cogent
economics
& finance

ECONOMETRICS | RESEARCH ARTICLE

Modern claim frequency and claim severity models: An application to the Russian motor own damage insurance market

Evgenii V. Gilenko[1]* and Elena A. Mironova[2]

Abstract: During 2012–2015, the motor insurance in Russia received considerable attention both from the parts of the Russian government and from the insurance business. This was caused, in particular, by significant losses from the side of insurance companies that occurred during 2012–2013. Experts explain these losses not only by the effects of inflation or by the changes in Russian insurance legislation, but also by the incomplete set of factors that has been used by insurance companies for tariff calculation. This research analyses the factors that influence claim frequency and claim severity in the Russian motor own damage (MOD) insurance to assess the efficiency of the existing set of factors used for MOD insurance tariff calculations. To this end, we employ the appropriate claim frequency and claim severity models on the data provided by one of the leading St. Petersburg (Russia) insurance companies for the period of 2012–2013. The results of our calculations, organized within a resampling framework, show that additional factors may indeed be worth taking into account in the MOD insurance tariff calculation.

ABOUT THE AUTHORS

Evgenii V. Gilenko graduated from the St. Petersburg State University (the Department of Economic Cybernetics) in 2002. He obtained his MA in Economics degree from the European University at St. Petersburg in 2004. He defended his PhD thesis at the St. Petersburg State University in 2005. Nowadays holds a position of an associate professor of the Department of Public Administration at the Graduate School of Management of the St. Petersburg State University. The fields of research interest: macroeconomics, the big data analysis.

Elena A. Mironova obtained a specialist degree in Mathematical Methods in Economics at the St. Petersburg State University, Russia, in 2014, and a master's degree in Actuarial Science and Mathematical Finance at the University of Amsterdam, the Netherlands, in 2015. At the moment holds a position of a data scientist in a Dutch consultancy firm. Her research interests lie in the fields of applied econometrics, nonlife insurance, the big data analysis.

Evgenii V. Gilenko

PUBLIC INTEREST STATEMENT

During 2012–2015, the motor insurance in Russia received considerable attention both from the parts of the Russian government and from the insurance business. This was caused, in particular, by significant losses from the side of insurance companies that occurred during 2012–2013. Experts explain these losses not only by the effects of inflation or by the changes in Russian insurance legislation, but also by the incomplete set of factors that has been used by insurance companies for tariff calculation. This research analyses the factors that influence claim frequency and claim severity in the Russian motor own damage (MOD) insurance to assess the efficiency of the existing set of factors used for MOD insurance tariff calculations. The results of our calculations show that additional factors may indeed be worth taking into account in the MOD insurance tariff calculation.

cogent · oa

Subjects: Microeconomics; Econometrics; Insurance

Keywords: generalized linear models; hurdle model; gamma-distribution; motor insurance; claim frequency; claim severity

1. Introduction

Over the period from 2012 to 2015, the motor insurance market in Russia was influenced by significant changes both in motor insurance legislation and in consumer protection laws, as well as the macroeconomic situation in the country.

First of all, since the decisions of the Russian Supreme Court Plenum which were made in 2012–2013, motor insurance services have been covered by the Consumer Protection Law. By the end of 2013, these decisions resulted in a dramatic increase in the legal costs of Russian motor insurance companies because in most cases court decisions were made in favor of customers. Moreover, according to these changes in legislation, insurers had to pay greater amounts of penalties. The fact that motor insurance services were now covered by the Consumer Protection Law implied compensation of a fine in the amount of 50% of collected premium and (until 2014) of the penalty in addition to the insurance payments.

Thus, in 2013, the number of claims paid in compulsory motor third party liability (CMTPL) insurance increased by 25.7% over the previous year, while in 2012 this figure was only 8.9% (Russian Association of Motor Insurers [RAMI], 2014). Consequently, in the early 2014, the CMTPL loss ratio in several Russian regions exceeded 100% (Rating Agency 'Expert' [RAEX], 2014a).

It is worth noting here that the CMTPL and MOD insurance markets in Russia differ significantly. On the one hand, third party liability insurance is compulsory for every driver; therefore, the CMPTL tariffs and the limits of indemnity, as well as insurance directives, are set by the Central Bank and are heavily regulated. On the other hand, MOD is a voluntary type of insurance, and insurers have the freedom to change their tariffs whenever it is necessary. The MOD market in Russia is much smaller in terms of the number of policies than the CMPTL market but is also much more flexible and more exposed to the changes both within and outside the industry. However, many insurers prefer to offer both CMPTL and MOD products which strengthens to the connection between the two insurance markets.

