"ALEXANDRU IOAN CUZA" UNIVERSITY OF IASI
FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION
DOCTORAL SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION

DOCTORAL THESIS

PRICING MODELS APPLIED IN NON-LIFE INSURANCE

ABSTRACT

SCIENTIFIC COORDINATOR
Prof. Vehiclemen PINTILESCU

PhD STUDENT
DAVID Mihaela

IAŞI, 2015

INTRODUCTION

Beck (2001) asserts that we live in a society that fears any uncertain event and for which all changes are considered risk-taking. Numerous studies, in particular on the principle of precaution, show that the individuals want to live in a safe society. This feeling of uncertainty and fear leads individuals to show special attention to the advantages of safety.

In a civilization of risk (Ladeg, 1981), the request for insurance has grown, having in response the guarantee of the financial security against possible losses. Thus, the emergence and development of the insurance related to the pressing need to protect man and his assets against the risks he faces. The main role of insurance is to provide a means for transferring (whole or in part) on payment of an insurance premium, the economic impact that is involved by these uncertain events.

The probabilistic nature of risks and their quantification have led to actuarial science, which is based on probability theory and statistics. So, the task of risk assessment lies primarily on actuaries who have developed over time various models through which they tried to establish a link between the occurrence of risks and the need to know how they manifest. Econometric modeling is designed to describe this connection, to determine the probability of risks and to assess their economic impact on the insurance company, and thus determining insurance premiums that reflect the seriousness of the risks.

In the work vehicleried out by insurance companies, the need to apply different premiums or tarifs depending on the degree of risk is highlighted by the presence of heterogeneity within the insurance portfolio, which leads to the appearance of asymmetric information. This means the effect of applying the same price for the entire portfolio, which involves providing adverse risks (at a lower price compared to their actual price) and, contrary, to discourage average risk insurance. This scenario can lead to a spiral effect, which means that the insurer can keep a disproportionate number of 'bad' risks in the portfolio and, as a consequence, has to continuously increase the insurance premiums.

The principle underlying the calculation of differentiated insurance premium in the insurance portfolio is represented by a pricing process that involves several steps. Thus, the acceptance of risk by the insurance company is followed by the analysis of a priori pricing meaning the segmentation or classification of all of the risks in terms of factors of influence, so that each group with identical hazards will have the same premium. In this stage of the analysis, the actuary determines the impact of the observable factors on the insured risk and the correlations between data. This step enables obtaining the elements of the calculation of pure insurance premium, the estimated frequency and the estimated cost of claims declared by the insured.

A major criticism to the approach of pricing through risk classification is given by the impossibility of integration in the premium calculation of some information on the insured that cannot be seen by the insurer and may represent significant risk factors. In this context, actuarial literature introduces the a posteriori stage in the pricing process, being represented by the theory of credibility. This way, the predictive power of the history of individual policyholders is assessed, integrating the retrospective component in the calculation of the insurance premium. In other words, the a posteriori pricing analysis allows the correction and adjustment of the a priori pricing in order to obtain a reasonable risk shootout. One of the commercial versions of this theory is the bonus-malus system, through which the insured's past experience is considered with regard to risk production. If, during the stage of the a priori pricing, the insurers have the freedom to set premiums based on appropriate methodology and risk factors they consider relevant, the bonus-malus system is required by the legislation of each country and must be respected by each insurance company without any change in its implementation. The bonus-malus system is understood both as an integral part of the pricing meant to combat the problem of asymmetric information, but also as a way to reduce competition between insurance companies.

On these coordinates, an issue of academic research is the construction and analysis of best econometric models to estimate the frequency and cost of claims, according to the available information.

The applicability in general insurance of the actuarial science has a rich and long history. Casualty Actuarial Society Institution of USA was established in 1914 and currently includes 6,700 members. Also, actuarial science institutions from countries such as Belgium, France, Canada, and beyond, operate in a well-defined framework, responding to requirements imposed by the insurance pricing issues. In this regard, there are numerous thematic publications exclusively from actuarial research area. Out of these, we mention the Casualty Actuarial Society Publications, The North American Actuarial Journal, and Bulletin Français d'Actuariat. Therefore, analyzing the foreign literature, one can observe over time the outstanding contribution of researchers to understand the specifics and functionality of the actuarial methods, and also the actuaries` effort to adapt and develop new models for risk assessment taking into account the requirements and challenges of the evolving insurance market.While foreign reference literature presents an impressive list of works and trends in order to improve the pricing methods applied in insurance, there is a lack of such studies in Romanian literature. This finding is explained by the fact that the insurance market is underdeveloped, especially the actuarial science as a field of research and activity in Romania. Romanian Actuarial Association was established only in 2000 and its aim is the recognition, support and promotion of the actuarial profession in Romania, but research in this area does not have yet a well-defined structure. Consequently, the practical utility and especially the complexity and timeliness of the analyzed phenomenon explain the urgent need and importance of the introduction and development of actuarial research in Romania.

