MULTIVARIATE MODELLING OF MOTOR THIRD PARTY LIABILITY INSURANCE CLAIMS

Aivars Spilbergs[1], Andris Fomins[1], Māris Krastiņš[1]

[1] *BA School of Business and Finance, Riga, Latvia*

EUROPEAN JOURNAL OF BUSINESS SCIENCE AND TECHNOLOGY

Volume 8 Issue 1
ISSN 2694-7161
www.ejobsat.com

ABSTRACT

The aim of the study is to identity the main factors that affect claims amount paid by insurers in case of road accidents and to predict losses from valid third-party liability insurance (MTPLI) policies until their expiration. Such an assessment is essential to adequately cover MTPLI policies and ensure the sustainable development of insurance companies.

The geography of the study covers the MTPLI market of Europe in the main areas, but a deeper analysis of the impact of various factors, interactions, and interrelationships in MTPLI product is focused on Latvian market data due to availability of high-quality primary data.

The research is based on the analysis of primary Latvian MTPLI policies data of more than 128,000 road traffic accidents that have occurred during the time period from 2014 till 2020. Risk driver selection was performed based on the existing scientific studies and correlation analysis of the sample set. Both linear and nonlinear forms of relationships were used for modelling. A multivariate modeling was used to identify significant risk factors and to quantify their impact on loss of incidents.

Statistical stability of the models was tested using chi-squared, t-tests and p-values. Validation of models calibrated where done using prediction errors measurements: mean square error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) assessment both within sample and out of sample technics. The results indicated that the driver's behavior (penalties and Bonus-Malus) as well as vehicle parameters (weight and age), had significant impacts on crash losses.

KEY WORDS

road traffic accidents, risk drivers, non-life insurance, MTPL insurance, private insurance, passenger cars, Bonus-Malus system, MTPL insurance claims paid, multivariate modelling

JEL CODES

G22, C38

SPILBERGS, Aivars, FOMINS, Andris, and KRASTIŅŠ, Māris. 2022. Multivariate Modelling of Motor Third Party Liability Insurance Claims. *European Journal of Business Science and Technology,* 8 (1): 5–18.
ISSN 2694-7161, DOI 10.11118/ejobsat.2022.002.

1 INTRODUCTION

Road safety is not only a public health concern, but a seroius socioeconomic issue as well. The World Health Organization (WHO) estimates that about 1.3 million people lose their lives annually in road traffic accidents and millions of people sustaining injuries (World Health Organization, 2021). Road traffic accidents globally are the main cause of death among young people aged up to 29 years (World Health Organization, 2021). Although the number of road traffic fatalities in EU has dropped significantly over the past 20 years, about 22,800 people died in road traffic accidents in the EU in 2019: 44.2% of those killed were car drivers or passengers; 20.2% were pedestrians (Eurostat, 2021).

Motor premiums accounted for 36% of all property and casualty (P&C) insurance in Europe in 2019 and totaled €147 billion, therefore is Europe's most widely purchased P&C insurance line (Insurance Europe, 2021). That has been demonstrated by various studies of European Motor insurance market and has high degree of actuality for insurance business, as well as for the researchers.

The European insurance market is not homogeneous. Significant differences in the development of this market segment can be observed, including opposite trends. While in the Baltic states (Latvia and Estonia) a decrease in insurance premiums is observed in 2020, Slovakia and the Czech Republic show significant increase of premiums. If we look at the stability of the trends, the Baltic region stands out too, in a relatively short period of time, completely opposite trends can be observed – both in Latvia and Estonia there are periods of the largest increase in premiums (more than 15% in 2017–2019), but in the following period the largest decrease in premiums can be observed (10% in Latvia in 2020). Similar trends could be observed during the 2008 crisis in the Baltic insurance markets – from record market growth to record decrease, especially in the Latvian market (Insurance Europe, 2021).

Portugal, Finland, Sweden, Slovakia, and the Czech Republic show a positive trend in increase of insurance premiums for motor insurance, compared to the opposite situation that can be observed in Greece, Poland, Italy, Estonia. Very strong negative trend has been observed in Latvia, 10% decrease in 2020 compared to the period of 2017–2019. As for the share of insurance claims, historical fluctuations in the opposite directions can be observed in almost all European countries. If the period 2017–2019 clearly showed an increase in insurance claims paid (growth more than +4%), then for most European countries 2020 brought an even greater reduction in claims (more than 5%; see Fig. 2), apart from Latvia and Bulgaria, where we can observe a continuing increase in claims (Insurance Europe, 2021).

Analyzing the motor claims paid (Fig. 2), we can observe a decrease in claims paid in 2020 in most EU countries, although an opposite trend can be observed in Latvia and Bulgaria, where motor claims have increased in 2020 compared to the period of 2017–2019. Thus, it is obvious that the trend observed in countries such as Bulgaria and Latvia is definitely negative for the participants of the insurance market – insurance companies. In Europe overall, most of the countries show a decline in insurance premiums (negative trend for the insurance industry). At the same time a decrease in the claims (which has a positive effect on the insurance industry as a whole) can be observed. The decrease in claims is larger than the decrease in insurancepremiums, that means increase in profitability of insurance companies. In this respect Latvia stands out as well, where both the decrease in insurance premiums and the increase in claims clearly indicates negative trend in the insurance market (FKTK, 2021).

Taking into account the specifics of the study, the MTPLI claims paid amounts per road traffic accident in Latvia during 2014–2019 was analyzed, see Fig. 3.

