

# Utilization of Regression and Correlation Analysis for Optimal Composition of Factors Affecting Premium Rates of Compulsory Contractual Insurance

Alena Košťálová

Department of Communications  
University of Žilina, FPEDAS  
Žilina, Slovak Republic

Email: alena.kostalova@fpedas.uniza.sk

Michal Pavličko

Dept. of Quantitative Methods and Economic Informatics  
University of Žilina, FPEDAS  
Žilina, Slovak Republic

Email: michal.pavlicko@fpedas.uniza.sk

**Abstract**—The paper is dedicated to the proposal of optimal composition of factors to determine the premium rates of the compulsory contractual motor third party liability insurance in the selected insurance companies. The main aim of this paper is to find dependence of the premium rates on each factor by usage of regression and correlation analysis. The results of analysis show the optimal factors for selected companies.

**Keywords**—insurance; compulsory contractual insurance; motor third party liability insurance; regression analysis; correlation analysis.

## I. INTRODUCTION

Cars became an integral part of everybody's life. They help us to save our time and they bring us comfort and independence. On the other hand, rising amount of vehicles increases the traffic density that also affects rising risk of accidents. Insurance companies offer some products for reduction of negative effects that can occur by driving a car. Compulsory contractual motor third party liability insurance and voluntary motor hull insurance belongs among such products. Every car owner has to sign a contract with chosen insurance company in a case of compulsory insurance. Main task of this product is to ensure accident compensation caused by car owner to the third party. The insurance companies take different factors into account when they calculate the insurance costs. Many of these factors don't need to affect insurance costs but they can have some psychological effect on potential client, who can believe that the more factors are taken into account by insurance cost calculation the more suitable product is made "just for him". This paper is dedicated to looking for optimal composition of the factors that affect the amount of insurance costs of compulsory contractual motor third party liability insurance by means of correlation and regression analysis.

## II. THEORETICAL BACKGROUND

The insurance can be defined as the relationship between the insurer and the insured. One contracting party (insurer) guarantees to the other contracting party (insured or policyholder) that it will compensate insurance costs of the

Grant support: 1/0721/15 – Research on the impact of postal services and telecommunication convergence on regulatory approaches in the postal sector.

claim in the case of occurrence of insurance event under previously agreed conditions. [3]

The literature sources can provide some classification of insurance according to different criteria such as insurance type, level of the risk and obligation of insurance formation. According to obligation of insurance formation, the insurance can be compulsory and contractual. Compulsory insurance is such an insurance type, that conditions and insurance liability is set by the law. Contractual insurance can be divided into:

- Compulsory contractual insurance – is such an insurance type, that obligation of insurance formation is set by the law, but insurance conditions are dependent on insurance contract.
- Voluntary contractual insurance – is voluntary type of insurance. Acceptation of insurance contract and its insurance conditions isn't compulsory. [3]

### A. Compulsory contractual insurance against civil liability in respect of the use of motor vehicles

Typical sign of this type of insurance is that insurance cover is paid to the damaged party as the third party, who is neither insured party nor policyholder. If insured party causes the damages to the third party and the third party claims compensation costs for accident event then the insured party have to contact his insurance company which finally decides about covering of compensation claim. [3]

Until the year 2001, Slovenská poisťovňa was monopoly provider of the compulsory contractual insurance against civil liability in respect of the use of motor vehicles also known as the motor third party liability insurance (thereinafter MTPL) in Slovak Republic. The situation was changed after adoption of law no. 381/2001 of MTPL. The monopoly of Slovenská poisťovňa was replaced by the system of compulsory contractual insurance and free choice of insurance provider. The law specifies the entities that must conclude MTPL (holder of vehicle registered in documents of domestic motor vehicle, owner or provider of domestic vehicle, tenant of domestic vehicle, driver of foreign vehicle).

This type of insurance doesn't relate to damage compensation of vehicle of insured party if accident is caused

by insured party. Voluntary motor hull insurance is used for covering of damage compensation in such cases. This type of insurance is also used as liability insurance in foreign countries if insured party asks and gets so-called “green card” from insurance company.

MTPL covers the damage compensation (healthcare costs of injured party or expenses in case of death, expenses caused by damaging, theft, destruction or loss of things, objective costs related to legal representation of claim administration, claims of loss of profit) that is caused by motor vehicle operation of insured party to damaged third party.

### III. INSURANCE BUSINESS IN SLOVAK REPUBLIC

The National Bank of Slovakia is supervising the insurance market of Slovak republic. The main activity of insurance market supervision is mainly observing of the Insurance Act or any other valid legislative rules and conditions resulting from legislative norms. [4]

The Insurance Act and the Civil Code are fundamental legislation that contains the characteristics of entities that provides insurance services, type of business and specifications of insurance market supervision and regulation. Furthermore, it regulates the rights and obligations of each participating party, the manner of management of commercial insurance companies and other important facts. [1]

Insurance regulation together with supervision represents an interference of economy operation of insurance company. Supervisors have such rights and permissions that authorize them to suggest the solutions for commercial insurance companies in the case if they are acting incorrectly. They also make suggestions to remove discrepancies with the legislation. In developed economics of Western Europe countries, the supervision of insurance market is performed by several institutions.