Considering the extent of the problem resulting from this research field, this research aims to define criteria for analysis and assessment models used in insurance pricing. In this respect, it outlines a number of objectives specific to the research topics:

1. critical analysis of pricing models applied in insurance;
2. the identification and classification of premium risk factors;
3. analysis of the characteristics of each specific application model based on the nature of the analyzed data;

4. highlighting the national and European regulatory requirements and their impact on the calculation of premiums;
5. formulating some proposals on prices applied in insurance.

According to these objectives, the work is structured in two parts, followed by conclusions and bibliography. The first part contains three chapters and deals with conceptual and methodological aspects of pricing in insurance. The second part of the paper consists of four chapters and is an empirical study on the analysis applied in insurance pricing.

CHAPTERS PRESENTATION

Chapter I, entitled *"The concept of insurance pricing"*, aims at presenting a brief history on the need and the emergence of insurance and introducing basic concepts of insurance, general insurance in particular. An essential part of this chapter is to define the concept of insurance pricing. We concentrate on the connection, established in the literature, between charging and asymmetric information. In insurance pricing there is a procedure for determining the appropriate risk premium that the insured individual presents. Developing this idea, we can say that pricing insurance is understood as the set of methods by which prices are determined by the insured or premiums paid to the insurance company in exchange for the transfer of risk.

The literature presents two types of insurance pricing: *a priori* and *a posteriori pricing*.

The main idea in a priori pricing is to classify the insured risks in several categories, so that in each group they can be considered equivalent and belonging to the same law. In other words the a priori pricing permits grouping the risks in certain pricing classes, each class including policyholders with identical risks.

During the stage of a priori pricing, one does not know all the factors influencing the insured risk's occurrence and, as such, their impact cannot be assessed. In this respect, the heterogeneity still present at the pricing levels should not be attributed to chance, but

must be considered, to a certain extent, as the result of the influence of factors unobserved a priori. *A posteriori pricing* represents the way of recovering the information which was inaccessible a priori, by integrating individual policyholders` history.

Asymmetric information, between the two parties involved in the insurance, occurs when the insurer fails to accurately assess the risk of the insured. Two of the most analyzed aspects of the asymmetric information in economic literature are the *adverse selection* and the *moral hazard*.

The adverse selection describes a situation where the decision of individuals to ensure is positively correlated with their level of risk, and the insured cannot introduce this correlation in the insurance premium calculation. This is due to the private information on the insured, to which the insurer does not have access, or due to legal rules that prevent insurers to use certain types of information to determine the price of insurance.

The problem of the *moral hazard* arises when, after the contract of insurance, the policyholders' attitude changes, meaning that they become indifferent to risk, knowing that once insured they will not have to pay the cost of claims.

The literature shows that the effects of adverse selection justify the a priori pricing analysis, while pricing a posteriori is rather attributed to the elimination of the effects of moral hazard. These correlations must be regarded as constitutive elements in establishing a fair pricing structure. In this regard, the main purpose of pricing is the accurate individual risk assessment, so policyholders pay an insurance premium corresponding to the frequency and severity of the reported risks.

Chapter II, entitled *"Econometric models of insurance pricing"*, presents the econometric models used in insurance pricing, as a starting point for determining the insurance premium. Insurance premium assessment is a complex issue and it is a difficult task because it requires the establishment of a concrete analysis approach, a framework for conducting precise sequencing and selection of techniques or tools used. In this area, the actuary mission is to estimate a model describing as realistically as possible how the

premium of the insurance contract is influenced by risk factors. Considering the established distinction between the a priori and a posteriori pricing, actuarial investigations were directed to find suitable methods or tools for each of the two types of pricing applied in insurance. In response to this challenge, Generalized Linear Models are introduced as a tool for the analysis of a priori pricing and the bonus-malus system as the baseline for a posteriori pricing analysis.

In this regard, the first part of the chapter analyzes the literature and presents the methodology regarding the main models for estimating the frequency and cost of claims in order to determine pure insurance premium. From the class of Generalized Linear Models, there are highlighted models that allow estimating the *frequency of claims* (the Poisson model, the quasi-Poisson model, Negative Binomial, Hurdle, Zero-Inflated Poisson and Zero-Inflated Negative Binomial) and models to estimate the *cost of claims* (Gamma, Inverse-Gaussian and log-normal models). This approach generally allows an analysis of the risk level of an insurance portfolio, accurately determining the appropriate pure premium. The second part of this chapter defines the application of the bonus-malus system as the baseline for the a posteriori pricing analysis. Through this system, the first reference is adjusted according to the assessment of the past claims produced by the insured. Finally, the aim of integrating a priori and a posteriori pricing techniques in a common risk assessment mechanism is evidenced by obtaining a fair insurance premium.