As shown in Fig. 3, the amount of claims paid per RTA in Latvia is increasing, and there is a steady upward trend in claims (approx. 20% increase 2019 vs. 2014). The above trend points

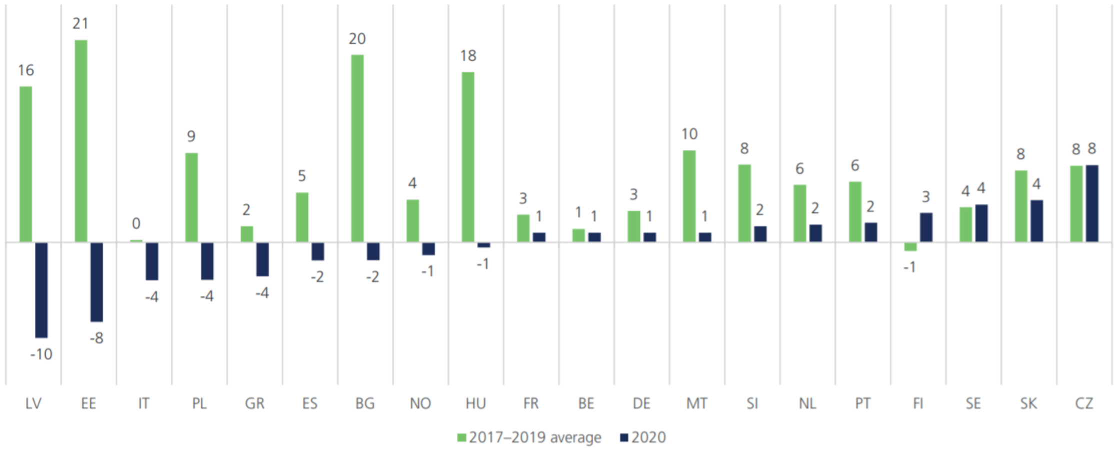

Fig. 1: Year-to-year change in motor premiums in Europe, 2020 compared to 2017–2019 average (%). Source: Insurance Europe, 2021.

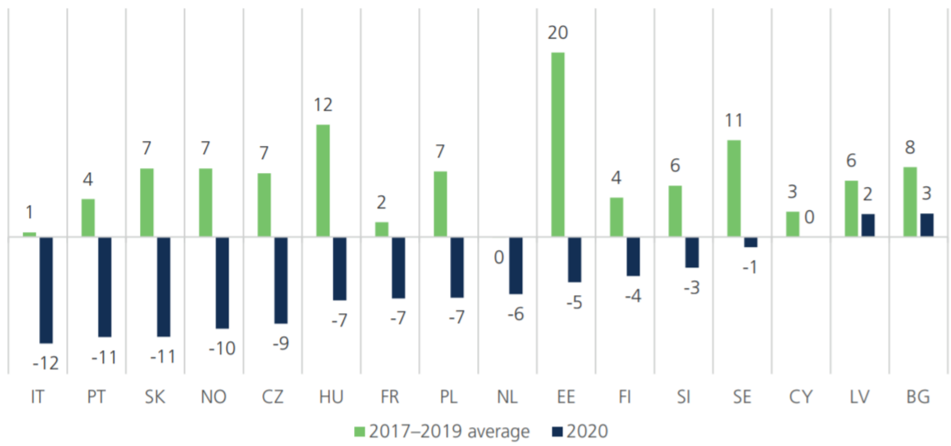

Fig. 2: Year-to-year change in motor claims in Europe, 2020 compared to 2017–2019 average (%). Source: Insurance Europe, 2021.

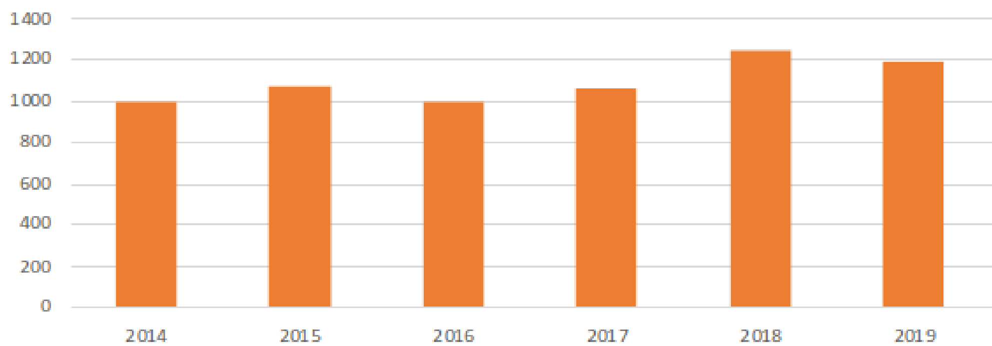

Fig. 3: MTPLI claims paid amounts in Latvia per road traffic accident, EUR. Source: Created by authors based on MIB data.

to the need for a more accurate and detailed analysis of the causes of insurance costs.

We can conclude that the market situation is unstable, and an important task for insurers is to diversify clients even more carefully, by analyzing factors that affect the risk level of each specific group of clients, which, in turn, will increase the overall profitability of the company's insurance portfolio. In this study the authors evaluate the influence of different factors on the size of MTPLI claims, which will help to understand the main risk factors for this insurance product. The authors performed analysis of claim severity in motor third party liability insurance by using the general linear model.

Generally two approaches are being used to determine net premiums in non-life insurance. Either the target variable is equal to the net premium (euros of loss per exposure) or it is separately modelled the claims frequency (number of claims per exposure) and the claim severity (average loss per claim).

2 LITERATURE REVIEW

2.1 Risk Factors Influence on Claim Severity in Road Traffic Accidents

One of the insurers' main concerns is establishing a tariff structure that distributes these claims and losses among policy holders most equitably and reasonably. This task of determining the pure premium belongs predominantly to actuaries who evaluate the probability of the risk occurrence, determine the risk factors in order to establish commensurate tariffs for each class so that everyone and each pay premium that, in one way or the other, reflects their riskiness.

Insurance principle of sharing danger or risk among a group of individuals requires grouping individual risks into various categories or classes with a homogeneous set of characteristics. Such classification usually boils down to different classes for which members of a given class share the same set of risk characteristics; in that respect, each class will have a certain number of insurance claims and accumulated losses and pay the same premium rate. To reasonably estimate this premium and price insurance policies, insurers must predict the expected loss accurately, often referred to as the "pure premium" (El Kassimi and Zahi, 2021).