Changes in legislation and increasing losses in the CMPTL market forced some motor insurance companies to leave the loss-generating regions of the Russian Federation. Other motor insurance companies had to compensate their CMTPL losses by adjusting the tariffs in other types of motor insurance, in particular, by motor own damage (MOD) insurance, as well as impose additional insurance products to their customers, which consequently caused a negative reaction from the customers' part.

Simultaneously, since the middle of 2012, there was an extended discussion on the necessity to increase the CMTPL limits of indemnity that remained unchanged since their first introduction in 2002. Finally, in October 2014, the CMTPL limits of indemnity were increased by more than 200% in order to take into account the risen prices for spare parts and maintenance and repair services (223-FZ, 2014).

Correspondingly, Russian motor insurance companies lobbied increasing the CMTPL tariffs that nowadays are rigid and set in a centralized manner by the Central Bank of Russia that was set in charge of the CMTPL insurance in Russia in September 2013. It is only in October 2014 that the CMTPL tariffs were also increased initially by 30% and then increased again by almost 60% in April 2015.

But a sharp depreciation of the Russian national currency (ruble) at the end of 2014 worsened the situation in the Russian automotive market dramatically. Vehicles and repair parts of foreign production became much more expensive, thus increasing the expenses of Russian motor insurance companies. Moreover, the currency depreciation caused a significant decrease in the demand for new cars in the first half of 2015 with sharpest monthly drop in March, 2015, by 42.5% YoY (Association of European Businesses [AEB], 2015). This virtually ceased the inflow of new clients to the motor insurance companies. As a result, in the first quarter of 2015, the total premium income from MOD policies decreased by12%YoY (Media Information Group "Insurance Today" [MIGIT], 2015).

The rest of the paper is organized as follows. Section 2 presents the motivation and the aim of the research. Section 3 provides a formal representation of the hurdle and Gamma-distribution generalized linear models, as well as the description of the resampling procedure used for calculations in this research. Section 4 focuses on the sample and variables selection issues. Section 5 presents the results of calculations. Section 6 concludes.

2. Research motivation and aim

The circumstances described above, along with low concentration and high competition levels in the Russian motor insurance market, pushed the motor insurance companies into a very difficult position in the early 2015. Most of the motor insurance companies had to resort to, in fact, the only available way to compensate their losses—increase the prices of MOD policies significantly. In some cases, the increase in MOD policy prices even in some top-ranking motor insurance companies made up about 50% (RAEX, 2014b).

Thus, in this research, we focus on the MOD insurance in the Russian motor insurance market. As it was shown above, the prominent increase in the prices of policies in the MOD insurance market was influenced by a number of factors, including changes in the CMPTL insurance market (although the CMTPL prices themselves are rigid and strictly regulated by the Central Bank of Russia). Nevertheless, we share the opinion of some experts in the Russian insurance business that the best way to avoid MOD policies over-pricing is a more accurate risk assessment and taking into account an extended range of possible factors influencing claim frequency and claim severity in Russian MOD insurance market.

Nowadays, in Russian MOD insurance, the following factors are *typically* taken into account: driver's attributes (experience, age, claim records); vehicle's characteristics (brand, model, year of production, whether it was bought on credit, etc.); policy conditions (franchise, vehicle's value).

Sometimes Russian motor insurance companies consider *additional* factors such as driver's gender (which is legal in Russia, as compared, for example, to the European countries where it is forbidden to differentiate drivers by gender) and vehicle's mileage (which is known as "pay-as-you-drive' insurance in the United Kingdom and the USA).

It is worth mentioning that before the ban on gender pricing for insurance products under the EU Gender Directive took effect on 21 December 2012, the European insurance companies were readily including the driver's gender in motor insurance tariff calculations, since it is quite well known that, despite the fact that female drivers drive the car more cautiously, they tend to have higher frequency of claims, but less severe losses (Mann, 2013; Miller, 2015). Thus, before 21 December 2012, there was a tariff discount for female drivers, which was cancelled after that date due to the requirements of the EU Gender Directive (Association of British Insurers [ABI], 2010).

Thus, a better understanding of the actual impact of all these factors on claim frequency and claim severity could help Russian motor insurance companies improve their financial position in the current critical situation in the Russian economy. It is important to stress that Russian motor insurance companies nowadays having flexibility in changing the MOD tariffs use it to compensate, in particular, for the negative situation in the CMTPL insurance market.

cogent ·· economics & finance

Thus, the ultimate aim of this research is to determine the factors that can be added to the set of factors used in motor own damage tariff calculation of a modern Russian motor insurance company. To this end, we apply appropriate statistical models of both claim frequency and claim severity classes to a modern MOD insurance data set, provided by one of the motor insurance companies operating in St. Petersburg (Russia).