Chapter III, entitled "*Criteria for analysis and assessment of econometric pricing models"*, highlights the best utility of this work since aims to address issues raised by the complexity of pricing analysis in the context of insurance. In this respect, analyzing empirical studies in the literature, one can distinguish four criteria for analysis and evaluation of these models, related to: the homogeneity of insurance portfolio, to the distribution form of claim cost, the choice and use of risk factors and the robustness of the applied models. All these criteria are not presented selectively, but in a synthetic and analytical manner.

Homogeneity of the portfolio

One of the most important criteria of data analysis in insurance and of evaluation of econometric models appropriate to the nature of the data, is the *homogeneity of insurance portfolio*. It should be noted that the homogeneity of the portfolio can be analyzed in the frequency of claims, which means that only the count data estimation models are considered.

To clearly present the criteria regarding the homogeneity of the portfolio, the limits of the Poisson model in general insurance are discussed. One of the most important limits, with a major impact on the quality of claim frequency modeling results is the equidispersion hypothesis, which assumes that the conditional mean and variance in the frequency of claims are equal. In general insurance, this assumption is frequently violated leading to the appearance of overdispersion, which involves a greater variance from the mean.

Literature identifies three main causes of overdispersion. Firstly, the high level of dispersion is the effect of unobserved heterogeneity in the analyzed data. One of the most important implications of unobserved heterogeneity is the phenomenon of zero inflation often seen during the analysis of general insurance data. Secondly, overdispersion may occur because the generation of the first event may be different from that which determines subsequent events. Third, overdispersion can be caused by the failure of the hypothesis of independence of the dependent variables (the events occurred), which is implicit for the Poisson model. In this context, the literature introduces various methods for detecting overdispersion. The paper presents only some of these methods, which are most commonly found in the literature.

Finally, the empirical studies on the application of count models, alternative to the Poisson model, allowing overdispersion correction data for insurance are discussed. The link between the count models is the presence of overdispersion in the analyzed data, indicating the lack of homogeneity of the portfolio. Poisson model of alternative models meet the limits imposed by the practical application of their insurance and their complexity is increasing as the portfolio is heterogeneous. In other words, the model chosen for

determining the insurance premium must respond to the problems posed by a heterogeneous portfolio, taking into account the nature of overdispersion and its causes.

The shape of the distribution of the cost of claims

Given the nature of data on the amount of claims recorded after the insured risk, there looms another criterion of pricing analysis. This criterion is expressed based on the shape that follows the distribution of the cost of claims in non-life insurance.

In a study on the cost of claims from motor third party liability insurance, Jong and Heller (2013) offer explanations of aberrant observations or outliers in the database, representing the exceptional claims produced as a consequence of the insured risk. The authors underline that, on the one hand, these values have a great influence on the results obtained by applying the method of the least squares. On the other hand, their removal from the database is not a suitable solution because it would have implications for the reality of the studied phenomenon and therefore on the calculation of the insurance premium. In this regard, the literature points out that the Gamma model properties are sensitive to extreme values and thus emphasized the advantage of the log-normal model, which assumes that by applying the logarithm to the costs, the importance of the values related to exceptional claims is reduced. The inverse-Gaussian model also answers the Gamma model restrictions on high costs of claims, but its use is more suitable for data recorded in fire insurance branch.

The usefulness of these models to determine the average cost of claims in insurance permits improving the quality of the calculation model for the insurance premiums and achieving results consistent with the nature of the data and type of risk insured.

Risk factors

In determining insurance premiums, risk factors or independent variables are generally qualitative, so continuous variables are introduced into the analysis after grouping them in different classes.

In the price analysis, the value of empirical studies for insurance companies is supported through the formulation and evaluation, consistent with the specificity of the studied portfolio, some theoretical assumptions regarding important risk factors. On this basis, the insurer may segment the portfolio in homogeneous risk groups, so that the premiums requested from the policyholder correspond to the degree of risk. Also, based on certain risk factors, the insurance company can develop pricing policy through the implementation of mechanisms to encourage policyholders to become more cautious and responsible drivers.

In literature there are numerous classifications of important factors for determining the risk degrees of the policyholders. In most of the empirical studies, the independent variables are divided into three categories: the characteristics of the insured, the characteristics of the insured asset and the characteristics of the insurance contract.

Robustness of the models

The criterion of the robustness of the models requires, on the one hand, a comparative analysis of the proposed models based on tests appropriate to their nature (nested or non-nested models), and on the other hand, refers to the application of tests for measuring the quality of the models by comparing the estimated values to those observed in the data analysis. Regarding the latter, literature introduces a test only for the count-type models applied to estimate the frequency of claims, but not for models that estimate the cost of claims.