In estimating claims distributions for motor insurance, the cost of claims is often associated with two components: the probability of an accident and the amount of claim per accident, if it occurs. Claims frequency and severity components are being termed by actuaries. This is the traditional way of decomposing this so-called "two-part" data, where one can think of a zero as arising from a vehicle without a claim. This approach allows to incorporate multiple claims per vehicle (Frees and Valdez, 2008).

Denuit and Lang have observed, that technical profit margins in European motor third party liability insurance (MTPLI) markets have been very small. In this situation the actuaries have to design a tariff structure that will fairly distribute the burden of claims among policyholders (Denuit and Lang, 2004).

According to Oh et al. (2020), the insurance premium based on a Bonus-Malus system can be viewed as a discretely approximated analogy of the Bayesian premium. Traditionally, insurers have been using the frequency driven Bonus-Malus systems, which ignore the claim severity information. More sophisticated Bonus-Malus systems are based on both frequency and severity of claim and adopt the assumption of independence between frequency and severity. This is often invalid in real-life insurance practice (Oh et al., 2020).

Denuit and Lang see MTPLI rate-making as classifying policies by their risk characteristics. They imply that some risk characteristics are observable and are typically seen as non-random covariates, and some are unobservable and must be seen as unknown parameters (Denuit and Lang, 2004).

Šoltés, Zelinová and Bilíková in their research have revealed the impact of factors, that have the major impact on claim severity in road traffic accidents. According to their study engine power and engine displacement are strongly correlated factors, that have the highest impact on claim severity. Significant interaction between the age group of the driver and the place of driver's residence was confirmed. The results of their empirical study show that a substantial increase of claims paid occurs only with higher power vehicles. Drivers age is second significant factor, that leaves an impact on insurance claims paid. The highest average severity of road traffic accidents was found among the drivers in the age group over 60 whose place of residence is certain regional cities. They showed 7.8-fold higher average severity of road traffic accidents, with other variables fixed, as compared to owners under the age of 50 living in other regional towns of Czech Republic which were the least risky in terms of claim severity (Šoltés et al., 2019).

The study of Adanu, Smith, Powell and Jones indicated that drivers under 30 had a 0.46 probability of being at-fault in any crash. This group of drivers had 0.45 probability of the crash they have been involved as being serious. At-fault male drivers had probabilities of being involved in a crash or a serious injury crash of 0.54 and 0.52, respectively. The results of their study suggest that the risk of these drivers' groups being at-fault for crashes and being involved in a crash with serious injury are parallel (Adanu et al., 2017).

Employment status of the driver has a significant impact on probability of being involved in a road traffic accident and on its severity as well. According to the results of the study of Adanu et al. (2017), the unemployed drivers had a probability of 0.23 of being at-fault in a crash and the probability that they were at-fault in a serious crash was 0.57, thus probability of an unemployed driver being at-fault for a serious crash was 1.32 times higher than an employed, self-employed, or retired driver (Adanu et al., 2017).

The probability of distracted or intoxicated driver being at-fault for a serious injury crash was 1.89 and 1.14 times higher, respectively. Driver, who does not observe the speed limit, increased the risk of a crash being serious by 2.17. The probability of 0.86 that a crash involving a driver not wearing a seatbelt was serious suggests that the risk of serious injury in this type of crashes almost six times higher than average (Adanu et al., 2017).

The gender of drivers has fixed effect on the occurrence of serious injury crashes. The study of Adanu et al. (2017), shows that the impact of such factors as of age, employment status, and race of at-fault drivers show varied impact across regions. Lack of seatbelt usage, distracted driving and drugs and alcohol usage also showed varied impact across regions. Regional parameters affect the probability of serious injury crashes significantly. Credit score correlates to driving behavior significantly as well (Adanu et al., 2017).

Szymańska concluded, that the use of the age of the insured as one of the ratemaking variables can be justified. Insured persons aged from 18 to 25 cause on average more damage per year with higher average value, and therefore should pay higher premiums. The contribution range is between 140% and 150% of the basic premium. The insured under the age of 25 pay 300% of the basic premium if they buy insurance for the first time in this company, and 200% of the basic premium on the continuation of insurance. People aged from 25 to 28, cause on average more damage, slightly above the average value. Their contribution rate should be 119%. Another group that should have raised contributions are the insured at the age from 43 to 53. Insured persons in this age group should pay premiums increased by 4%. Persons aged 28–43 and over 53, on the other hand, could have a small discount (Szymańska, 2017).

Ayuso et al. (2010) concluded that the cost of accidents depends substantially on the number of bodily injuries. There are some traffic violations that are associated with a higher probability of serious or fatal accidents. For every possible combination of traffic violations, probability that the accident is slight, serious or fatal can be predicted. The costliest traffic violation is exceeding the speed limit, which increases the expected cost of an accident by two-thirds, compared to accidents that do not involve any traffic violations (Ayuso et al., 2010).

In the case of independent, net premiums are proportional to the total claim frequency and total claims severity and are inversely proportional to the age of the policy. This means that the more frequent the claim frequency and the larger the severity of the claim, the higher the net premium while the larger the policy age, the smaller the net premium. Insured persons that historically have large claim severity are charged higher net premiums than persons that have a small claim severity. This shows that net earned premiums are fairer than the net premiums on the classic Bonus-Malus System (Pratama et al., 2020).

3 METHODOLOGY AND DATA

3.1 Data

Calculations of motor insurance premiums are based on detailed statistical analyses of large data bases maintained by insurance companies, recording individual claim experience. The actuarial evaluation relies on a statistical model incorporating all the available information about the risk.

The data on MTPLI used in this research is obtained from Motor Insurers' Bureau of Latvia (MIB). Our study sample comprises of a total more than 128 thousand road traffic incidents resulted in claims to MTPLI policies issuer in Latvia during time period 2014–2020 in relation to passenger cars owned by private persons. Based on the policy parameters available in the MIB database, the following parameters were selected for further research (Fig. 4).

