	Categorical Predictors	*# of Class*	*Multi-level factors*
FREQUENCY MODEL	*Payment_Frequency*	4	1, 2, 4®, 12
	B-M Class	7	*B*0®, *B*1–*B*3, *B*4, *B*5, *B*6, *B*7, *M*1–*M*2
	Region	5	*R*01–*R*04–*R*06–*R*09–*R*11, *R*02–*R*05–*R*10, *R*03®, *R*07, *R*08
	Policyholder_Age	9	18–23, 24–27, 28–31, 32–37®, 38–44, 45–53, 54–61, 62+, *LE*
	Vehicle_Age	9	0, 1, 2, 3, 4®, 5, 6, 7, 8+
	Vehicle_Power	3	0–76®, 77–112, 133+
	TSI	6	0–5 000, 5 001–10 000®, 10 001–15 000, 15 001–25 000, 25 001–35 000, 35 001+
	Vehicle_Engine_Volume	5	0–1 354, 1 355–1 397®, 1 398–1480, 1 481–1 750, 1 751+
	Mileage_per_Year	3	0–15 000®, 15 001–30 000, 30 001+
	Deductible	4	*No Deductible*, ≤ 1%®, ≤ 2%, > 2%
SEVERITY MODEL	*B-M Class*	6	*B*0®, *B*1–*B*2–*B*3, *B*4, *B*5, *B*6–*B*7, *M*1–*M*2
	Region	4	*R*01–*R*07–*R*08, *R*03–*R*04®, *R*02–*R*05–*R*10, *R*06–*R*09–*R*11
	Policyholder_Age	8	18–26®, 27–32, 33–37, 38–45, 46–55, 56–61, 62+, *LE*
	Vehicle_Age	6	0, 1, 2, 3, 4, 5+®
	Vehicle_Power	4	0–80®, 81–95, 96–124, 125+
	TSI	5	0–5 000, 5 001–10 000®, 10 001–15 000, 15 001– 25 000, 25 001+
	Mileage_per_Year	4	0–5 000, 5 001–10 000, 10 001–13 000®, 13 001+

Source: Own construction

2.2 Model building and validation

In previous chapter we started with descriptive statistics on the motor hull portfolio and explanatory data to gain insight on behaviour of the dataset with respect to the number of claims and its subsets with respect to the explanatory variables. As already stated in the last chapter, we will only use 10 predictors in our tarification model; an intercept will be included.

The most frequently used is the backward elimination process, where on intends to reduce the saturated model to a complete model, meaning a model with the best explanatory terms. To begin, all possible variables are included in the model and then the stepwise terms are excluded, every time the term which p-value is bigger than a 5% significance level. The other option is to use the Wald test to check the statistical significance of predictors.

Following Table 3 shows performed Wald test to check relevance of the explanatory variables for final proposed risk models to explain the response variable. It was tested based on the relation (6). Variables *Vehicle_Weight* and *Policyholder_entity* were excluded from both frequency and severity models. Moreover, variables Deductible and *Payment_Frequency* were also removed from the severity model because a p-value of it is lower than a predefined level 5%. These variables do not improve significantly the quality of this model.

Table 3 Wald test of significance of explanatory variables for risk models

Predictors	FREQUENCY_MODEL			SEVERITY_MODEL		
	df	Chisq	Pr(>Chisq)	df	Chisq	Pr(>Chisq)
Intercept	1	702.712	< 2.2e-16	1	17 802.802	< 2.2e-16
Payment Frequency	3	56.432	3.397e-12	-	-	-
B-M Class	6	577.505	< 2.2e-16	5	57.727	3.580e-11
Region	4	327.139	< 2.2e-16	3	16.361	0.0009
Policyholder Age	8	208.724	< 2.2e-16	7	64.065	2.317e-11
Vehicle Age	8	205.724	< 2.2e-16	5	129.799	< 2.2e-16
Vehicle Power	2	24.171	5.64e-06	3	37.222	4.130e08
TSI	5	106.849	< 2.2e-16	4	195.684	< 2.2e-16
Vehicle Engine Vol.	4	23.757	8.934e-05	-	-	-
Mileage per Year	2	60.868	6.062e-14	3	40.405	8.744e-09
Deductible	3	1 621.859	6.854e-12	-	-	-

Source: Own construction, customized in R

When the models were constructed and parameters were estimated (column *Estimate* in Table 4), their significance was tested by Wald test (column p-value in Table 4) defined by (7).