In scope of economic integration of the European Union the insurance business is very hard to regulate and the unification of the principles and laws of each country is very complicated. There can be mentioned two extremes as an example; the first one is German system where the supervision is very precisely specified and as an opposite there is very liberal concept of supervision in England and Nederland. [1]

### IV. OBJECTIVE AND METHODOLOGY

The aim of presented paper lies in suggestion of optimal composition of the factors that have the greatest impact on the premium rates of MTPL in selected insurance companies. Methods of regression and correlation analysis are used to achieve this aim.

The dependence of the premium rates on individual factors is modulated on selected insurance companies such as Allianz Slovenská poisťovňa a. s. (thereinafter ALLNZ), Kooperatíva poisťovňa a. s. (thereinafter KOOP) and Wüstenrot poisťovňa, a. s. (thereinafter WU).

In order to fulfil the stated objective, it was needed to create a scenario of modulated clients with different parameters such as the trade-mark of vehicles, cylinder capacity in  $\text{cm}^3$ , engine

power in kW, the age of the policyholder and production year of the vehicle

Equal input conditions were ensured for such specified modulated clients (annually payed premium, i.e. once per year, the vehicle is used solely for the normal operation, no participation on damages, i.e. the amount of premium isn't reduced by co-participation).

Subsequently, by means of possibility to contract MTPL through webpages of selected insurance companies the amount of premium was investigated. These data served as an input for the implementation of the survey of the dependence of the premium rates on individual factors.

The research, examination and evaluation of dependence, especially causal dependence among two or more statistical signs, are very important tasks of statistics. The statistical dependence arises in a case when one phenomenon is the cause that results in formation of the second phenomenon. The statistical methods such as regression and correlation analysis are the most used methods for such a survey. [5]

#### A. Regression analysis

Regression expresses mutual relationship among two or more variables, especially numeric ones, by using a model characterizing their interdependence. [5]

Two kinds of variables exist in the regression model. [6]

Dependent variables – are representing the numeric variables, which dependence on other variables is examined. These variables are referred to as  $Y$ , and its observations are referred to as  $y_i$ , where  $i = 1, 2, \dots, n$ , where  $n$  is the number of observations in the sample.

Independent variables – are such variables, which are expected to cause changes in a dependent variable  $Y$  and they are used to estimate values of dependent variable. They are referred to as  $X_1, X_2, \dots, X_k$ , and their observations are referred to as  $x_{i1}, x_{i2}, \dots, x_{ik}$ , where  $i = 1, 2, \dots, n$ .

The regression analysis is divided according to the number of dependent and independent variables. According to the number of dependent variables there exist one-equation models with one dependent variable is denoted  $Y$  and multi-equation models with two or more dependent variables. According to the number of independent variables there exists simple linear regression with one independent variable is denoted  $X$  and multiple linear regression with two or more independent variables that affect the dependent variable is denoted  $Y$ . There is an assumption about existence of causal dependency between dependent variable and independent variables. It means that one event or a group of events, known as the cause, generate another event, so called the effect. [6]

In the simple linear regression model, the dependent variable  $y$  is expressed as a function of independent variable  $x$ . Regression line of variable  $y$  in respect to variable  $x$  can be entered in following format: [7]

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (1)$$

where:

- $y_i$  – represents the  $i^{\text{th}}$  value of estimated variable,
- $\beta_0, \beta_1$  – are unknown parameters of regression model, so called regression intercept,
- $x_i$  – represents the  $i^{\text{th}}$  value of the independent variable,
- $\varepsilon_i$  – is called error term, disturbance or noise that refers to the error of the  $i^{\text{th}}$  measurement.

The multiple linear regression model with several independent variables can be characterized with following formula: [6]

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_3 x_{i3} + \varepsilon_i, \quad (2)$$

where:

- $\beta_0, \beta_1, \dots, \beta_k$  – are unknown parameters of regression model,
- $x_{ij}$  – represents the  $i^{\text{th}}$  value of independent variable  $X_j$  in the  $i^{\text{th}}$  observation ( $i = 1, 2, \dots, n$ ),

The quality of the model is described by various characteristics such as the Residual Sum of Squares – RSS (3) and the Coefficient of Determination –  $R^2$  (4) that represents the ratio of the overall model fit. [6]

