YEREVAN STATE UNIVERSITY

Gulyan Anahit Gevorg

Martingale Approach to Bonus-Malus Systems

Dissertation

08.00.08-"Mathematical Economics" Specialty

Submitted for the degree of candidate of Physico-Mathematical sciences

Scientific adviser: Candidate of Physico-Mathematical sciences

Chitchyan R.N.

Yerevan 2016

Table of Contents

Introduction

<u>Topicality of the Dissertation:</u> There are many types of Bonus-Malus Systems (BMS) used in the world. They have various applications. Insurance companies operating in some countries are highly recommended to apply the same BMS for a particular class of insurance such as Compulsory Third Party Liability (CMTPL) insurance in Armenia. The most developed countries in this area, mainly European ones, entitled full or partial freedom to insurance companies by developing a highly competitive market. The common characteristic of these systems is that the transition of a policyholder from one BMS class to another is described by the number of claims incurred by that policyholder. In applying this rule many problems are arise which cannot be solved by the experience rating methods used up to now. The application of martingale theory in this field gives new opportunities to introduce more efficient systems. The topicality of this dissertation is conditioned by the study of the current issues of BMS's and finding new ways to solve them.

<u>The Objective and Issues:</u> The main aim of this dissertation is the construction of new BMSs, which will include the number of claims and aggregate claim amount components as a posteriori risk classification. The systems constructed here must be financially balanced and the future malus of a policyholder must be proportional to the loss incurred by insurance company because of him. On the basis of the given dissertation the following problems have to be studied:

- Analyzing currently operating BMSs; considering the approaches of solving or eliminating their current issues
- Suggesting new systems satisfying to optimal BMS definition
- Testing suggested systems by using Armenian CMTPL insurance market data

<u>Research Methods and Informational Backgrounds:</u> Two new BMS models are presented in the following dissertation. On the basis of the first model's construction stands the

martingale approach. For the second model the Markov process have been applied where the parameters of the model were estimated with the Expectation Maximization (EM) algorithm.

The data used in this dissertation has been gained from “IngoArmenia” Insurance CJSC, from the official site of the “Armenian Motor Insurance Bureau” and from the professional literature sources.

The databases were analyzed with the help of Easy-Fit, SPSS and MS Excel software packages.

<u>Scientific Novelty:</u>

1. An alternative model for BMS was proposed where a necessary and sufficient condition was found out for the premiums of the insurance policies’ portfolio to form a martingale series.

2. It was shown that the proposed model can reach to a stationary state.

3. An upper bound for the ruin probability in the alternative BMS was found out with the help of martingales and supermartingales.

4. It was stated the claim amount below which the “bonus hunger” phenomenon arises in the alternative BMS model

5. A generalized BMS model was proposed where the transition of the policyholder among BMS classes described by his/her current class, by number of claims and by aggregate claim amount

6. Estimates for generalized BMS model parameters were stated with the help of “hidden” Markov models and change of measure

7. For the claims number and aggregate claim amount random variables hypothesis for distributions were made on the basis of data received from insurance company.

8. Comparative analyzes of the current BMS with BMSs proposed in this dissertation were done.

The all results presented in the dissertation are new.

Practical and Theoretical Significance: The main results of this research have theoretical and practical character as well. The models presented in this work can be used by insurance companies as well as by the supervisory and decision-making participants of the insurance market for the research, strategic and commercial purposes.

Approbation: The main results of the dissertation have been presented in the scientific seminars held in the department of Actuarial Mathematics and Risk Management of YSU, in the IX International Academic Congress "Contemporary Science and Education in Americas, Africa and Eurasia", Rio De Janeiro, Brazil, 18-20 July, 2015. The results have been discussed at the "IngoArmenia" insurance company and at the Central Bank of Armenia.

The main results of the dissertation are presented in 4 (four) articles, 3 (three) of which were published in the journals accepted by Supreme Certifying Commission (SCC) of Armenia and one was included in the SCOPUS database.

The Structure and the Content of the Dissertation: The dissertation is stated in 109 pages (the appendix is excluded); consists of an introduction, four main chapters, conclusion, appendix and the list of 134 cited references.

Short Description of the Dissertation

The first chapter of the dissertation, which is called **The Contemporary Research Perspectives of the Bonus-Malus Systems,** is devoted to the international experience of functioning BMS's, to the ways of their study, to their current issues and to the problem-solving methods and approaches applied up to now. The first chapter presents also the necessity of this dissertation; the main characteristics of the models introduced in this chapter which allow eliminating current BMS's some problems.

In the second chapter called **BMS Study under the Financial Balance Principle** a new BMS model constructed on the basis of "optimal" BMS principle is suggested. Under the "optimal" BMS principle the maluses and bonuses provided by an insurance company during an insurance year are financially balanced. In this model for the calculation of the insurance premium the yearly aggregate claim amount paid by company is used. Parameters of the model are selected in such a way that the premiums collected by an insurance company will form a martingale series. It is suggested to apply a critical claim amount method for the evaluation of the model's parameters. For the model presented in the second chapter of this work the upper bound for the ruin probability of an insurance company is also derived with the help of martingales.

In the third chapter of this dissertation called **BMS Analysis Using Hidden Markov Models** another new BMS model is introduced, where the next BMS class of a policyholder is determined by the number of his claims and by the aggregate claim amount. The real number of claims and aggregate claim amount of a policyholder are considered as "hidden" Markov processes. The parameters of the model are estimated with the Expectation-Maximization (EM) algorithm.

The last chapter is devoted to the testing of new models presented in this dissertation by a comparative analysis with the BMS used in Armenian CMTPL insurance.

CHAPTER1

THE CONTEMPORARY RESEARCH PERSPECTIVES OF THE BONUS-MALUS SYSTEMS

Introduction

Bonus-Malus System (BMS) is a tool used by insurance companies to "penalize" the policyholders responsible for one or more claims by a premium surcharge (malus) and to "reward" the policyholders who had a claim-free year by awarding discount of the premium (bonus).

Describing BMS one should mention their main characteristics, which are:

- BMS classes (with the help of them one recognizes the amount of future premium)
- The beginner's class (new policyholder join to the system from that class)
- The transition rules (the pre-defined conditions in case of which the policyholder moves from its current class to another)

The BMSs were used from 1950's. They can be applied in different areas of insurance but they appear mainly in motor transport insurance (CASCO) and in motor third party liability insurance for road vehicle (MTPL).

The Section of International Actuarial Association for **A**ctuarial **ST**udies **I**n **N**on-life insurance (ASTIN) was created in 1957. In those years at ASTIN's conferences great attention was paid to the problem of "fairly constructed premium". To solve this problem mathematically the policies with no-claim-discount system were considered. Actually the no-claim-discount system is the special case of BMS.

1.1. Bonus-Malus Systems as Markov Chains

The first theoretical and practical works on construction of BMS belongs to Bichsel [1] and Buhlmann [2], who consider the problem under the game theory framework. Derron [3] states “that a subsequent adjustment of premiums according to the past claim record can be a suitable way of obtaining a fair premium”. Gurtler [4], [5], [6] introduced a standard for evaluating the fairness of a premium. Derron [3] extends and complements the results obtained by Gurtler. Welten [7] points out that the bonus a policyholder obtains usually consists of at least three components depending on the length of time, which precedes the current insurance period: a component concerning the individual claim frequency, an individual random factor and a collective random factor. The last two components tend to zero for increasing length of time. The sum of these last two components, called “unearned bonus”, should be taken into account by insurance companies in the short run, and would lead to a bonus reserve.

Many of the BMS in practice follow a Markov chain consisting of a finite number of classes any of which corresponds to some percentage of the base premium. The premium can be reviewed upward or downward depending on a policyholder’s past record of reported accidents and in accordance with transition rules (see for instance [8], [9], [10] and [11]). To get the next class occupied by a policyholder it is enough to have information on its current class and the number of claims made by him during the current period. This come to show that the BMS can be described as Markov chain: the future (the class occupied in $t+1$ year) depends on present (on the current class and on the number of claims during the current period) and does not depend on past (the full history on claims number and occupied classes in $1,2,\dots,t-1$ periods).

One of the characteristic features of Markov chain is its transition matrix. In the case of BMS it is the probability matrix of policyholder’s transition from one class to another which are explored in a number of works (see for example [12], [13]): The transition of policyholder from one state to another depends on the number of claims which is a random variable and represents a great interest from mathematical point of view. The probabilistic distributions of claim counts and their forecasting methods can be found in [14], [15], and [16]. The simplest and the most popular distribution considered is the Poisson distribution to which devoted the works [10] and [17] .

To identify whether some probability law is applicable to claims distribution, Gosslaux and Lemaire [18] examined six different observed claim histories of automobile insurance third party liability (TPL) portfolios received from five countries. They fitted the Poisson distribution, the Negative Binomial distribution, the Generalized Geometric distribution and the mixed Poisson distribution to each of them by the method of Moments and the Maximum Likelihood method[1]. They concluded that there is no single probability distribution law providing a good fit to all of them. Moreover, there was at least one example where each model gets rejected by a Chi-square test (at the level 10%). After them Kestemont and Paris [19] defined a large class of probability distributions by using mixtures of Poisson processes and developed an efficient method of estimating its parameters. For the six data sets mentioned above they proposed a law depending on three parameters which always gives extremely good fits. Willmot [20] showed that the Poisson-Inverse Gaussian distribution merit consideration as a model for the claims distribution as it fits to data well. Furthermore, this law enjoys abundance of convenient mathematical properties. The author compared the Poisson-Inverse Gaussian distribution with Negative Binomial one and concluded that the fits are better with the Poisson-Inverse Gaussian in all the six cases studied in [18]. The Poisson-Inverse Gaussian distribution was investigated in [21] and [22] as well. Ruohonen [23] considered a model for the claim number processes. The model is the weighed Poisson process with a three-parameter Gamma distribution as the structure function and is compared with the two-parameter Gamma model which gives the Negative Binomial distribution. He fitted his model to some data that can be found in the actuarial literature and the results were satisfying. Panjer [24] proposed to model the number of automobile claims with Generalized Poisson-Pascal distribution which includes three parameters (in fact it is the Hofmann distribution). The fits obtained were satisfactory, too. Note that the Poisson-Inverse Gaussian, the Polya-Appeli and the Negative Binomial distributions are special cases of this distribution. Consul [25] tried to fit the same six data sets by the Generalized Poisson law. Although the Generalized Poisson distribution is not rejected by a chi-square test, the fits obtained in [19], for instance, are always much better. Moreover, Elvers [26] proved that the Generalized Poisson distribution did not fit very well the data observed in automobile TPL insurance portfolio. According to Elvers' note, the distribution hypothesis was in almost every case rejected by a chi-square test. Later Ter Berg [27] considered

[1] Some of the listed discrete distributions are presented in Appendix, Table A.1

slightly different model which involves the Generalized Poisson as well. He introduced a log-linear model, which is able to combine explanatory variables. The results were satisfactory.

Denuit [28] concludes that the Poisson-Goncharov distribution introduced in Lefèvre & Picard [29] is suitable to describe the claims count distribution in automobile insurance. Using approptiate evaluation methods he shows that the Poisson-Goncharov distribution is a good fit for the six data sets considered in [18] as well as for the other data sets achievable in actuarial literature. As a special case of probabilistic distribution introduced in [19], the authors of [16] investigated ordinar Poisson, Poisson-Inverse Gaussian and Negative Binomial distributions.

It was initially assumed that the claim process is homogeneous. In contrast to that assumption the experience shows that automobile insurance portfolios are manifold. The fact that in a portfolio the policyholders are drivers of different personalities and professional background, it forces researches to conclude that the claim processes are heterogeneous. In this case the mixture distributions are applicable where the "mixing" distribution becomes a tool for risk heterogeneity assessment. It assumes in mixture distributions that the number of claims follows to some distribution which parameter is not determined in advance and is random variable. The first reference to this statistical method is given by Albrecht [30]. Lemaire [31] uses the Gamma distribution for that parameter and get the Negative Binomial distribution for claims number process. This property was mentioned also in [15], [32], [33] and [34]: The BMS construction in that case is very simple. Trembley [21] used the Inverse Gaussian distribution to random parameter and get the Poisson-Inverse Gaussian distribution for claims number. It seems that in the above mentioned work [21] the construction of the model is very complicated where the use of modified Bessel functions is needed, but in fact it is not necessary. In [35]the automobile insurance portfolio assessment done with the help of a three-parameter distribution encompassing the Negative Binomial and the Poisson-Inverse Gaussian distributions, which was also discussed in [19] and is known as Hofmann distribution. With the help of Simar's [36] "Non-Parametric Maximum Likelihood Estimation" (NPMLE) method in [35] the non-parametric estimation used also to the claims number process. It is assumed that the claims number follows to the mixture Poisson distribution, but the "mixing" distribution is not known in advance. The estimation is given with the help of Maximum Likelihood method. Although the NPMLE is a power method for claims number assessment, but it is not applicable to BMS construction. The disadvantage of that method is in its discrete nature. To eliminate this fact in [37] the smoothing of the method described in [36] was made with Gamma kernel. A similar work has been done in [38] where the Tucker-Lindsey moment estimation was smoothed with lognormal kernel.

To analyze BMS with Markov chain transition matrix on a long time period one should discuss its stationarity problem. The solution of this problem is used by the actuaries of insurance companies for the long term forecasting. With the help of stationary distribution it is possible to identify the distribution of policyholders' among BMS classes. Each BMS class corresponds to some premium level, so the total amount of future premium income can be evaluated. Several algorithms have been proposed in order to compute the stationary distribution of the policyholders' in a given BMS. Dufresne [39] proposed a very nice technique requiring independence between the annual numbers of accidents per policyholder. Although the application of the Poisson process is possible, unfortunately, this independence assumption rules out all the mixed Poisson distributions. We should note that the Dufresne's method would not be applicable to BMS with non-uniform penalties per claim while the technique described in [35] remains applicable for all BMS. Dufresne adapted the reasoning to the mixed Poisson case in his work [40], but at the cost of many numerical difficulties. On the basis of the model offered in [39], the authors of [35] search the stationary distribution with the non-parametric NPMLE method. This method was used also in [9] for the TPL insurance case.

It is possible that the system could not reach to its stationary state at all or the rate of convergence to stationary state may be slow in comparison to the typical sojourn time of a customer in the portfolio. This problem was discussed in [17], [41], [42] considering the Bayesian view of premium calculation which can be found in [16], [43], [44].

In most analysis of BMS it is assumed that the claim frequency of an individual policyholder remains constant during the time. This assumption implies to constant transition probabilities and makes it possible to model BMS as a homogeneous Markov chain [10]. However it is known that the claim frequency may be time dependent for many reasons.

Moreover, actuarial tools used to evaluate BMS such as the mean first passage times and stationary probabilities are not necessarily monotonic functions of claim frequency. That is why it is often difficult to determine these measures' reaction to changes in claim frequency. In order to mitigate the assumption of a constant claim frequency in [45] is the concept of ergotic Markov Set-Chains defined by Hartfiel [46] used. This method signifies a specific generalization of the idea of classical Markov chains, to analyze the consequences of claim frequency changes, to evaluate the possible change intervals of the transition probabilities, the stationary distributions as well as the mean first passage time.

There are BMSs which are not Markovian. One of such systems is Belgian BMS which is a "dying" one. From 2004 Belgian companies have complete freedom of using their own BMSs. The Markov property disturbed due to the special bonus rule sending the policyholder in the malus zone to initial class after four claim-free years. The works referred to this special bonus rule and to Belgian BMS are [10], [47], [48], [49]. Lemaire [10] proposed to split the classes from 16 to 21 into subclasses. This method gives an opportunity to get free from the special bonus rule and consider the model as a Markov chain again.

1.2. Risk Classification in BMS

One of the main tasks of an actuary is to design a tariff structure that fairly distributes the loading of current claims among its portfolio of current policyholders. If the insurance portfolio consist of heterogeneous risks, then it is fair to split the portfolio into homogeneous groups of policies with policyholders belonging to the same group paying the same premium.

The classification variables used to partition the portfolio into homogeneous groups are called *a priori* classification variables. Their values can be determined at the start of the insurance. In TPL insurance, for instance, the commonly used classification variable are the age, gender, type and use of car, occupation, residential address, marital status, etc. To identify the a priori classification variables The Generalized linear models are used. The examples of their use in actuarial literature are [50], [51], [52].

In the models which are only the function of time and of past number of accidents the characteristics of each individual are not taken into consideration. As mentioned in Dionne and Vanasse [11], the premiums do not vary simultaneously with other variables that affect the claim frequency distribution. One of the most interesting examples is the age variable. Suppose that age has a negative effect on the expected number of claims. It would imply that insurance premiums should decrease with age. Even though age is a statistically significant variable and premium tables are functions of time, the premium derived from BMS based only on the *a posteriori* criteria and not on age factor.

In most practical situations some of important a priori classification factors such as driver's knowledge of rules of the road, reflexes, driving style etc. cannot be taken into account. So, even after the a priori classification variables have been chosen, tariff groups may still be heterogeneous. In this case it remains to believe that these characteristics are revealed by the number and sizes of claims reported by the policyholders over the incoming insurance periods.

It is interesting to mention that in North America, emphasis has traditionally been laid on a priori ratings using many classifying variables, while in Europe much importance was placed on the a posteriori evaluation of policyholders and just a few a priori classifying variables were used. From 1994, however, European Union (EU) introduced a complete rating freedom and each insurance company now is free to set up its own rates, select its own a priori and a posteriori classification variables, and finally, design its own BMS. Most of the companies in EU have taken advantage of this freedom by introducing more rating factors.

Thus, the next premium in BMS adjusted taking into account the individual claim experience of a policyholder. So, it turns out that the policyholders are classified by a priori classification variables initially and after each insurance period they are reclassified by BMS rules according to their claim experience.

The a posteriori rate-marking is an efficient way of classifying policyholders into groups according to their risk. As noted by Lemaire [10], if insurers are allowed to use only one rating variable, it should be a merit rating variable, as merit rating variables are the best predictors of the claim frequency of a policyholder. Besides encouraging policyholders to carefully driving, merit rating systems aim to better evaluate individual risks, so that every policyholder in the long run will pay a premium corresponding to his own claim frequency. Such systems are called experience rating, merit rating, no-claim discount or bonus-malus systems.

Dionne and Vanasse in [11] and [53] presented a BMS that integrates a priori and a posteriori information on an individual basis. This BMS is derived as a function of the number of accidents, of the years that the policyholder is in the portfolio and of the individual characteristics which are significant for the claim frequency. Picech [54] and Sigalotti [55] derived a BMS that incorporates the a posteriori and the a priori rate-marking, with the engine power as the single a priori rating variable. Sigalotti developed a recursive procedure to compute the sequence of increasing equilibrium premiums needed to balance out premiums income and expenditures compensating for the premium decrease created by the BMS transition rules. Picech developed a heuristic method to build a BMS that approximates the optimal merit-rating system. Taylor [56] developed the setting of a Bonus-Malus scale where some rating factors are used to recognize the differentiation of underlying claim frequency by experience, but only to the extent that this differentiation is not recognized within base premiums. Pinquet [57] developed the design of optimal BMS from different types of claims, such as claims at fault and claims not at fault.

Another possible approach is the a priori construction of some risk homogeneous classes and then to build up a bonus-malus table for each class. Following this purpose, Lazar & all [58] have estimated the number of claims as the dependent variable with the Poisson regression model. Between the significance explicative variables there could be noted: the gender of the insured person, the brand of the car, the cylinder capacity of the engine, and the age of the car. Taking all this into consideration, the research made in [34] states that the ignorance of the claim size in a posteriori tariff systems is not always justified, even though the risk classification is homogeneous.

1.3. BMS Current Issues

Currently operating BMSs are multifarious as well as their analysis tools, comparison benchmarks and current issues.

For the diversity of BMSs' such as presented in [59] there must be tools to make a comparison analysis among them. Loimaranta [60] develops formulas for some asymptotic properties of BMS, where the systems are understood as Markov chains. He introduces the quantities like efficiency of a BMS, discriminatory power of bonus rules and minimum variance bonus scale. The last one gives an asymptotic solution for the problem to find locally "the best" bonus scales for given bonus rules. On the basis of methods introduced in [60] Vepsalainen [61] studied the BMSs used in Denmark, Norway, Sweden, Finland, Switzerland and West Germany. Lemaire [62] defines an efficiency concept for a BMS, which differs from the concept given in [60]. De Pril [63] presented a more general concept of efficiency, which includes both earlier ones as special cases. On the basis of the above mentioned analysis tools Lemaire and Zi [59] compare 30 various systems of bonus-malus used in different countries. Lemaire summarizes the previous researches in [10] and [64]. He shows five measures of BMS efficiency

- ✓ the relative stationary average level
- ✓ the coefficient of variation of the insured's premium
- ✓ elasticity of the mean asymptotic premium with respect to the claim frequency
- ✓ the average optimal retention
- ✓ the rate of convergence of BMS

From practical point of view it is well known that the existing BMSs possess several considerable disadvantages which are difficult or even impossible to handle within the traditional theory of experience rating [65]. Therefore it is necessary to examine them from different point of view. The existing systems are based on the following characteristic: the claim amounts are omitted as a posterior tariff criterion. This characteristic leads to the following disadvantages:

i. Regarding an occurred claim, the future loss of bonus will in many cases exceed the claim amount.

ii. In many cases it gives the policyholder a filling of unfairness especially when a policyholder makes a small claim and the other one a large; they have the same penalty within the same risk group.

iii. The consequence of (i and ii) is the well-known bonus hunger behavior of policyholders.

iv. Bonus hunger behavior leads to asymmetric information between policyholders, insurers and regulators.

Many authors have focused on the disadvantages mentioned above. The aim of these authors has not been to solve or eliminate the disadvantages, but rather to take them into the modeling account in connection with the mathematical optimization of the BMS.