The estimated regression models in Table 4 will be discussed in section 3 but let us first consider the degree of multicollinearity. In our observational study we have many explanatory variables where some relations among them may imply perfect linear combinations with other predictors. In practise, presence of the multicollinearity, regression estimates are unstable and have high standard errors. Variable has a little partial effect because it is predicted well by others. Excluding a nearly redundant predictor can help to reduce standard errors of other estimated effects. To identify potential problem of the collinearity among the explanatory variables we chose according to (Agresti, 2015) variance inflation factors (*VIF*) which measure the inflation in the variances of parameter estimates due to collinearities in the model. A VIF_j of 1 means that there is no correlation among the j-th predictor and remaining predictor variables, and hence the variance of β_j is not inflated at all. These calculations are straightforward and easily comprehensible; if the value of *VIF* is higher than 5 there is a problematic multicollinearity.

In case of this empirical study, the backward selection of variables could produce inconsistent results, variance partitioning analyses may be unable to identify unique sources of the variation, or the parameter estimates may include substantial amounts of uncertainty. In our proposed risk models, we didn't find any *VIF* value higher than 5, that is, no issue with this task.

3 RESULTS AND DISCUSSION

In this part, we present the process results of establishing the risk premium. We follow the standard process in GLMs analysis by separate analyses for the claim frequency and the claim severity. The authors (Ohlsson and Johansson, 2010) state some logical reasons for this separation. In our dataset D, we have an information about the number of claims and the claim costs on policy level with the duration of policy in force measured in years. In the Table 3 are presented the estimated regression coefficients (designated as Estimate) for each category of both proposed risk models, that includes all effects that explain the variation of the claim frequency and costs.

To illustrate, we give an interpretation of the value denoted as $e^{Estimate}$ shown in Table 4, for example, within the *Policyholder* age variable for the *Frequency model*. From the data in this table, we find that the age of the vehicle owner is a significant factor affecting the frequency or the expected number of claims during the year, and as the age of the owner decreases this frequency. Based on the relations (3) and (4) it is possible to formulate the following statements. The most risk category in the policyholder age is between the ages of 18 and 23 ($e^{Estimate} = e^{0.2959} = 1.3443$. For which the expected (average) number of claims during the year is 34.43% greater than in the reference category of 32 to 37 years, and up to 68.16% (1.3443 / 0.7994) greater than in the least risk category 62+. The above statements are based on the assumption that the other factors incorporated in the regression frequency model are at the same level (ceteris paribus). If the owner of the vehicle is a legal entity (LE), the expected number of claims during the year is approximately at category of 28 to 31 years, more precisely 8.2% higher than in the reference category.

Table 4 Analysis of parameter estimates in the risk models

	FREQUENCY MODEL					SEVERITY MODEL				
Predictor	Categories	Estimate	Std. Error	p-value	$e^{Estimate}$	Categories	Estimate	Std. Error	p-value	$e^{Estimate}$
	Intercept	–1.0413	0.0393	0.0000	0.3530	**Intercept**	7.6252	0.0571	0.0000	2 049.1903
Policyholder Age	**18–23**	0.2959	0.0880	0.0008	1.3443	**18–26**	**0.0000**	-	-	**1.0000**
	24–27	0.1791	0.0436	0.0000	1.1961	**27–32**	–0.1674	0.0539	0.0019	0.8459
	28–31	0.0609	0.0326	0.0413	1.0628	**33–37**	–0.2645	0.0539	0.0000	0.7676
	32–37	**0.0000**	-	-	**1.0000**	**38–45**	–0.3526	0.0553	0.0000	0.7029
	38–44	–0.1745	0.0305	0.0000	0.8399	**46–55**	–0.2965	0.0544	0.0000	0.7434
	45–53	–0.1881	0.0312	0.0000	0.8285	**56–61**	–0.2761	0.0594	0.0000	0.7587
	54–61	–0.1948	0.0316	0.0000	0.8230	**62+**	–0.3645	0.0631	0.0000	0.6945
	62+	–0.2239	0.0405	0.0000	0.7994	**LE**	–0.3072	0.0542	0.0000	0.7355
	LE	0.0788	0.0289	0.0064	1.0820					