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

The selection of model ( $M_1$ ) or sub-model ( $M_2$ ) is implemented on the bases of the  $F$ -test of the sub-model. Let the model  $M_1$  to has  $k_1$  of the estimated parameters and let the sub-model  $M_2$  to has  $k_2$  estimated parameters, while  $k_2 < k_1$ . The comparative characteristic is the  $F$ -statistic with 95% percentile of  $F$ -distribution with  $k_1 - k_2$  and  $n - k_1$  degrees of freedom. Corresponding quantile is denoted  $F_{0.95}(k_1 - k_2; n - k_1)$ .  $F$ -statistic is calculated with following formula: [6]

$$F = \frac{\frac{RSS_2 - RSS_1}{k_1 - k_2}}{\frac{RSS_1}{n - k_1}}, \quad (5)$$

where  $RSS_1$  is residual sum of squares in model  $M_1$  and  $RSS_2$  is residual sum of squares in sub-model  $M_2$ . If the proposition  $F > F_{0.95}(k_1 - k_2; n - k_1)$  is true then we incline to the model  $M_1$  else we incline to the model  $M_2$ .

The analysis begins with inclusion of all possible quantitative factors into the most numerous model and then one

by one, the unimportant factors are continually eliminated according to the results of  $F$ -test sub-model. The elimination of factors lasts until the momentum when all remaining factors are important, i.e. statistically significant (stepwise regression, backward elimination). [6]

### B. Correlation analysis

The main task of correlation analysis is the determination of regression function quality and the power of dependence. The stronger is the relationship between dependent and independent variables and also the better is the regression function the more concentrated are the empiric values around the values calculated by the regression function. The Pearson product-moment correlation coefficient, denoted  $r_{xy}$ , is usually used for measurement of the linear regression. [5]

The correlation coefficient can gain values in interval from  $-1$  to  $+1$  (marginal values included). If the correlation coefficient equals to  $+1$  it means total positive linear dependence. If the correlation coefficient equals to  $-1$  it means total negative correlation, in other words there is total inverse dependence between the variables. If it equals to  $0$  there doesn't exist the linear dependence between the variables. If the values of correlation coefficient are getting close to  $-1$  or  $+1$  it means there exists weaker or stronger dependence. If the values are getting close to  $0$  it doesn't need to mean there is no dependence between events but it means there is no linear dependence, on the other hand there can exists strong nonlinear dependence. For the correlation coefficient applies  $r_{yx} = r_{xy}$  and  $r_{xy}^2 = R^2$ , where  $R^2$  is the coefficient of determination of the simple linear regression model. [5]

Regression and correlation analysis evaluates mutual relationship (correlation) between dependent and independent variables. They are in the relationship of free causal dependence. [7]

## V. RESULTS AND DISCUSSION

The content of this part of our paper is examination of the dependence of MTPL insurance cost ( $y$ ) on several quantitative factors ( $x_i$ ) which affect premium rates of individual selected insurance companies by the simple and multiple regression analysis. The parameter estimations such as the interval estimations with 95% reliability, the statistical significance and the coefficient of determination for the power of dependence were determined for each model. On the basis of already mentioned methodology (previous chapter), there were suggested the following optimal factors, it means such factors that significantly affect premium rates of MTPL in selected insurance companies.

### A. The examination of dependence of premium rates of MTPL on individual factors for ALLNZ company

The quantitative factors, that affect premium rates of MTPL in ALLNZ company, are: cylinder capacity in  $\text{cm}^3$ ; engine power in kW; the age of the policyholder and production year of the vehicle. At the beginning, there was designed multiple linear regression model describing the dependence of insurance price ( $y$ ) on all considered factors together ( $x_1$  – cylinder capacity in  $\text{cm}^3$ ;  $x_2$  – engine power in kW;  $x_3$  – the age of the

policyholder;  $x_4$  – production year of the vehicle). The dependence of insurance price on listed factors can be modulated by multiple linear model  $M_4$ .

$$y \approx 154.45 + 55.64x_1 + 0.29x_2 - 3.28x_3 - 0.06x_4 \text{ (€)} \quad (6)$$

Results of the  $F$ -test showed that the parameters  $\beta_1$ ,  $\beta_2$  and  $\beta_4$  are insignificant because their  $F$ -statistic values didn't exceed the critical values. The most insignificant parameter of them was  $\beta_4$  parameter because of the lowest value of  $F$ -statistic. After that the model  $M_4$  were reduced to sub-model  $M_3$ .

$$y \approx 153.80 + 54.66x_1 + 0.31x_2 - 3.28x_3 \text{ (€)} \quad (7)$$

Next, the sub-model  $M_3$  showed insignificant parameters  $\beta_1$  and  $\beta_2$ , while the most insignificant parameter was  $\beta_2$ , therefore the model  $M_3$  were reduced to its sub-model  $M_2$ .

$$y \approx 154.07 + 71.67x_1 - 3.37x_3 \text{ (€)} \quad (8)$$

The sub-model  $M_2$  can be considered as the final one according to the fact that all remaining parameters were statistically significant. There can be stated that the optimal factors for the calculation of premium rate in ALLNZ company are the cylinder capacity in  $\text{cm}^3$  and the age of the policyholder.