Chapter IV, entitled *"Data and methodology of research"*, defines the structural framework of the empirical study. It involves, firstly, a clear statement of objectives and assumptions based on criteria defined in the first part of the paper. After that, two automobile insurance portfolios belonging to an insurance company in Romania and an insurance company in France are presented. The chapter ends with the methodological approach to the research.

The fundamental objectives of the empirical study on the two insurance companies in France and Romania are:

- Analysis and evaluation of the econometric models applied in insurance pricing based on the four criteria defined in the previous chapter: the homogeneity of the portfolio, the distribution form's of the claim cost, the risk factors and the robustness of the models;
- Proposals on the pricing framework in general insurance;
- Advice on pricing policies of insurance companies.

In order to fulfill these objectives, frame of research is defined on the following assumptions:

- Homogeneity of the portfolio is evidenced by the degree of dispersion of the analyzed data;
- The estimation of the frequency and cost of claims is done based on the models that best respond to the specificity of the analyzed portfolio;
- Transformation of independent variables is supported by a more realistic representation of the differences on the riskiness of the insured, leading to segmentation of the portfolio in classes of homogeneous risk;
- Adverse selection is captured by the inclusion of some important factors that indicate additional information about the evolution of vehicle accidents per insured and moral hazard is confirmed by the factors that indicate information on policyholders' degree of aversion to risk.

The phenomenon under investigation concerns the branch of vehicle liability insurance, which covers claims to the insured vehicle belonging to third parties. The proposed methods can be addressed in other branches of insurance (CASCO insurance, fire insurance, theft insurance, travel insurance etc.), taking into account the peculiarities of their contracts.

For the first study, we have had a database with 982109 observations representing motor third party liability insurance policies concluded for a period of three years, between 2010 and 2012. These policies constitute the automobile insurance portfolio belonging to an insurance company which is active in Romania. Based on observations on the analyzed portfolio, one can divide the

factors into three categories depending on: the insured, the insured vehicle and the insurance contract. The first category includes *the age of the insured*, the *county* where the insurance was concluded, and the *county of residence* of the insured. The second category includes certain characteristics considered relevant to the insured goods: *the type of vehicle, maximum authorized mass, engine capacity, engine power, number of seats and fuel used*. Characteristics of insurance contracts' form the third category of factors and include the *year of the insurance contract, duration of the insurance contract and the bonus-malus coefficient*.

For the second empirical study, we have had a database containing 150100 records over the period 2007-2009. The database units are represented by motor third party liability insurance policies that constitute the portfolio of an insurance company in France. Except the dependent variables, *frequency* and *cost of claims*, describing the insured risk, other variables are known a priori by the insurer, being used to "customize" the profile of each insured. These variables also reflect certain characteristics of the insured (*age, occupation*), the asset provided (*type, category, group, the purpose of the vehicle, the age of the vehicle and the use of a GPS device*) and the insurance contract (*the year when the insurance contract was concluded, the bonus-malus coefficient, the CASCO insurance*).

Chapter V, *"Empirical study on pricing of vehicle insurance in Romania"*, and **Chapter VI**, "*Empirical study on pricing of vehicle insurance in France"*, illustrate the stages of the pricing process in the two portfolios of vehicle insurance, closely following the methodological approach on achieving the objectives and verifying the hypotheses of the research. In this part of the empirical study, the focus is on detecting the features of the risk factors considered, but also on how the proposed econometric models respond to the specific issues raised by the two vehicle insurance portfolios studied. In this regard, the chapters are structured as follows: vehicle insurance portfolio analysis; vehicle claims frequency and cost estimation; determining risk factors of pure insurance premium; presentation and interpretation of bonus-malus system code provided the insurance legislation in Romania and France.

Chapter VII, entitled *"Comparative analysis of empirical studies"*, is dedicated to the comparative analysis of results for the two insurance companies. Thus, certain conclusions can be drawn regarding the specific models to estimate the two components of pure insurance premium.

Modeling claims frequency

The results from the application of models to estimate the frequency of vehicle claims are summarized in the table below. All these observations are also made on the findings in the literature.