Table 4 (continuation)

	FREQUENCY MODEL					SEVERITY MODEL				
Predictor	**Categories**	**Estimate**	**Std. Error**	**p-value**	$e^{Estimate}$	**Categories**	**Estimate**	**Std. Error**	**p-value**	$e^{Estimate}$
Vehicle Age	**0**	–0.5923	0.0674	0.0000	0.5531	**0**	–0.5263	0.0764	0.0000	0.5908
	1	–0.3148	0.0380	0.0000	0.7299	**1**	–0.4359	0.0445	0.0000	0.6467
	2	–0.1184	0.0339	0.0005	0.8883	**2**	–0.2655	0.0398	0.0000	0.7668
	3	–0.0890	0.0285	0.0018	0.9148	**3**	–0.0955	0.0336	0.0045	0.9089
	4	**0.0000**	-	-	**1.0000**	**4**	–0.0721	0.0293	0.0137	0.9304
	5	0.0692	0.0276	0.0122	1.0717	**5+**	**0.0000**	-	-	**1.0000**
	6	0.1832	0.0332	0.0000	1.2011					
	7	0.2757	0.0415	0.0000	1.3175					
	8+	0.1879	0.0492	0.0001	1.2067					
Payment Frequency	**1**	–0.1271	0.0216	0.0000	0.8806	n. s.				
	2	–0.0743	0.0277	0.0073	0.9284					
	4	**0.0000**	-	-	**1.0000**					
	12	0.1429	0.0406	0.0004	1.1536					
Vehicle Power	**0–76**	**0.0000**	-	-	**1.0000**	**0–80**	**0.0000**	-	-	**1.0000**
	77–112	0.1065	0.0245	0.0000	1.1124	**81–95**	0.0862	0.0303	0.0045	1.0900
	113+	0.1974	0.0484	0.0000	1.2182	**96–124**	0.1581	0.0407	0.0001	1.1713
						125+	0.3687	0.0647	0.0000	1,4459
TSI	**0–5000**	–0.2373	0.0314	0.0000	0.7888	**0–5 000**	–0.3070	0.0322	0.0000	0.7357
	5001–10 000	**0.0000**	-	-	**1.0000**	5001–10 000	**0.0000**	-	-	**1.0000**
	10 001–15 000	0.2017	0.0266	0.0000	1.2235	**10 001–15 000**	0.1964	0.0300	0.0000	1.2170
	15 001–25 000	0.2083	0.0407	0.0000	1.2316	**15 001–25 000**	0.3630	0.0484	0.0000	1.4376
	25 001–35 000	0.3328	0.0756	0.0000	1.3949	**25 001+**	0.8070	0.0881	0.0000	2.2412
	35 001+	0.3576	0.1014	0.0004	1.4299					
Engine Volume	**0–1354**	0.0835	0,0292	0.0043	1.0871	n.s.				
	1 355–1 397	**0.0000**	-	-	**1.0000**					

Table 4 (continuation)