"Deductibles"

To diminish some of the disadvantages (i-iv) Holtan [65] suggested the use of very high deductibles that may be borrowed by the policyholder to the insurance company. He assumes a constant deductible for all policyholders, so it is independent of the level they occupy in BMS. Although this approach is technically acceptable, it causes considerable practical problems. Based on this in [66] suggested using deductibles depending on the BMS class of the policyholder. This approach leads to competition in insurance market. The insurer is not inclined to leave the company after a claim as the deductible is payable after the claim submitting.

"Bonus hunger"

Among the disadvantages (i-iv) the huge share belongs to the problem which in 1960 Philipson [67] called “hunger for bonus”. Alting von Geusau [68] investigates “to what extent it is possible to develop a theoretical framework to test that a no-claim-discount-system will prevent the insured from submitting small claims to the insurance company”, and “that the insured who has just lost his no-claim discount will use every possibility for submitting claims with in his mind the idea that in this way he will earn back his higher non-reduced premium” [69].

Grenander [70] derives equations to determine a rule of the form “pay the damage if its amount is smaller than a critical value and claim it otherwise”. However, the equations are generally difficult to solve, and it is not proved that they really determine an optimal policy in the sense that the total expected discounted cost of premiums and payments during a long future planning period is minimized. Haehling von Lanzenauer [71], [72], [73] analyses the problem on the assumption that a policyholder can cause at most one accident per year.

De Leve and Weeda [74] suggest a mathematical model, where the policyholder can cause more than one claim in a year but after making one claim during an insurance period, the policyholder is placed in the class with the highest premium. The model called generalized Markov programming. It yields an optimal strategy, which is a function $s(t)$ that minimizes the expected costs for the policy holder. The function $s(t)$ is such that "if at any time t an accident occurs with damage s and no damages have been claimed since the last payment of premium, then s should be claimed if $s > s(t)$". In this approach the decision depends on the point of time during the year and the premium paid at the beginning of that year. Weeda [75] extends the analysis of the same model to case where the damage distribution is given by an arbitrary distribution and focuses on the theoretical aspects of the derived iteration scheme. However, although the model is continuous with respect to time axis, he considers discrete time for computational purposes.

Martin-Lof [76] shows that a decision rule of the form formulated in [70] is optimal in the sense that it minimizes total expected costs. The decision rule is derived by applying the general theory of Markov decision processes, which find an optimal control iteratively by using dynamic programming. In that work, however, the analysis was restricted to the case where the policyholder takes a decision only at the end of an insurance period for the total amount of damage sustained during that insurance period.

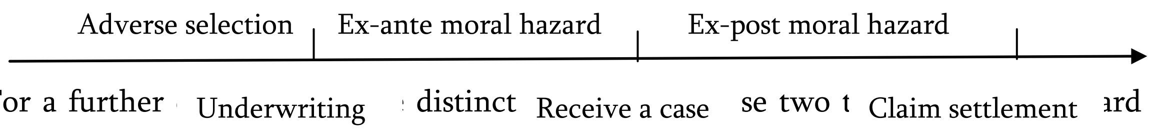
Haehling von Lanzenauer and Lundberg [77] develop a model which can be used in deriving the distribution of the number of claims for insurances with merit-rating structures. The problem is formulated and solved as a regular Markov process with the claim behavior integrated in the analysis. Haehling von Lanzenauer [78] develops an optimal decision rule for situations where the policyholder takes a decision more than once a year, which is valid for any merit-rating system. He splits up a year into a number of periods, which results in a discrete model in which the optimal critical claim size can be determined by dynamic programming.

However, this derivation of an optimal critical claim size is incomprehensible. Lemaire [79] derives an algorithm for obtaining the optimal strategy for a policyholder. In his model the policyholder remains always insured (the so-called infinite horizon model) which leads to a critical claim size which is independent of the year in which the accident takes place. Also, in order to compute the optimal policy, he uses policy iteration, which is very time-consuming, whenever the state space is large. He applies this algorithm in [62] to compare BMS used in Norway, Denmark, Finland, Sweden, Switzerland and West Germany.

Hastings [80] presents a simple model based on a typical British policy, assuming that the number of accidents is Poisson and the amount of damage is Negative Exponentially distributed. He assumes an optimal critical claim size, which is constant throughout the year, independent of the number of claims already made during the year and of the time until the next premium payment. He determines optimal critical claim sizes, which minimize the long-run average costs of premiums and repairs. The problem is formulated as a Markov decision problem and is solved by dynamic programming. Almost all studies mentioned above have in common that they assume a discrete time axis. De Pril [81] gives a formulation based on a continuous time, where the optimal critical claim size can be determined by solving a set of recurrent differential equations. However, for solving these equations, one should make a discretization of the model, which gives rise to the same results as in [78]. Norman and Shearn [82]build on Hastings' model, where they drop the restriction of a constant optimal critical claim size. Moreover, they present a much simpler state description than the one used by Haehling von Lanzenauer in [78]. The optimal decision rule has been compared with rules of thumb that appear to produce remarkably good results. Tijms [83] gives a model that is equal to the model presented in [82]. Kolderman and Volgenant [84] present a continuous model based on generalized Markov programming, applicable to BMS used by Dutch motor insurance companies. However, in the computational part of their study they use discrete time for numerical reasons. Lemaire [31] describes a simple model with the assumption that all claims are reported in the middle of the insurance period. Menist and Volgenant [85]compute the optimal critical claim size by considering the difference between the expected costs in case of claiming and that of not claiming damage. They restrict the analysis to a finite time horizon. Dellaert at al. [69] restrict their work on TPL insurance, where there is no deductible, i.e. the total amount of damage is covered. The analysis is based on the suggestions of [82], where the time is discrete. This assumption is not very restrictive in practical situations as a policyholder generally allowed some time (at least 24 hours) to decide to claim or not to claim for damage. Dellaert at al. [86] consider the optimal behavior of a policy holder having vehicle damage insurance. Recurrent systems of formulas are derived in each of last two articles. The solution of systems gives the optimal amount limit above which the policyholder will claim the damage to the insurance company.

"Information Asymmetry"

As a consequence of bonus hunger, adverse selection and moral hazard phenomena, information nonconformity among insured, insurers and their regulators arises. In the case of motor insurance the owner generally knows his driving level, the car's performance and other status. It is easy to obtain a car's information, while it is hard to recognize a car owner's driving level directly. It is necessary to classify the customers' risk and charge a reasonable premium, but the underwriting cost will be higher if the insurer wants to identify the risk more closely. Because of information asymmetry, speculative psychology high-risk costumers are more willing to insure in real life, compared to low-risk populations. This is adverse selection. The insurer therefore tends to develop a higher premium, low-risk populations (quality customers) will eventually be "driven out" of the auto insurance market, and the market efficiency will be reduced. After an insurance contract become effective, the insurer cannot be aware of the owner's driving level and his cautious attitude. In case of an accident, the driver's derogation measures cannot be observed either. The accident due to owner's negligence and the fraud after the accident will increase the claims costs, so the insurer also faces moral hazard. Moral hazard can be subdivided into two types' *ex-ante* and *ex-post*. Ex-ante moral hazard occurs before an accident. It concerns the impact of the policyholder's actions on the probability of occurrence and severity upon occurrence of the insured event. For example, the policyholder would demonstrate ex-ante moral hazard if they smoked in bed only after the purchase of fire insurance. Ex-post moral hazard occurs during or after an accident, and affects the severity of the claim. For example, the policyholder would demonstrate ex-post moral hazard if they did not pick up their wallet as they escaped from their burning house, which they would have done if they had not been covered by the fire insurance policy. Here is a concise figure of that:



For a further … distinct … se two t … ard see [87].

Rothschild and Stiglitz [88] and Stiglitz [89] discussed how to design an optimal insurance contract to deal with adverse selection and moral hazard. Dionne and Lasserre [90], Cooper and Hayes [91] discussed the multi-period insurance contract, and pointed out the experience ratemaking and risk classification can solve information asymmetry. There are a lot of researches to answer the question: Do adverse selection and moral hazard exist in the auto insurance market? Kim. H, Kim. D, Im and Hardin [92] and Cohen and Siegelman [93] proved that moral hazard exists in South Korea's auto insurance market. Unlike the studies of Dahlby [94] and [95], Puelz and Snow [96] and Cohen [97], where the existence of a coverage-risk correlation is suggested, Chiappori and Salanie [98] for French, Dionne et al [99] for Quebec and Saito [100] for Japanese automobile insurance markets use more refined methods and cannot reject a zero correlation between higher insurance coverage and more accidents in their data and cannot find evidence for asymmetric information. By analyzing the above mentioned results for Danish insurance market, Donnelly at all [101] prove that the existence of adverse selection depends on the research method and statistical tests.

"Classification by Claim Severity"

The main reason for BMS disadvantages (i) and (ii) is the application of big maluses for claims with small severity. A reliance or "sense of fair-dealing" to the BMS will arise when the "punishment" of a policyholder as a malus is proportional to the loss incurred by insurance company because of him. This leads to BMS construction with taking into account the claim severity as well.

One of the first models of BMS designed to take severity into consideration is Picard [102]. Picard generalized the Negative Binomial model in order to take into account the subdivision of claims into two categories, small and large losses. In order to separate large from small losses, two options could be used:

- ✓ The losses under a limiting amount are regarded as small and the remainder as large.
- ✓ Subdivision of accidents in those that caused property damage and those that cause bodily injury, penalizing more severely the policyholders who had a bodily injury accident.

Pinquet [51] designed an optimal BMS which makes allowance for the severity of the claims in the following way: starting from a rating model based on the analysis of number of claims and of costs of claims, two heterogeneity components are added. They represent unobserved factors that are relevant for the explanation of the severity variables. The costs of claims are supposed to follow Gamma or Lognormal distribution. The rating factors, as well as the heterogeneity components are included in the scale parameter of the distribution. Considering that the heterogeneity also follows a Gamma or Lognormal distribution, a credibility expression is obtained which provides a predictor for the average cost of claim for the following period. Frangos and Vrontos [103] assumed that the number of claims is distributed according the Negative Binomial distribution and the losses of the claims are distributed according to the Pareto distribution, and they have expanded the frame that Lemaire [10] used to design an optimal BMS based on the number of claims. Applying Bayes' theorem the posterior distribution of the mean claim frequency and the posterior distribution of the mean claim size given the information about the claim frequency history and the claim size history for each policyholder for the time period he is in the portfolio have been found out. For more on this subject see Vrontos [104]. In [103] the development of a generalized BMS is presented, which integrates the a priori and the a posteriori information on an individual basis. In this generalized BMS the premium is a function of the years that the policyholder is in the portfolio, of his number of accidents, of the size of loss that each of these accidents incurred, and of the significant a priori rating variables for the number of accidents and for the size of loss that each of these claims incurred. Pitrebois at al. [105] suggested introduction of claim amount in the model via premium adjustment factor, which is calculated according to credibility techniques. Bonsdorff [106] discussed some asymptotic properties of Bonus-Malus systems based on the number and on the size of the claims.

***

Analyzing BMS current issues in the following thesis suggested to leave traditional methods discussed up to now and construct systems, which are basically different from those ones. The systems constructed here are financially balanced and the future malus of a policyholder is proportional to the loss incurred by insurance company because of him.

CHAPTER 2

BMS STUDY UNDER THE FINANCIAL BALANCE PRINCIPLE

Introduction

One of the main actuarial principles is the presupposition of financially balanced insurance product. After having analyzed 30 BMSs from around the world Lemaire and Zi [59] concluded that at the end of 30 years the average premium level of a policyholder will be about 40%-70% of the starting (base) premium. Base premium is calculated by the financial balance principle. So, in long time period a financial imbalance arises, which will eventually result in the detriment of the company. Mentioned disbalanced condition can lead to serious financial consequences, even to a crash of the company. For the solution of this problem Sammaritini [107] suggested to permit a driver to move to a lower class only if the claim frequency of his preceding class is lower than a fixed value. Coene and Doray [108] presented 3 hypothetical BMSs, where the financially balanced system is achieved by using the premiums as parameters of the model. The premium for each class of BMS will be determined in such a way that the total premiums received is at least equal to 100% of the initial premium after established number of years.

Generalizing the principle of financially balanced BMSs Lemaire [10] defines the concept of optimal BMS, which has two conditions.

BMS is called optimal if it is:

- Financially balanced for the insurer, i.e. the total amount of bonuses is equal to the total amount of maluses
- Fair for the policyholder, i.e. each policyholder pays a premium proportional to the risk that he imposes to the pool.

In the view of probability theory, the bonuses and maluses provided by the insurer are random variables, while still the detailed definition of "the total amount of bonuses (or maluses)" remains uncertain. Therefore we offer the following statement of the "optimal" BMS:

"Financially balanced for the insurer, i.e. the expected value of total amount of bonuses is equal to the expected value of total amount of maluses."

In other words, this statement can be interpreted as "the expectation of the BMS total premiums collected by an insurance company remains constant".

One of the processes satisfying to this condition is the martingale series widely known in probability theory.

2.1. Notations and Definitions

Here are some notations and definitions from probability theory (see for instance [109]) used for the model construction.

Suppose that all the observations made on a probability space $(\Omega, \mathcal{F}, P)$, where Ω is the set of elementary outcomes ω, $\mathcal{F}$ is a σ-algebra of subsets of Ω and P is a given probability measure on $\mathcal{F}$.

Time and dynamics have a significant role for the model construction, so consider that a series of σ-algebra $\{\mathcal{F}_n\}_{n\geq 0}$ is given:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \ldots \subseteq \mathcal{F}$$

This non-descending σ- algebra series, otherwise called filtration, is interpreted as follows: the n-th member of the series $\mathcal{F}_n$-is the known information about the observation up to time n.

So, suppose that the basic probabilistic model is $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathrm{P})$ filtered probability space.

Definition 2.1: *Let $(\Omega, \mathcal{F})$ be some measurable space and $(R, \mathcal{B}(R))$ a borelian space with a system of $\mathcal{B}(R)$ borelian sets. A real function $\xi = \xi(\omega)$ given on $(\Omega, \mathcal{F})$ is called $\mathcal{F}$-mesurable function or a **random variable**, if for any $B \in \mathcal{B}(R)$*

$$\{\omega: \xi(\omega) \in B\} \in \mathcal{F}$$

Or, it is the same as $\xi^{-1}(B) \equiv \{\omega: \xi(\omega) \in B\}$ is a measurable set in Ω.

Definition 2.2: *Let $X_0, X_1, \dots$ be a series of random variables given on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, P)$. If X_n is $\mathcal{F}_n$-measurable for any $n \geq 0$, then we will say that $X = (X_n, \mathcal{F}_n)_{n\geq 0}$ collection or just $X = (X_n, \mathcal{F}_n)$-is a **stochastic series**.*

Definition 2.3: *If for $X = (X_n, \mathcal{F}_n)$ stochastic series X_n is $\mathcal{F}_{n-1}$-measurable as well, it will be written as $X = (X_n, \mathcal{F}_{n-1})$ assuming $\mathcal{F}_{-1} = \mathcal{F}_0$ and X will be called **predictable series**.*

Definition 2.4: *Let $X_n: \Omega \to \mathbb{R}$, then $X = (X_n, \mathcal{F}_n)$ stochastic series will be called a **martingale**, if for any $n \geq 0$:*

$$E|X_n| < \infty \tag{2.1.1}$$

$$E(X_{n+1}|\mathcal{F}_n) = X_n \tag{2.1.2}$$

From properties of conditional expectation it is obvious that the second property of martingale definition can be rephrased by the following formula:

$$\int_A X_{n+1} dP = \int_A X_n dP$$

for any $n \geq 0,\ A \in \mathcal{F}_n$ and specially if $A = \Omega$ then it can be written that

$$EX_n = EX_{n-1} = \dots = EX_1 = X_0$$

Definition 2.5: *Let $X_n: \Omega \to \mathbb{R}$, then $X = (X_n, \mathcal{F}_n)$ stochastic series will be called a **submartingale (supermartingale)**, if for any $n \geq 0$:*

$$E|X_n| < \infty$$

$$E(X_{n+1}|\mathcal{F}_n) \geq (\leq) X_n$$

Theorem 2.1: *(Doob's theorem on submartingales[2]): If* $X = (X_n, \mathcal{F}_n)_{n\geq 1}$ *is a submartingale with* $\sup_n E|X_n| < \infty$*, then there exists* $\lim X_n = X_\infty$ *a.s. with* $E|X_\infty| < \infty$.

Corollary 2.1: *If* $X = (X_n, \mathcal{F}_n)_{n\geq 1}$ *is a nonnegative martingale, then there exist* $\lim_{n\to\infty} X_n$ *a.s.*

Theorem 2.2: *(Markov's theorem[3]) The law of large numbers is applicable for series of random variables* Z_k*, with any type of dependence, if the quantity* $\frac{Var\sum_{k=1}^{n} Z_k}{n^2} \to 0$ *(a.s.) as* $n \to \infty$.

Definition 2.6: *A random variable* $\tau = \tau(\omega)$*, which takes values from* $\{0,1,\dots,+\infty\}$ *is called a* ***Markov time*** *with respect to a filtration* $\mathcal{F}_n$*, if for any* $n \geq 0$

$$\{\tau = n\} \in \mathcal{F}_n$$

In the case of $P(\tau < \infty) = 1$ *the Markov time* τ *will be called a* ***stopping time***.

Theorem 2.3*: (Doob's theorem on optional stopping time[4]) Let for a martingale (submartingale, supermartingale) series* $X = (X_n, \mathcal{F}_n)_{n\geq 1}$ τ *is a stopping time with respect to* $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$ *with* $E\tau < \infty$*, and for any* $n \geq 0$ *and some constant* C *holds*

$$E(|X_{n+1} - X_n||\mathcal{F}_n) \leq C$$

then $E|X_\tau| < \infty$ *and*

$$EX_\tau = EX_0$$

$$EX_\tau \geq EX_0\ (for\ submartingale)$$

[2] The proof of the theorem and its corollary are given in [109](pg.688)

[3] The proof can be found in [134](pg.294)

[4] The proof is given in [109](pg.662)

$EX_\tau \leq EX_0\ (for\ supermartingale)$

2.2. Presentation of BMS as Martingale Series

Let us consider a portfolio of an insurance product. Suppose that a series of independent and identically distributed random variables $Y_1, Y_2, \dots$ are yearly aggregate claim losses of that portfolio, given on a $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathrm{P})$ filtered probability space where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma\{Y_1, Y_2, \dots Y_n\}$. And suppose that $Y_1, Y_2, \dots$ random variables are so that $EY < \infty$ condition is satisfied [110].

Let's denote $P_0, P_1, \dots$ as random variables, which describe yearly aggregate premium charge for that portfolio, where $P_0 = const$ is given and the other members of that series are defined by the following formula:

$$P_n = (1 - \alpha_n)P_{n-1} + \beta_n Y_n \qquad n \geq 1 \tag{2.2.1}$$

where

P_n-is an aggregate premium collected for n-th year of the portfolio.

Y_n –is an aggregate claim loss for the given portfolio within $(n-1; n)$ time interval. It is necessary to note that Y_n is independent of P_{n-1} for all n, $(n \geq 1)$.

$\alpha = (\alpha_n, \mathcal{F}_{n-1})_{n\geq 1}$ is a predictable series with $\alpha_n \in (0,1)$, which will be called *a series of bonus factors.*

$\beta = (\beta_n, \mathcal{F}_{n-1})_{n\geq 1}$ is also a predictable series with $\beta_n \in (0,1)$, which will be called *a series of malus factors.*

Lemma 2.1: *The series $P = (P_n, \mathcal{F}_n)$ constructed by formula (2.2.1), where α_n and β_n are $\mathcal{F}_{n-1}$-measurable, is a martingale if and only if:*

$$\frac{\alpha_n}{\beta_n} = \frac{EY}{P_{n-1}} \tag{2.2.2}$$

Proof:* *Necessary: Let series α_n and β_n are $\mathcal{F}_{n-1}$-measurable and the series constructed by (2.2.1) is a martingale. It means that the conditions of *Definition 2.4* are satisfied

$$P_{n-1} = E(P_n|\mathcal{F}_{n-1}) = E((1-\alpha_n)P_{n-1} + \beta_n Y_n|\mathcal{F}_{n-1}) = (1-\alpha_n)P_{n-1} + \beta_n EY$$

Here the independence of Y_n's and the properties of conditional expectation (see [109], pg.270) were used.

For the (2.1.2) condition of martingale definition it is necessary that:

$$\beta_n = \frac{\alpha_n P_{n-1}}{EY}$$

It is obvious that this result is equivalent to (2.2.2).

Sufficiency: Let's α_n and β_n are $\mathcal{F}_{n-1}$-measurable series and the relationship (2.2.2) holds. Let construct a series of P_n according to the formula (2.2.1) and show that $P = (P_n, \mathcal{F}_n)$, where $\mathcal{F}_n = \sigma\{Y_1, Y_2, \dots Y_n\}$ is a martingale.