Policies which had missing values or values out of expected range were excluded from the analysis. No further elimination of outliers was undertaken, since we argue that their effect on the results of regression analysis is negligible due to the large sample size. A more detailed description of the parameters of MTPLI policies included in the study with descriptive statistics is provided in Tab. 1.

Due to the large data set selected (more than 128 thousand incidents) and to ensure a comprehensive validation of the calibrated models, it was divided into two parts, i.e., training sample of policies (70%) for modeling and 30% for out of sample model validation. We used stratified sampling to achieve an even distribution of accident events.

Tab. 1: Descriptive statistics of the initial MTPLI policies sample (based od MIB data)

Variable group	Variable	Observations	Min	Max	Median	Mean	1st Quantile	3rd Quantile
Car holder	Age	128,144	18	96	41	43.1	31	54
	Experience	128,144	0	73	16	19.79	8	25
	Gender	128,144	Categorical					
Driver behavior	Penalties	128,144	0	18	0	0.67	0	0
	BM class	128,144	1	17	9	8.99	7	11
Vehicle	Brand	128,144	Categorical					
	Age	128,144	0	58	13	12.96	10	17
	Engine cap.	128,144	599	7,536	1,984	2,084	1,781	2,401
	Power	128,144	51	478	96	101.8	77	120
	Weight	128,144	960	6,577	2,000	2,043.1	1,795	2,195
Policy	Maturity	128,144	Categorical					

		EUR mio.	Min	Max	Median	Mean	1st Quantile	3rd Quantile
Claims	Amount	129.80	50	864,057	516.12	1,150.76	276.35	1,079.67

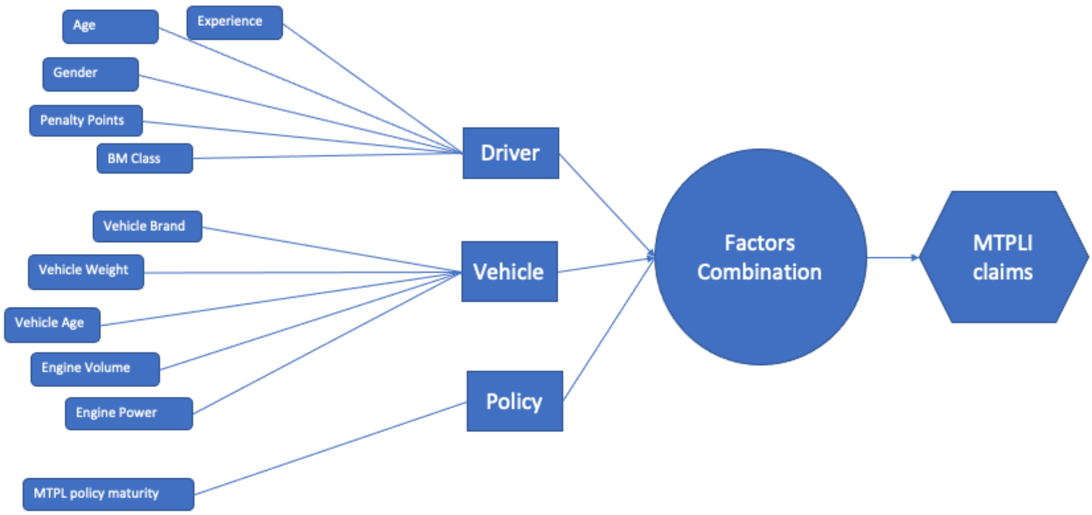

Fig. 4: Parameters selected for the research of MTPLI claims.
Source: Created by authors based on MIB data.

3.2 Methodology

Fig. 5 illustrates the process of claims paid risk assessment, regression models calibration and validation.

Based on the results of the previous studies, authors put forward the following hypotheses:

1. Multivariable regression makes it possible to significantly improve the degree of model's determination for road traffic accidents likelihood evaluation;

2. The grouping of independent variables values significantly improves the adequacy of road traffic accidents risk assessment models.

Over the last few decades, generalized linear models have been widely used in insurance claims valuation practice (Haberman and Renshaw, 1996; de Jong and Heller, 2008; Kaas et al., 2008; Frees, 2010; Agresti, 2015).

As claims paid data are positive values and skewed to the right, we applied log transformation of the losses and thus, the relationships between the MTPLI claims amounts and the selected independent variables can be expressed in the following form:

$$\begin{aligned}\log \mathrm{CP} = \beta_0 &+ \beta_1\, x_{\mathrm{POLM}} + \beta_2\, x_{\mathrm{CDG}} + \\ &+ \beta_3\, x_{\mathrm{CDA}} + \beta_4\, x_{\mathrm{CDE}} + \\ &+ \beta_5\, x_{\mathrm{BM}} + \beta_6\, x_{\mathrm{CDP}} + \\ &+ \beta_7\, x_{\mathrm{VA}} + \beta_8\, x_{\mathrm{VCA}} + \\ &+ \beta_9\, x_{\mathrm{VPO}} + \beta_{10}\, x_{\mathrm{BR}} + \\ &+ \beta_{11}\, x_{\mathrm{VWE}},\end{aligned} \tag{1}$$

where CP – claims paid amount per MTPLI policy; β_0 – intercept or constant of model; $\beta_1, \ldots, \beta_n$ – vector of regression coefficients; x_{POLM} – MTPLI policy maturity; x_{CDG} – MTPLI policy holder gender; x_{CDA} – MTPLI policy holder age; x_{CDE} – MTPLI policy holder experience; x_{BM} – bonus-malus of MTPLI policy holder; x_{CDP} – penalties of MTPLI policy holder; x_{VA} – vehicle age; x_{VCA} – vehicle engine capacity; x_{VPO} – vehicle engine power; x_{BR} – vehicle brand; x_{VWE} – vehicle weight.

3.3 Simulations and Grouping of Variable Values

Training sample and R version 4.0.5 was used for associations analysis (R Core Team, 2021). During the study, univariate linear and polynomial models were first calibrated using ungrouped factor values.