	FREQUENCY MODEL					SEVERITY MODEL				
Predictor	Categories	Estimate	Std. Error	p-value	$e^{Estimate}$	Categories	Estimate	Std. Error	p-value	$e^{Estimate}$
Engine Volume	**1 398–1 480**	0.0744	0.0328	0.0234	1.0772					
	1 481–1 750	0.0476	0.0287	0.0467	1.0488			n.s.		
	1 751+	0.1437	0.0337	0.0000	1.1545					
Mileage per Year	**0–15 000**	**0.0000**	-	-	**1.0000**	**0–5 000**	–0.1688	0.0533	0,0015	0.8447
	15 000–30 000	0.1141	0.0198	0.0000	1.1209	**5 001–10 000**	–0.1523	0.0332	0,0000	0.8587
	30 000–inf	0.5379	0.0898	0.0000	1.7124	**10 001–13 000**	**0.0000**	-	-	**1.0000**
						13 001+	–0.1276	0.0244	0.0000	0.8802
B-M Class	**B0**	**0.0000**	-	-	**1.0000**	**B0**	**0.0000**	-	-	**1.0000**
	B1-B3	–0.2408	0.0216	0.0000	0.7860	**B1-B2-B3**	–0.1618	0.0265	0.0000	0.8506
	B4	–0.3893	0.0283	0.0000	0.6775	**B4**	–0.1988	0.0346	0.0000	0.8197
	B5	–0.5957	0.0345	0.0000	0.5512	**B5**	–0.1848	0.0414	0.0000	0.8313
	B6	–0.6677	0.0485	0.0000	0.5129	**B6-B7**	–0.2349	0.0556	0.0000	0.7906
	B7	–0.9646	0.1316	0.0000	0.3811	**M1-M2**	–0.1374	0.0531	0.0096	0.8716
	M1-M2	0.1419	0.0439	0.0012	1.1525					
Deductible	**No Deductible**	0.7058	0.0209	0.0000	2.0255					
	<=1%	**0.0000**	-	-	**1.0000**			n.s.		
	<=2%	–0.2518	0.0291	0.0000	0.7774					
	>2%	–0.8400	0.1099	0.0000	0.4317					
Region	**R_A**	–0.4376	0.0247	0.0000	0.6456	**R_C**	–0.0972	0.0249	0.0001	0.9074
	R_B	–0.1266	0.0260	0,0000	0.8811	**R_D**	**0.0000**	-	-	**1.0000**
	R03	**0.0000**	-	-	**1.0000**	**R_E**	–0.0710	0.0319	0.0258	0.9315
	R07	–0.0718	0.0271	0.0080	0.9307	**R_F**	–0.0623	0.0309	0.0438	0.9396
	R08	–0.1905	0.0265	0.0000	0.8265					

Legend: *R_A – R01-R04-R06-R09-R11, R_B – R02-R05-R10, R_C – R01-R07-R08, R_D – R03-R04, R_E – R02-R05-R10, R_F – R06-R09-R11, n. s. – non-significant.*

Source: Own construction, customized in R

Similarly, we can analyse and interpret the expected (average) severity in the context of individual variables. As an example, let's take a situation for the variable vehicle power (*Vehicle_Power*), which is given in the kilowatts (kW). The most risk category in terms of vehicle power consists of vehicles with an engine power of more than 125kW ($e^{Estimate} = e^{0.3687} = 1.4459$). For which the expected (average) severity per year and per policy is 44.59% greater than in the reference category with engine power up to 80kW, provided that the other factors incorporated in the severity regression model are at the same level (ceteris paribus).

The both final risk models introduced in the Table 4 represent the best choice among the other proposed ones. Determining appropriate model is crucial in the regression modelling and the emphasis is on simplicity. In this section, the models with different risk factors are compared based on the analysis of deviance and AIC and BIC, see Table 5.

The several predictive models for frequency and severity has been proposed and tested to find suitable subset of variables in the data set resulting for the best performing model. All predictors in the frequency and severity in MODEL 1 (full model) were processed as categorical variables. Using the stepwise regression with the backward selection strategy the variables *Vehicle_Weight* and *Policyholder_entity* were iteratively removed as least contributive predictors. Afterwards it was tested MODEL 2 for the frequency and severity without these two insignificant variables. In case of the severity MODEL 2 it has been excluded also the variable *Deductible*. According to the results of the analysis of deviance, AIC and BIC, the best model for the claim frequency and severity was chose as MODEL 2 in the both cases.

Table 5 The analysis of deviance, AIC and BIC

Criterion	FREQUENCY		SEVERITY	
	MODEL 1	MODEL 2	MODEL 1	MODEL 2
Deviance	53 517.93	53 518.26	11 877	11 734
AIC	71 380.00	71 375.00	199 914	199 850
BIC	71 842.14	71 808.19	200 284	199 984

Source: Own construction

Regarding to the descriptive data analysis provided in the section 2.1 the real data is not normal distributed, that is, we cannot use ordinary linear regression model. The linear regression model assumes that the outcome of response variable can be expressed by a weight sum of the selected variables with an individual error that follows a normal distribution. Simple weight sum is too restrictive for many real prediction problems. The outcome given the features might have a non-Gaussian distribution, the features might interact and the relationship between the features and the outcome might be nonlinear. This paper deals with estimation of the annual claim frequency and severity in the motor hull insurance based on generalized linear models.