Calculate the conditional expectation by putting (2.2.2) in (2.2.1):

$$E(P_n|\mathcal{F}_{n-1}) = E\left(P_{n-1} - \alpha_n P_{n-1} + \frac{\alpha_n P_{n-1}}{EY_n} Y_n \middle| \mathcal{F}_{n-1}\right) = P_{n-1} - \alpha_n P_{n-1} + \frac{\alpha_n P_{n-1}}{EY_n} EY = P_{n-1}$$

∎

Remark 2.1: *The relationship (2.2.2) can be considered as loss ratio of the given insurance portfolio.*

2.3. Financial Stability Coefficient Estimates

Offering a new insurance product the insurance company wishes to have a financially stable model. For that purpose it states its strategy for that risk portfolio and defines a premium level. The aggregate premium received for that portfolio must be sufficient to cover some level of aggregate claim with appropriate probability which is defined in the company's strategy. This means that the company states some Y_c critical value of aggregate claim and some ε probability and defines the aggregate premium P_n so that it is greater than Y_c critical value with $1-\varepsilon$ probability. This method is called a quantile method. Mathematically it is expressed as:

$$\mathrm{P}(P_n > Y_c) = 1 - \varepsilon \qquad (2.3.1)$$

Finding α_n and β_n presented in (2.2.1) is our main purpose but Lemma 2.1 gives us only their relationship (2.2.2). Suppose that the distribution function of aggregate claim is given $Y \sim F_Y(x)$. Let's find α_n and β_n with the help of expressions (2.2.2), (2.3.1) and financially stable and optimal BMS concepts. First of all let's reform the left side of formula (2.3.1).

$$P(P_n > Y_c) = P((1-\alpha_n)P_{n-1} + \beta_n Y_n > Y_c) = P\left(Y_n > \frac{Y_c - (1-\alpha_n)P_{n-1}}{\beta_n}\right)$$
$$= 1 - F_Y\left(\frac{Y_c - (1-\alpha_n)P_{n-1}}{\beta_n}\right)$$

It is obvious that

$$F_Y\left(\frac{Y_c - (1-\alpha_n)P_{n-1}}{\beta_n}\right) = \varepsilon$$

Using inverse distribution function we get:

$$\frac{Y_c - (1-\alpha_n)P_{n-1}}{\beta_n} = F_Y^{-1}(\varepsilon)$$

Substituting (2.2.2) and making some rearrangements we get:

$$\beta_n = \frac{Y_c - P_{n-1}}{F_Y^{-1}(\varepsilon) - EY} \qquad (2.3.2)$$

and

$$\alpha_n = \frac{Y_c - P_{n-1}}{F_Y^{-1}(\varepsilon) - EY} \cdot \frac{EY}{P_{n-1}} \qquad (2.3.3)$$

Hereafter we take the absolute values for α_n and β_n to avoid negative results.

__Lemma2.2:__ The coefficients α_n and β_n have finite limits as $n \to \infty$.

Proof: The nonnegative martingale $(P_n, \mathcal{F}_n)$ given with (2.2.1) satisfies to the conditions of the *Corollary 2.1*, so there exists finite $\lim_{n\to\infty} P_n = P_\infty$ a.s. It is not difficult to note that $|Y_c - P_\infty| < \infty$ and for sufficient small ε the condition $F_Y^{-1}(\varepsilon) - EY \neq 0$ is satisfied. So, we have

$$\lim_{n\to\infty} \beta_n = \lim_{n\to\infty} \left|\frac{Y_c - P_{n-1}}{F_Y^{-1}(\varepsilon) - EY}\right| = \left|\frac{Y_c - P_\infty}{F_Y^{-1}(\varepsilon) - EY}\right| \triangleq \beta < \infty$$

and

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \left|\frac{Y_c - P_{n-1}}{F_Y^{-1}(\varepsilon) - EY}\right| \cdot \frac{EY_n}{P_{n-1}} = \left|\frac{Y_c - P_\infty}{F_Y^{-1}(\varepsilon) - EY}\right| \cdot \frac{EY}{P_\infty} \triangleq \alpha < \infty \qquad \blacksquare$$

2.4. Upper Bound for the Probability of Ruin in the Alternative BMS

Lemma 2.2 leads us to a conclusion that starting from some time n the bonus and malus coefficients will not depend on time and we can consider the following model

$$P_k = (1-\alpha)P_{k-1} + \beta Y_k, \qquad k \geq n \tag{2.4.1}$$

Consider a BMS portfolio where the capital amount at a time i is decreased by the total amount of claims for time interval $(i-1, i)$ and increased by premiums collected at the time i. In addition to the assumptions of independence and identical distribution for Y_k, here we assume also that $VarY < \infty$.

The surplus process of the portfolio is then defined by

$$U_n = u + \sum_{k=1}^{n} P_k - \sum_{k=1}^{n} Y_k \tag{2.4.2}$$

where $u = U_0$ is the initial capital of the portfolio, P_k is the aggregate premium defined with (2.4.1) and Y_k is the aggregate claim amount of the portfolio for time interval $(k-1, k)$. In addition to the independence of Y_k's, here we suppose also that they are identically distributed random variables (i.i.d.) and independent of P_{k-1} as described for model (2.2.1). In addition to that, assume also that $VarY < \infty$.

Now recall some definitions from ruin theory (see for instance [15], [111]).

<u>Definition 2.7:</u> *The event that* U *ever falls below zero is called ruin:*

$$Ruin = \{U_n < 0 \; for \; some \; n\}.$$

<u>Definition 2.8:</u> *The time* $\tau(u)$ *when the process falls below zero for the first time is called ruin time:*

$$\tau(u) = inf\{n > 0; U_n < 0\}.$$

The probability of ruin is then given by

$$\psi(u) = P\left(\bigcup_{n\geq 0}\{U_n < 0\}|U_0 = u\right) = P\left(\inf_{n\geq 0} U_n < 0\,|U_0 = u\right) = P(\tau(u) < \infty)$$

Write

$$Z_k = Y_k - P_k = Y_k - (1-\alpha)P_{k-1} - \beta Y_k = (1-\beta)Y_k - (1-\alpha)P_{k-1} \qquad (2.4.3)$$

This variable shows the net loss of the portfolio at time k.

The total net loss of the portfolio up to time n is defined as

$$S_n = Z_1 + \cdots + Z_n, \;\; n \geq 1, \;\; S_0 = 0 \qquad (2.4.4)$$

So, for the probability of ruin we have the following equivalent expression:

$$\psi(u) = P\left(\inf_{n\geq 1}(-S_n) \leq -u\right) = P\left(\sup_{n\geq 1} S_n > u\right)$$

We know nothing about the independence of Z_k's, but we can calculate their variation. From (2.4.1) P_k can be expressed as:

$$P_k = (1-\alpha)^k P_0 + (1-\alpha)^{k-1}\beta Y_1 + \cdots + (1-\alpha)\beta Y_{k-1} + \beta Y_k \qquad (2.4.5)$$

So, using the independence of Y_k's we get

$$Var(P_k) = Var((1-\alpha)^k P_0 + (1-\alpha)^{k-1}\beta Y_1 + \cdots + (1-\alpha)\beta Y_{k-1} + \beta Y_k)$$
$$= Var(Y)\beta^2 \frac{1-(1-\alpha)^{2k}}{1-(1-\alpha)^2}$$

Using the relationship $(2.4.3)$ and independence of Y_k and P_{k-1}we get:

$$Var(Z_k) = (1-\beta)^2 Var(Y) + (1-\alpha)^2 Var(P_{k-1})$$
$$= Var(Y)\left((1-\beta)^2 + (1-\alpha)^2\beta^2 \frac{1-(1-\alpha)^{2(k-1)}}{1-(1-\alpha)^2}\right)$$

For $Var(S_n)$ we can write:

$$Var(S_n) = Var\left(\sum_{k=1}^{n} Z_k\right) = \sum_{k=1}^{n} Var(Z_k) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} cov(Z_i Z_j)$$

From $(2.4.3)$ and $(2.4.5)$ we conclude that

$$Var(S_n) \sim A \cdot n \cdot Var(Y) + B \cdot (1-(1-\alpha)^{C\cdot n}) \qquad (2.4.6)$$

where $A, B, C < \infty$ are constants.

In the case of $(2.4.6)$ using *Theorem 2.2* we have:

$$\frac{Var(S_n)}{n^2} \sim \frac{A \cdot Var(Y)}{n} + \frac{B \cdot (1-(1-\alpha)^{Cn})}{n^2} \xrightarrow[n\to\infty]{} 0$$

So, S_n satisfies to the law of large numbers, that is $\frac{S_n}{n} \to EZ$, which in particular implies, that $S_n \to +\infty$ or $S_n \to -\infty$ according to the sign of EZ. Hence, if $EZ \geq 0$, ruin is unavoidable in the case of any starting capital u.

<u>Preposition 2.1:</u> *If EY is finite and the condition*

$$EZ = (1-\beta)EY - (1-\alpha)P_0 \geq 0$$

holds then, for any fixed $u > 0$, ruin occurs with probability 1.

The insurance company should choose the bonus and malus coefficients in such a way, that $EZ < 0$. In this case the company avoids ruin occurring with probability 1 and may hope to have $\psi(u) < 1$.

Definition 2.9: *(Net Profit Condition): The process Z satisfies to the net profit condition (NPC), if*

$$EZ = (1-\beta)EY - (1-\alpha)P_0 < 0 \tag{2.4.7}$$

Taking the expectations in (2.4.5), recalling $P_0 = EP_k$ martingale property and using the i.i.d. property of Y_k's we have:

$$P_0 = (1-\alpha)^k P_0 + \frac{1-(1-\alpha)^k}{\alpha}\,\beta EY,$$

which gives:

$$P_0 = \frac{\beta EY}{\alpha}$$

By inputting it in the condition (2.4.7), we get the following NPC

$$\alpha < \beta.$$

This result is in accordance with the interpretation of $\frac{\alpha}{\beta}$ as the loss ratio of the portfolio (see *Remark 2.1*).

Using (2.4.5) for S_n we can write:

$$
\begin{aligned}
S_n = \sum_{k=1}^{n} Y_k - \sum_{k=1}^{n} P_k \\
&= (Y_1 + Y_2 + \cdots + Y_n) \\
&- \left(P_0 \sum_{k=1}^{n} (1-\alpha)^k + \beta Y_1 \sum_{k=1}^{n} (1-\alpha)^{k-1} + \beta Y_2 \sum_{k=2}^{n} (1-\alpha)^{k-2} + \cdots + \beta Y_n \right) \\
&= (Y_1 + Y_2 + \cdots + Y_n) \\
&- \left(P_0 \frac{(1-\alpha)(1-(1-\alpha)^n)}{\alpha} + \beta Y_1 \frac{1-(1-\alpha)^n}{\alpha} + \cdots + \beta Y_n \frac{1-(1-\alpha)}{\alpha} \right) \\
&= -P_0 \frac{(1-\alpha)(1-(1-\alpha)^n)}{\alpha} - \left(\frac{\beta}{\alpha} - 1\right) Y_1 - \cdots \left(\frac{\beta}{\alpha} - 1\right) Y_n \\
&+ \frac{\beta}{\alpha}((1-\alpha)^n Y_1 + \cdots + (1-\alpha) Y_n) \\
&= -C_n - \left(\frac{\beta}{\alpha} - 1\right)(Y_1 + Y_2 + \cdots + Y_n) + \frac{\beta}{\alpha}((1-\alpha)^n Y_1 + \cdots + (1-\alpha) Y_n)
\end{aligned}
$$

Denote

$$M_n = Y_1 + Y_2 + \cdots + Y_n$$

and

$$G_n = (1-\alpha)^n Y_1 + \cdots + (1-\alpha) Y_n$$

So, we have

$$S_n = -C_n - \left(\frac{\beta}{\alpha} - 1\right) M_n + \frac{\beta}{\alpha} G_n$$

Denote $\varphi_X(t) = Ee^{tX}$ the moment generating function (m.g.f.) of the random variable X.

Lemma 2.3: *For any $\gamma > 0$, and for any $\alpha, \beta \in (0,1)$ satisfying the NPC, the sequence*

$$M'_n = \frac{e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)M_n}}{\varphi_Y^n\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)}$$

is a martingale.

Proof: Note that $M_{n+1} = M_n + Y_{n+1}$, Calculate

$$E(M'_{n+1}|Y_1,Y_2,\dots,Y_n) = E\left(\frac{e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)M_{n+1}}}{\varphi_Y^{n+1}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)}\middle|Y_1,Y_2,\dots,Y_n\right)$$

$$= E\left(\frac{e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)(M_n+Y_{n+1})}}{\varphi_Y^{n+1}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)}\middle|Y_1,Y_2,\dots,Y_n\right)$$

$$= \frac{e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)M_n}}{\varphi_Y^{n}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)} E\left(\frac{e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y_{n+1}}}{\varphi_Y\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)}\middle|Y_1,Y_2,\dots,Y_n\right) = M'_n\frac{E\,e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y_{n+1}}}{\varphi_Y\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)}$$

$$= M'_n$$ ∎

Lemma 2.4: *For any $\gamma > 0$, such that $\varphi_Y(\gamma) < \infty$ and for any $\alpha, \beta \in (0,1)$ satisfying the NPC, the sequence*

$$G'_n = \frac{e^{\gamma\frac{\beta}{\alpha}G_n}}{\varphi_Y^n\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}$$

is a supermartingale.

Proof: Note that

$$\begin{aligned} G_{n+1} &= (1-\alpha)^{n+1}Y_1 + \dots + (1-\alpha)^2 Y_n + (1-\alpha)Y_{n+1} \\ &= (1-\alpha)((1-\alpha)^n Y_1 + \dots + (1-\alpha)Y_n) + (1-\alpha)Y_{n+1} = (1-\alpha)(G_n + Y_{n+1}) \end{aligned}$$

Calculate

$$E(G'_{n+1}|Y_1, Y_2, \dots, Y_n) = E\left(\frac{e^{\gamma\frac{\beta}{\alpha}G_{n+1}}}{\varphi_Y^{n+1}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle|Y_1, Y_2, \dots, Y_n\right)$$

$$= E\left(\frac{e^{\gamma\frac{\beta}{\alpha}(1-\alpha)(G_n+Y_{n+1})}}{\varphi_Y^{n+1}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle|Y_1, Y_2, \dots, Y_n\right)$$

$$= \frac{e^{\gamma\frac{\beta}{\alpha}(1-\alpha)G_n}}{\varphi_Y^{n}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)} E\left(\frac{e^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y_{n+1}}}{\varphi_Y\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle|Y_1, Y_2, \dots, Y_n\right)$$

$$= \frac{\left(e^{\gamma\frac{\beta}{\alpha}G_n}\right)^{(1-\alpha)}}{\varphi_Y^{n}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)} \cdot \frac{E\, e^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y_{n+1}}}{\varphi_Y\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}$$

The second multiplier is 1 according to the definition of m.g.f. The denominator of the first multiplier is the same as the denominator of the G'_n and we have that $(1-\alpha) \in (0,1)$, so, the nominator is less than the nominator of G'_n, so we get:

$$E(G'_{n+1}|Y_1, Y_2, \dots, Y_n) \leq G'_n$$

which is the definition of the supermartingale. ∎

__Lemma 2.5:__ For any $\gamma > 0$, such that $\varphi_Y(\gamma) < \infty$ and for any $\alpha, \beta \in (0,1)$ satisfying the NPC, the sequence

$$S'_n = \frac{e^{\gamma S_n}}{\varphi_Y^n\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^n\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)} \tag{2.4.8}$$

is a supermartingale.

Proof: Note that

$$S_{n+1} = -C_{n+1} - \left(\frac{\beta}{\alpha} - 1\right) M_{n+1} + \frac{\beta}{\alpha} G_{n+1}$$

$$= -(C_n + (1-\alpha)^{n+1}) - \left(\frac{\beta}{\alpha} - 1\right)(M_n + Y_{n+1}) + \frac{\beta}{\alpha}(1-\alpha)(G_n + Y_{n+1})$$

Calculate

$$E(S'_{n+1}|Y_1, Y_2, \dots, Y_n) = E\left(\frac{e^{\gamma S_{n+1}}}{\varphi_Y^{n+1}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^{n+1}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle| Y_1, Y_2, \dots, Y_n\right)$$

$$= E\left(\frac{e^{\gamma\left(-(C_n+(1-\alpha)^{n+1})-\left(\frac{\beta}{\alpha}-1\right)(M_n+Y_{n+1})+\frac{\beta}{\alpha}(1-\alpha)(G_n+Y_{n+1})\right)}}{\varphi_Y^{n+1}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^{n+1}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle| Y_1, Y_2, \dots, Y_n\right)$$

$$= \frac{e^{\gamma\left(-(C_n+(1-\alpha)^{n+1})-\left(\frac{\beta}{\alpha}-1\right)M_n+\frac{\beta}{\alpha}(1-\alpha)G_n\right)}}{\varphi_Y^{n}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^{n}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)} E\left(\frac{e^{\gamma\left(-\left(\frac{\beta}{\alpha}-1\right)Y_{n+1}+\frac{\beta}{\alpha}(1-\alpha)Y_{n+1}\right)}}{\varphi_Y\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle| Y_1, Y_2, \dots, Y_n\right)$$

Here γ, $(1-\alpha)$ and $\left(\frac{\beta}{\alpha}-1\right)$ are positive according to their definition and NPC, so it can be proved that $Cov\left(e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y}, e^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}\right) \le 0$ (see for instance [112], [113]). For the second multiplier we write:

$$E\left(\frac{e^{\gamma\left(-\left(\frac{\beta}{\alpha}-1\right)Y_{n+1}+\frac{\beta}{\alpha}(1-\alpha)Y_{n+1}\right)}}{\varphi_Y\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle| Y_1, Y_2, \dots, Y_n\right) = \frac{E\left(e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y} \cdot e^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}\right)}{Ee^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y}Ee^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}}$$

$$= \frac{Cov\left(e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y}, e^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}\right) + Ee^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y}Ee^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}}{Ee^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y}Ee^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}}$$

$$= 1 + \frac{Cov\left(e^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y}, e^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}\right)}{Ee^{-\gamma\left(\frac{\beta}{\alpha}-1\right)Y}Ee^{\gamma\frac{\beta}{\alpha}(1-\alpha)Y}} \le 1$$

Using Lemmas 2.3 and 2.4, for the first multiplier we have:

$$\frac{e^{\gamma\left(-\left(C_n+(1-\alpha)^{n+1}\right)-\left(\frac{\beta}{\alpha}-1\right)M_n+\frac{\beta}{\alpha}(1-\alpha)G_n\right)}}{\varphi_Y^n\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^n\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)} = \frac{e^{\gamma S_n}}{e^{\gamma((1-\alpha)^{n+1}+\alpha G_n)}\varphi_Y^n\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^n\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}$$

$$\leq \frac{e^{\gamma S_n}}{\varphi_Y^n\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^n\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)} = S'_n$$

The inequality holds as $\gamma((1-\alpha)^{n+1}+\alpha G_n) > 0$.

So, we get

$$E(S'_{n+1}|Y_1, Y_2, \dots, Y_n) \leq S'_n \qquad \blacksquare$$

Now we can find an upper bound for the probability of ruin in the model (2.4.2).

<u>Theorem 2.4:</u> *If for some $\gamma > 0$, the process S'_n given by (2.4.8) is a supermartingale, where $S_n \to -\infty$ as $n \to \infty$, then*

$$\psi(u) \leq \frac{e^{-\gamma u}}{E\left(\left.\frac{e^{-\gamma U_{\tau(u)}}}{\varphi_Y^{\tau(u)}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^{\tau(u)}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\right|\tau(u) < \infty\right)} \tag{2.4.9}$$

<u>Proof:</u> Doob's *Theorem 2.3* on optional stopping time for the supermartingale S'_n at time $\tau(u)\wedge T = \min(\tau(u), T)$ satisfies the inequality

$$ES'_0 \geq ES'_{\tau(u)\wedge T}$$

We cannot use the stopping time $\tau(u)$ directly because $P(\tau(u) = \infty) > 0$ and also because the conditions of the optional stopping theorem present a problem; however, using $\tau(u)\wedge T$ invokes no problems because $\tau(u)\wedge T$ is bounded by T.

Using the condition $S_0 = 0$, we get:

$$1 = ES'_0 \geq ES'_{\tau(u)\wedge T} = E\left(S'_{\tau(u)}; \tau(u) \leq T\right) + E\left(S'_T; \tau(u) > T\right) \tag{2.4.10}$$

The second term in (2.4.10) converges to 0, as $T \to \infty$ due to the condition $S_n \to -\infty$ as $n \to \infty$.

Recall that $U_n = u - S_n$, and $\psi(u) = P(\tau(u) < \infty)$. For (2.4.10) write

$$1 \geq E\left(S'_{\tau(u)}; \tau(u) < \infty\right) = E\left(\frac{e^{\gamma S_{\tau(u)}}}{\varphi_Y^{\tau(u)}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^{\tau(u)}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}; \tau(u) < \infty\right)$$

$$= e^{\gamma u} E\left(\frac{e^{-\gamma U_{\tau(u)}}}{\varphi_Y^{\tau(u)}\left(-\gamma\left(\frac{\beta}{\alpha}-1\right)\right)\varphi_Y^{\tau(u)}\left(\gamma\frac{\beta}{\alpha}(1-\alpha)\right)}\middle| \tau(u) < T\right) P(\tau(u) < \infty)$$

So, we have the statement of the theorem. ∎

2.5. Bonus Hunger Behavior

One of the consequences of the bonus–malus systems is the willingness of the insured to overtake the small claims on its own debit and not claim them, to keep the reduced premium payment. This phenomenon is called the "hunger for bonus" and is discussed in the Chapter 1 of this dissertation. Some insurance companies actually enable policyholders to buy the bonus by paying extra premium. Thus the bonus–malus system becomes also a strong marketing instrument.

While analyzing the effect of "hunger for bonus" it is necessary to answer the following question: "When the policyholder will not report the claim"? The answer is: "He will not report the claim if its amount is less than discounted value of all future premium reductions" [114]. The premium reductions are clearly given by difference between premiums paid in the different bonus classes. In general it can be expressed as follows.