#### B. The examination of dependence of premium rates of MTPL on individual factors for KOOP company

There is only one quantitative factor affecting insurance premium of MTPL in KOOP company and it is the cylinder capacity in  $\text{cm}^3$ . There was designed simple linear regression model because of examination of just one factor that affects premium rates. The dependence of insurance price ( $y$ ) on the cylinder capacity in  $\text{cm}^3(x)$  modulates following regression line.

$$y \approx 14.48 + 54.27x_1 \text{ (€)} \quad (9)$$

Next, the significance of the model was tested by  $F$ -test (5% level of significance). The  $F$ -statistic exceeded the critical value and proved the model of dependence of insurance price on the cylinder capacity in  $\text{cm}^3$ . This is the optimal factor for calculation of premium rates in KOOP company.

#### C. The examination of dependence of premium rates of MTPL on individual factors for WU company

Two quantitative factors, that affect the premium rates of MTPL in WU company, are: engine power in kW and the age of the policyholder

At the beginning, the multiple linear regression model was designed for dependence of the insurance price ( $y$ ) on both considered factors ( $x_1$  – engine power in kW;  $x_2$  – the age of the policyholder).

$$y \approx 103.58 + 1.45x_1 - 2.13x_2 \text{ (€)} \quad (10)$$

According to the fact, that both estimated parameters  $\beta_1$  and  $\beta_2$  are statistically significant, this model can be considered as the final one. In WU company, the optimal factors for insurance premium calculation of MTPL are the engine power in kW and the age of the policyholder.

## VI. CONCLUSION

The result of multiple regression model for ALLNZ company is the fact that just two parameters are statistically significant such as the cylinder capacity in  $\text{cm}^3$  and the age of the policyholder. Other two parameters are statistically insignificant therefore they don't affect calculation of the premium rates. The linear model explains just 53% of variability in the insurance price.

On the other hand, in the case of KOOP company, there exists just one factor affecting the insurance price and it is the cylinder capacity in  $\text{cm}^3$ . The linear regression model explains approximately 85% of variability in the insurance price.

Third examined entity was WU company. The premium rates are affected by two factors such as the engine power in kW and the age of the policyholder. Both of these factors are statistically significant. The linear regression model explained the dependence of insurance price on these two factors which have approximately 57% impact to variability of premium rates.

The comparison of estimated regression coefficient in final multiple regression models to the age of the policyholder in WU company and ALLNZ company brings very interesting finding. The results show that increased age of the policyholder causes greater fall of insurance price in ALLNZ company than in WU company.

The more detailed analysis of the regression models showed another interesting finding. The factor of the age of policyholder is statistically more significant than the cylinder capacity in  $\text{cm}^3$  in ALLNZ company. On the other hand, the factor of age of the policyholder is statistically less significant than the factor of engine power in kW in WU company. It means, that the age of policyholder affects the premium rates of MTPL more than the power of engine in ALLNZ company, while in WU company it is reversed.

As it was mentioned in the introduction of our paper, different factors are taken into account by the insurance companies, but some of the factors are used just in the scope of competitive battle for customers and they needn't affect the premium rates significantly. This was also confirmed by examination of optimal composition of factors in one of three monitored insurance companies.

## REFERENCES

- [1] V. Čejková, D. Martinovičová, S. Nečas, "Poistný trh", Bratislava: Iura Edition, 2011, pp. 222, ISBN 978-80-8078-399-0.
- [2] A. Košťálová, M. Kováčiková, L. Švábová, "Poistovníctvo", Žilina: EDIS, 2008, pp. 207, ISBN 978-80-8070-805-4.
- [3] A. Majtánová et al., "Poistovníctvo", Bratislava: Iura Edition, 2009, pp. 327, ISBN 978-80-8078-260-3.
- [4] V. Čejková, D. Martinovičová, "Poistovníctvo: Úvod do teórie", Bratislava: Merkury, 2007, pp. 138, ISBN 978-80-89143-60-3.
- [5] D. Chlebková, "Ekonomická štatistika", Žilina: EDIS, 2009, pp. 153, ISBN 978-80-554-0058-7.
- [6] E. Šoltés, "Regresná a korelačná analýza s aplikáciami", Bratislava: Iura Edition, 2008, pp. 287, ISBN 978-80-8078-163-7.
- [7] V. Palacková et al., "Štatistika pre ekonómov", Bratislava: Iura Edition, 2003, pp. 358, ISBN 80-89047-74-2.