We try to achieve better understanding the relation of the frequency and severity on the presented risk factors. The empirical study results are represented in the Table 4. This particular case study shows that the variables *Vehicle_Weight* and *Policyholder_entity* and *Deductible* have no statistical significance for the annual claim analysis. Based on the principle of simplicity we used the analysis of deviance to choose suitable model. In fact, this model is quite simple, what is very important and useful in the actuarial practice.

To better demonstration of achieved results from the Table 4, it is computed random policyholder profile to set the risk premium, see Table 6.

Table 6 Motor hull insurance: the model results for random selected potential customer profile

Policyholder's properties		Frequency			
	Risk profile	Reg. coeff	$e^{\hat{\beta}_{freq_j}}$	Reg. coeff	$e^{\hat{\beta}_{freq_j}}$
Intercept	1	–1.0413	0.3530	7.6252	2 049.1903
Payment Frequency	12	0.1429	1.1536	0.0000	1.0000
Policyholder Age	28	0.0609	1.0628	–0.1674	0.8459
Vehicle Age	0	–0.5923	0.5531	–0.5263	0.5908
B-M Class	B0	0.0000	1.0000	0.0000	1.0000
Region	R2	–0.1266	0.8811	–0.0710	0.9315
Vehicle Engine Volume	1 420	0.0744	1.0772	0.0000	1.0000
Vehicle Power	78.6	0.1065	1.1124	0.0000	1.0000
TSI	17 300	0.2083	1.2316	0.3630	1.4376
Deductible	<=1%	0.0000	1.0000	0.0000	1.0000
Mileage per Year	7800	0.0000	1.0000	–0.1523	0.8587
$\Pi e^{\hat{\beta}_j}$	×	×	**0.3112**	×	**1 177.6**

Source: Own construction

The frequency model predicts the number of claims for the different categories of the policyholders. General form of this model (see Table 4) is given by:

$$\hat{y}_f = e^{-1.0413} \cdot (e^{0.2959})^{ph\_age\ 18\text{–}23} \cdot (e^{0.1791})^{ph\_age\ 24\text{–}27} \cdot \ldots \cdot (e^{-0.0718})^{region\mathrm{R07}} \cdot (e^{-0.1905})^{region\mathrm{R08}}.$$

The expected claim frequency (the average number of the claims during the year) is then determined for some client with the properties listed in the Table 6 according to the formula:

$$\hat{y}_f = 0.3530 \cdot 1.1536 \cdot 1.0628 \cdot 0.5531 \cdot 1 \cdot 0.8811 \cdot 1.0772 \cdot 1.1124 \cdot 1.2316 \cdot 1 \cdot 1 = 0.3112.$$

The similar form can be expressed for the severity model which predicts the claim costs per policy where the various properties of the policyholder are taken into consideration:

$$\hat{y}_s = e^{7.6252} \cdot (e^{-0.1674})^{ph\_age\ 27\text{–}32} \cdot \ldots \cdot (e^{-0.0623})^{region R\_F}.$$

The expected severity during the year per policy, is then determined for the client with the properties listed in the Table 6 according to the formula:

$$\hat{y}_s = 2\,049.1903 \cdot 1 \cdot 0.8459 \cdot 0.5908 \cdot 1 \cdot 0.9315 \cdot 1 \cdot 1 \cdot 1.4376 \cdot 1 \cdot 0.8587 = 1\,177.6.$$

According to the Formula (1) we can calculate the risk premium for some client as:

RiskPremium = 0.3112 · 1177.6 = 366.5005.

To sum it up, it is proposed GLMs approach to investigate the risks connected with non-life policy. Based on the risk models from section 3.2, estimated premium for the specific risk profile of policyholder is EUR.

CONCLUSION

Motor hull insurance is one of the most widespread insurance in many countries and lots of data is disponible. Process of the setting the price is often difficult exercise since there are many different explanatory variables available. It is also very important that the rating system for set the risk premiums is treated carefully by company. Policyholders may leave when they are overcharged or in the contrary very low price may attract bad risks.