Our research distinguishes itself in the following features.

(1) The research is carried out under the modern conditions of the Russian motor insurance market which, to our knowledge, has not been extensively studied yet with the application of the specified statistical models.

(2) Although we use the traditional hurdle and Gamma-distribution generalized linear models, we run calculations within a stratified resampling framework that allows providing the necessary robustness check of the obtained results.

3. Research methodology

Two principal models constitute the basis of this research. In this section, we briefly consider the traditional hurdle model and the GLM with a Gamma-distribution (hereafter called GLM-Gamma-distribution model for simplicity). These models represent, correspondingly, the classes of count models (used for claim frequency analysis) and the GLM models (used for claim severity analysis).

3.1. The hurdle model

There are several models that can be used to run the claim frequency analysis. The main feature of the traditional *hurdle model* is that it describes the link between regressors and the count dependent variable that has a large number of observations with zero values which is crucial in insurance (Mullahy, 1986). The model basically consists of two components: the component for zero values and the component for positive count values (starting with 1).

The zero values component is described by a binomial or a truncated count values distribution. The count component of the model is based on a zero-truncated Poisson or a negative binomial distribution and is bound by 1 from below (a negative binomial specification is considered in this research). The two components of the model may include different sets of regressors. Thus, the hurdle model combines two models: a logit-regression for separation of the zero-values component and a Poisson or a negative binomial model bound by the unit value within the count component.

A formal setup (probability density functions) of the standard hurdle model can be presented as follows (Mullahy, 1986):

$$f_{hurdle}(y;x,z,\eta,\gamma) = \begin{cases} f_{zero}(0;z,\gamma), & \text{if } y = 0, \\ \frac{(1-f_{zero}(0;z,\gamma))\cdot f_{count}(y;x,\eta)}{(1-f_{count}(0;x,\eta))}, & \text{if } y > 0, \end{cases} \tag{1}$$

where x, z are the sets of regressors for each of the two components (index i is discarded for the sake of simplicity); η, γ are the vectors of coefficients' estimates in the corresponding component; $f_{zero}(0; z, \gamma)$ is the zero-values component, bound from the right by $y = 1$; $f_{count}(y; x, \eta)$ is the count values component, bound from the left by $y = 1$.

The parameters of (1) are estimated by the maximum-likelihood method. It is noteworthy that the two components are estimated separately and independently. In order to see whether the suggested model is relevant, the Wald test of joint equality of the coefficients in both components is conducted. The null hypothesis is $\eta_j = \gamma_j$ for every coefficient (indexed j). Depending on the resulting *p*-value, a conclusion can be made whether separate processes need to be considered for zero values and count values of the dependent variables. The result of this test could also be interpreted as a general goodness of fit indicator: if the null hypothesis is rejected, then the hurdle model fits the

data adequately. For the details of application of the hurdle model see, for example, Zeileis, Kleiber, and Jackman (2008).

3.2. The GLM-Gamma distribution model

Claim severity modeling can be carried out by the usage of several possible models that differ in their assumptions on the distribution of claims. Such distributions are the Poisson distribution, the Gamma distribution, and others (Ohlsson, Johansson, & John, 2010). It should be noted that a claim severity distribution implies positive continuous values. Moreover, the claim severity distribution should be right-skewed, since large claims usually are not frequent to appear, while a bigger number of relatively small claims is common in motor insurance (Frees & Valdez, 2008).

Thus, as it is typically assumed in the literature, the cost of an individual claim is Gamma distributed with probability density function (Ohlsson et al., 2010):

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} exp(-\beta x); \; x > 0. \tag{2}$$

where $\alpha > 0$ is a shape parameter and $\beta > 0$ is a scale parameter. It should be noted that a sum of independent Gamma distributed variables also has a Gamma distribution. Let X be the sum of ω claims (independent $G(\alpha, \beta)$-distributed random variables), then $X \sim G(\omega\alpha, \beta)$. Then, the probability density function for $Y = X/\omega$ is computed as follows (index i is omitted for the sake of simplicity):

$$f_Y(y) = \omega f_X(\omega y) = \frac{(\omega\beta)^{\omega\alpha}}{\Gamma(\omega\alpha)} y^{\omega\alpha-1} exp(-\omega\beta y); y > 0. \tag{3}$$

Thus, $Y \sim G(\omega\alpha, \omega\beta)$ with expected value equal to α/β. It can be shown that the Gamma distribution specification is a special case of the exponential dispersion models of the generalized linear models. The parameters and coefficients of the generalized linear model based on the Gamma distribution can be estimated by the maximum likelihood method. For the details on the theoretical aspects of this model refer to Ohlsson et al. (2010).