Fig. 5: Process of claims paid risk assessment, regression models calibration and validation.

Tab. 2: Statistics of initial models (based on simulation results)

Model type	**AIC**	**BIC**	**Residual deviance**	***F*-statistic**	***p*-value**
Univariate linear	118,452	118,548	18,868	145.7	<2.2E−16
Univariate polynomial	117,481	117,577	18,707	109.2	<2.2E−16
Multivariate linear	94,886	94,914	15,125	62.08	<2.2E−16
Multivariate polynomial	94,815	94,853	15,113	45.48	<2.2E−16

Calibrated model's performance was evaluated using traditional statistics such as Akaike information criterion (AIC; see Akaike, 1974), Bayesian information criterion (BIC; see Schwarz, 1978) and residual deviance was used for evaluation of selected models (Hosmer et al., 2013). These statistics are relative measures that evaluates model fit and penalizes overfit. Lower residual, AIC and BIC values indicate better model fit to the data.

From univariate models, as one can see from Tab. 2, the best fit was polynomial model with driver penalties as independent variable. Among the multivariate models, the best fit was also shown by polynomial model.

In the next step, in order to ensure a better fit of the model, the grouping of independent variables values was performed, taking into account the results of previous research, the experience of industry experts, and the correlations shown by actual claims paid amount statistics.

As can be seen from Tab. 3, the age distribution of MTPL policyholders is fairly even, while that of drivers with experience over 12 years is almost two-thirds. There is a visible tendency for the average amount of compensation paid in case of RTA to decrease as the age and experience of the driver increases.

Tab. 4 summarizes the impact of drivers' behavioral indicators on the average level of compensation paid.

Tab. 3: Descriptive statistics of independent variables – car driver age and experience (based on MIB data)

Variable	n	**Average claims paid (EUR)**
Age group (years)		
0–27	20,097	1,073.55
28–37	33,942	992.61
38–47	27,376	1,003.17
48–57	22,136	998.49
58+	24,593	990.61
Experience (years)		
0–2	11,205	1,050.51
3–5	10,554	1,074.08
6–8	11,135	999.73
9–11	13,168	1,016.03
12+	82,082	993.84

Tab. 4: Descriptive statistics of independent variables – car driver behaviour (based on MIB data)

Variable	n	**Average claims paid (EUR)**
Penalties (points)		
0	97,301	966.55
1–3	22,554	1,169.86
4–6	5,439	1,010.53
7–9	2,237	1,143.07
10+	613	1,157.44
Bonus-Malus		
1–3	1,135	1,465.05
4–6	29,018	1,115.56
7–9	45,182	987.34
10–12	35,732	949.86
13+	17,077	972.60

Tab. 5: Risk groups by vehicle brand (based on MIB data)

Risk group	**Brand**
1	Buick, Cadillac, Daewoo, Daihatsu, Daimler, GAZ, Lamborghini, Maserati, MG, Moskvich, Oldsmobile, Proton, Puch, Rolls-Royce, Saturn, Vauxhall, ZAZ
2	Austin, Citroen, Honda, Kia, Lancia, MCC, Mercury, Mitsubishi, Pontiac, Saab, Toyota
3	Alfa Romeo, Audi, Chevrolet, Chrysler, Dacia, Dodge, Fiat, Ford, Hyundai, Infiniti, Jeep, LADA, Land Rover, Lexus, Mazda, Mercedes Benz, Mini, Nissan, Opel, Peugeot, Porsche, Renault, Seat, Skoda, Smart, Ssang Yong, Suzuki, Tesla, Volvo, VW
4	BMW, Hummer, Isuzu, Jaguar, GMC, Plymouth, Rover
5	Acura, Bentley, Lincoln, Subaru, Range Rover

In the absolute majority (76%) of RTA cases, car holders did not have penalty points and the average amount of compensation paid was the lowest in this group. As the penalty points, which characterize the behavior of drivers, increases, there is a tendency to increase the average amount of paid compensation. Similar associations are formed between the Bonus-malus class and the MTPL claims paid – as the BM class improves, the average amount of claims paid decreases.

In order to assess the impact of the car brand on the amount of claims paid as a result of the RTA, it was divided into five groups according to the average amount of claims paid over the period from 2014–2020, see Tab. 5.

Tab. 6 summarizes the results of grouping vehicle parameters and the average amount of claims paid for each group.

As can be seen in Tab. 6, as a result of grouping all five car parameters, there is a clear trend: as the car age, brand risk group, weight, engine capacity and power increases, the average amount of claims paid increases.

The ANOVA test was used to evaluate the grouping results and it showed the difference of average claims paid in the groups as statistically significant (p-values < 0.05).

For regression analysis, the values of regressors x_{CDA}, x_{CDE}, x_{BM} x_{CDP}, x_{VA}, x_{VCA}, x_{VPO} and x_{VWE} were set equal to the means of the respective group, but x_{CDG} and x_{BR} were treated as categorical variables.

Tab. 6: Descriptive statistics of independent variables – vehicle parameters (based on MIB data)

Variable	n	**Average claims paid (EUR)**
Vehicle age (years)		
0–4	8,199	978.49
5–9	23,544	974.63
10–14	47,404	1,007.89
15–19	34,483	1,018.70
20+	14,514	1,055.41
Vehicle brand		
Group 1	100	728.01
Group 2	23,270	943.91
Group 3	88,381	1,003.14
Group 4	13,736	1,115.93
Group 5	2,657	1,192.75
Vehicle weight (kgs)		
<1,600	8,963	961.57
1,600–1,899	39,408	978.36
1,900–2,199	47,826	1,001.93
2,200–2,499	15,426	1,053.61
2,500+	16,521	1,080.38
Engine (cc)		
<1,400	9,494	958.97
1,400–1,799	27,762	977.18
1,800–2,199	50,396	992.19
2,200–2,599	21,823	1,022.86
2,600+	18,669	1,105.39
Power (kW)		
<60	8,769	972.54
60–94	51,948	982.66
95–129	44,521	1,006.34
130–164	14,563	1,058.80
165+	8,343	1,126.25

4 RESULTS

4.1 Statistical Test Results

A stepwise variables selection procedure was applied to find 'best fit' model using grouped values of variables. Using the function `step()` in R package (R Core Team, 2021), various combinations of variables were simulated and the three best according AIC criteria model's were evaluated in depth. Model '1' includes all 11 investigated variables, Model '2' does not include the variable 'Experience', and Model '3' does not include the variable 'Age'.