Criteria for analysis	**Romanian insurance portfolio**	**French insurance portfolio**
Uniformity of portfolio	The equidispersion hypothesis imposed by the application of Poisson model is not validated.	
Causes of the heterogeneity of the portfolio	- the real overdispersion caused by the zero values inflation is identified.	- the real overdispersion is found due to the inflation of zero and to the unnobserved heterogeneity.
Overdispersion correction		
The quasi-Poisson model	- is built on different risk factors compared with Poisson model, leading to different results and enabling a concrete comparison between the two regressions.	- is built on the same risk factors compared with Poisson model, leading to similar results, which does not improve the analysis.
The Negative Binomial models (NB1 and NB2)	- are built on different independent variables compared with Poisson model and the results justify the presence of overdispersion.	- are built on the same risk factors like the Poisson model and the results justify the presence of overdispersion.
	- both NB1 and NB2 models are not validated, meaning that they canot be compared with the proposed *count* models, and the obtained results cannot be taken into consideration.	- both the NB1 and NB2 models are validated, both indicating a greater predictive power compared to Poisson model.
The Hurdle model	- the probability that the	- the two phases of the

	insured may not have any claim is explained by different factors than those influencing the probability that the policyholders declare at least one accident for which they are responsible.	model, the logistic regression and the Poisson regression, are built on the same independent variables; - the signs of regression coefficients for the logistic model are different from those for the Poisson model, which suggests that the results are consistent.
	The models explain correctly the conection between the claims frequency and the significant risk factors.	
	- the hurdle model is considerably better than the quasi-Poisson and Poisson models to estimate the frequency of vehicle insurance claims.	- the hurdle model is particularly chosen instead of the Poisson regression and NB1 is preferred to NB2.
The Zero-Inflated models	- allow the shootout of the policyholders who have claims, but have not declared it to the insurance company from those who have no claims.	
	- the two phases of the ZIP model (logistic regression and Poisson) and of the ZINB model (logistic regression and NB2) are built on different risk factors.	- the stage corresponding to the Poisson model (in the case of the ZIP model) and of the NB2 model (in the case of the ZINB model) are built on the same risk factors as for the Poisson regression, respectively NB2.
	They are statistically validated and are chosen over other suggested regression models.	
	- the ZIP model is considered the best to correct overdispersion caused by zero inflation.	- ZINB model is considered the best to correct overdispersion caused by the zero inflation and by the unobserved heterogeneity.
The robustness of models	- the ZIP model fulfills all the criteria to be considered the most robust of all the proposed models.	- the ZINB model fulfills all the conditions to be considered the most robust model of all seven models proposed to estimate the frequency of vehicle claims.
Pure insurance premium calculation	Based on the ZIP model (in the case of the Romanian portfolio) and on the ZINB model (in the case of the French portfolio) the estimated average vehicle claims frequency is	

	determined, which is the first component of the pure insurance premium.

Modeling the cost of claims

Analog to the presentation of the results obtained for the count models, the findings for the three models which estimate the cost of claims are shown in the table below.

Criteria for analysis	**Romanian insurance portfolio**	**French insurance portfolio**
The shape of the distribution of the claim cost		
The Gamma model	Is significant for both analyzed portfolios.	
The Inverse-Gaussian model	- is not validated, which means that it doesn`t explain correctly the conection between the cost of claims and the considered risk factors considered; - the results cannot be included in the analysis and the comparison to other patterns is not possible.	- the model explains correctly the conection between the cost of claims and the risk factors that significantly influence the variation in the cost. - is built on different risk factors compared with the Gamma and log-normal models.
The log-normal model	- is significant in explaining the variation in the cost of vehicle claims based on the risk factors included, which differ from the risk factors which are significant for the Gamma model or the Inverse-Gaussian model.	
The robustness of models	- the log-normal model is considerably better compared to the Gamma model, responding better to more extreme values in the two portfolios on the occurrence of certain insured risks.	- the Gamma model is more appropriate, compared to the Inverse-Gaussian model, to estimate the cost of vehicle claims; - the Gamma and the Inverse Gaussian models are rejected in favor of the log-normal.
Pure insurance premium calculation	Based on log-normal model (for both analyzed portfolios) the estimated cost of vehicle claims is determined, which is the second component of the pure insurance premium.	

Risk factors for the pure insurance premium

Empirical evidence on the influence of risk factors on the insurance premium is summarized in the table below.

Criteria for analysis	**Romanian insurance portfolio**	**French insurance portfolio**
Transformation of the independent variables	Is supported by a more realistic representation of the differences on the riskiness of the insured, resulting in the portfolio segmentation in homogeneous risk classes.	
Risk factors common to both portfolios		
Year of observation	Justifies the application of price analysis to determine the insurance premium from one year to another.	
Age of the insured as quantitative variable	- risk increases with age.	- the frequency of claims decreases with age.
Age of the insured as qualititative variable	Enables obtaining more homogeneous groups of policyholders according to the different degrees of risk .	
	- defining three age groups of policyholders	- defining five age groups of policyholders
	- elderly insured are the most risky group of customers for the company in comparison with the rest of the portfolio.	- beginner policyholders are riskier than other groups; - elderly insured present a higher degree of risk compared to other age groups, except the beginners.
Duration of the contract	- represents the insurance and can be for 6 months or 12 months , as requested by the insured.	- the contract can last for a period from 1 to 15 years.
	there is a decrease in the frequency of claims as the insurance contract period increases.	
Risk factors found in other empirical studies	- age of the insured - type of vehicle - maximum authorized mass - engine capacity - engine power - fuel used - duration of the contract.	- age of the insured - type of vehicle - group of vehicle - purpose of the vehicle - presence of a GPS device - bonus-malus coefficient - the duration of the insurance contract - Casco insurance policy