3.3. The resampling procedure

In order to provide additional validity to the research, a resampling procedure is employed as a robustness check of the results. Each of the two models (the hurdle model and the GLM-Gamma distribution model) is estimated on 1,000 randomly chosen subsamples.

Each random subsample is obtained in two steps. First, a hierarchical cluster analysis is performed. The whole data set is split into several clusters. Second, randomly assigned 80% of observations from each of the clusters are included into the training subsample. In total, each subsample contains 2,920 observations. Thus, the resampling procedure contributes to a better level of diversification and mixture of observations within one sub-sample.

4. Research sample and hypotheses

4.1. Sample selection

The data on MOD insurance policies were provided for this research by one of the largest motor insurance companies based in St. Petersburg, Russia.[1] The policies were issued in 2012–2013. The period for the policies was chosen intentionally since it is in 2012 that the changes in the Russian motor insurance market started, thus, we decided to choose the research period appropriately.

The initial number of policies (observations) was 4,237. The final sample was formed based on the following criteria.

- **Insurance contract termination:** the initial sample contained terminated policies that were not included into the final sample because termination should not be considered as a typical situation forthe insurance company.
- **Insurance contract renewal:** the initial sample contained only 58 policies that were renewed by existing clients of the company; therefore, these policies are excluded since new clients might behave differently from existing clients.
- **The country of vehicle's production:** the initial sample contained only 153 policies with the insured vehicles of domestic (Russian) manufacture, which were thus excluded from consideration.
- **The region of prior use ofthe insured vehicles:** the policies with the region of prior use other than St. Petersburg were excluded in the final sample.
- **The sum insured:** in the initial sample, the sum insured ranged from 63,000 to 22.5 million rubles (approximately from 1,600 to 563,000 euros, as of May 2012). Extremely low and high values of the insured sums (outliers) were excluded from consideration. The final range of sum insured is from 110,000 to 4.0 mln rubles (approximately from 2,800 to 100,000 euros, as of May 2012).
- **The type of insured event:** in this research, we only consider the losses that happened *due to the driver's fault*. In particular, this research focuses on the losses that happened due to a road traffic accident or a collision with a stationary object. This choice is supported by the fact that it is quite difficult to describe the dependency between driver's or vehicle's characteristics and frequency/severity of other types of claims. For instance, it would be rather difficult to conclude that during a storm trees fall more often on "Opel" vehicles than on "Toyota" vehicles.
- **Loss sum:** extremely large losses (greater than 230,000 rubles or 5,750 euros (as of May, 2012)) were detected as outliers and thus excluded from consideration.

After applying these criteria, the final sample for the research comprised 3,649 policies (observations).

4.2. Variables description

4.2.1. The dependent variables
Number of claims was selected as the dependent variable for the hurdle model. At least one loss was claimed in 716 policies (19.6% of the final sample) and 2,933 policies caused no losses (80.4% of the final sample). The number of policies with more than two claims was less than 20, so at further steps we considered only three categories of policies: with no losses, with 1 loss, and with 2 or more losses. The distribution of the policies across these categories can be seen in Table 1.

Average claim size (within one policy) is the dependent variable in the GLM-Gamma distribution model (see Figure 1). In order to be consistent with the usage of the Gamma distribution, we ran a formal test of average claim size fit for the Gamma-distribution (Villaseñor & González-Estrada, 2015). With the test statistic V = 2.076 and the *p*-value = 0.142, we do not reject the null hypothesis that the Gamma-distribution is well suited for the *average claim size* variable, which is in line with the modern literature (Resti, Ismail, & Jamaan, 2013).

Table 1. The distribution of the number of losses			
	0 losses	**1 loss**	**More than 2 losses**
Number of policies	2,933	567	149

cogent – economics & finance

Figure 1. The distribution of the average claims in the whole sample.
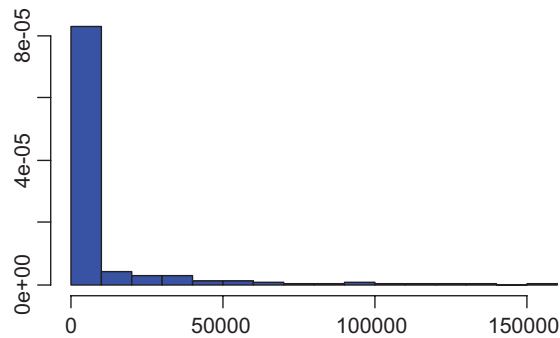
Table 2. Descriptive statistics of the independent variables					
Name	**Mean**	**SD**	**Median**	**Min**	**Max**
Franchise	0.19	0.69	0.00	0.00	4.00
Gender	0.56	0.50	1.00	0.00	1.00
ClassN	5.15	2.87	4.00	1.00	10.0
Credit	0.12	0.33	0.00	0.00	1.00
CarAge	1.27	1.59	1.00	0.00	7.00
Experience	9.71	8.12	8.00	0.00	46.0
SumIns	704,247.9	455,854.6	575,870.0	130,000.0	3,800,000.0