We have discussed in the paper the use of generalized linear models in actuarial practise which represent a suitable tool to predict key ratios, like the claim frequency, claim severity and the risk premium. GLMs are very effective because they are fairly accurate and are easy to explain to the layman in terms of the effect of each rating factor. Classification of the observed losses according to the appropriate risk factors is very important in determining how accurate the rating system is, the risk factors tells us exactly which level of which risk factor causes the biggest loss – should be charged the highest risk premium and which causes the smallest loss should be the lowest premium. The core concept of GLMs is to keep the weighted sum of features but allow non-Gaussian outcome distributions and connect the expected mean and the weighted sum through a possibly non-linear function.

At the first stage, the frequency of claims is estimated using the Poisson regression. In the next stage, the severity is determined by Gamma model where the log-link function is defined in both cases. The risk premium can be then expressed as the product of the expected claim counts and average cost per claim. Since all the weights are in the exponential function, the effect interpretation is not additive, but multiplicative. The regression coefficients as resulting from the frequency-severity model presented in the Table 4 can be also not continuous or their progress is not smooth enough which can be caused by inadequate accuracy, but also the data that does not have the behave how we would be expected. In practice this happen very often, when some factors really reflect an increasing or decreasing risk.

Apart from the general risk factors as *Policyholder age*, *Vehicle age*, *TSI*, etc…, we tend to classify the observed losses according to the Bonus-Malus system variable. This system leads to a discount – bonus in risk premium. When the claims have occurred the premium increases as the consequence of it – malus, see Table 5.

We processed a dataset with n = 91 685 policies. According to descriptive analyses provided in the initial section of the empirical study we see, that histogram of the claim frequency and claim severity is strongly right-skew, see the Figure 1 and Figure 2. It follows from this that ordinary linear regression is not fully suitable. The policyholders are divided into the groups based on the risk factors, see Table 2. According to these 10 risk factors, we get 192 000 groups. Exposure, total number of claims and total claim amount is known for each group. The variables *Vehicle_Weight* and *Policyholder_entity* are statistically insignificant and rejected at significance level of 0.05 in the risk model. The next variable Deductible is rejected just for claim severity model.

The actuaries should be aware of the so-called "one-dimensional analysis" and should not be tempted to stop the analysis in finding the averages of responses caused by each risk factor in our portfolio. The reason is very justified, these risk factors are very likely to be correlated.

We try to find the suitable GLMs for the claim frequency and claim severity in term of the risk factors. The models with different risk factors are constructed and compared each other using the analysis of deviance AIC and BIC criterion. The best risk models are those that have the lowest decision criterions compared to others that is MODEL 2 in both cases, see Table 5.

ACKNOWLEDGMENT

The paper was supported by a grant agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic VEGA No. 1/0120/18 – *Modern risk management tools in the internal models of insurance companies in Solvency II*.

The paper was supported by a grant agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic VEGA No. 1/0647/19 – *Modern tools for managing and modelling of risks in non-life insurance*.

The paper was supported by the internal grant project I-19-102-00 of the University of Economics in Bratislava for young pedagogical staff, scientific and PhD students entitled *Modern stochastic methods applied in tourism in Slovak Republic*.

References

AGRESTI, A. *Foundations of linear and generalized linear models.* New York: John Wiley & Sons, 2015.

ALLISON, P. D. *Logistic regression using SAS: Theory and application.* 2nd Ed. North Carolina: SAS Institute, 2012.

ANDERSON, D., FELDBLUM, S., MODLIN, C., SCHIRMACHER, D., SCHIRMACHER, E., THANDI, N. *A Practitioner's Guide to Generalized Linear Models: A foundation for theory, interpretation and application.* 3rd Ed. Towers Watson, 2007.

CHARPENTIER, A. AND DENUIT, M. *Mathématiques de l'Assurance Non-Vie, Tome II: Tarification et provisionnement.* Paris: Economica, 2005.

DAVID, M. Auto insurance premium calculation using generalized linear models. *Procedia Economics and Finance*, 2015, 20(1), pp. 147–156.

DE JONG, P. AND HELLER, G. Z. *Generalized linear models for insurance data.* Cambridge: Cambridge University Press, 2008.