Let's consider a policyholder, who is in the portfolio for $n-1$ years and has just had an accident. We denote P'_{n+k} for a premium after k years in the case that the policyholder will not claim this accident and will not have any accident during next k years. Using formula (2.2.1) for $\beta_j = 0, j = n, n+1, \dots, n+k$, we get:

$$P_n^{'} = (1-\alpha_n)P_{n-1}$$

$$P_{n+1}^{'} = (1-\alpha_{n+1})P_n^{'} = (1-\alpha_{n+1})(1-\alpha_n)P_{n-1}$$

...

$$P_{n+k}^{'} = (1-\alpha_{n+k})P_{n+k-1}^{'} = (1-\alpha_{n+k})\cdot \ldots \cdot (1-\alpha_n)P_{n-1} = P_{n-1}\prod_{j=0}^{k}(1-\alpha_{n+j})$$

Let $P_{n+k}^{''}$ be a premium after k years in the case that the policyholder will claim this accident and will not have any accident during next k years. Applying (2.2.1) for this premium, where $\beta_n \neq 0$ and $\beta_j = 0, j = n+1, \ldots, n+k$, we have:

$$P_n^{''} = (1-\alpha_n)P_{n-1} + \beta_n Y_n$$

$$P_{n+1}^{''} = (1-\alpha_{n+1})P_n^{''} = (1-\alpha_{n+1})(1-\alpha_n)P_{n-1} + (1-\alpha_{n+1})\beta_n Y_n$$

...

$$P_{n+k}^{''} = (1-\alpha_{n+k})P_{n+k-1}^{''} = (1-\alpha_{n+k})\cdot \ldots \cdot (1-\alpha_n)P_{n-1} + (1-\alpha_{n+k})\cdot \ldots \cdot (1-\alpha_{n+1})\beta_n Y_n$$
$$= P_{n-1}\prod_{j=0}^{k}(1-\alpha_{n+j}) + \beta_n Y_n \prod_{j=1}^{k}(1-\alpha_{n+j})$$

The policyholder will claim for the mentioned accident, if the claim amount is greater than u_n, where

$$u_n = \beta_n Y_n + \sum_{m=1}^{k}(P_{n+m}^{''} - P_{n+m}^{'})v^m = \beta_n Y_n\left(1 + \sum_{m=1}^{k} v^m \prod_{j=1}^{m}(1-\alpha_{n+j})\right) \qquad (2.5.1)$$

here v is the financial discount factor.

So, the probability of claiming an accident is:

$$p_n(u_n) = P(Y_n > u_n) = 1 - P(Y_n \le u_n) = 1 - F_Y(u_n) \qquad (2.5.2)$$

where $F_Y(\cdot)$ is the claim amount distribution function.

2.6. Some Examples

Example 1

Suppose that the yearly aggregate claims of an insurance company are distributed exponentially with λ rate: $Y \sim Exp(\lambda)$: Let's find the coefficients α_n and β_n, as well as the probabilities of claiming and ruin.

The characteristics of an exponential distribution needed here are the distribution function $F_Y(x) = 1 - e^{-\lambda x}$, the expectation $E(Y) = \frac{1}{\lambda}$, the inverse distribution function $F^{-1}(\varepsilon, \lambda) = -\frac{ln(1-\varepsilon)}{\lambda}$ and m.g.f. $\varphi_Y(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}, t < \lambda$. Substituting these expressions in (2.3.2), (2.3.3),(2.4.9) and (2.5.2) we accordingly get:

$$\beta_n = \frac{\lambda(P_{n-1} - Y_c)}{1 + \ln(1-\varepsilon)}$$

$$\alpha_n = \frac{P_{n-1} - Y_c}{P_{n-1}(1 + \ln(1-\varepsilon))}$$

$$\psi(u) \leq \frac{e^{-\gamma u}}{E\left(\left.\frac{\left(\lambda + \gamma\left(\frac{\beta}{\alpha} - 1\right)\right)^{\tau(u)} \left(\lambda - \gamma\frac{\beta}{\alpha}(1-\alpha)\right)^{\tau(u)} e^{-\gamma U_{\tau(u)}}}{\lambda^{2\tau(u)}}\right| \tau(u) < \infty\right)}$$

and

$$p_n(u_n) = e^{-\lambda u_n}$$

Example 2

Suppose that the yearly aggregate claims of an insurance company have a Pareto distribution with parameters μ and λ ($Y \sim Pareto(\mu, \lambda)$): For finding α_n, β_n and $p_n(u_n)$ we need the distribution function $F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\mu}$, the inverse of that distribution function which have the following view: $F^{-1}(\varepsilon, \mu, \lambda) = \lambda\left((1-\varepsilon)^{-\frac{1}{\mu}} - 1\right)$ and the expectation, which is expressed as $E(Y) = \frac{\lambda}{\mu-1}$. Putting the mentioned above expressions in (2.3.2), (2.3.3) and (2.5.2) we get:

$$\beta_n = \frac{Y_c - P_{n-1}}{\lambda\left((1-\varepsilon)^{-\frac{1}{\mu}} - \frac{\mu}{\mu-1}\right)}$$

$$\alpha_n = \frac{Y_c - P_{n-1} + \frac{\lambda}{\mu-1}}{P_{n-1}\lambda\left((\mu-1)(1-\varepsilon)^{-\frac{1}{\mu}} - \mu\right)}$$

$$p_n(u_n) = \left(\frac{\lambda}{\lambda + u_n}\right)^{\mu}$$

The probability of ruin is not presented here as the m.g.f. $\varphi_Y(t)$ of the Pareto distribution is not known for $t > 0$ cases.

CHAPTER 3

BMS ANALYSIS USING HIDDEN MARKOV MODELS

Introduction

In the world surrounding us we are able to see only one's final decision, but what stands up behind that decision, what could compel him to make the decision is not known. One of the theories exploring that "hidden" processes is called Hidden Markov Models (HMM). HMM is used in cryptanalysis, gesture recognition, partial discharge, bioinformatics especially in DNA recognition, time series' analysis and in other spheres. In preceding years HMM has a wide application in the different branches of artificial intelligence as well. Those branches are speech and handwriting recognition, computer translations and so on. HMM with its different types are also used in different models of financial and insurance analysis (see [115], [116], [117], [118]), particularly in BMS models like in [13], [119], [120].

In BMS models as a "hidden" process considered the real number of insurance accidents. The insurer knows only about those accidents which were claimed. The "hidden" accidents are not included in the statistics of claims. So, the insurer has an issue to evaluate the real number of accidents otherwise estimations received from statistics will be biased from real values.

In this chapter of the thesis an example of construction of a new BMS is considered, where the movement between classes is modeled as a discrete time Markov chain, which is dependent on the claim number and aggregate claim amount processes which are assumed to be a HMM.

3.1. An Extended BMS Model Description

Consider a set of L policyholders. Each policyholder belongs to one of a finite number C of classes (tariff groups) sorted by order; class 1 being the one with lowest premiums etc. That is, each premium depends on the class to which a policyholder belongs. Each year the class of a policyholder is determined on the basis of the class of the previous year, on the number of claims and on the aggregate claim loss reported during that year. If no claim has been reported, then the policyholder gets a bonus expressed in the lowering to a class with a lower premium or stay at the lowest premium class. Otherwise the policyholder may stay in the same class or gets maluses (penalized) by being shifted to a higher class with possibly higher premium. New policyholders are assigned to a certain class [121].

Let x_n^i, be the number of policyholders in a class i at time n, where $i = 1,2,\dots,C$. Then $X_n = (x_n^1, \dots, x_n^C)$ will be the distribution of policyholders among classes at time n. It is obvious that the state space $S_X = \{(x_n^1, \dots, x_n^C)\}$ of X_n process is finite.

We split the positive half of the real line into a convenient set of disjoint intervals $I_1, I_2, \dots, I_K$ and discuss the aggregate claim of each policyholder on those intervals. We will denote by H_n^i the number of policyholders whose aggregate claim in the n-th year falls in interval I_i , $i = 1,2 \dots, K$. So, $H_n = (H_n^1, \dots, H_n^K)$ row vector will show the distribution of policyholders among the reported aggregate claim intervals.

Consider the number of reported claims. Its state space will be the space of natural numbers {0, 1, 2 …}. Without any distortion we can suppose that it is limited by some number N. We will denote by N_n^i the number of policyholders who have reported i, $i = 0,1,\dots,N$ claims during the n-th year. Then $N_n = (N_n^0, \dots, N_n^N)$ row vector will be the distribution of policyholders among the reported claim numbers.

Assumptions underlying the model

- The processes X, H and N are Markov chains which, for technical reasons (that will become apparent later) and without loss of generality[5], accordingly live on the standard basis $\{e_1, \dots, e_{|S_X|}\}$, $\{f_1, \dots, f_{|S_H|}\}$ and $\{h_1, \dots, h_{|S_N|}\}$ in $\mathbb{R}^{|S_X|}$, $\mathbb{R}^{|S_H|}$ and

[5] Consider the process $\mathcal{X}_k$ with $S_{\mathcal{X}} = \{s_1, \dots, s_K\}$ finite state space. Let consider the function $\phi_k(s_i) = \begin{cases} 0, & \text{if } k \neq i \\ 1, & \text{if } k = i \end{cases}$. We will construct the process $X_k = (\phi_1(\mathcal{X}_k), \dots, \phi_K(\mathcal{X}_k))$. It is easy to note that for any k only one component of X_k is 1 and others are 0. So, instead of process $\mathcal{X}_k$ we will consider the process X_k which is generated by it and which state space is $S_X = \{e_1, \dots, e_K\}$ the space of unit column vectors.

respectively $\mathbb{R}^{|S_N|}$, where the i-th component of each vector e_i, f_i and h_i is 1 and others are 0 (see [122], pg. 5). $|S_X|$, $|S_H|$ and $|S_N|$ are the sizes of S_X, S_H and S_N sets accordingly.

- It is assumed that the movement between classes is based on the current class of the policyholder, on the number of claims and on the aggregate claim reported in the year. So the movement of process X_n between its states depends on levels of X_{n-1}, $H_{n-1}$and N_{n-1} and the transition matrix is not time-dependent.

- The next assumption refers to aggregate claim process. It is assumed that the aggregate claims are not independent, so the aggregate claim of a policyholder in any year depends on the aggregate claim and reported claims number of the previous year. We can conclude that the movement between states of process H_n is based on the values of H_{n-1} and N_{n-1} as well. The transition matrix is time-dependent and is not known in advance.

- The yearly reported claim numbers for each policyholder are also suggested dependent. It is assumed that reported claims number of a policyholder depends on the claims number reported last year and on the policyholder's current class of BMS. In other words, the movement from one state for process N_n to another is founded on N_{n-1} and on X_{n-1}. For this process also the transition matrix is assumed a stochastic one.

The time-dependence of transition matrices can be explained by change of policyholders' behavior year by year. They can make conclusions based on their insurance history and be more professional. As an example in motor insurance, the driver can be more careful and make fewer claims if he has many claims in the previous year. On the other hand, he can prefer to cover some small claims himself if he is on the higher bonus class.

Let $\mathcal{J}_n = \sigma\{X_k, H_{k-1}, N_{k-1}, k \le n\}$ be the complete filtration generated by processes X, H and N up to the n-th year.

<u>Definition 3.1:</u> (Markov property (see [109] pg.788, and [123] pg.57)): *A discrete-time stochastic process $\{Y_n\}$, with finite-state space $S = \{s_1, s_2, \dots, s_n\}$, defined on probability space $(\Omega, \mathcal{F}, P)$ is a Markov chain if*

$$P\left(Y_{n+1}=s_{i_{n+1}}\middle|Y_0=s_{i_0},\dots,Y_n=s_{i_n}\right)=P\left(Y_{n+1}=s_{i_{n+1}}\middle|Y_n=s_{i_n}\right),$$

for all $n\geq 0$ *and all states* $s_{i_0},s_{i_1},\dots,s_{i_{n+1}}\in S$.

The use of Markov property for process X gives:

$$P\left(X_n=e_j\middle|\mathcal{J}_{n-1},H_{n-1},N_{n-1}\right)=P\left(X_n=e_j\middle|X_{n-1},H_{n-1},N_{n-1}\right)$$

Conditional on the event $(X_{n-1}=e_i,H_{n-1}=f_k,N_{n-1}=h_m)$, the above mentioned expression may be rewritten as:

$$P\left(X_n=e_j\middle|X_{n-1}=e_i,H_{n-1}=f_k,N_{n-1}=h_m\right)\triangleq p_{j,ikm}$$

Write $B=\left(p_{j,ikm}\right)_{j,i=1\ \ k=1\ \ m=1}^{|S_X|\ \ |S_H|\ \ |S_N|}$. Then B is a $|S_X|\times|S_X||S_H||S_N|$ stochastic matrix of tensor mapping $\mathbb{R}^{|S_X||S_H||S_N|}$ into $\mathbb{R}^{|S_X|}$ and has the form

$$B=\begin{pmatrix} p_{1,111} & \cdots & p_{1,|S_X||S_H||S_N|)} \\ \vdots & \ddots & \vdots \\ p_{|S_X|,111} & \cdots & p_{|S_X|,|S_X||S_H||S_N|} \end{pmatrix} \tag{3.1.1}$$

where the sum of elements of each column is 1:

$$\sum_{j=1}^{|S_X|}p_{j,ikm}=1 \tag{3.1.1'}$$

<u>Definition 3.2:</u> *(Doob decomposition for discrete time, see* [123], *pg.57 Remark): A processes which is the sum of a predictable process and a martingale is called a semimartingale.*

<u>Definition 3.3:</u> *(see* [109], *pg.655): The stochastic process* $\xi=(\xi_n,\mathcal{J}_n)_{n\geq0}$ *is called a martingale difference if* $E|\xi_n|<\infty$ *for all* $n\geq0$ *and* $E(\xi_{n+1}|\mathcal{J}_n)=0$ (P-a.s.).

<u>Definition 3.4:</u> *(see* [123], *pg.58): Given two (column) vectors X and Y the tensor or Kroneker product* $X\otimes Y$ *is the column vector obtained by stacking the rows of the matrix* XY *, where* *is the transpose, which entries obtained by multiplying the* i*-th entry of* X *by the* j*-th entry of* Y.

Lemma 3.1: *The process X has the following semimartingale representation (or Doob decomposition):*

$$X_n = BX_{n-1} \otimes H_{n-1} \otimes N_{n-1} + V_n,$$

where the $|S_X| \times 1$ column vector V_n is an $\mathcal{I}_n$ martingale difference.

Proof: Using the fact that Markov chain holds to expression $E(X_n|\mathcal{I}_{n-1}, Z_{n-1}, N_{n-1}) = BX_{n-1} \otimes Z_{n-1} \otimes N_{n-1}$ and the following famous property of conditional expectation

$$E(X|\mathcal{A}_1) = E(E(X|\mathcal{A}_2)|\mathcal{A}_1) = E(E(X|\mathcal{A}_1)|\mathcal{A}_2)$$

where $\mathcal{A}_1 \subset \mathcal{A}_2$, it is easy to note that

$$\begin{aligned} E(V_n|\mathcal{I}_{n-1}) &= E(E(X_n - BX_{n-1} \otimes Z_{n-1} \otimes N_{n-1}|\mathcal{I}_{n-1}, Z_{n-1}, N_{n-1})|\mathcal{I}_{n-1}) \\ &= E(E(X_n|\mathcal{I}_{n-1}, Z_{n-1}, N_{n-1}) - E(BX_{n-1} \otimes Z_{n-1} \otimes N_{n-1}|\mathcal{I}_{n-1}, Z_{n-1}, N_{n-1})|\mathcal{I}_{n-1}) \\ &= E(BX_{n-1} \otimes Z_{n-1} \otimes N_{n-1} - BX_{n-1} \otimes Z_{n-1} \otimes N_{n-1}|\mathcal{I}_{n-1}) = 0 \end{aligned}$$

i.e. V is a sequence of martingale differences. ■

For the transition matrix of process H we will denote

$$q_{s,rl}(n) \triangleq P(H_n = f_s|H_{n-1} = f_r, N_{n-1} = h_l),$$

where index n indicates the time-dependence of H_n transition matrix.

The same analysis as for Markov chain X shows that the Markov chain H has representation

$$H_n = Q_n H_{n-1} \otimes N_{n-1} + W_n$$

where $Q_n = \left(q_{s,rl}(n)\right)_{s=1}^{|S_Z|}{}_{r=1}^{|S_H|}{}_{l=1}^{|S_N|}$ transition matrix has the following form:

$$Q_n = \begin{pmatrix} q_{1,11}(n) & \cdots & q_{1,|S_H||S_N|}(n) \\ \vdots & \ddots & \vdots \\ q_{|S_H|,11}(n) & \cdots & q_{|S_Z|,|S_H||S_N|}(n) \end{pmatrix} \quad (3.1.2)$$

Note that:

$$\sum_{s=1}^{|S_H|} q_{s,rl}(n) = 1 \qquad (3.1.2')$$

and W_n is a $|S_H| \times 1$ column vector, which is a martingale difference.

As for others, for the process N let denote

$$a_{u,vw}(n) \triangleq P(N_n = h_u | N_{n-1} = h_v, X_{n-1} = e_w)$$

Similarly, for the process N Doob decomposition will be

$$N_n = A_n N_{n-1} \otimes X_{n-1} + L_n$$

where $A_n = \left(a_{u,vw}(n)\right)_{u=1}^{|S_N|}{}_{v=1}^{|S_N|}{}_{w=1}^{|S_X|}$ matrix has the form

$$A_n = \begin{pmatrix} a_{1,11}(n) & \cdots & a_{1,|S_N||S_X|}(n) \\ \vdots & \ddots & \vdots \\ a_{|S_N|,11}(n) & \cdots & a_{|S_N|,|S_N||S_X|}(n) \end{pmatrix} \qquad (3.1.3)$$

and

$$\sum_{u=1}^{|S_N|} a_{u,v,w}(n) = 1 \qquad (3.1.3')$$

L_n is a $|S_N| \times 1$ column vector and is a martingale difference.

To eliminate the time-dependence of transition matrices Q_n and A_n, we will develop matrix-processes, for which transition matrices will not be time-dependent.

Consider the simplexes U and V with non-negative column vectors $U = \{(u_1, \dots, u_{|S_Z|})^T\}$ and respectively $V = \{(v_1, \dots, v_{|S_N|})^T\}$, where $u_i \geq 0$ $(v_j \geq 0)$ and $\sum_{i=1}^{|S_H|} u_i = 1$ $(\sum_{j=1}^{|S_N|} v_j = 1)$.

By definition, each column of Q_n and A_n is a point in U and respectively in V. Let's partition the sets U and V by the following way:

$$U = U_1 \cup U_2 \cup \dots \cup U_Q, \text{where } U_i \cap U_j = \emptyset, i \neq j$$

and

$$V = V_1 \cup V_2 \cup \dots \cup V_A, \text{where } V_i \cap V_j = \emptyset, i \neq j$$

Let

$$\Theta = (U_1, U_2, \dots, U_Q)^{|S_H||S_N|}$$

and

$$\Delta = (V_1, V_2, \dots, V_A)^{|S_X||S_N|}$$

That is Θ (accordingly Δ) is the Cartesian product of $|S_H||S_N|$ ($|S_X||S_N|$) copies of the ordered set $(U_1, U_2, \dots, U_Q)$ (accordingly $(V_1, V_2, \dots, V_A)$). In other words, it represents the set of $|S_H| \times |S_H||S_N|$ ($|S_N| \times |S_X||S_N|$) sized matrices generated by partition $U_1, U_2, \dots, U_Q$ $(V_1, V_2, \dots, V_A)$ of the set $U(V)$.

We define the following Markov chain $\breve{Q}_n$ (respectively$\breve{A}_n$) on the set Θ (Δ) as follows: if the first column of Q_n (A_n) is a point in U_{i_1} (accordingly in V_{i_1}), the second column in U_{i_2} (accordingly in V_{i_2}), …, the last column in $U_{i_{|S_H||S_N|}}$ (correspondingly in $V_{i_{|S_X||S_N|}}$), then $\breve{Q}_n = (U_{i_1}, \dots, U_{i_{|S_H||S_N|}})$ ($\breve{A}_n = (V_{i_1}, \dots, V_{i_{|S_X||S_N|}})$), i.e. $\breve{Q}_n$ ($\breve{A}_n$) keeps track only of the location of the columns of Q_n (A_n) in $(U_1, U_2, \dots, U_Q)$ (accordingly in $(V_1, V_2, \dots, V_A)$). So, with the help of $Q_n(A_n)$ the matrix $\breve{Q}_n$ ($\breve{A}_n$) can be defined identically.

Write $\Theta = \{\theta_1, \dots, \theta_{Q|S_H||S_N|}\}$ ($\Delta = \{\delta_1, \dots, \delta_{A|S_X||S_N|}\}$). We shall identify the ordered set Θ (correspondingly Δ) with the standard basis $\{b_1, \dots, b_{Q|S_H||S_N|}\}$ ($\{m_1, \dots, m_{Q|S_H||S_N|}\}$) of $\mathbb{R}^{Q|S_H||S_N|}$ ($\mathbb{R}^{A|S_X||S_N|}$). So, for matrices Q_n and $A_n$we develop processes $\breve{Q}_n$ and accordingly $\breve{A}_n$, for which transition matrices are not time-dependent and have the following form:

$$\breve{Q}_n = D\breve{Q}_{n-1} + R_n$$

where $D = \left(d_{qp}\right)_{q,p=1}^{Q|S_H||S_N|}$ with

$$d_{qp} \triangleq P\left(\breve{Q}_n = b_q \middle| \breve{Q}_{n-1} = b_p\right) \tag{3.1.4}$$

and

$$\sum_{q=1}^{Q|S_H||S_N|} d_{qp} = 1 \tag{3.1.4'}$$

$\{R_n\}$ is a sequence of martingale differences on σ-field $\sigma\{\breve{Q}_0, \breve{Q}_1, \dots, \breve{Q}_n\}$.

Similarly,

$$\breve{A}_n = K\breve{A}_{n-1} + T_n$$

where $K = \left(k_{qp}\right)_{q,p=1}^{A|S_X||S_N|}$ with

$$k_{qp} \triangleq P\left(\breve{A}_n = m_q \middle| \breve{A}_{n-1} = m_p\right) \tag{3.1.5}$$

and

$$\sum_{\varpi=1}^{A|S_X||S_N|} k_{\varpi\vartheta} = 1 \tag{3.1.5'}$$

$\{T_n\}$ is a sequence of martingale differences on σ-field $\sigma\{\breve{A}_0, \breve{A}_1, \dots, \breve{A}_n\}$.