4.2.2. The independent variables

The independent variables are *driving experience* (lExp),[2] *franchise* (Franchise), *driver's gender* (Gender), *vehicle's class* (ClassN), whether the vehicle is *bought on credit* (Credit), *the sum insured* (lSumIns), *the age of the vehicle* (CarAge). The choice of the independent variables was motivated, on the one hand, by the data available, but on the other hand, by the considerations of their expected influence on the dependent variables. The descriptive statistics for the independent variables are provided in Table 2.

4.3. Research hypotheses

For convenience, here we provide a brief description of the independent variables along with the expected influence both on claim frequency and claim severity. These expectations form the set of our research hypotheses.

Driving experience: Driving experience of "the worst" driver.[3] Ranges 0–46 years in the sample. Expected to have a negative influence on the claim frequency. The influence on the claim severity can be twofold: on the one hand, more experienced drivers should cause fewer losses, but, on the other hand, they may feel too self-confident and thus get into more severe car accidents.

Gender: The gender of "the worst" driver. Coded 0—female driver, 1—male driver. In this research, we are aimed at testing the hypothesis that female drivers tend to have more, but less severe claims than male drivers.

Franchise: Theamount of franchise deductible. Coded 0—0 rubles, 1—less than 3,500 rubles, 2—from 3,500 to 7,499 rubles, 3—from 7,500 to 12,499 rubles, 4—more than 12,500 rubles. We expect that franchise has a negative influence both on claim frequency and claim severity since the drivers who use it are expected to drive more cautiously. Besides, by its nature, the franchise itself means fewer losses to be claimed.

Class of the vehicle: A numeric representation of vehicles classification. Ranges from 1 to 10 according to the standard vehicle classification from A to S. The higher is the class the more expensive is the vehicle. Thus, the drivers of more expensive vehicles are expected to be more cautious.

Credit status of the vehicle: Coded 0 if the vehicle was not bought on credit, and 1 if the vehicle was bought on credit. We expect that the drivers whose vehicles were bought on credit tend to drive more cautiously, so that they have fewer and less severe claims.

Sum insured: The amount of coverage, usually equal to the market price of the vehicle. It can be expected that the more expensive is the vehicle, the more cautious is the driver, thus, the influence on claim frequency should be negative.[4]

Age of the vehicle: Rangesfrom 0 to 7 years.[5] Generally, it can be expected that the older is the vehicle, the less claims the driver will make to the insurance company due to the fact that it is more important for the driver not to spoil his or her records in the insurance company than to cover often a quite small damage.

A special note is worth making concerning the driver's age (ranged from 18 to 71 years in our sample). We do not include this variable as an independent variable into our models since it is highly correlated with the driver's experience. Still, we ran the same calculations in several age sub-groups. The results were quite similar to those obtained for the whole sample (as described below).

5. Results and discussion
The results of the application of the resampling procedure (see Section 3.3) to the hurdle[6] and GLM-Gamma distribution models are given in Table 3 and Figures 2 and 3. Table 3 contains the percentage of statistically significant (at the 10% level of significance) estimates of each of the variables in 1,000 runs of the resampling procedure, while Figures 2 and 3 represent the results graphically.[7] In the figures, the estimates of the coefficients are given along the *X*-axis, while the *p*-values of their individual significance tests are given along the *Y*-axis. This allows to look both at the signs and the distribution of statistical significance of the estimated coefficients in all of the 1,000 runs at the same time (at the 10% level of significance). The results are also summarized in Table 3.

Before we start the discussion of the obtained results, it is worth reminding that the observations in our sample are the claims of car accidents but not the actual number car accidents. This means that for some reasons the drivers may not claim some losses. For example, one of the 'natural' reasons is the franchise bought by a driver. Or, if the loss is really small, then the driver is tempted not to claim it in order not to spoil his or her records in the insurance company.