DOWLE, M., SRINIVASAN, A, SHORT, T., LIANOGLOU, S, SAPORTA, R., ANTONYAN, E. *data.table: Extension of Data. frame.* R package, 2015.

DRAPER, N. AND SMITH, H. *Applied Regression Analysis.* 2nd Ed. New York: Wiley, 1981.

DUAN, Z., CHANG, Y., WANG, Q., CHEN, T., ZHAO, Q. A Logistic Regression Based Auto Insurance Rate-Making Model Designed for the Insurance Rate Reform. *International Journal of Financial Studies*, 2018, 6(1), p. 18.

EDWARD, W. F. *Regression Modeling with Actuarial and Financial Applications.* Cambridge University Press, 2010.

EWALD, M. AND WANG, Q. *Predictive Modeling: A Modeler's Introspection – A Paper Describing How to Model and How to Think Like a Modeler* [online]. Schaumburg: Society of Actuaries, 2015. [cit. 9.9.2019] <*https://www.soa.org/globalassets/assets/files/research/projects/2015-predictive-modeling.pdf*>.

FORBES, C., EVANS, M., HASTINGS, N., PEACOCK, B. *Statistical distributions.* New York: John Wiley & Sons, 2011.

FOX, J. *Applied regression analysis and generalized linear models.* New York: Sage Publications, 2015.

FREES, E., LEE, G., YANG, L. Multivariate frequency-severity regression models in insurance. *Risks*, 2016, 4(1), p. 4.

GARRIDO, J., GENEST, C., SCHULZ, J. Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 2016, 70(1), pp. 205–215.

GOLDBURD, M., KHARE, A., TEVET, D. *Generalized linear models for insurance rating.* Casualty Actuarial Society, CAS Monographs Series 5, 2016.

HABERMAN, S. AND RENSHAW, A. E. Generalized linear models and actuarial science. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1996, 45(4), pp. 407–436.

HEBÁK, P., HUSTOPECKÝ, J., MALÁ, I. *Vícerozměrné statistické metody (2).* Prague: Informatorium, 2005.

KAFKOVÁ, S. AND KŘIVÁNKOVÁ, L. Generalized Linear Models in Vehicle Insurance. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 2014, 62(2), pp. 383–388.

KIM, K. AND TIMM, N. *Univariate and multivariate general linear models: theory and applications with SAS.* Boca Raton: Chapman and Hall/CRC, 2006.

LITTELL, C. L., STROUP, W. W., FREUND, R. J. *SAS for Linear Models.* 4th Ed. North Carolina: SAS Institute, 2010.

MCCULLAGH, P. AND NELDER, J. A. *Generalized linear models.* 2nd Ed. London: Chapman and Hall, 1989.

NELDER, J. A. AND WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 1972, 135(3), pp. 370–384.

OHLSSON, E. AND JOHANSSON, B. *Non-Life Insurance Pricing with Generalized Linear Models. EAA Lecture Notes.* Berlin: Springer, 2010.

R CORE TEAM. R: *A language and environment for statistical computing* [online]. R Foundation for Statistical Computing, Vienna, Austria, 2019. <http://www.R-project.org/>.

SHI, P., FENG, X., IVANTSOVA, A. Dependent frequency – severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 2015, 64(1), pp. 417–428.

VALECKÝ, J. Modelling Claim Frequency in Vehicle Insurance. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 2016, 64(2), pp. 683–689.

VALECKÝ, J. Calculation of solvency capital requirements for non-life underwriting risk using generalized linear models. *Prague Economic Papers*, 2017, 26(4), pp. 450–466.

VENABLES, W. N. AND RIPLEY, B. D. *Modern Applied Statistics with S.* 4th Ed. New York: Springer, 2002.

WOOLDRIDGE, J. M. *Introductory econometrics: a modern approach.* 5th Ed. Mason: South-Western, 2013.

XACUR, O. A. Q. AND GARRIDO, J. Generalised linear models for aggregate claims: to tweedie or not? *European Actuarial Journal*, 2015, 5(1), pp. 181–202.

XIE, S. AND LAWNICZAK, A. Estimating Major Risk Factor Relativities in Rate Filings Using Generalized Linear Models. *International Journal of Financial Studies*, 2018, 6(4), p. 84.