Risk factors which differ from other empirical studies	- the county - insured`s residence - number of seats in the vehicle.	- insured`s occupation - value of the vehicle
Finding the assymetry of the information		
The adverse selection	- the *period of insurance* risk factor captures additional information on development for each insured contingency, but to a limited extent.	- risk factors such as *insurance contract duration* and the *bonus-malus coefficient* permit registering the risk occurence evolution for each insured and partly capture the individual risk level.
The moral hazard	- the *period of insurance* risk factor indicates the policyholder`s degree of aversion to risk, but to a limited extent	- the risk factors *Casco optional insurance policy and the use of the GPS device* provide information on the policyholders' risk aversion and show how cautious and responsible they are on the insured risk.

Proposals for the analysis of pricing in general insurance

Starting from the main objective of the work and the results of the two empirical studies, this paper aims to propose a pricing scheme that can be applied by insurance companies to determine the insurance premium. So, integrating the four evaluation criteria for analysis and the econometric models in the process of pricing allows us to define the framework of analysis in several steps.

In the proposed pricing scheme, the stages are not defined in an arbitrary manner, but each stage refers to the nature of the data analyzed, the specificity of the insurance portfolio, but also to the functionality of the econometric models on insurance data.

Proposals for pricing policy of insurance companies

The usage, in the price analysis, of some risk factors such as the duration of the insurance contract, the bonus-malus coefficient, the

existence of a GPS device or an optional insurance policy in France, the period of insurance in the case of the Romanian portfolio and the age of the policyholders in both cases, provides additional information on the risk of policyholders. Considering the results obtained on the basis of these risk factors, insurance companies can make decisions to change the pricing policy either by implementing measures to prevent the occurrence of an insured risk or by stimulating policyholders to become more responsible drivers and remain loyal customers of the insurance company.

Utility of the research

The research is useful, on the one hand, in choosing the most appropriate econometric models on which to obtain an insurance premium according to the degree of risk insured. On the other hand, it is useful the formulation of proposals for adaptation to the specific data each company works, and for diminishing the occurrence of the risks insured and their integration into the pricing policies of insurance companies.

THE OVERALL CONCLUSION OF RESEARCH

In the thesis *"Pricing Models Applied in Non-Life Insurance"* we planned a concrete analysis framework that responds to the complexity of pricing in non-life insurance. Starting from the data in the empirical literature, we tried to define a set of criteria by which to attain problems posed by the pricing models applied to obtain insurance premium. In this regard, there are the homogeneity of insurance portfolio, the distribution of the cost of claims, peculiarities of certain risk factors and the robustness of the proposed models.

Given this analytical framework, we have tried to show that, to obtain a fair insurance premium, companies must take into account the specificity of the insurance portfolio, the peculiarities of the risk factors and how econometric pricing models respond to specific data analysis. Research results show that the pricing policy varies from

one insurance company to another at both methodological and legislative levels.

RESEARCH LIMITATIONS

The methodology established to meet the objectives proposed in this paper and to verify the hypotheses formulated on the basis of these objectives is not without some limitations.

Firstly, the transformation of independent variables was based on the graphic representation of the marginal effect of each variable on the frequency of vehicle claims. This technique entails some arbitrary that can influence the modeling and the results on the robustness of the models, but also the obtaining of homogeneous risk groups.

Secondly, given that one cannot make a distinction, from one year to another, between the new policyholders of the company and those who have renewed their insurance contract, the inclusion of the year of observation as a risk factor allows one only to find differences between the frequency and cost of claims from one year to another. It therefore cannot be applied to a panel analysis, but only to a cross-sectional analysis which does not permit whole capturing of the individual unobserved heterogeneity registered for each insured, in time.

Third, the lack of certain information on the behavior of the insured which are not known to the insurance companies (unobserved individual heterogeneity) is treated only by applying the zero-inflated models in the a priori stage, and the bonus-malus system in the a posteriori stage. Replacing the time factor by certain variables that provide additional information on the progress of the degree of risk of the policyholders may not fully reflect the reality of the studied phenomenon.