Table 3. The percentage of statistically significant estimates at the 10% significance level (out of 1,000 samples)		
Variable's name	**Hurdlemodel (zero component)**	**Hurdle model (count component)**	**GLM-Gamma distribution model**
Driving experience	*100.0%*	*94.1%*	6.2%
Franchise	*99.6%*	*15.3%*	0.7%
Class	*71.1%*	0.0%	*65.0%*
Gender	0.4%	*39.1%*	*7.7%*
Credit	0.0%	0.0%	*15.7%*
SumInsured	0.5%	0.0%	not included
CarAge	9.8%	0.0%	0.3%

Note: The SumInsured variable was not included into the GLM-Gamma distribution model since the average claimed loss is directly related to the sum insured, thus, the inclusion of sum insured would result in a biased estimation.

❋ cogent ·· economics & finance

Figure 2. The GLM-Gamma distribution model estimates (from 1,000 subsamples).

Notes: In the graphs, the estimates of the coefficients are given along the *X*-axis, while the *p*-values of their individual significance tests are given along the *Y*-axis. The dashed horizontal line depicts the 10% level of significance, while the solid vertical line depicts the mean value of the estimated coefficients.
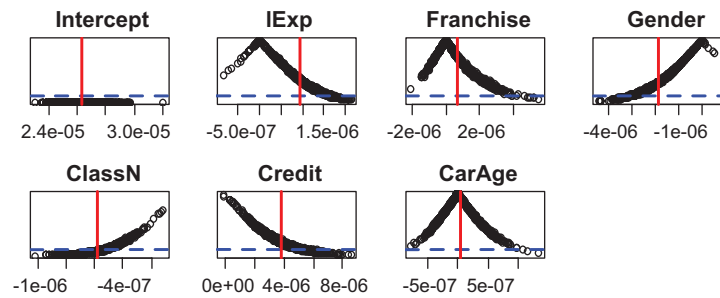

Figure 3. The hurdle model estimates (from 1,000 subsamples).

Notes: In the graphs, the estimates of the coefficients are given along the *X*-axis, while the *p*-values of their individual significance tests are given along the *Y*-axis. "0" and "1" at the end of the variable names represent the zero and the count components of the hurdle model correspondingly. The dashed horizontal line depicts the 10% level of significance, while the solid vertical line depicts the mean value of the estimated coefficients.
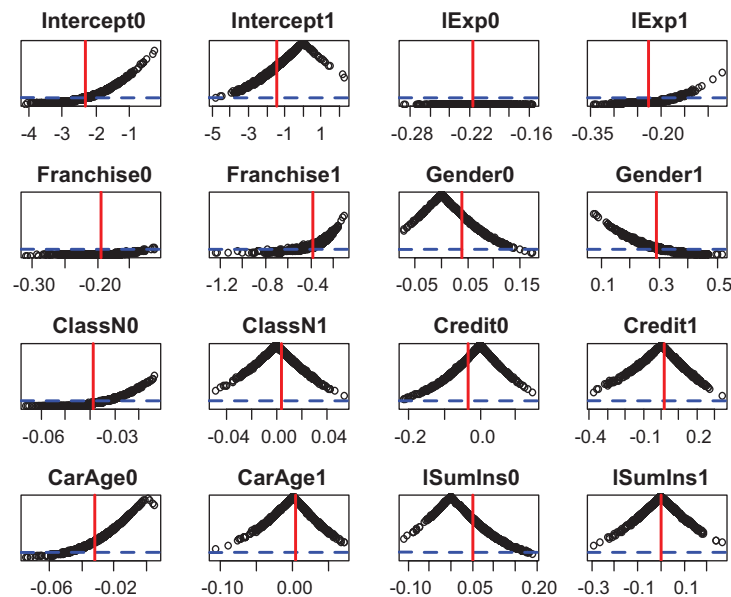
5.1. The results of the claim frequency analysis (the hurdle model, zero component)

For the zero component of the hurdle model, the results of calculations show that **driving experience, franchise,** and **the class of the vehicle** are the most "influential" factors in the sense that their coefficients are all negative and are statistically significant (at the 10% level of significance), correspondingly, in 100.0, 99.6, and 71.1% out of the 1,000 subsamples (see Table 3). Since the zero component of the hurdle model reflects the probability of claiming at least one car accident (as opposed to no claims at all), the three variables (quite expectedly) decrease this probability. Indeed, the more experience the driver has, the fewer car accidents (s)he is expected to get into, thus, fewer claims to cause. The larger the franchise is, the fewer claims there will be from the part of the driver. And the higher is the class of the vehicle, the more expensive the vehicle is, thus, it may indeed be expected that the drivers of such vehicles exercise more caution while driving.