Using the gouped values of the variables, the fit of the models improved, as can be seen in Tab. 7, where chi-squared test results, AIC and BIC for comparable model's is summarized.

Tab. 7: Chi-squared test results of comparable regression models (based on simulation results)

Model	**Residual df**	**Residual deviance**	**AIC**	**BIC**
Model 1	89,659	14,996	94,200.92	94,614.70
Model 2	89,663	14,997	94,199.77	94,575.94
Model 3	89,663	15,000	94,214.29	94,590.46

Tab. 8: LRT statistics of selected models (based on simulation results)

	Model 1			**Model 2**			**Model 3**		
Factor	**df**	**Deviance**	***p*-value**	**df**	**Deviance**	***p*-value**	**df**	**Deviance**	***p*-value**
POL_M	4	20.513	***	4	20.513	***	4	20.513	***
CD_Ge	2	14.172	***	2	14.172	***	2	14.172	***
CD_Age	4	5.369	***	4	5.369	***	na		
CD_Exp	4	1.931	*	na			4	3.077	**
CD_BM	4	15.505	***	4	15.963	***	4	15.935	***
CD_Pen	4	50.712	***	4	50.917	***	4	50.974	***
VE_Age	4	8.345	***	4	8.266	***	4	8.905	***
VE_Bra	4	3.521	***	4	3.539	***	4	3.606	***
VE_We	4	24.624	***	4	24.775	***	4	24.208	***
VE_Cap	4	4.279	***	4	4.308	***	4	4.251	***
VE_Pow	4	4.337	***	4	4.364	***	4	4.116	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tab. 9: VIF multicollinearity test results (based on simulation results)

	Model 1			**Model 2**			**Model 3**		
Factor	**GVIF**	**df**	**$GVIF^{\frac{1}{2df}}$**	**GVIF**	**df**	**$GVIF^{\frac{1}{2df}}$**	**GVIF**	**df**	**$GVIF^{\frac{1}{2df}}$**
POL_M	1.2711	4	1.0304	1.2663	4	1.0299	1.2483	4	1.0281
CD_Ge	1.1192	2	1.0285	1.1032	2	1.0248	1.0910	2	1.0220
CD_Age	2.8268	4	1.1387	1.3294	4	1.0362	na		
CD_Exp	2.8497	4	1.1399	na			1.3402	4	1.0373
CD_BM	1.3182	4	1.0351	1.2402	4	1.0273	1.3075	4	1.0341
CD_Pen	1.0625	4	1.0076	1.0614	4	1.0075	1.0569	4	1.0069
VE_Age	1.6250	4	1.0626	1.6170	4	1.0619	1.6127	4	1.0616
VE_Bra	1.4310	4	1.0458	1.4302	4	1.0457	1.4211	4	1.0449
VE_We	4.5494	4	1.2085	4.5367	4	1.2081	4.5282	4	1.2078
VE_Cap	9.4315	4	1.3238	9.4286	4	1.3237	9.4129	4	1.3235
VE_Pow	6.6353	4	1.2669	6.6304	4	1.2668	6.6140	4	1.2664

The grouping of variables provided an opportunity to improve the degree of claims amounts paid determination compared to ungrouped variables. According to the chi-squared test results the is statistically significant difference in performance of models with grouped variables values and ungrouped. The differences between the three best fit models with grouped variable values are not statistically significant.

Three best fit models were subject to in-depth evaluation: According to the statistics summarized in Tab. 7, it can be seen that all three selected models fit relatively equal: AIC prefers the sequence of models 2, 1 and 3; BIC – sequence 2, 3 and according residual deviance 1, 2 and 3.

Likelihood ratio test (LRT) was used to estimate factor's, included in models, statistical significance. The LRT test statistics summarized in Tab. 8 show that estimates of all independent variables included in the tree best fit models are statistically significant at least at $\alpha = 0.05$ level.

Judging by the effect of each variable on the total deviation, the car driver's penalty points (33.5%), car weight (16.2%) and MTPLI policy term (13.5%) have the greatest impact.

The multicollinearity test of the variables included in the model was performed using variance inflation factor (VIF) statistics $\mathrm{GVIF}^{\frac{1}{2\mathrm{df}}}$. As can be seen from Tab. 9, the test result does not show multicollinearity at the level of independent variable groups.

4.2 Model's Validation Results

After multivariate simulation and testing, it is common to compare a number of alternative models based on the training and on the test samples. For the training sample, the within-sample comparisons are typically based on k-fold cross-validation (Refaeilzadeh et al., 2016).

Tab. 10: Five- and Ten-fold cross validation results (based on simulation results)

Model	**Five-fold cross-validation**	**Ten-fold cross-validation**
Model 1	0.1503992	0.1504042
Model 2	0.1606585	0.1606482
Model 3	0.1680324	0.1679897

According to the obtained results, the stability of all three models is high and both results indicate preferences 1, 2 & 3.

Comparison among models using test data – out-of-sample – comparisons are important in MTPLI cause many of these models are used for predictive purposes such as assessment of premiums un reserves for new MTPLI policies. Out-of-sample measures compare held-out during model calibration observations to those predicted by the model. Traditionally, mean absolute errors (MAE), mean squared error (MSE) and the root-mean-square error (RMSE) have been used to summarize differences between these two (Oh et al., 2020; Yunos et al., 2019).

Similar to cross-validation results were obtained from analysis of MSE, RSME and MAE estimation results, see Tab. 11.