Fourth, there have been formulated certain conclusions about the asymmetric information on the impact of certain risk factors on the risk level of policyholders, delineating between policyholders with risk aversion and their counterparts, and between cautious policyholders and those unresponsive to the insured risk. In this context, asymmetric information is treated superficially. The issues

specific to this phenomenon may become in themselves the subject of a complex and extensive research work in the field of actuarial science research.

Fifth, not all models presented in the literature on estimating the frequency and cost of claims have been introduced in the paper. On the one hand, the research field is extensive, offering several lines of research or analysis to pricing in insurance. On the other hand, the complexity of models developed in the literature complicates their application to real data. However, the purpose of the work, namely to draw an analysis framework on insurance pricing is implemented. Starting from the proposed research the pricing models or theories presented in the literature can be developed and completed.

SELECTIVE BIBLIOGRAPHY

1. Agresti, A., *Foundations of Linear and Generalized Linear Models*, John Wiley & Sons, New Jersey, 2015
2. Allain, E., Brenac, T., *Modèles linéaires généralisés appliqués à l'étude des nombres d'accidents sur des sites routiers: le modèle de Poisson et ses extensions*, Recherche Transports Sécurité, Volume 72, 2001, pp. 3-18
3. Antonio, K., Valdez, E.A., *Statistical concepts of a priori and a posteriori risk classification in insurance*, Advances in Statistical Analysis, Volume 96, Issue 2, 2012, pp. 187-224
4. Boucher, J.P., Denuit, M., Guillen, M., *Risk classification for claims counts - A comparative analysis of various zero-inflated mixed Poisson and hurdle models*, North American Actuarial Journal, Volume 11, Issue 4, 2007, pp. 110-131
5. Boucher, J.P., Denuit, M., *Crédibilité linéaire bivariée utilisant le nombre de périodes avec réclamations: modèles de Poisson, modèles à barrières et modèles gonflés de zéros*. Assurances et gestion des risques, Volume 75, 2008a, pp. 487-520
6. Boucher, J.P., Denuit, M., *Credibility premium for the zero-inflated Poisson model and new hunger for bonus interpretation*. Insurance: Mathematics and Economics, Volume 42, Issue 2, 2008b, pp. 727-735
7. Boucher, J.P., Denuit, M., Guillen, M., *Models of insurance claim counts with time dependence based on generalisation of Poisson and Negative Binomial Distributions*, Advancing the Science of Risk Variance, Vol. 2, Issue 1, 2008a, pp. 135-162
8. Boucher, J.P., Denuit, M., Guillen, M., *Modeling of insurance claim counts with hurdle distribution for panel data*, Advances in Mathematical and Statistical Modeling, Statistics for Industry and Technology, 2008b, pp. 45-59
9. Boucher, J.P., Guillen, M., *A survey on models for panel count data with applications to insurance*, Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales, Vol. 103, Issue 2, 2009, pp. 277-295
10. Boucher, J.P., Denuit, M., Guillen, M., *Number of accidents or number of claims? An approach with zero-inflated Poisson*

models for panel data, Journal of Risk and Insurance, Volume 76, Issue 4, 2009, pp. 821-846
11. Boucher, J.P., Pérez-Marín, A.M., Santolino, M., *Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident*, Anales del Instituto de Actuarios Españoles, Volume 19, 2013, pp. 135-154
12. Cameron, A.C., Trivedi, P. K., *Econometric models based on count data. Comparisons and applications of some estimators and tests*, Journal of Applied Econometrics, Vol. 1, 1986, pp. 29-53
13. Cameron, A.C., Trivedi, P.K., *Regression-based tests for overdispersion in the Poisson model*, Journal of Econometrics, Vol. 46, Issue 3, 1990, pp. 347-364
14. Cameron, A.C., Trivedi, P.K., *Count data models for financial data*, Handbook of Statistics, Volume 14, Statistical Methods in Finance, 1996, pp. 363-392, Amsterdam, North-Holland
15. Cameron, A.C., Trivedi, P.K., *Regression Analysis of Count Data*, Econometric Society Monograph, Cambridge University Press, 1998
16. Cameron, A.C., Trivedi, P.K., *Essentials of Count Data Regression*. In A Companion to Theoretical Econometrics, Malden MA: Blackwell Publishing Ltd. Chapter 15, 1999
17. Cameron, A.C., Trivedi, P.K., *Regression Analysis of Count Data*, 2nd ed., Cambridge University Press, Cambridge, 2013
18. Charpentier, A., Denuit, D., *Mathématiques de l'assurance non-vie*, Tome I: *Principe fondamentaux de théorie du risque*, Economica, Paris, 2004
19. Charpentier, A., Denuit, M., *Mathématiques de l'assurance non-vie*, Tome II: *Tarification et provisionnement*, Economica, Paris, 2005
20. Denuit, M., *Tarification automobile sur données de panel*, ASTIN Bulletin, Volume 36, Issue 1, 2003
21. Denuit, M., Lang, S., *Nonlife Ratemaking with Bayesian GAM's*, Insurance: Mathematics and Economics, Volume 35, Issue 3, 2004, pp. 627-647.
22. Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R., *Actuarial Theory for Dependent Risks. Measures, Orders and Models*, John Wiley & Sons, London, 2005