So, we have the following BMS model:

$$\begin{cases} X_n = BX_{n-1} \otimes H_{n-1} \otimes N_{n-1} + V_n \\ H_n = Q_n H_{n-1} \otimes N_{n-1} + W_n \\ N_n = A_n N_{n-1} \otimes X_{n-1} + L_n \\ \breve{Q}_n = D\breve{Q}_{n-1} + R_n \\ \breve{A}_n = K\breve{A}_{n-1} + T_n \end{cases} \tag{3.1.6}$$

It must be noted that the constructed model is a revised and an extended one described in [13]. It has the following peculiarities:

- The model, developed here is “policyholder-oriented”, that is the considered group of policyholders is divided into subgroups from 3 different points of view:
  - ✓ partition by levels of BMS
  - ✓ partition by groups of aggregate claim amounts
  - ✓ partition by number of reported claims

  In [13], the partition is applied to two different events: first of all the group of policyholders is divided on subgroups by BMS levels and the other partition applied to the set of reported claims, which are sub grouped by claim amount.

- The process Z_n is the distribution of claim numbers by claim amount intervals in [13], so $\sum_{j=1}^{K} Z_n^j$ represents the total number of claims and it means that a policyholder, who makes more than one claim during the entire year, can appear in different groups of claim amounts simultaneously. In the model presented in this paper, H_n is the distribution of policyholders among aggregate claim amount groups, so $\sum_{j=1}^{K} H_n^j$ represents the number of policyholders and it means that policyholder's location within the aggregate claim amount intervals can be identified uniquely.
- In comparison with [13], where the transition between BMS levels depends on the reported claim numbers, in the model, presented in this paper, the above-mentioned transition depends on the aggregate claim amount, reported by policyholder as well.

3.2. Hidden Markov Models. Change of Measure

Hidden Markov model is a general tool for representing probability distributions over sequences of observations. The hidden Markov model gets its name from two defining properties. Let us denote the observation at time n by the random variable Y_n. First, it assumes that the observation at time n was generated by some process $\tilde{F}_n$, whose state is hidden from the observer. Second, it assumes that the state of this hidden process satisfies the Markov property: that is, given the value of $\breve{F}_{n-1}$, the current state $\breve{F}_n$ is independent of all the states prior to $n-1$. The outputs Y_n have the same property. More detail about hidden Markov models see for instance [122] and [124].

The processes $\breve{Q}_n$ and $\breve{A}_n$, mentioned above, are hidden Markov models. To estimate their states recursively, we define a measure $\tilde{P}$, on the measurable space $(\Omega, \mathcal{F})$, under which processes X, H and N are sequences of statistically independent and identically distributed random variables. The existence of measure $\tilde{P}$ is provided by the following theorem of Radon-Nicodym.

__Theorem 3.1:__ (Radon-Nicodym): If P and $\tilde{P}$ are two probability measures on $(\Omega, \mathcal{G})$ such that for each $C \in \mathcal{G}$, $\tilde{P}(C) = 0$ implies $P(C) = 0$ $(P \ll \tilde{P})$, then there exists a nonnegative 53unction Λ, such that $P(L) = \int_L \Lambda d\tilde{P}$ for all $L \in \mathcal{G}$.

The function $\Lambda = \frac{dP}{d\tilde{P}}\Big|_{\mathcal{G}}$ is called the Radon Nicodym derivative.

The probability measure P is referred to as the "real world" measure, under which we have the relations (3.1.6).

Suppose that under the measure $\tilde{P}$ the processes X, H and N are i.i.d sequences with the following distributions:

$$\tilde{P}\left(X_n = e_j \middle| X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \triangleq \tilde{p}_{j,ikm}$$

$$\tilde{P}(H_n = f_s | H_{n-1} = f_r, N_{n-1} = h_l) \triangleq \tilde{q}_{s,rl}(n)$$

$$\tilde{P}(N_n = h_u | N_{n-1} = h_v, X_{n-1} = e_w) \triangleq \tilde{a}_{u,vw}(n)$$

where the corresponding matrices are

$$\tilde{B} = \left(\tilde{p}_{j,ikm}\right)_{j,i=1}^{|S_X|} {}_{k=1}^{|S_H|} {}_{m=1}^{|S_N|},$$

$$\tilde{Q}_n = \left(\tilde{q}_{s,rl}(n)\right)_{s=1}^{|S_H|} {}_{r=1}^{|S_H|} {}_{l=1}^{|S_N|}$$

and

$$\tilde{A}_n = \left(\tilde{a}_{u,vw}(n)\right)_{u=1}^{|S_N|} {}_{v=1}^{|S_N|} {}_{w=1}^{|S_X|}.$$

Under the measure $\tilde{P}$ the dynamics of $\breve{Q}_n$ and $\breve{A}_n$ remains unchanged.

Define $$\Lambda_n = \prod_{t=1}^{n} \lambda_t \tag{3.2.1}$$

where

$$\lambda_t = \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_t, e_j\rangle\langle X_{t-1}, e_i\rangle\langle H_{t-1}, f_k\rangle\langle N_{t-1}, h_m\rangle} \times \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} \left(\frac{q_{s,rl}(n)}{\tilde{q}_{s,rl}(n)}\right)^{\langle H_t, f_s\rangle\langle H_{t-1}, f_r\rangle\langle N_{t-1}, h_l\rangle} \times \prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} \left(\frac{a_{u,vw}(n)}{\tilde{a}_{u,vw}(n)}\right)^{\langle N_t, h_u\rangle\langle N_{t-1}, h_v\rangle\langle X_{t-1}, e_w\rangle} \tag{3.2.2}$$

where $\langle\ ,\ \rangle$ denotes the usual scalar product and $\Lambda_0 = \lambda_0 = 1$.

Let $\mathcal{G}_n$ be the complete filtration generated by $\{X_k, H_{k-1}, N_{k-1}, \tilde{Q}_k, \tilde{A}_k, k \le n\}$.

<u>Lemma 3.2:</u> *The sequence of random variables* $\{\Lambda_n\}_{n\ge 0}$ *is a* $\{\mathcal{G}_n, \tilde{P}\}$ *martingale with expectation* $\tilde{E}(\Lambda_n) = 1$.

<u>Proof:</u> We have to show that

$$\tilde{E}(\Lambda_n | \mathcal{G}_{n-1}) = \Lambda_{n-1}$$

From (3.2.1) we have $\Lambda_n = \Lambda_{n-1}\lambda_n$. As Λ_{n-1} is $\mathcal{G}_{n-1}$ measurable; it is sufficient to show that

$$\tilde{E}(\lambda_n | \mathcal{G}_{n-1}) = 1.$$

Using (3.2.2) representation of λ_n and properties of conditional expectation, let's calculate

$$\tilde{E}(\lambda_n|\mathcal{G}_{n-1}) = \tilde{E}\big(\tilde{E}(\lambda_n|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)|\mathcal{G}_{n-1}\big)$$
$$= \tilde{E}\left(\tilde{E}\left(\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle}\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)^{\langle H_n, f_s\rangle}\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u\rangle}|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1}\right.\right.$$
$$\left.\left.= f_k, N_{n-1} = h_m\right)|\mathcal{G}_{n-1}\right)$$

Under the measure $\tilde{P}$ all processes are independent, so the expectation of the product can be written as the product of expectations, where each item equals to 1:

$$\tilde{E}\left(\tilde{E}\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle}\middle|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\middle|\mathcal{G}_{n-1}\right)$$
$$= \tilde{E}\left(\sum_{j=1}^{|S_X|}\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\tilde{P}(X_n = e_j|X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)\middle|\mathcal{G}_{n-1}\right)$$
$$= \sum_{j=1}^{|S_X|} p_{j,ikm} = 1$$

as well as

$$\tilde{E}\left(\tilde{E}\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)^{\langle H_n, f_s\rangle}\middle|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\middle|\mathcal{G}_{n-1}\right)$$
$$= \tilde{E}\left(\sum_{s=1}^{|S_H|}\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\tilde{P}(H_n = f_s|H_{n-1} = f_k, N_{n-1} = h_m)\middle|\mathcal{G}_{n-1}\right) = \sum_{s=1}^{|S_H|} q_{s,km}(n)$$
$$= 1$$

and finally

$$\tilde{E}\left(\tilde{E}\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u\rangle}\middle|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\middle|\mathcal{G}_{n-1}\right)$$
$$= \tilde{E}\left(\sum_{u=1}^{|S_N|}\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)\tilde{P}(N_n = h_u|X_{n-1} = e_i, N_{n-1} = h_m)\middle|\mathcal{G}_{n-1}\right) = \sum_{s=1}^{|S_N|} a_{u,mi}(n)$$
$$= 1 \qquad \blacksquare$$

According to Radon-Nicodym theorem and Kolmogorov's extension theorem (see [122], Appendix A), with the help of measure $\tilde{P}$ we can define the "real world" measure P as follows:

$$\left.\frac{dP}{d\tilde{P}}\right|_{\mathcal{G}_n} \triangleq \Lambda_n \qquad (3.2.3)$$

Theorem 3.2: *Under probability measure* P, *as defined from* $\tilde{P}$ *via* (3.2.3) *the dynamics* (3.1.6) *hold.*

Proof: Let's present the generalized version of Bayes' theorem (see [123], pg. 132) for the processes X_n, H_n and N_n , which are taking the following forms accordingly with the help of *Lemma 3.2*:

$$\begin{aligned}
&P\left(X_n = e_j | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \\
&\quad = \frac{\tilde{E}\left(\langle X_n, e_j\rangle \Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right)}{\tilde{E}\left(\Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right)} = \\
&\quad = \frac{\Lambda_{n-1}\tilde{E}\left(\langle X_n, e_j\rangle \lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right)}{\Lambda_{n-1}} \\
&\quad = \tilde{E}\left(\langle X_n, e_j\rangle \lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \qquad (3.2.4')
\end{aligned}$$

$$\begin{aligned}
&P\left(H_n = f_s | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \\
&\quad = \frac{\tilde{E}\left(\langle H_n, f_s\rangle \Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right)}{\tilde{E}\left(\Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right)} \\
&\quad = \tilde{E}\left(\langle H_n, f_s\rangle \lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \qquad (3.2.4'')
\end{aligned}$$

$$\begin{aligned}
&P\left(N_n = h_u | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \\
&\quad = \frac{\tilde{E}\left(\langle N_n, h_u\rangle \Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right)}{\tilde{E}\left(\Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right)} \\
&\quad = \tilde{E}\left(\langle N_n, h_u\rangle \lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \qquad (3.2.4''')
\end{aligned}$$

Note that under the condition $[\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m]$ and $\langle X_n, e_j\rangle$, formula (3.2.2) takes the form

$$\lambda_n = \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)^{\langle H_n, f_s\rangle}\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u\rangle} \qquad (3.2.2')$$

For the condition $[\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m]$ with $\langle H_n, f_s\rangle$, formula (3.2.2) takes the form

$$\lambda_n = \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle} \left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right) \left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u\rangle} \qquad (3.2.2'')$$

as well as under the condition $[\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m]$ and $\langle N_n, h_u\rangle$, formula (3.2.2) takes the form:

$$\lambda_n = \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle} \left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)^{\langle H_n, f_s\rangle} \left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right) \qquad (3.2.2''')$$

So, (3.2.4') with (3.2.2') gives:

$$\begin{aligned}
&\tilde{E}\left(\langle X_n, e_j\rangle \lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \\
&\quad = \tilde{E}\left(\langle X_n, e_j\rangle \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)^{\langle H_n, f_s\rangle}\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u\rangle} \middle| \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1}\right. \\
&\quad \left. = f_k, N_{n-1} = h_m\right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{s=1}^{|S_H|} \langle X_n, e_j\rangle \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u\rangle} \\
&\qquad \cdot \tilde{P}(H_n = f_s | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m) \\
&\qquad = \sum_{u=1}^{|S_N|}\sum_{s=1}^{|S_H|} \langle X_n, e_j\rangle \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right) \\
&\qquad \cdot \tilde{P}(H_n = f_s, N_n = h_u | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m) \\
&\qquad = \sum_{u=1}^{|S_N|}\sum_{s=1}^{|S_H|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right) \\
&\qquad \cdot \tilde{E}\left(\langle X_n, e_j\rangle\langle H_n, f_s\rangle\langle N_n, h_u\rangle | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \\
&\qquad = \sum_{u=1}^{|S_N|}\sum_{s=1}^{|S_Z|} p_{j,ikm} q_{s,km}(n) a_{u,mi}(n) = p_{j,ikm}
\end{aligned}$$

where the distributions of processes X_n, H_n, N_n under the measure $\tilde{P}$ and properties (3.1.2') and (3.1.3') have been used.

The (3.2.4 ′) with (3.2.2 ′) and use of properties (3.1.1′) and (3.1.3′) gives:

$$\tilde{E}(\langle H_n, f_s\rangle\lambda_n|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)$$
$$= \sum_{j=1}^{|S_X|}\sum_{u=1}^{|S_N|}\langle H_n, f_s\rangle\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)$$
$$\cdot \tilde{P}(X_n = e_j, N_n = h_u|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)$$

$$= \sum_{j=1}^{|S_X|}\sum_{u=1}^{|S_N|}\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)$$
$$\cdot \tilde{E}(\langle X_n, e_j\rangle\langle H_n, f_s\rangle\langle N_n, h_u\rangle|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)$$
$$= \sum_{j=1}^{|S_X|}\sum_{u=1}^{|S_N|} p_{j,ikm}q_{s,km}(n)a_{u,mi}(n) = q_{s,km}(n)$$

And (3.2.4 ′′) with (3.2.2 ′′), (3.1.1′) and (3.1.2′) gives:

$$\tilde{E}(\langle N_n, h_u\rangle\lambda_n|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)$$
$$= \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|}\langle N_n, h_u\rangle\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)$$
$$\cdot \tilde{P}(X_n = e_j, H_n = f_s|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)$$
$$= \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|}\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)$$
$$\cdot \tilde{E}(\langle X_n, e_j\rangle\langle H_n, f_s\rangle\langle N_n, h_u\rangle|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)$$
$$= \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|} p_{j,ikm}q_{s,km}(n)a_{u,mi}(n) = a_{u,mi}(n) \qquad \blacksquare$$

Theorem 3.3: *Joint distribution of processes H_n and N_n expressed via their marginal distributions is as follows:*

$$P(H_n = f_s, N_n = h_u|\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m) = q_{s,km}(n)a_{u,mi}(n).$$

It means that under the measure P processes H_n and N_n are statistically independent ones.

Proof: Using again the generalized Bayes' theorem, for the joint distribution, we can write:

$$
\begin{aligned}
P(H_n = f_s, N_n = h_u | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m) \\
&= \frac{\tilde{E}(\langle H_n, f_s\rangle\langle N_n, h_u\rangle\Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)}{\tilde{E}(\Lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)} \\
&= \tilde{E}(\langle H_n, f_s\rangle\langle N_n, h_u\rangle\lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m)
\end{aligned} \tag{3.2.5}
$$

In this case, under the conditions $[\mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m]$, $\langle H_n, f_s\rangle$ and $\langle N_n, h_u\rangle$ for (3.1.6) we have:

$$
\lambda_n = \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle} \left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right),
$$

so,(3.2.5) takes the form:

$$
\begin{aligned}
\tilde{E}(\langle H_n, f_s\rangle\langle N_n, h_u\rangle\lambda_n | \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m) \\
&= \sum_{j=1}^{|S_X|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)\left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)\left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right) \\
&\cdot \tilde{E}\left(\langle X_n, e_j\rangle\langle H_n, f_s\rangle\langle N_n, h_u\rangle \middle| \mathcal{G}_{n-1}, X_{n-1} = e_i, H_{n-1} = f_k, N_{n-1} = h_m\right) \\
&= \sum_{j=1}^{|S_X|} p_{j,ikm} q_{s,km}(n) a_{u,mi}(n) \\
&= q_{s,km}(n) a_{u,mi}(n) \qquad \blacksquare
\end{aligned}
$$

Define the measure valued process $g_n(\omega, \varpi)$, which is unnormalized conditional probability under measure $\tilde{P}$

$$
g_n(\omega, \varpi) = \tilde{E}\left(\Lambda_n \langle \breve{Q}_n, b_\omega\rangle\langle \breve{A}_n, m_\varpi\rangle | \mathcal{J}_n\right) \tag{3.2.6}
$$

Using the generalized Bayes' theorem, the normalized conditional probability takes the form:

$$
\begin{aligned}
P\left(\breve{Q}_n = b_\omega, \breve{A}_n = m_\varpi | \mathcal{J}_n\right) &= \frac{\tilde{E}\left(\Lambda_n \langle \breve{Q}_n, b_\omega\rangle\langle \breve{A}_n, m_\varpi\rangle | \mathcal{J}_n\right)}{\tilde{E}(\Lambda_n | \mathcal{J}_n)} \\
&= \frac{g_n(\omega, \varpi)}{\tilde{E}\left(\Lambda_n \sum_{\upsilon=1}^{Q|S_H||S_N|}\langle \breve{Q}_n, b_\upsilon\rangle \sum_{\vartheta=1}^{A|S_X||S_N|}\langle \breve{A}_n, m_\vartheta\rangle \middle| \mathcal{J}_n\right)} = \frac{g_n(\omega, \varpi)}{\sum_{\upsilon=1}^{Q|S_H||S_N|}\sum_{\vartheta=1}^{A|S_X||S_N|} g_n(\upsilon, \vartheta)}
\end{aligned}
$$

Theorem 3.4: *The unnormalized probability $g_n(\omega, \varpi)$ satisfies the recursion:*

$$g_n(\omega,\varpi) = \langle \frac{B}{\tilde{B}} X_{n-1} \otimes H_{n-1} \otimes N_{n-1}, X_n \rangle \times \sum_{s=1}^{|S_H|} \langle \breve{Q}_\omega H_{n-1} \otimes N_{n-1}, f_s \rangle$$
$$\times \sum_{u=1}^{|S_N|} \langle \breve{A}_\varpi N_{n-1} \otimes X_{n-1}, h_u \rangle \times \sum_{\upsilon=1}^{Q|S_H||S_N|} \sum_{\vartheta=1}^{A|S_X||S_N|} d_{\omega\upsilon} k_{\varpi\vartheta} g_{n-1}(\upsilon,\vartheta),$$

where $g_0(\omega,\varpi)$ is the initial joint probability of $\breve{Q}_n$ and $\breve{A}_n$.

Proof: In view of (3.2.1) and (3.2.2) and using the distributions of H_n and N_n under measure $\tilde{P}$ for (3.2.6) we write:

$$\begin{aligned} g_n(\omega,\varpi) &= \tilde{E}\left(\Lambda_n \langle \breve{Q}_n, b_\omega\rangle \langle \breve{A}_n, m_\varpi\rangle | \mathcal{I}_n\right) \\ &= \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \Bigg(\Lambda_{n-1} \langle \breve{Q}_n, b_\omega\rangle \langle \breve{A}_n, m_\varpi\rangle \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle \langle X_{n-1}, e_i\rangle \langle H_{n-1}, f_k\rangle \langle N_{n-1}, h_m\rangle} \\ &\times \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} \left(\frac{q_{s,rl}(\omega)}{\tilde{q}_{s,rl}(\omega)}\right)^{\langle H_{n-1}, f_r\rangle \langle N_{n-1}, h_l\rangle} \times \prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} \left(\frac{a_{u,vw}(\varpi)}{\tilde{a}_{u,vw}(\varpi)}\right)^{\langle N_{n-1}, h_v\rangle \langle X_{n-1}, e_w\rangle} \\ &\times \tilde{P}(H_n = f_s, N_n = h_u | \mathcal{I}_n) \Bigg) \end{aligned} \tag{3.2.7}$$

Now recall the dynamics (3.1.6) for the processes $\breve{Q}_n$ and $\breve{A}_n$, that is $\langle \breve{Q}_n, b_\omega\rangle = d_{\omega\upsilon} \langle \breve{Q}_{n-1}, b_\upsilon\rangle$, and $\langle \breve{A}_n, m_\varpi\rangle = k_{\varpi\vartheta} \langle \breve{A}_{n-1}, m_\vartheta\rangle$, and change the probability with expectation, for (3.2.7) we get:

$$\begin{aligned} g_n(\omega,\varpi) = \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \Bigg(& \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle \langle X_{n-1}, e_i\rangle \langle H_{n-1}, f_k\rangle \langle N_{n-1}, h_m\rangle} \\ & \times \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} q_{s,rl}(\omega)^{\langle H_{n-1}, f_r\rangle \langle N_{n-1}, h_l\rangle} \times \prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} a_{u,vw}(\varpi)^{\langle N_{n-1}, h_v\rangle \langle X_{n-1}, e_w\rangle} \\ & \times \tilde{E}\left(\Lambda_{n-1} d_{\omega\upsilon} \langle \breve{Q}_{n-1}, b_\upsilon\rangle k_{\varpi\vartheta} \langle \breve{A}_{n-1}, m_\vartheta\rangle | \mathcal{I}_{n-1}\right) \Bigg) \end{aligned}$$

$$= \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n, e_j\rangle\langle X_{n-1}, e_i\rangle\langle H_{n-1}, f_k\rangle\langle N_{n-1}, h_m\rangle}$$
$$\times \sum_{s=1}^{|S_Z|} \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} q_{s,rl}(\omega)^{\langle H_{n-1}, f_r\rangle\langle N_{n-1}, h_l\rangle} \sum_{u=1}^{|S_N|} \left(\prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} a_{u,vw}(\varpi)^{\langle N_{n-1}, h_v\rangle\langle X_{n-1}, e_w\rangle} \right)$$
$$\times \sum_{v=1}^{Q|S_H||S_N|} \sum_{\vartheta=1}^{A|S_X||S_N|} d_{\omega v} k_{\varpi\vartheta} \tilde{E}\left(\Lambda_{n-1}\langle \breve{Q}_{n-1}, b_v\rangle\langle \breve{A}_{n-1}, m_\vartheta\rangle | \mathcal{I}_{n-1}\right)$$

It must be noted that the last equation holds due to distributions of $\breve{Q}_n$ and $\breve{A}_n$ under the measure $\tilde{P}$. In view of (3.2.6) for $n-1$, we get the proof of the theorem. ■

3.3. Predictions for Claim numbers and Aggregate Claim Amounts

It is important for insurance companies to predict the possible number and amount of future claims m years into the future.