Tab. 11: MSE, RMSE and MAE of bet fit models (based on simulation results)

	Within-sample			**Out-of-sample**		
Model	**MSE**	**RMSE**	**MAE**	**MSE**	**RMSE**	**MAE**
Model 1	0.1671767	0.4088725	0.3282445	0.1669904	0.4086446	0.3279425
Model 2	0.1671895	0.4088881	0.3282511	0.1670094	0.4086679	0.3279522
Model 3	0.1672165	0.4089212	0.3282862	0.1670052	0.4086627	0.3279751

5 DISCUSSION AND CONCLUSIONS

When studying the European motor insurance market, we can observe great volatility both between the reference periods and different, often opposite trends within the same period between countries. In general, there is a downward trend in the market for both insurance premiums and insurance claims, which is generally a negative trend, but MTPLI markets maintain stable. On the other hand, some trends are even dangerous for market stability: for example, in case of Latvia, where insurance premiums decrease and insurance claims increase in the same time period, which can potentially endanger market stability in the future. These signals increase the pressure on insurers to look for profit-seeking segments and to identify segments from which insurers should opt out or make tariff adjustments.

The results of the study confirmed that a multivariate regression model with grouped variable values offers the potential to improve significantly the ability to explain the effects of independent variables (car drivers behavior, age and experience, Bonus-Malus, vehicle parameters and MTPLI policy maturity) on claims paid amount in case of RTA, which is largely consistent with the studies of Klein et al. (2014), Charpentier et al. (2016), Adanu et al. (2017) and Šoltés et al. (2019).

The most important factor group that influences the amount of MTPLI claims paid was the car driver's behavior, which was assessed according to the current car driver penalty points and Bonus-Malus, total determining 43% of claims paid amount variation. Car driver penalties determine about 33% of the variation of losses. Assuming that the other variables in model '1' are constant, one additional penalty point causes an increase in MTPLI losses by about 11 EUR. This conclusion is largely consistent with the results of the study on road traffic regulations violations (Ayuso et al., 2010).

The second single most important factor influencing the amount of MTPLI claims paid is the weight of the vehicle, determining ca. 16% of the variation of losses – when the weight increases by 100 kg, the amount of compensation increases by an average of 13.5 EUR, assuming that the other variables are constant.

The third important factor was the maturity of the MTPLI policy, determining 13.4% of the variation of losses – for policies with a maturity of less than 3 months, the losses are 53% higher than the average. There are few studies on the effect of this factor on the amount of claims paid.

Bonus-Malus is also an important loss driver – for classes below 8 on the scale, the amount of losses increases rapidly as the score decreases and this is largely in line with, for example, Frangos and Karlis (2004), Klein et al. (2014) or Charpentier et al. (2016).

Similar to, for example Frangos and Karlis (2004) or Adanu et al. (2017), the claims paid in case of RTA caused by men are 9.2% higher than by women.

Car driver age – losses caused by drivers under the age of 28 are 6.5% higher than the average, which largely coincides with studies by other authors, such as National Safety Council (2022), Adanu et al. (2017), or Szymańska (2017). On the other hand, no increase in induced losses was observed for elderly car drivers – contrary to the suggestions of other studies, such as Denuit and Lang (2004), Frangos and Karlis (2004) or Klein et al. (2014).

Vehicle age – as the car age increases by one year, the average amount of MTPL insurance indemnities increases by about 3.8 EUR. A similar conclusion follows from studies by other authors, such as Denuit and Lang (2004).

Similar to other studies such as Klein et al. (2014) or Šoltés et al. (2019), the effect of engine power is not linear – when the power increases above 100 kW, the amount of losses caused by RTA increases much faster.

As can be seen in Fig. 6, the impact of certain factors on the amount of claims paid is different from the impact on the likelihood of a RTA.

The statistical tests performed within the framework of the research allow concluding that both hypotheses are confirmed:

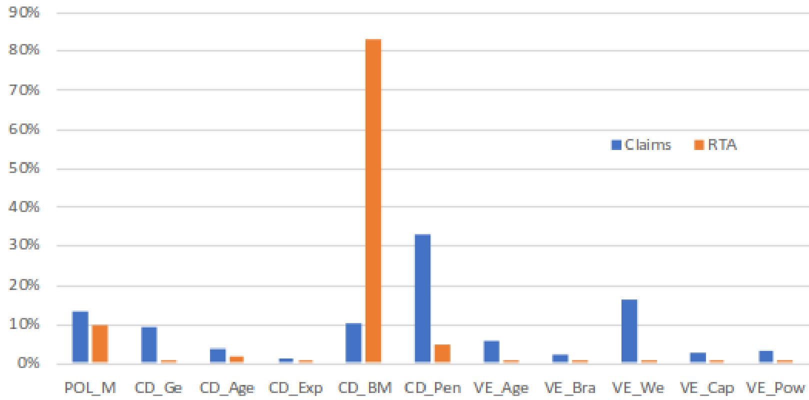

Fig. 6: Weights of factors determining the variation of claims paid and RTA likelihood

1. multivariable regression makes it possible to significantly improve the degree of model's determination for road traffic accidents likelihood evaluation;
2. the grouping of independent variables values significantly improves the adequacy of road traffic accidents risk assessment models.

The results of the research are significant for a reasonable assessment of MTPLI risk factors and pricing of policies to ensure the sustainable development of the insurance business. Product managers of three Latvia's leading transport insurance companies were involved in the approbation of the obtained results and they confirmed its practical applicability.

6 ACKNOWLEDGEMENT

This particular study has been funded by the Research Grant Fund of BA School of Business and Finance. The primary data for the study were provided by the Motor Insurers' Bureau of Latvia.

7 REFERENCES

Adanu, E. K., Smith, R., Powell, L. and Jones, S. 2017. Multilevel Analysis of the Role of Human Factors in Regional Disparities in Crash Outcomes. *Accident Analysis and Prevention*, 109, 10–17. DOI: 10.1016/j.aap.2017.09.022.