23. Denuit, M., *An actuarial analysis of the French bonus-malus system*, Scandinavian Actuarial Journal, Vol. 2006, Issue 5, 2006, pp. 247-264
24. Denuit, M., Maréchal, X., Pitrebois, S., Walhin, J. F., *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales*, Wiley, New York, 2007
25. Dobson, A.J., Barnett, A., *An Introduction to Generalized Linear Models*, 3rd Edition, CRC Press, 2008
26. Frees, E.W., *Regression modeling with actuarial and financial applications*, Cambridge University Press, Cambridge, 2010
27. Frees, E.W., Valdez, E. A., *Hierarchical Insurance Claims Modeling*, Journal of the American Statistical Association, Volume 103, Issue 484, 2008, pp. 1475-1469
28. Gourieroux, C., Jasiak, J., *Heterogeneous INAR(1) Model with Application to Vehicle Insurance*, Insurance: Mathematics and Economics Vol. 34, Issue 2, 2004, pp. 177-192
29. Greene, W.H., *Econometric Analysis*, Prentice Hall, 2002
30. Greene, W.H., *Functional forms for the negative binomial model for count data*, Economics Letters, Elsevier, Volume 99, Issue 3, 2008, pp. 585-590
31. Hilbe, J.M., *Negative Binomial Regression*, 2nd edition, Cambridge University Press, New York, 2011
32. Hilbe, J.M., *Modeling Count Data*, Cambridge University Press, New York, 2014
33. Ismail, N., Jemain, A.A., *Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models*, Casualty Actuarial Society Forum, 2007
34. Ohlsson, E., Johansson, B., *Non-Life Insurance Pricing with Generalized Linear Models*, Springer - Verlag, Berlin, 2010
35. Jong, P., Heller, G.Z., *Generalized Linear Models for Insurance Data*, 5th Edition, Cambridge University Press, New York, 2013
36. Jorgensen, B., *The Theory of Dispersion Models*, Chapman and Hall, London, 1997
37. Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M., *Modern Actuarial Risk Theory Using R*, Second Edition, Springer - Verlag, Berlin, 2008

38. Lambert, D., *Zero-Inflated Poisson regression, with an application to defects in Manufacturing*, Technometrics, Volume 34, Issue 1, 1992, pp. 1-14
39. McCullagh, P., Nelder, J.A., *Generalized Linear Models*, Second Edition, Chapman and Hall, London, 1989
40. Nelder, J.A., Wedderburn, R.W.M., *Generalized Linear Interactive Models*, Journal of the Royal Statistical Society, 1972, pp. 370-384
41. Nelder, J.A., Pregibon, D., *An Extended Quasi-Likelihood Function*, Biometrika, Volume 74, 1987, pp. 221-232
42. Ohlsson, E., Johansson, B., *Non-life insurance pricing with Generalized Linear Models*, Springer Verlag, 2010
43. Pinquet, J., *Une analyse des systèmes bonus-malus en assurance automobile*, Assurances, Volume 67, Issue 2, 1999, pp. 241-249
44. Pinquet, J., *Essais en économétrie de l'assurance non-vie*, Habilitation à diriger des recherches, discipline: Sciences Economiques, 2010
45. Vasechko, O.A., Grun-Réhomme, M., Benlagha, N., *Modélisation de la fréquence des sinistres en assurance automobile*, Bulletin Français d'Actuariat, Volume 9, Issue 18, 2009, pp. 41-63
46. Vasechko, O.A., Grun-Réhomme, M., Albizzati, M.O., *Les jeunes conducteurs : surprimes ou fidélisation?*, Working Papers ERMES from ERMES, University Paris 2, 2010
47. Vasechko, O.A., Grun-Réhomme, M., *L'impact de la sinistralité passée sur la sinistralité future (2): une modélisation des classes de risques*, Insurance and Risk Management, Volume 79(3-4), October 2011- January 2012, pp. 279-311
48. Vuong, Q., *Likelihood ratio tests for model selection and non-nested hypotheses*, Econometrica, Volume 57, Issue 2, 1989, pp. 307-33
49. Whitney, A.W., *The theory of experience rating*, Proceedings of the Casualty Actuarial Society, Vol. 4, 1918, pp. 274-292
50. Yip, K. Yau, K., *On Modeling Claim Frequency Data in General Insurance with Extra Zeros*, Insurance: Mathematics and Economics, Volume 36, Issue 2, 2005, pp.153-163