As a result of generalized Bayes' theorem, the joint probability distribution of the claim numbers and aggregate claim amounts reported by a policyholder has the form:

$$P(H_n = f_s, N_n = h_m | \mathcal{I}_n) = \frac{\tilde{E}(\langle H_n, f_s\rangle\langle N_n, h_m\rangle\Lambda_n | \mathcal{I}_n)}{\tilde{E}(\Lambda_n | \mathcal{I}_n)}$$

where the denominator is constant in the case of fixed n.

Having the information up to time n, the above-mentioned distribution's behavior at the end of current year is presented by the following Lemma:

Lemma 3.3: *The unnormalized joint conditional probability distribution of reported claim numbers and aggregate claim amounts has the form:*

$$\tilde{E}(H_n N_n \Lambda_n | \mathcal{I}_n)$$
$$= \langle \frac{B}{\tilde{B}} X_{n-1} \otimes H_{n-1} \otimes N_{n-1}, X_n\rangle$$
$$\times \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} f_s \langle \breve{Q}_\omega H_{n-1} \otimes N_{n-1}, f_s\rangle h_u \langle \breve{A}_\varpi N_{n-1}$$
$$\otimes X_{n-1}, h_u\rangle g_n(\omega, \varpi)$$

Proof: Using (3.2.1) and (3.2.2) and the distributions of H_n and N_n under measure $\tilde{P}$, we have

$$\tilde{E}(H_n N_n \Lambda_n | \mathcal{J}_n) = \sum_{s=1}^{|S_H|} f_s N_n \Lambda_n \tilde{P}(H_n = f_s | \mathcal{J}_n) = \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} f_s h_u \tilde{E}(\langle H_n, f_s \rangle \langle N_n, h_u \rangle \Lambda_n | \mathcal{J}_n)$$

$$= \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} f_s h_u \tilde{E}\left(\langle H_n, f_s \rangle \langle N_n, h_u \rangle \Lambda_n \sum_{\omega=1}^{Q|S_H||S_N|} \langle \breve{Q}_n, b_\omega \rangle \sum_{\varpi=1}^{A|S_X||S_N|} \langle \breve{A}_n, m_\varpi \rangle \middle| \mathcal{J}_n \right)$$

$$= \sum_{s=1}^{|S_H|} f_s \sum_{u=1}^{|S_N|} h_u \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \langle H_n, f_s \rangle \langle N_n, h_u \rangle \langle \breve{Q}_n, b_\omega \rangle \langle \breve{A}_n, m_\varpi \rangle \Lambda_{n-1}$$
$$\times \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}} \right)^{\langle X_n, e_j \rangle \langle X_{n-1}, e_i \rangle \langle H_{n-1}, f_k \rangle \langle N_{n-1}, h_m \rangle}$$
$$\times \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} \left(\frac{q_{s,rl}(\omega)}{\tilde{q}_{s,rl}(\omega)} \right)^{\langle H_{n-1}, f_r \rangle \langle N_{n-1}, h_l \rangle}$$
$$\times \prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} \left(\frac{a_{u,vw}(\varpi)}{\tilde{a}_{u,vw}(\varpi)} \right)^{\langle N_{n-1}, h_v \rangle \langle X_{n-1}, e_w \rangle} \tilde{P}(H_n = f_s, N_n = h_m | \mathcal{J}_n) =$$

$$= \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}} \right)^{\langle X_n, e_j \rangle \langle X_{n-1}, e_i \rangle \langle H_{n-1}, f_k \rangle \langle N_{n-1}, h_m \rangle}$$
$$\times \sum_{s=1}^{|S_H|} f_s \sum_{u=1}^{|S_N|} h_u \sum_{\omega=1}^{Q|S_H||S_N|} \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} q_{s,rl}(\omega)^{\langle H_{n-1}, f_r \rangle \langle N_{n-1}, h_l \rangle}$$
$$\times \sum_{\varpi=1}^{A|S_X||S_N|} \prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} a_{u,vw}(\varpi)^{\langle N_{n-1}, h_v \rangle \langle X_{n-1}, e_w \rangle}$$
$$\times \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \tilde{E}(\langle H_n, f_s \rangle \langle N_n, h_u \rangle \langle \breve{Q}_n, b_\omega \rangle \langle \breve{A}_n, m_\varpi \rangle \Lambda_{n-1} | \mathcal{J}_n)$$

We assume that the processes under the consideration are sequences of independent random variables under measure $\tilde{P}$, so instead of condition $\mathcal{J}_n$ we can write $\mathcal{J}_{n-1}$, and then we will get the following result:

$$
\begin{aligned}
&\tilde{E}(H_n N_n \Lambda_n | \mathcal{J}_n) \\
&\qquad = \langle \frac{B}{\tilde{B}} X_{n-1} \otimes H_{n-1} \\
&\qquad \otimes N_{n-1}, X_n \rangle \sum_{s=1}^{|S_H|} f_s \sum_{u=1}^{|S_N|} h_u \sum_{\omega=1}^{Q|S_H||S_N|} \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} q_{s,rl}(\omega)^{\langle H_{n-1}, f_r \rangle \langle N_{n-1}, h_l \rangle} \\
&\qquad \times \sum_{\varpi=1}^{A|S_X||S_N|} \prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} a_{u,vw}(\varpi)^{\langle N_{n-1}, h_v \rangle \langle X_{n-1}, e_w \rangle} \\
&\qquad \times \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \frac{\tilde{E}\left(\langle H_n, f_s \rangle \langle N_n, h_u \rangle \langle \breve{Q}_n, b_\omega \rangle \langle \breve{A}_n, m_\varpi \rangle \Lambda_n \middle| \mathcal{J}_{n-1} \right)}{\tilde{E}(\lambda_n | \mathcal{J}_{n-1})}
\end{aligned}
$$

Using the proof of *Lemma 3.2* and the independence under the measure $\tilde{P}$ we can change $\mathcal{J}_{n-1}$ by $\mathcal{J}_n$ in the condition, so we get the statement of *Lemma 3.3*. ■

We wish to get a formula for the m step prediction of the joint distribution function of claim numbers and aggregate claim amounts given the history up to time n. First, we will find out it for 2 step prediction and then extend it to m steps.

Consider the unnormalized conditional joint distribution of $\langle H_n, f_{s_0} \rangle$, $\langle H_{n+1}, f_{s_1} \rangle$, $\langle N_n, h_{u_0} \rangle$ and $\langle N_{n+1}, h_{u_1} \rangle$.

$$
\begin{aligned}
&\tilde{E}\left(\langle H_n, f_{s_0} \rangle \langle H_{n+1}, f_{s_1} \rangle \langle N_n, h_{u_0} \rangle \langle N_{n+1}, h_{u_1} \rangle \Lambda_{n+1} \middle| \mathcal{J}_n \right) \\
&= \sum_{\omega_0, \omega_1=1}^{Q|S_H||S_N|} \sum_{\varpi_0, \varpi_1=1}^{A|S_X||S_N|} \sum_{j_1=1}^{|S_X|} \tilde{E}\left(\langle H_n, f_{s_0} \rangle \langle H_{n+1}, f_{s_1} \rangle \langle N_n, h_{u_0} \rangle \langle N_{n+1}, h_{u_1} \rangle \langle X_{n+1}, e_{j_1} \rangle \right. \\
&\cdot \left. \langle \breve{Q}_n, b_{\omega_0} \rangle \langle \breve{Q}_{n+1}, b_{\omega_1} \rangle \langle \breve{A}_n, m_{\varpi_0} \rangle \langle \breve{A}_{n+1}, m_{\varpi_1} \rangle \Lambda_n \lambda_{n+1} \middle| \mathcal{J}_n \right) \\
&= \sum_{\omega_0, \omega_1=1}^{Q|S_H||S_N|} \sum_{\varpi_0, \varpi_1=1}^{A|S_X||S_N|} \sum_{j_1=1}^{|S_X|} \prod_{i=1}^{|S_X|} \left(\frac{p_{j_1, i s_0 u_0}}{\tilde{p}_{j_1, i s_0 u_0}} \right)^{\langle X_n, e_i \rangle} \left(\frac{q_{s_1, s_0 u_0}(\omega_1)}{\tilde{q}_{s_1, s_0 u_0}(\omega_1)} \right) \prod_{w=1}^{|S_X|} \left(\frac{a_{u_1, u_0 w}(\varpi_1)}{\tilde{a}_{u_1, u_0 w}(\varpi_1)} \right)^{\langle X_n, e_w \rangle} \\
&\cdot \tilde{E}\left(\langle H_n, f_{s_0} \rangle \langle H_{n+1}, f_{s_1} \rangle \langle N_n, h_{u_0} \rangle \langle N_{n+1}, h_{u_1} \rangle \langle X_{n+1}, e_{j_1} \rangle \right. \\
&\cdot \left. \langle \breve{Q}_n, b_{\omega_0} \rangle \langle D\breve{Q}_n, b_{\omega_1} \rangle \langle \breve{A}_n, m_{\varpi_0} \rangle \langle K\breve{A}_n, m_{\varpi_1} \rangle \Lambda_n \middle| \mathcal{J}_n \right)
\end{aligned}
$$

By using the Markov property of $\breve{Q}_n$ and properties of conditional expectation, it is easy to note that

$$\tilde{E}\left(\langle \breve{Q}_{n+1}, b_{\omega_1}\rangle | \mathcal{I}_n\right) = \sum_{\omega_1=1}^{Q|S_H||S_N|} \langle \breve{Q}_{n+1}, b_{\omega_1}\rangle \tilde{P}\left(\breve{Q}_{n+1} = b_{\omega_1} | \breve{Q}_n = b_{\omega_0}\right) = d_{\omega_1\omega_0}$$

A similar relationship is valid for the process $\breve{A}_n$.

Note that $H_n, H_{n+1}, N_n, N_{n+1}$ and X_{n+1} are not included in $\mathcal{I}_n$. Using their distributions under $\tilde{P}$,we have:

$$\tilde{E}\left(\langle H_n, f_{s_0}\rangle\langle H_{n+1}, f_{s_1}\rangle\langle N_n, h_{u_0}\rangle\langle N_{n+1}, h_{u_1}\rangle\Lambda_{n+1}\middle|\mathcal{J}_n\right)$$
$$= \sum_{\omega_0,\omega_1=1}^{Q|S_H||S_N|} q_{s_1 s_0 u_0}(\omega_1) \sum_{\varpi_0,\varpi_1=1}^{A|S_X||S_N|} \sum_{j_1=1}^{|S_X|} \prod_{i=1}^{|S_X|} \left(p_{j_1, i s_0 u_0}\, a_{u_1,u_0 i}(\varpi_1)\right)^{\langle X_n, e_i\rangle} d_{\omega_1\omega_0}$$
$$\cdot\, \tilde{E}\left(\langle H_n, f_{s_0}\rangle\langle N_n, h_{u_0}\rangle\middle|\mathcal{J}_n\right) g_n(\omega_0, \varpi_0) \tag{3.3.1}$$

Now applying the concept of mathematical induction to formula (3.3.1), for m steps we have the following theorem:

Theorem 3.5: *The m step prediction for joint probability distribution of the processes H_n and N_n given the information $\mathcal{J}_n$ has the form*

$$\tilde{E}\left(\Lambda_{n+m} \prod_{t=0}^{m} \langle H_{n+t}, f_{s_t}\rangle\langle N_{n+t}, h_{u_t}\rangle \middle| \mathcal{J}_n\right)$$
$$= \sum_{\omega_0,\omega_1\ldots\omega_m=1}^{Q|S_H||S_N|} q_{s_1 s_0 u_0}(\omega_1) \sum_{\varpi_0,\varpi_1\ldots\varpi_m=1}^{A|S_X||S_N|} \sum_{j_1,j_2,\ldots,j_m=1}^{|S_X|} \prod_{i=1}^{|S_X|} \left(p_{j_1, i s_0 u_0}\, a_{u_1,u_0 i}(\varpi_1)\right)^{\langle X_n, e_i\rangle}$$
$$\times \prod_{t=1}^{m} p_{j_{t+1}, j_t s_t u_t} \cdot q_{s_{t+1}, s_t u_t}(\omega_{t+1}) \cdot a_{u_{t+1}, u_t j_t}(\varpi_{t+1}) d_{\omega_t \omega_{t-1}} k_{\varpi_t \varpi_{t-1}}$$
$$\cdot\, \tilde{E}\left(\langle H_n, f_{s_0}\rangle\langle N_n, h_{u_0}\rangle\middle|\mathcal{J}_n\right) g_n(\omega_0, \varpi_0). \tag{3.3.2}$$

3.4. Evaluation of the Coefficients of Transition Matrices in Extended Bonus-Malus System

Consider the Extended model of BMS (3.1.6). The system is described with the following set of parameters satisfying the conditions (3.1.1′), (3.1.2′), (3.1.3′), (3.1.4′) and (3.1.5′) :

$$\theta := \left\{\begin{array}{lll} p_{j,ikm}, & i,j = \overline{1,|S_X|}, \quad k = \overline{1,|S_H|}, & m = \overline{1,|S_N|}, \\ q_{s,rl}(n), & s,r = \overline{1,|S_H|}, & l = \overline{1,|S_N|}, \\ a_{u,vw}(n), & u,v = \overline{1,|S_N|}, & w = \overline{1,|S_X|}, \\ d_{\omega v}, & \omega, v = \overline{1, Q|S_H||S_N|}, & \\ k_{\varpi\vartheta}, & \varpi,\vartheta = \overline{1, A|S_X||S_N|} & \end{array}\right\}$$

Our purpose is the estimation of the model parameters and it is presented here in two methods [125].

1. Estimation with EM algorithm. One of the best methods of HMM's coefficient estimation is the EM algorithm which is described detailed in [122].

We have to define a new set of parameters $\hat{\theta}$

$$\hat{\theta} := \left\{ \begin{array}{llll} \hat{p}_{j,ikm}(n), & i,j = \overline{1,|S_X|}, & k = \overline{1,|S_H|}, & m = \overline{1,|S_N|}, \\ \hat{q}_{s,rl}(n), & s,r = \overline{1,|S_H|}, & l = \overline{1,|S_N|}, & \\ \hat{a}_{u,vw}(n), & u,v = \overline{1,|S_N|}, & w = \overline{1,|S_X|}, & \\ \hat{d}_{\omega v}(n), & \omega, v = \overline{1,Q|S_H||S_N|}, & & \\ \hat{k}_{\varpi\vartheta}(n), & \varpi,\vartheta = \overline{1,A|S_X||S_N|} & & \end{array} \right\},$$

which maximizes the conditional pseudo log-likelihood functions (3.4.5) and satisfies to conditions (3.1.1′), (3.1.2′), (3.1.3′), (3.1.4′) and (3.1.5′).

Define

$$
\begin{aligned}
&{}^{X}\mathcal{J}_n^{j,ikm} \triangleq \textstyle\sum_{t=1}^{n} \langle X_t, e_j\rangle\langle X_{t-1}, e_i\rangle\langle H_{t-1}, f_k\rangle\langle N_{t-1}, h_m\rangle \\
&{}^{H}\mathcal{J}_n^{s,rl} \triangleq \textstyle\sum_{t=1}^{n} \langle H_t, f_s\rangle\langle H_{t-1}, f_r\rangle\langle N_{t-1}, h_l\rangle \\
&{}^{N}\mathcal{J}_n^{u,vw} \triangleq \textstyle\sum_{t=1}^{n} \langle N_t, h_u\rangle\langle N_{t-1}, h_v\rangle\langle X_{t-1}, e_w\rangle \qquad (3.4.1) \\
&{}^{\breve{Q}}\mathcal{J}_n^{\omega v} \triangleq \textstyle\sum_{t=1}^{n} \langle \breve{Q}_t, b_\omega\rangle\langle \breve{Q}_{t-1}, b_v\rangle \\
&{}^{\breve{A}}\mathcal{J}_n^{\varpi\vartheta} \triangleq \textstyle\sum_{t=1}^{n} \langle \breve{A}_t, m_\varpi\rangle\langle \breve{A}_{t-1}, m_\vartheta\rangle
\end{aligned}
$$

Each figure in (3.4.1) for the corresponding process mentioned on the left top angle represents the number of jumps of the process from one state to another up to time n.

Each figure in the next set of notations shows the number of occasions up to time n for which the corresponding Markov chain was in the mentioned state:

$$
\begin{aligned}
&{}^{X}\mathcal{O}_n^{ikm} \triangleq \textstyle\sum_{t=1}^{n} \langle X_{t-1}, e_i\rangle\langle H_{t-1}, f_k\rangle\langle N_{t-1}, h_m\rangle \\
&{}^{H}\mathcal{O}_n^{rl} \triangleq \textstyle\sum_{t=1}^{n} \langle H_{t-1}, f_r\rangle\langle N_{t-1}, h_l\rangle \\
&{}^{N}\mathcal{O}_n^{vw} \triangleq \textstyle\sum_{t=1}^{n} \langle N_{t-1}, h_v\rangle\langle X_{t-1}, e_w\rangle \qquad (3.4.2)
\end{aligned}
$$

$$^{\check{Q}}\mathcal{O}_n^{v} \triangleq \sum_{t=1}^{n}\langle \check{Q}_{t-1}, b_v\rangle$$

$$^{\check{A}}\mathcal{O}_n^{\vartheta} \triangleq \sum_{t=1}^{n}\langle \check{A}_{t-1}, m_\vartheta\rangle$$

It is obvious that the figures in (3.4.1) and (3.4.2) are random variables.

Remark 3.1: *For each process holds the relationship* $\sum_{i=1}^{|S_*|} {}^*\mathcal{T}_n^{i*} = {}^*\mathcal{O}_n^*$.

To replace parameters θ by $\hat{\theta}$ in (3.1.6) we define the following likelihood function:

$$\begin{aligned}\Gamma_n = &\prod_{t=1}^{n}\prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_H|}\prod_{m=1}^{|S_N|}\left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)^{\langle X_t,e_j\rangle\langle X_{t-1},e_i\rangle\langle H_{t-1},f_k\rangle\langle N_{t-1},h_m\rangle}\\ &\times\prod_{r,s=1}^{|S_H|}\prod_{l=1}^{|S_N|}\left(\frac{\hat{q}_{s,rl}(n)}{q_{s,rl}(n)}\right)^{\langle H_t,f_s\rangle\langle H_{t-1},f_r\rangle\langle N_{t-1},h_l\rangle}\\ &\times\prod_{u,v=1}^{|S_N|}\prod_{w=1}^{|S_X|}\left(\frac{\hat{a}_{u,vw}(n)}{a_{u,vw}(n)}\right)^{\langle N_t,h_u\rangle\langle N_{t-1},h_v\rangle\langle X_{t-1},e_w\rangle}\times\prod_{\omega,v=1}^{Q|S_H||S_N|}\left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)^{\langle\check{Q}_t,b_\omega\rangle\langle\check{Q}_{t-1},b_v\rangle}\\ &\times\prod_{\varpi,\vartheta=1}^{A|S_X||S_N|}\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right)^{\langle\check{A}_t,m_\varpi\rangle\langle\check{A}_{t-1},m_\vartheta\rangle}\end{aligned}\tag{3.4.3}$$

where in the case of $\theta = 0$ we take $\hat{\theta} = 0$ and $\frac{\hat{\theta}}{\theta} = 1$:

It is not difficult to show that Γ_n is a martingale-measure, so according to the Radon-Nycodim theorem there exists a measure $P_{\hat{\theta}}$ so that $\left.\frac{dP_{\hat{\theta}}}{dP_\theta}\right|_{\mathcal{G}_n} = \Gamma_n$ holds.

Lemma 3.4: *Under the measure* $P_{\hat{\theta}}$ *the analogue of the system (3.1.6) holds for parameter set* $\hat{\theta}$.

Proof: Let $\mathcal{G}_n$ be the complete filtration generated by $\{X_k, H_{k-1}, N_{k-1}, \tilde{Q}_k, \tilde{A}_k, k \le n\}$. We have to show that under the measure $dP_{\hat{\theta}}$ the following relationships are hold:

$$E_{\hat{\theta}}\left(\langle X_{n+1}, e_j\rangle|\mathcal{G}_n\right) = \hat{p}_{j,ikm}(n)$$

$$E_{\hat{\theta}}\left(\langle H_{n+1}, f_s\rangle|\mathcal{G}_n\right) = \hat{q}_{s,rl}(n)$$

$$E_{\hat{\theta}}\left(\langle X_{n+1}, h_u \square\rangle \middle| \mathcal{G}_n\right) = \hat{a}_{u,vw}(n)$$

$$E_{\hat{\theta}}\left(\langle \breve{Q}_{n+1}, b_\omega \rangle \middle| \mathcal{G}_n\right) = \hat{d}_{\omega v}(n)$$

$$E_{\hat{\theta}}\left(\langle \breve{A}_{n+1}, m_\varpi \rangle \middle| \mathcal{G}_n\right) = \hat{k}_{\varpi\vartheta}(n)$$

Note that in the case of information up to time n, (3.4.3) gets the form:

$$\Gamma_{n+1} = \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)^{\langle X_{n+1}, e_j\rangle} \left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)^{\langle H_{n+1}, f_s\rangle} \left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right)^{\langle N_{n+1}, h_u\rangle} \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)^{\langle \breve{Q}_{n+1}, b_\omega\rangle} \left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right)^{\langle \breve{A}_{n+1}, m_\varpi\rangle}$$

According to the Bayes' theorem

$$E_{\hat{\theta}}(*\,|G_n) = \frac{E(*\cdot \Gamma_{n+1}|G_n)}{E(\Gamma_{n+1}|G_n)} \tag{3.4.4}$$

where $(*)$ is the scalar product of any of our process with according unit basis $(\langle \text{process}, \text{basis}\rangle)$.