Agresti, A. 2015. *Foundations of Linear and Generalized Linear Models*. Wiley & Sons. ISBN 978-1-118-73003-4.

Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19 (6), 716–723. DOI: 10.1109/TAC.1974.1100705.

Ayuso, M., Guillén, M. and Alcañiz, M. 2010. The Impact of Traffic Violations on the Estimated Cost of Traffic Accidents with Victims. *Accident Analysis and Prevention*, 42 (2), 709–717. DOI: 10.1016/j.aap.2009.10.020.

Charpentier, A., David, A. and Elie, R. 2016. *Optimal Claiming Strategies in Bonus Malus Systems and Implied Markov Chains* [online]. Available at: `https://ssrn.com/abstract=2790583`. DOI: 10.2139/ssrn.2790583.

de Jong, P. and Heller, G. Z. 2008. *Generalized Linear Models for Insurance Data*. Cambridge University Press. ISBN 978-0-521-87914-9.

Denuit, M. and Lang, S. 2004. Non-Life Rate-Making with Bayesian GAMs. *Insurance: Mathematics and Economics*, 35 (3), 627–647. DOI: 10.1016/j.insmatheco.2004.08.001.

El Kassimi, F. and Zahi, J. 2021. Non-Life Insurance Ratemaking Techniques: A Literature Review of the Classic Methods. *International Journal of Accounting, Finance, Auditing, Management and Economics*, 2 (1), 344–361. DOI: 10.5281/zenodo.4474479.

Eurostat. 2021. *Road Accident Fatalities – Statistics by Type of Vehicle* [online]. Brussels. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_accident_fatalities_-_statistics_by_type_of_vehicle.

FKTK. 2021. *Statistics – Insurance* [online]. Riga. Available at: https://www.fktk.lv/en/statistics/insurance/.

Frangos, N. and Karlis, D. 2004. Modelling Losses Using an Exponential-Inverse Gaussian Distribution. *Insurance: Mathematics and Economics*, 35 (1), 53–67. DOI: 10.1016/j.insmatheco.2004.04.005.

Frees, E. W. and Valdez, E. A. 2008. Hierarchical Insurance Claims Modeling. *Journal of the American Statistical Association*, 103 (484), 1457–1469. DOI: 10.1198/016214508000000823.

Frees, E. W. 2010. *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press. ISBN 978-0-521-76011-9.

Haberman, S. and Renshaw, A. E. 1996. Generalized Linear Models and Actuarial Science. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 45 (4), 407–436. DOI: 10.2307/2988543.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. 2013. *Applied Logistic Regression*. 3rd ed. John Wiley & Sons. DOI: 10.1002/9781118548387.

Insurance Europe. 2021. *European Insurance: Preliminary Figures 2020* [online]. Brussels. Available at: https://www.insuranceeurope.eu/news.

Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. 2008. *Modern Actuarial Risk Theory: Using R*. 2nd ed. Springer. ISBN 978-3-540-70992-3. DOI: 10.1007/978-3-540-70998-5.

Klein, N., Denuit, M., Lang, S. and Kneib, T. 2014. Nonlife Ratemaking and Risk Management with Bayesian Generalized Additive Models for Location, Scale, and Shape. *Insurance Mathematics and Economics*, 55 (C), 225–249. DOI: 10.1016/j.insmatheco.2014.02.001.

National Safety Council. 2022. *Age of Driver* [online]. Available at: https://injuryfacts.nsc.org/motor-vehicle/overview/age-of-driver/.

Oh, R., Kim, J. H. T. and Ahn, J. Y. 2020. Designing a Bonus-Malus System Reflecting the Claim Size Under the Dependent Frequency-Severity Model. *Probability in the Engineering and Informational Sciences*, 1–25. DOI: 10.1017/S0269964821000188.

Pratama, A. S., Nurrohmah, S. and Novita, M. 2020. Determination of Net Premium Rates on Bonus-Malus System Based on Frequency and Severity Distribution. *Journal of Physics: Conference Series*, 1442, 012034. DOI: 10.1088/1742-6596/1442/1/012034.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing* [online]. Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Refaeilzadeh, P., Tang, L. and Liu, H. 2016. Cross-Validation. In Liu, L. and Özsu, M. T. (eds.). *Encyclopedia of Database Systems*. Springer, New York, NY. DOI: 10.1007/978-1-4899-7993-3_565-2.

Schwarz, G. 1978. Estimating the Dimension of a Model. *The Annals of Statistics*, 6 (2), 461–464.

Szymańska, A. 2017. The Application of Bühlmann-Straub Model to the Estimation of the Net Premium Rate Depending on the Age of the Insured in the Motor Third Liability Insurance. *Statistics in Transition New Series*, 18 (1), 151–165. DOI: 10.21307/stattrans-2016-063.

Šoltés, E., Zelinová, S. and Bilíková, M. 2019. General Linear Model: An Effective Tool for Analysis of Claim Severity in Motor Third Party Liability Insurance. *Statistics in Transition New Series*, 20 (4), 13–31. DOI: 10.21307/stattrans-2019-032.

World Health Organization. 2021. *Road Traffic Injuries* [online]. Available at: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

Yunos, Z. M., Shamsuddin, S. M., Sallehuddin, R. and Alwee, R. 2019. Hybrid Predictive Modelling for Motor Insurance Claim. *IOP Conference Series: Materials Science and Engineering*, 551, 012075. DOI: 10.1088/1757-899X/551/1/012075.

AUTHOR'S ADDRESS

Aivars Spilbergs, Department of Economics and Finance, BA School of Business and Finance, 161 Kr. Valdemara str., Riga, Latvia, e-mail: aivars.spilbergs@ba.lv

Andris Fomins, Department of Economics and Finance, BA School of Business and Finance, 161 Kr. Valdemara str., Riga, Latvia, e-mail: andris.fomins@ba.lv

Māris Krastiņš, Department of Economics and Finance, BA School of Business and Finance, 161 Kr. Valdemara str., Riga, Latvia, e-mail: maris.krastins@ba.lv