The influence of other factors is mostly insignificant, although, the *CarAge* variable tends to have the negative sign of its coefficient, which speaks in favor of the fact that as the vehicle's age grows, the drivers tend to claim the cars accidents more rarely, which is quite understandable. The drivers of older vehicles are more concerned about their records in the insurance company than about making their vehicles look perfect.

5.2. The results of the claim frequency analysis (the hurdle model, count component)

The count component of the hurdle model reflects the probability of getting into more than one car accident (as opposed to getting into only one). The most influential factors are **driving experience** (with negative estimates of the coefficient), **franchise** (with negative estimates of the coefficient) and **driver's gender** (with positive estimates of the coefficient). See Table 3 for the details.

While the results for the two former factors (driving experience and franchise) are intuitively explainable in the same manner as before, it is worth noting that the fraction of cases where franchise has a significant influence dropped dramatically from 99.6% (for the zero component) to 15.3% (for the count component). This is not surprisingly, since for 2 or more accident the franchise amount may not be enough to cover the losses, and thus, the driver has to address to the insurance company.

The result for driver's gender reflects the fact that female drivers (coded as 0 in the sample) do tend to have lower probability of getting into more than one car accident as compared to male drivers (coded as 1 in the sample). This is one of the results sought for in this research.

Unfortunately, all the other regressors are not influential in the count component. Noteworthy, in the count component the class of the vehicle has no significant influence (as compared to the zero component). This may mean that after the first car accident the class of the vehicle stops determining the driver's manner of driving.

5.3. The results of the claim severity analysis (the GLM-Gamma distribution model)

The results for the GLM-Gamma distribution model show that the three most influential factors are **the class of the vehicle**, its **credit status** and the **driver's gender**, although the latter is statistically significant in a quite small number of cases (only 7.7%).

The class of the vehicle negatively influences the average claim size. This may be due to the fact that the drivers of more expensive cars (the cars of higher classes) tend to claim small losses more often as these drivers might care more about their cars.

The influence of the credit status is mostly positive which can be explained by the fact that, again, typically more expensive vehicles are bought on credit, and thus, such cars require higher compensations.

The influence of the driver's gender is mostly negative which means that in the studied sample female drivers cause and claim more severe losses. But the actual situation here is ambiguous since it may be the case that female drivers just tend not to claim small losses, while, as some motor insurance experts point out, male drivers in Russia tend to claim (sometimes using cheating schemes) even those losses that were caused by these drivers themselves.

5.4. Discussion of the results

Due to the space limitations, here we provide only a brief discussion of the results obtained. Let us recall that the aim of this research is to determine the factors that could be added to the set of factors used in motor own damage tariff calculation of a modern Russian motor insurance company. Specifically, the factor of interest in the current research is the driver's gender.

The results of calculations for the hurdle model show that female drivers do tend to drive more cautiously and have smaller probability of getting into more than 1 car accident. The results for the GLM-Gamma distribution model are more controversial and show that female drivers tend to cause higher payments for vehicles damages. But, as it was mentioned above, this may be the result of a specific behavior of female drivers towards the insurance company: as experts have it, female drivers tend not to claim small damages of their vehicles and not to cheat with the payments for them.

Some vagueness of the results of the models estimation within the resampling framework, in our opinion, may also be caused by the lack of some data. Our calculations would definitely benefit from a bigger sample, as well as from inclusion of additional factors such as, for example, the form of compensation (payment to the insured or arrangements for repair services) as well as vehicle's mileage which cannot be fully replaced by the vehicle's age.[8]

6. Conclusion

Nowadays, considering all the relevant factors for motor insurance tariff calculation is of crucial importance for a modern insurance company. This is why the research aimed at determining such factors should provide insurance companies with information on how to extend the set of factors used in the tariff calculations.

Despite the above-mentioned limitations, our research shows that for an insurance company in Russia it would be beneficial to provide discounts for female drivers on the motor own damage insurance tariffs, which is still not wide-spread in the Russian motor insurance market. The application of the appropriate statistical models helps in achieving this aim.

It should be stressed that the insurance company that provided the data for the current research was one of the first in Russia to introduce such *a discount for female drivers*. As of now, according to the available information, the company benefited from this, while initially it was introduced only as a marketing tool. In our opinion, this is in line with the major contribution of this paper to the practitioners: in the modern Russian economic conditions, a detailed accounting for all relevant rating factors will help an insurance company remain viable.