The denominator of $(3.4.4)$ is:

$$E(\Gamma_{n+1}|\mathcal{G}_n) = \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\sum_{\omega=1}^{Q|S_H||S_N|}\sum_{\varpi=1}^{A|S_X||S_N|} \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right) \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right)\cdot P\left(H_{n+1} = f_s, N_{n+1} = h_u, \breve{Q}_{n+1} = b_\omega, \breve{A}_{n+1} = m_\varpi \middle| \mathcal{G}_n\right) = 1$$

were the distributions of model processes and $(3.1.1'), (3.1.2'), (3.1.3'), (3.1.4')$ and $(3.1.5')$ conditions for the parameter set $\hat{\theta}$ were used to get the result.

Now calculate the numerator of $(3.4.4)$ for each process which gives the statement of the Lemma:

$$
\begin{aligned}
&E\left(\langle X_{n+1}, e_j\rangle \Gamma_{n+1} \middle| \mathcal{G}_n\right) \\
&\quad = \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} \langle X_{n+1}, e_j\rangle \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right) \\
&\quad \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right) \cdot P\left(H_{n+1} = f_s, N_{n+1} = h_u, \breve{Q}_{n+1} = b_\omega, \breve{A}_{n+1} = m_\varpi \middle| \mathcal{G}_n\right) \\
&\quad = \sum_{s=1}^{|S_H|} \sum_{u=1}^{|S_N|} \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right) \\
&\quad \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right) \cdot q_{s,km}(n) a_{u,mi}(n) d_{\omega v}\, k_{\varpi\vartheta} E\left(\langle X_{n+1}, e_j\rangle \middle| \mathcal{G}_n\right) = \hat{p}_{j,ikm}(n)
\end{aligned}
$$

$$
\begin{aligned}
&E\left(\langle H_{n+1}, f_s\rangle \Gamma_{n+1} \middle| \mathcal{G}_n\right) \\
&\quad = \sum_{j=1}^{|S_X|} \sum_{u=1}^{|S_N|} \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} \langle H_{n+1}, f_s\rangle \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right) \\
&\quad \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right) \cdot P\left(X_{n+1} = e_j, N_{n+1} = h_u, \breve{Q}_{n+1} = b_\omega, \breve{A}_{n+1} = m_\varpi \middle| \mathcal{G}_n\right) \\
&\quad = \sum_{j=1}^{|S_X|} \sum_{u=1}^{|S_N|} \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right) \\
&\quad \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right) \cdot p_{j,ikm} a_{u,mi}(n) d_{\omega v}\, k_{\varpi\vartheta} E\left(\langle H_{n+1}, f_s\rangle \middle| \mathcal{G}_n\right) = \hat{q}_{s,km}(n)
\end{aligned}
$$

$$
\begin{aligned}
&E\left(\langle N_{n+1}, h_u\rangle \Gamma_{n+1} \middle| \mathcal{G}_n\right) \\
&\quad = \sum_{j=1}^{|S_X|} \sum_{s=1}^{|S_H|} \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} \langle N_{n+1}, h_u\rangle \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right) \\
&\quad \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right) \cdot P\left(X_{n+1} = e_j, H_{n+1} = f_s, \breve{Q}_{n+1} = b_\omega, \breve{A}_{n+1} = m_\varpi \middle| \mathcal{G}_n\right) \\
&\quad = \sum_{j=1}^{|S_X|} \sum_{s=1}^{|S_H|} \sum_{\omega=1}^{Q|S_H||S_N|} \sum_{\varpi=1}^{A|S_X||S_N|} \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right) \\
&\quad \times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right) \cdot p_{j,ikm} q_{s,km}(n) d_{\omega v}\, k_{\varpi\vartheta} E\left(\langle N_{n+1}, h_u\rangle \middle| \mathcal{G}_n\right) = \hat{a}_{u,mi}(n)
\end{aligned}
$$

$$E\left(\langle \breve{Q}_{n+1}, b_\omega\rangle \Gamma_{n+1} \middle| \mathcal{G}_n\right)$$
$$= \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\sum_{\varpi=1}^{A|S_X||S_N|} \langle \breve{Q}_{n+1}, b_\omega\rangle \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right)$$
$$\times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right)\cdot P\left(X_{n+1} = e_j, H_{n+1} = f_s, N_{n+1} = h_u, \breve{A}_{n+1} = m_\varpi \middle| \mathcal{G}_n\right)$$
$$= \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\sum_{\varpi=1}^{A|S_X||S_N|} \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right)$$
$$\times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right)\cdot p_{j,ikm}\, q_{s,km}(n)\, a_{u,mi}(n)\, k_{\varpi\vartheta} E\left(\langle \breve{Q}_{n+1}, b_\omega\rangle \middle| \mathcal{G}_n\right)$$
$$= \hat{d}_{\omega v}(n)$$

$$E\left(\langle \breve{A}_{n+1}, m_\varpi\rangle \Gamma_{n+1} \middle| \mathcal{G}_n\right)$$
$$= \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\sum_{\omega=1}^{Q|S_H||S_N|} \langle \breve{A}_{n+1}, m_\varpi\rangle \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right)$$
$$\times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right)\cdot P\left(X_{n+1} = e_j, H_{n+1} = f_s, N_{n+1} = h_u, \breve{Q}_{n+1} = b_\omega \middle| \mathcal{G}_n\right)$$
$$= \sum_{j=1}^{|S_X|}\sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\sum_{\omega=1}^{Q|S_H||S_N|} \left(\frac{\hat{p}_{j,ikm}(n)}{p_{j,ikm}}\right)\left(\frac{\hat{q}_{s,km}(n)}{q_{s,km}(n)}\right)\left(\frac{\hat{a}_{u,mi}(n)}{a_{u,mi}(n)}\right)$$
$$\times \left(\frac{\hat{d}_{\omega v}(n)}{d_{\omega v}}\right)\left(\frac{\hat{k}_{\varpi\vartheta}(n)}{k_{\varpi\vartheta}}\right)\cdot p_{j,ikm}\, q_{s,km}(n)\, a_{u,mi}(n)\, d_{\omega v}\, E\left(\langle \breve{A}_{n+1}, m_\varpi\rangle \middle| \mathcal{G}_n\right)$$
$$= \hat{k}_{\varpi\vartheta}(n)$$

■

Theorem 3.6: *The new estimates of parameter set $\hat{\theta}$ given the observations up to time n are given by*

$$\hat{p}_{j,ikm} = \frac{{}^{X}\hat{\mathcal{T}}_n^{j,ikm}}{{}^{X}\hat{\mathcal{O}}_n^{ikm}}, \qquad \hat{q}_{s,rl}(n) = \frac{{}^{H}\hat{\mathcal{T}}_n^{s,rl}}{{}^{H}\hat{\mathcal{O}}_n^{rl}}, \qquad \hat{a}_{u,vw}(n) = \frac{{}^{N}\hat{\mathcal{T}}_n^{u,vw}}{{}^{N}\hat{\mathcal{O}}_n^{vw}},$$

$$\hat{d}_{\omega v}(n) = \frac{{}^{\breve{Q}}\hat{\mathcal{T}}_n^{\omega v}}{{}^{\breve{Q}}\hat{\mathcal{O}}_n^{v}}, \qquad \hat{k}_{\varpi\vartheta}(n) = \frac{{}^{\breve{A}}\hat{\mathcal{T}}_n^{\varpi\vartheta}}{{}^{\breve{A}}\hat{\mathcal{O}}_n^{\vartheta}}$$

where ${}^{*}\hat{\mathcal{T}}_n^{*} = E({}^{*}\mathcal{T}_n^{*}|.)$ *And* ${}^{*}\hat{\mathcal{O}}_n^{*} = E({}^{*}\mathcal{O}_n^{*}|.)$.

Proof: To estimate the parameters we use (3.4.1) and (3.4.3). For $\hat{p}_{j,ikm}$ we take the following form:

$$\ln \Gamma_n = \sum_{i,j=1}^{|S_X|} \sum_{k=1}^{|S_H|} \sum_{m=1}^{|S_N|} \sum_{t=1}^{n} \langle X_t, e_j \rangle \langle X_{t-1}, e_i \rangle \langle H_{t-1}, f_k \rangle \langle N_{t-1}, h_m \rangle \left(\ln \hat{p}_{j,ikm} - \ln p_{j,ikm} \right) + R$$
$$= \sum_{i,j=1}^{|S_X|} \sum_{k=1}^{|S_H|} \sum_{m=1}^{|S_N|} {}^X\mathcal{J}_n^{j,ikm} \ln \hat{p}_{j,ikm} + R_1$$

where R_1 is independent of $\hat{p}_{j,ikm}$.

And for others we take accordingly the forms:

$$\ln \Gamma_n = \sum_{r,s=1}^{|S_H|} \sum_{l=1}^{|S_N|} \sum_{t=1}^{n} \langle H_t, f_s \rangle \langle H_{t-1}, f_r \rangle \langle N_{t-1}, h_l \rangle \left(\ln \hat{q}_{s,rl} - \ln q_{s,rl} \right) + R_2$$
$$= \sum_{r,s=1}^{|S_H|} \sum_{l=1}^{|S_N|} {}^H\mathcal{J}_n^{s,rl} \ln \hat{q}_{s,rl}(n) + R_2$$

where R_2 is independent of $\hat{q}_{s,rl}(n)$.

$$\mathrm{Ln}\, \Gamma_n = \sum_{u,v=1}^{|S_N|} \sum_{w=1}^{|S_X|} \sum_{t=1}^{n} \langle N_t, h_u \rangle \langle N_{t-1}, h_v \rangle \langle X_{t-1}, e_w \rangle \left(\ln \hat{a}_{u,vw}(n) - \ln a_{u,vw}(n) \right) + R_3$$
$$= \sum_{u,v=1}^{|S_N|} \sum_{w=1}^{|S_X|} {}^N\mathcal{J}_n^{u,vw} \ln \hat{a}_{u,vw}(n) + R_3$$

where R_3 is independent of $\hat{a}_{u,vw}(n)$.

$$\mathrm{Ln}\, \Gamma_n = \sum_{\omega,v=1}^{Q|S_H||S_N|} \sum_{t=1}^{n} \langle \breve{Q}_t, b_\omega \rangle \langle \breve{Q}_{t-1}, b_v \rangle \left(\ln \hat{d}_{\omega v}(n) - \ln d_{\omega v} \right) + R_4$$
$$= \sum_{\omega,v=1}^{Q|S_H||S_N|} {}^{\breve{Q}}\mathcal{J}_n^{\omega v} \ln \hat{d}_{\omega v}(n) + R_4$$

where R_4 is independent of $\hat{d}_{\omega v}(n)$.

$$\operatorname{Ln}\Gamma_n = \sum_{\varpi,\vartheta=1}^{A|S_X||S_N|}\sum_{t=1}^{n}\langle \breve{A}_t, m_\varpi\rangle\langle \breve{A}_{t-1}, m_\vartheta\rangle\left(\ln \hat{k}_{\varpi\vartheta}(n) - \ln k_{\varpi\vartheta}\right) + R_5$$

$$= \sum_{\omega,\upsilon=1}^{Q|S_H||S_N|} {}^{\breve{A}}\mathcal{T}_n^{\varpi\vartheta} \ln \hat{k}_{\varpi\vartheta}(n) + R_5$$

where R_5 is independent of $\hat{k}_{\varpi\vartheta}(n)$.

For the conditional expectations we have

$$E(\ln\Gamma_n|\mathcal{G}_n) = \begin{bmatrix} \sum_{i,j=1}^{|S_X|}\sum_{k=1}^{|S_H|}\sum_{m=1}^{|S_N|} {}^{X}\hat{\mathcal{T}}_n^{j,ikm} \ln \hat{p}_{j,ikm} + \hat{R}_1, \ \ for\ \hat{p}_{j,ikm} \\ \sum_{r,s=1}^{|S_H|}\sum_{l=1}^{|S_N|} {}^{H}\hat{\mathcal{T}}_n^{s,rl} \ln \hat{q}_{s,rl}(n) + \hat{R}_2, \ \ for\ \hat{q}_{s,rl}(n) \\ \sum_{u,v=1}^{|S_N|}\sum_{w=1}^{|S_X|} {}^{N}\hat{\mathcal{T}}_n^{u,vw} \ln \hat{a}_{u,vw}(n) + \hat{R}_3, \ \ for\ \hat{a}_{u,vw}(n) \\ \sum_{\omega,\upsilon=1}^{Q|S_H||S_N|} {}^{\breve{Q}}\hat{\mathcal{T}}_n^{\omega\upsilon} \ln \hat{d}_{\omega\upsilon}(n) + \hat{R}_4, \ \ for\ \hat{d}_{\omega\upsilon}(n) \\ \sum_{\omega,\upsilon=1}^{Q|S_H||S_N|} {}^{\breve{A}}\hat{\mathcal{T}}_n^{\varpi\vartheta} \ln \hat{k}_{\varpi\vartheta}(n) + \hat{R}_5, \ \ for\ \hat{k}_{\varpi\vartheta}(n) \end{bmatrix} \tag{3.4.5}$$

where $\hat{p}_{j,ikm}$, $\hat{q}_{s,rl}(n)$, $\hat{a}_{u,vw}(n)$, $\hat{d}_{\omega\upsilon}(n)$ and $\hat{k}_{\varpi\vartheta}(n)$ must satisfy the analogue of $(3.1.1'), (3.1.2'), (3.1.3'), (3.\ 1.4'), (3.1.5')$, as well.

Note that the conditional expectations also satisfy the relationship

$$\sum_{i=1}^{|S_*|} {}^{*}\hat{\mathcal{T}}_n^{i*} = {}^{*}\hat{\mathcal{O}}_n^{*} \tag{3.4.6}$$

Our purpose is to maximize (3.4.5) each time taking into account the constraints $(3.1.1'), (3.1.2'), (3.1.3'), (3.1.4')$ or $(3.1.5')$. Write λ_* for the Lagrange multiplier and put

$$L(\hat{\theta},\lambda_*) = \left[\begin{array}{l} \sum_{i,j=1}^{|S_X|}\sum_{k=1}^{|S_H|}\sum_{m=1}^{|S_N|} {}^{X}\hat{\mathcal{T}}_n^{j,ikm} \ln \hat{p}_{j,ikm} + \hat{R}_1 + \lambda_X\left(\sum_{j=1}^{|S_X|}\hat{p}_{j,ikm} - 1\right) \\ \sum_{r,s=1}^{|S_H|}\sum_{l=1}^{|S_N|} {}^{H}\hat{\mathcal{T}}_n^{s,rl} \ln \hat{q}_{s,rl}(n) + \hat{R}_2 + \lambda_H\left(\sum_{s=1}^{|S_H|}\hat{q}_{s,rl}(n) - 1\right) \\ \sum_{u,v=1}^{|S_N|}\sum_{w=1}^{|S_X|} {}^{N}\hat{\mathcal{T}}_n^{u,vw} \ln \hat{a}_{u,vw}(n) + \hat{R}_3 + \lambda_N\left(\sum_{u=1}^{|S_N|}\hat{a}_{u,vw}(n) - 1\right) \\ \sum_{\omega,v=1}^{Q|S_H||S_N|} {}^{\breve{Q}}\hat{\mathcal{T}}_n^{\omega v} \ln \hat{d}_{\omega v}(n) + \hat{R}_4 + \lambda_{\breve{Q}}\left(\sum_{\omega=1}^{Q|S_H||S_N|}\hat{d}_{\omega v}(n) - 1\right) \\ \sum_{\omega,v=1}^{Q|S_H||S_N|} {}^{\breve{A}}\hat{\mathcal{T}}_n^{\varpi\vartheta} \ln \hat{k}_{\varpi\vartheta}(n) + \hat{R}_5 + \lambda_{\breve{A}}\left(\sum_{\varpi=1}^{Q|S_H||S_N|}\hat{k}_{\varpi\vartheta}(n) - 1\right) \end{array}\right.$$

Differentiating in λ_* and in $\hat{\theta}$ we get the result by equating the derivatives to 0:

$$\begin{cases} \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \hat{p}_{j,ikm}} = \dfrac{{}^{X}\hat{\mathcal{T}}_n^{j,ikm}}{\hat{p}_{j,ikm}} + \lambda_X = 0 \\ \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \lambda_X} = \sum_{j=1}^{|S_X|}\hat{p}_{j,ikm} - 1 = 0 \end{cases} \quad \begin{cases} \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \hat{q}_{s,rl}(n)} = \dfrac{{}^{H}\hat{\mathcal{T}}_n^{s,rl}}{\hat{q}_{s,rl}(n)} + \lambda_H = 0 \\ \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \lambda_H} = \sum_{j=1}^{|S_X|}\hat{q}_{s,rl}(n) - 1 = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \hat{a}_{u,vw}(n)} = \dfrac{{}^{N}\hat{\mathcal{T}}_n^{u,vw}}{\hat{a}_{u,vw}(n)} + \lambda_N = 0 \\ \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \lambda_N} = \sum_{j=1}^{|S_X|}\hat{a}_{u,vw}(n) - 1 = 0 \end{cases} \quad \begin{cases} \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \hat{d}_{\omega v}(n)} = \dfrac{{}^{\breve{Q}}\hat{\mathcal{T}}_n^{\omega v}}{\hat{d}_{\omega v}(n)} + \lambda_{\breve{Q}} = 0 \\ \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \lambda_{\breve{Q}}} = \sum_{j=1}^{|S_X|}\hat{d}_{\omega v}(n) - 1 = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \hat{k}_{\varpi\vartheta}(n)} = \dfrac{{}^{\breve{A}}\hat{\mathcal{T}}_n^{\varpi\vartheta}}{\hat{k}_{\varpi\vartheta}(n)} + \lambda_{\breve{A}} = 0 \\ \dfrac{\partial L(\hat{\theta},\lambda_*)}{\partial \lambda_{\breve{A}}} = \sum_{j=1}^{|S_X|}\hat{k}_{\varpi\vartheta}(n) - 1 = 0 \end{cases}$$

Note that $\lambda_* = -\,{}^{*}\hat{O}_n$. For example from the first equation ${}^{X}\hat{\mathcal{T}}_n^{j,ikm} = -\lambda_X\hat{p}_{j,ikm}$, taking the sums of both sides we get: $\sum_{j=1}^{|S_X|} {}^{X}\hat{\mathcal{T}}_n^{j,ikm} = -\lambda_X\sum_{j=1}^{|S_X|}\hat{p}_{j,ikm}(n)$, taking into account (3.1.1′) and (3.4.6) for the process X, the optimal estimate for $\hat{p}_{j,ikm}(n)$ is

$\hat{p}_{j,ikm}(n) = \frac{X_{\hat{\mathcal{J}}_n}^{j,ikm}}{X_{\hat{\mathcal{O}}_n}^{ikm}}$. So we get the statement of the *Theorem 3.6.*

■

2. Parameter Estimation with Recursion. For getting the parameter estimation in the previous part we have to do some prior assumptions on the probability distribution of parameter set θ, but if we have the initial distributions the recursive estimation of the parameters can be done. The new estimate in this case is presented as the previous estimate corrected with the new information.

We assume that θ takes values in some set $\Theta \in \mathbb{R}^l$. Suppose we have the measure $\tilde{P}$ under which the processes of the system (3.1.6) are i.i.d.. According to Theorem 3.2 we define the "real world" measure P under which the system (3.1.6) holds. Consider the unnormalized joint conditional density:

$$\alpha_n(j,\theta) = \tilde{E}\big(\Lambda_n \langle X_n, e_j \rangle I(\theta \in d\theta)\big|\mathcal{J}_n\big)$$

where $I(A)$ is the indicator function of A.

The normalized joint conditional density is:

$$f_n(j,\theta) = \frac{\alpha_n(j,\theta)}{\sum_{j=1}^{|S_X|} \int_\Theta \alpha_n(j,u)du}$$

Assume given the initial density $f_0(\cdot)$.

Theorem 3.7: *The unnormalized joint conditional density $\alpha_n(j,\theta)$ satisfies the recursion:*

$$\alpha_n(j,\theta) = \sum_{i=1}^{|S_X|}\sum_{k=1}^{|S_H|}\sum_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right) \alpha_{n-1}(i,\theta)$$

Proof: Suppose $g(\theta)$ is any real valued Borel function on Θ. Then $\tilde{E}\big(\Lambda_n \langle X_n, e_j \rangle g(\theta)\big|\mathcal{J}_n\big) = \int_\Theta \alpha_n(j,u)g(u)du.$

On the other hand using the distribution of X_n under the measure $\tilde{P}$ we get:

$$\tilde{E}\left(\Lambda_n \langle X_n, e_j \rangle g(\theta) \middle| \mathcal{I}_n\right) = \sum_{j=1}^{|S_X|} \Lambda_n g(\theta) \tilde{E}\left(\langle X_n, e_j \rangle \middle| \mathcal{I}_n\right)$$

Substitute the dynamics of X_n from (3.1.6):

$$\alpha_n(j,\theta) = \sum_{i,j=1}^{|S_X|} \sum_{k=1}^{|S_H|} \sum_{m=1}^{|S_N|} \Lambda_n g(\theta) \tilde{E}\left(p_{j,ikm} \langle X_{n-1}, e_i \rangle \langle H_{n-1}, f_k \rangle \langle N_{n-1}, h_m \rangle \middle| \mathcal{I}_n\right)$$

Using the second relationship of (3.1.6), the distribution of H_n as well as the formulas (3.2.1) and (3.2.2) we get:

$$\alpha_n(j,\theta) = \sum_{i=1}^{|S_X|} \sum_{k=1}^{|S_H|} \sum_{m=1}^{|S_N|} \tilde{E}\left(\langle X_{n-1}, e_i \rangle \frac{\langle H_n, f_s \rangle}{q_{s,km}(n)} \Lambda_{n-1} g(\theta) \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right) \right.$$
$$\left. \times \prod_{s=1}^{|S_H|} \left(\frac{q_{s,km}(n)}{\tilde{q}_{s,km}(n)}\right)^{\langle H_n, f_s \rangle} \prod_{u=1}^{|S_N|} \left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u \rangle} \middle| \mathcal{I}_n \right) =$$

$$= \sum_{i=1}^{|S_X|} \sum_{k=1}^{|S_H|} \sum_{m=1}^{|S_N|} \tilde{E}\left(\langle X_{n-1}, e_i \rangle \Lambda_{n-1} g(\theta) \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right) \prod_{u=1}^{|S_N|} \left(\frac{a_{u,mi}(n)}{\tilde{a}_{u,mi}(n)}\right)^{\langle N_n, h_u \rangle} \middle| \mathcal{I}_n \right)$$
$$= \sum_{i=1}^{|S_X|} \sum_{k=1}^{|S_H|} \sum_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right) \tilde{E}\left(\langle X_{n-1}, e_i \rangle \Lambda_{n-1} g(\theta) | \mathcal{I}_{n-1}\right)$$
$$= \sum_{i=1}^{|S_X|} \sum_{k=1}^{|S_H|} \sum_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right) \int_{\Theta} \alpha_{n-1}(i,u) g(u) du$$

As g is arbitrary we see the result. ∎

Write $g_n(\omega, \varpi, \theta)$ for the unnormalized joint conditional density of processes $\breve{Q}_n$ and $\breve{A}_n$:

$$g_n(\omega,\varpi,\theta) = \tilde{E}\left(\Lambda_n \langle \breve{Q}_n, b_\omega \rangle \langle \breve{A}_n, m_\varpi \rangle I(\theta \in d\theta) \middle| \mathcal{I}_n\right) \tag{3.4.7}$$

The normalized joint conditional density is:

$$f_n(\omega,\varpi,\theta) = \frac{g_n(\omega,\varpi,\theta)}{\sum_{\omega=1}^{Q|S_H||S_N|}\sum_{\varpi=1}^{A|S_X||S_N|}\int_\Theta g_n(\omega,\varpi,u)\,du}$$

Assume given the initial density $f_0(\cdot)$.