Author details
Evgenii V. Gilenko[1]
E-mail: e.gilenko@gsom.pu.ru
ORCID ID: http://orcid.org/0000-0001-6539-0212
Elena A. Mironova[2]
E-mail: e.mironova1504@gmail.com
[1] The Graduate School of Management of the St. Petersburg State University, Volkhovskiy per., 3, St, Petersburg, Russia.
[2] The Amsterdam School of Economics, The University of Amsterdam, Johan Huizingalaan 763A, 1066 Amsterdam, VH, The Netherlands.

Notes
1. The name of this company remains undisclosed according to the company's confidentiality policy.
2. We took the logarithms of driving experience and sums insured to bring these variables to the scales comparable with the scales of other independent variables.
3. "The worst" driver is the driver with the most loss-generating combination of gender, age and driving experience (according to the company's data).
4. The *SumInsured* variable was not included into the GLM-Gamma distribution model since the average claimed loss is directly related to the sum insured, thus, the inclusion of sum insured would result in a biased estimation.
5. Vehicles older than 7 years are not insured by the motor own damage insurance in Russia.
6. The results of the Wald test for correctness of the hurdle specification were positive in all 1,000 runs.
7. In Figure 3 "0" and "1" at the end of the variable names represent the zero and the count components of the hurdle model correspondingly.
8. Unfortunately, the company that provided the data for this research did not keep records on the form of compensation and vehicles' mileages in a suitable format.

References
223-FZ. (2014). *On the amendments to the Federal Law "On the obligatory insurance of civil liabilities of owners of transport vehicles" and individual legislative acts of the Russian Federation* (Russian Federal Law # 223-FZ, July 21, 2014). Retrieved March 13, 2017, from http://english.garant.ru/mon/archive/2014/07/23/
ABI. (2010). *The use of gender in insurance pricing*. Association of British Insurers, (Research paper #24.) Retrieved March 13, 2017, from http://www.oxera.com/Oxera/media/Oxera/The-use-of-gender-in-insurance-pricingpdf?ext=.pdf
AEB. (2015). *March car market decrease is 42.5%*, Association of European Businesses, (Press release, April 2015). Retrieved March 13, 2017, from http://www.aebrus.ru/upload/iblock/ace/eng_car-sales-in-march-2015.pdf
Frees, E. W., & Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association, 103*, 1457–1469. doi:10.1198/016214508000000823.
Mann, N. (2013). *Gender directive 'has already pushed up women's car insurance prices'*. The Actuary: The magazine of the actuarial profession, (16.01.2013). Retrieved March 13, 2017, from http://www.theactuary.com/news/2013/01/gender-directive-has-already-pushed-up-womens-car-insurance-prices/#sthash.nTl3CgEz.dpuf
MIGIT. (2015). *Insurance market dynamics 2015*, Media-Information Group "Insurance Today". Retrieved March 13, 2017, from http://www.insur-info.ru/statistics/analytics/

Miller, P. (2015). *Overview of the European motor insurance markets*. SCOR Global P&C Focus, (February 2015). Retrieved March 13, 2017, from http://www.scor.com/images/stories/pdf/library/focus/FocusPC_Motors.pdf

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics, 33*, 341–365. http://dx.doi.org/10.1016/0304-4076(86)90002-3

Ohlsson, E., Johansson, B., & John, S. (2010). *Non-life insurance pricing with generalized linear models*. Berlin Heidelberg: Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-10791-7

RAEX. (2014a). *Russian insurers running at a loss in 2014: A milestone passed*. Rating Agency "Expert". Retrieved March 13, 2017, from http://raexpert.org/rankingtable/?table_folder=/insurance/itogi_2014/Tab15

RAEX. (2014b). *Compulsory motor insurance market in Russia*. Rating Agency "Expert". Retrieved March 13, 2017, from http://raexpert.org/related_activities/researches/insurance/

RAMI. (2014). *Annual report 2014*. The Russian Association of Motor Insurers. Retrieved March 13, 2017, from http://www.autoins.ru/media/C2CBF8C8-A0AC-4229-8C85-6561F40AC1C9/6456C143-276C-4DF6-8953-256B701AFE91/RAMI_Annual_2014_eng.pdf

Resti, Y., Ismail, N., & Jamaan, S. (2013). Estimation of claim cost data using zero adjusted gamma and inverse Gaussian regression models. *Journal of Mathematics and Statistics, 9*, 186–192. doi:10.3844/jmssp.2013.186.192

Villaseñor, J. A., & González-Estrada, E. (2015). A variance ratio test of fit for Gamma distributions. *Statistics and Probability Letters, 96*, 281–286. doi:10.1016/j.spl.2014.10.001

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software, 27*, 1–25. doi:10.18637/jss.v027.i08

Cogent Economics & Finance (ISSN: 2332-2039) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com