Theorem 3.8: The unnormalized joint conditional density $g_n(\omega,\varpi,\theta)$ satisfies the recursion

$$g_n(\omega,\varpi,\theta) = \langle \frac{B}{\breve{B}} X_{n-1} \otimes H_{n-1} \otimes N_{n-1}, X_n\rangle \times \sum_{s=1}^{|S_H|} \langle \breve{Q}_\omega H_{n-1} \otimes N_{n-1}, f_s\rangle$$
$$\times \sum_{u=1}^{|S_N|} \langle \breve{A}_\varpi N_{n-1} \otimes X_{n-1}, h_u\rangle \times \sum_{\upsilon=1}^{Q|S_H||S_N|} \sum_{\vartheta=1}^{A|S_X||S_N|} d_{\omega\upsilon} k_{\varpi\vartheta} g_{n-1}(\upsilon,\vartheta,\theta)$$

Proof: Suppose g is any real valued Borel function on Θ. Then

$$\tilde{E}\big(\Lambda_n\langle \breve{Q}_n, b_\omega\rangle\langle \breve{A}_n, m_\varpi\rangle g(\theta)\big|\mathcal{J}_n\big) = \int_\Theta g_n(\omega,\varpi,u)\,g(u)\,du$$

Using $(3.2.1)$ and $(3.2.2)$ as well as the distributions of processes H_n and N_n under the measure $\tilde{P}$, for $(3.4.7)$ we have:

$$
\begin{aligned}
g_n(\omega,\varpi,\theta) &= \tilde{E}\left(\Lambda_n\langle \breve{Q}_n, b_\omega\rangle\langle \breve{A}_n, m_\varpi\rangle g(\theta)\middle|\mathcal{I}_n\right)\\
&= \sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\Bigg(\Lambda_{n-1}\langle \breve{Q}_n, b_\omega\rangle\langle \breve{A}_n, m_\varpi\rangle g(\theta)\\
&\quad\times \prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_H|}\prod_{m=1}^{|S_N|}\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle H_{n-1},f_k\rangle\langle N_{n-1},h_m\rangle}\\
&\quad\times \prod_{r,s=1}^{|S_H|}\prod_{l=1}^{|S_N|}\left(\frac{q_{s,rl}(n)}{\tilde{q}_{s,rl}(n)}\right)^{\langle H_{n-1},f_r\rangle\langle N_{n-1},h_l\rangle}\times\prod_{u,v=1}^{|S_N|}\prod_{w=1}^{|S_X|}\left(\frac{a_\varpi(u,v,w)}{\tilde{a}_\varpi(u,v,w)}\right)^{\langle N_{n-1},h_v\rangle\langle X_{n-1},e_w\rangle}\\
&\quad\times \tilde{P}(H_n=f_s, N_n=h_u|\mathcal{I}_n)\Bigg)\\
&= \sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\Bigg(\prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_H|}\prod_{m=1}^{|S_N|}\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle H_{n-1},f_k\rangle\langle N_{n-1},h_m\rangle}\\
&\quad\times \prod_{r,s=1}^{|S_H|}\prod_{l=1}^{|S_N|}\left(\frac{q_\omega(s,r,l)}{\tilde{q}_\omega(s,r,l)}\right)^{\langle H_{n-1},f_r\rangle\langle N_{n-1},h_l\rangle}\times\prod_{u,v=1}^{|S_N|}\prod_{w=1}^{|S_X|}\left(\frac{a_\varpi(u,v,w)}{\tilde{a}_\varpi(u,v,w)}\right)^{\langle N_{n-1},h_v\rangle\langle X_{n-1},e_w\rangle}\\
&\quad\times \tilde{E}\left(\Lambda_{n-1}\langle \breve{Q}_n, b_\omega\rangle\langle \breve{A}_n, m_\varpi\rangle\langle H_n,f_s\rangle\langle N_n,h_u\rangle g(\theta)|\mathcal{I}_n\right)\Bigg)
\end{aligned}
$$

Using the dynamics of the processes $\breve{Q}_n$ and $\breve{A}_n$, that is $\langle \breve{Q}_n, b_\omega\rangle = d_{\omega v}\langle \breve{Q}_{n-1}, b_v\rangle$, and $\langle \breve{A}_n, m_\varpi\rangle = k_{\varpi\vartheta}\langle \breve{A}_{n-1}, m_\vartheta\rangle$, we get:

$$
\begin{aligned}
g_n(\omega,\varpi,\theta) &= \sum_{s=1}^{|S_H|}\sum_{u=1}^{|S_N|}\Bigg(\prod_{i,j=1}^{|S_X|}\prod_{k=1}^{|S_H|}\prod_{m=1}^{|S_N|}\left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}}\right)^{\langle X_n,e_j\rangle\langle X_{n-1},e_i\rangle\langle H_{n-1},f_k\rangle\langle N_{n-1},h_m\rangle}\\
&\quad\times \prod_{r,s=1}^{|S_H|}\prod_{l=1}^{|S_N|} q_\omega(s,r,l)^{\langle H_{n-1},f_r\rangle\langle N_{n-1},h_l\rangle}\times\prod_{u,v=1}^{|S_N|}\prod_{w=1}^{|S_X|} a_\varpi(u,v,w)^{\langle N_{n-1},h_v\rangle\langle X_{n-1},e_w\rangle}\\
&\quad\times \tilde{E}\left(\Lambda_{n-1}d_{\omega v}\langle \breve{Q}_{n-1}, b_v\rangle k_{\varpi\vartheta}\langle \breve{A}_{n-1}, m_\vartheta\rangle g(\theta)|\mathcal{I}_{n-1}\right)\Bigg)
\end{aligned}
$$

$$
= \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}} \right)^{\langle X_n, e_j \rangle \langle X_{n-1}, e_i \rangle \langle H_{n-1}, f_k \rangle \langle N_{n-1}, h_m \rangle}
$$

$$
\times \sum_{s=1}^{|S_Z|} \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} q_\omega(s,r,l)^{\langle H_{n-1}, f_r \rangle \langle N_{n-1}, h_l \rangle} \sum_{u=1}^{|S_N|} \left(\prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} a_\varpi(u,v,w)^{\langle N_{n-1}, h_v \rangle \langle X_{n-1}, e_w \rangle} \right)
$$

$$
\times \sum_{v=1}^{Q|S_H||S_N|} \sum_{\vartheta=1}^{A|S_X||S_N|} d_{\omega v} k_{\varpi\vartheta} \tilde{E}\left(\Lambda_{n-1} \langle \breve{Q}_{n-1}, b_v \rangle \langle \breve{A}_{n-1}, m_\vartheta \rangle g(\theta) | \mathcal{J}_{n-1} \right) =
$$

$$
= \prod_{i,j=1}^{|S_X|} \prod_{k=1}^{|S_H|} \prod_{m=1}^{|S_N|} \left(\frac{p_{j,ikm}}{\tilde{p}_{j,ikm}} \right)^{\langle X_n, e_j \rangle \langle X_{n-1}, e_i \rangle \langle H_{n-1}, f_k \rangle \langle N_{n-1}, h_m \rangle}
$$

$$
\times \sum_{s=1}^{|S_Z|} \prod_{r,s=1}^{|S_H|} \prod_{l=1}^{|S_N|} q_\omega(s,r,l)^{\langle H_{n-1}, f_r \rangle \langle N_{n-1}, h_l \rangle} \sum_{u=1}^{|S_N|} \left(\prod_{u,v=1}^{|S_N|} \prod_{w=1}^{|S_X|} a_\varpi(u,v,w)^{\langle N_{n-1}, h_v \rangle \langle X_{n-1}, e_w \rangle} \right)
$$

$$
\times \sum_{v=1}^{Q|S_H||S_N|} \sum_{\vartheta=1}^{A|S_X||S_N|} d_{\omega v} k_{\varpi\vartheta} \int_\Theta g_{n-1}(v, \vartheta, u) g(u) du
$$

As g is arbitrary we have the statement of the theorem. ■

3.5. An Application of the Extended Model

A practical example of the model described above is presented in this section. Let take a BMS with "C" levels of premium where the first level presents the lowest premium and the last level corresponds to the highest possible premium. Let take $I_1 = [0, a)$, $I_2 = [a, b)$ and $I_3 = [b, +\infty)$ intervals for possible aggregate claim amounts, where $0 < a < b < +\infty$. We also assume that a policyholder can make 0, 1 or 2 claims per year (the claim numbers greater than 2 are considered as 2, so we write "2+", that is two or more claims). The transition rules of a policyholder between the premium levels are given by the Table3.1.

Number of claims	Aggregate Claim Interval	Level (+/-)
0		-1
1	I_1	+1
1	I_2	+3
1	I_3	+5

2+	I_1	+2
2+	I_2	+4
2+	I_3	+6

Table 3.1

As an example we will explain the 5-th row of the table as follows: if the policyholder makes two or more claims during the year and the total loss of the company from that claims belongs to the second interval (that is to interval $[a, b)$), then the policyholder will be penalized by 4 levels from his/her current level for the next year. It is obvious that in case of no claim the probability of having 0 amount of loss is 1.

If the policyholder is on the first level and makes no claim in the current year, then he/she stays at the same level on the next year. On the other side, if there is not sufficient number of levels to go up (for example if the policyholder is on the “C-1” level and makes “2+” claims, then he/she must be shifted by at least 2 levels up, but only 1 level remains to reach the maximum premium), then the policyholder moves to the highest level.

From Table 3.1 we can conclude that the number of BMS levels must be about or even more than 20. Otherwise, soon after the start (in few years) the majority of policyholders will be concentrated about the highest level of premium. Considering for example 20 levels of BMS and 50 policyholders, we get the cardinality $|S_X| = 20^{50} \approx 1.13 * 10^{65}$. It is obvious that it would be too hard to work with a vector of that size in practice.

When a policyholder changes his/her level of BMS, the process X_n changes its state, so instead of transition matrix of X_n it is enough to analyze the one step transition matrix of a policyholder between BMS levels which in the case of Table 3.1 is the following $C \times C$ matrix:

$$B = \begin{pmatrix} p_0 & p_0 & 0 & & & 0 & 0 \\ p_1q_1 & 0 & p_0 & & & 0 & 0 \\ p_2q_1 & p_1q_1 & 0 & & & 0 & 0 \\ p_1q_2 & p_2q_1 & p_1q_1 & & & 0 & 0 \\ p_2q_2 & p_1q_2 & p_2q_1 & & & 0 & 0 \\ p_1q_3 & p_2q_2 & p_1q_2 & & & 0 & 0 \\ p_2q_3 & p_1q_3 & p_2q_2 & \ddots & \ddots & 0 & 0 \\ 0 & p_2q_3 & p_1q_3 & & & 0 & 0 \\ 0 & 0 & p_2q_3 & & & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & & & p_0 & 0 \\ 0 & 0 & 0 & & & 0 & p_0 \\ 0 & 0 & 0 & & & 1-p_0 & 1-p_0 \end{pmatrix}$$

Note that the i-th column represents the probabilities of movements from the current level i to the level which corresponds to the number of row of the matrix. Here p_k, $k = 0,1,2$ is the probability of having k claims (in case of $k = 2$ it is the probability of having 2 or more claims) and q_m, $m = 1,2,3$ is the probability that the aggregate claim amount of a policyholder who made k claims belongs to the I_m interval.

In case of insurance event during the next year the policyholder will decide: “to claim” or “not to claim” and this will the “Hidden Process”.

To get a more flexible BMS the decision-makers have to think about the values of k, m and C.

CHAPTER 4

COMPARATIVE ANALYSIS OF THE BONUS-MALUS SYSTEMS

Introduction

Compulsory motor third party liability insurance (CMTPL) has been introduced in Armenian insurance market since 2011. Two years after that the Belgian BMS with 22 classes was applied to CMTPL insurance. BMS classes are described in Table 4.1. In 2013 only the bonus was used and then in 2014 the malus was added as well.

BMS Class	Premium as Percentage of Base Premium
Class 22	**200%**
Class 21	180%
Class 20	160%
Class 19	152%
Class 18	144%
Class 17	136%
Class 16	128%
Class 15	120%
Class 14	116%
Class 13	112%
Class 12	108%

BMS Class	Premium as Percentage of Base Premium
Class 11	104%
Class 10	**100%**
Class 9	97%
Class 8	94%
Class 7	91%
Class 6	88%
Class 5	85%
Class 4	82%
Class 3	75%
Class 2	65%
Class 1	**50%**

Table 4.1

Transition rules of BMS used in RA are as follows:

The use of Bonus-Policyholder gets bonus with 1 class if he had a CMTPL insurance policy for 345 days and had no claim during that period.

The use of Malus-Policyholder gets malus with 4 classes for each accident claimed to insurance company.

According to the official data the loss ratio of the CMTPL insurance in RA was 53% on December 31 of 2015, but it was fluctuated about 60%-75.4% during 2011-2014 years. More detailed information on CMTPL insurance and BMS in Armenia can be found on the official site of Armenian Motor Insurance Bureau [126].

In this chapter of the dissertation the analysis based on data provided by an insurance company holding 22% of CMTPL insurance policies (hereafter we will call it “Company”) is presented. The main purpose is the testing of Alternative and Extended BMSs’ proposed in this dissertation and their comparative analysis with BMS used in RA (current BMS) [127].

4.1. Data Description

Based on the legal restrictions, the company has provided depersonalized data with preservation of the personal data privacy law's statutes for this analysis.

Though, all companies in Armenia use the same BMS for CMTPL insurance policies, the movement percentage of policyholders between companies is high. The movement depends on the company's rating as well as on the personal experience and opinion of the policyholder. So, after the end of the contract a certain group of policyholders' change the company. This circumstance does not constitute an obstacle to the new company to determine the policyholder's BMS next class, as all data concerning to CMTPL insurance and BMS are coordinated in one database which is achievable for all companies eligible to sign a CMTPL insurance contract.

Not a long time has passed from BMS introduction, so having the current BMS class of a policyholder one can restore his claim history with high accuracy. For the claim amount of past years the average claim amount for the corresponding claim number was used.

The company provided CMTPL insurance database for 01.01.2011-31.12.2015 period. As a result of preliminary analysis, the data concerning to 2011 and 2012 was ruled out, because the roughly statistics are diverge from those of 2013-2015. That period coincides with the BMS introduction. So, the experience approves the change of policyholders' behavior after BMS introduction.

On the next stage of the analysis the information on policies signed shorter than one year was excluded. This is done due to regulator factors of premiums used for policies with short time (see [128]). This regulator factors follow some distortion of the average premium.

So, for the analysis 178,094 CMTPL insurance policy data was used. Some results have been gotten with differentiating the policy sign year and some other ones without differentiation.

Table 4.2 shows the percentage of analyzed data of the insurance market in accordance with the policy sign year.

Policy sign year	Percentage of analyzed data from insurance market
2013	14.4%
2014	15%
2015	10.8%
Percentage of all analyzed policies to all signed in the insurance market in 2015	38.9%

Table 4.2

4.2. Goodness of Fit Tests for Claims Number and Claim Amount Variables

To determine the BMS behavior and forecast premium amount to be collected, it is very important to model the number of claimed accidents to insurance company and the amount lost by the company due to a policyholder. Claims number and claim amount are random variables for insurance company. For modeling each of them some probability distributions are used. Those distributions are listed and characterized in [14], [15], [16]. To accept or reject any distribution one must state a statistical hypothesis. The statistical hypotheses were checked with Kolmogorov-Smirnov and χ^2 statistics, which description and applications are described in [129], [130], [131], [132]. Calculations of the statistics were made with the help of computer packages Easy-Fit, SPSS and MS Excel.

For the claims number variable the Poisson, Negative Binomial, Binomial, Geometric and Hypergeometric distributions were checked. All distributions were rejected with 10% of significance except of Negative Binomial one. It was accepted with the same level of confidence for 2013, 2014, 2015 and aggregate data.

Here the goodness of fit test for the claim number distribution is presented. Let denote ξ the random variable representing the claim number of a policyholder in a year. We are suggesting that its distribution belongs to the class of the second-type Negative Binomial distributions' (NB2). The probability mass function of that distribution is

$$p_{NB2}(x,k,p) = P(\xi = x) = \frac{\Gamma(k+x)}{\Gamma(x+1)\Gamma(k)} p^k (1-p)^x, \quad x = 0,1,2,\dots \tag{4.2.1}$$

where

$$\Gamma(y) = \int_0^\infty t^{y-1} e^{-t} dt, \quad y > 0$$

In this case the statistical hypothesis can be formulated as follows:

$$H:\ \xi \sim F_\Theta(x) \tag{4.2.2}$$

where $\Theta = (k, p)$ is the group of parameters of the distribution (4.1).

In the case of the Negative Binomial Type 2 distribution it is expedient to estimate model parameters with the method of moments. The expectation and the variation of the tested distribution are:

$$E(X) = \frac{k(1-p)}{p}; \quad Var(X) = \frac{k(1-p)}{p^2}$$

According to the method of moments, the following system of equations must be solved to find the estimates $\hat{k}$ and $\hat{p}$:

$$\begin{cases} \dfrac{k(1-p)}{p} = \bar{x} \\ \dfrac{k(1-p)}{p^2} = s^2 \end{cases} \tag{4.2.3}$$

here $E(\xi)$ and $Var(\xi)$ are calculated from the sample data.

The solution of (4.2.3) is:

$$\hat{k} = \frac{\bar{x}^2}{s^2 - \bar{x}}; \quad \hat{p} = \frac{\bar{x}}{s^2} \tag{4.2.4}$$

For the discrete distributions the Pearson's χ^2 goodness measure is usable (see for instance [130], pg. 301-302). The method is based on the Pearson-Fisher's theorem, which can be formulated for the discrete random variable by the following way: if the hypothesis (4.2.2) is true, then the statistics given by (4.2.5) formula converges to the distribution $\chi^2(s-k-1)$, when $n \to \infty$.

$$\chi^2_{stat} = \sum_{j=1}^{s} \frac{\left(v_j - np_j(\tilde{\Theta})\right)^2}{np_j(\tilde{\Theta})} \tag{4.2.5}$$

where s is the number of the possible x_j values of the random variable ξ, v_j is the frequencies of the values x_j in the sample, $(s-k-1)$ is the degrees of freedom for the χ^2 distribution and $p_j(\tilde{\Theta}) = P_{\tilde{\Theta}}(\xi = x_j)$, where $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$.

The hypothesis (4.2.2) must be rejected with the significance level ε, if $\chi^2_{stat} > \chi^2_{\varepsilon}(s-k-1)$, where $\chi^2_{\varepsilon}(s-k-1)$ is the corresponding percentile point of the distribution χ^2 with $(s-k-1)$ degrees of freedom. Otherwise, the hypothesis (4.2.2) is not rejected.

For the goodness of fit test we use all statistical data received for the period 2013-2015 years. There were data on about 430,000 policies, which is large enough to say that n converges to infinity and the described procedure for goodness of fit test is applicable. The possible values of random variable ξ and their frequencies are presented in the Table 4.3

x_j	v_j
0	407145
1	21168
2	1578
3	200

Table 4.3

Here $\bar{x} = \frac{\sum_{j=0}^{s} x_j v_j}{\sum_{j=0}^{s} v_j} = 0.08534$ and $s^2 = \frac{\sum_{j=0}^{s} x_j^2 v_j}{\sum_{j=0}^{s} v_j} - \bar{x}^2 = 0.09589$. By inputting these values in (4.2.4) we get:

$$\hat{k} = 0.504222; \ \hat{p} = 0.896787 \tag{4.2.6}$$

Using Table 4.3 as well as formulas (4.2.1) and (4.2.4) – (4.2.6) for the Pearson's statistics we get the value $\chi^2_{stat} = 3.26$. For the given data in Table 4.3, $s = 4$ and the number of estimated parameters is $k = 2$. For the statistics (4.2.5) calculated by data given in

Table 4.3, in the case of $\varepsilon = 5\%$ the percentile point $\chi^2_{0.05}(1) = 3.841$ must be considered. We have $\chi^2_{stat} = 3.26 < 3.841 = \chi^2_{0.05}(1)$, so the hypothesis $(4.2.2)$ does not rejected.

The results of goodness of fit tests for years 2013, 2014, 2015 and aggregate data with model parameters are presented in Table 4.4:

Year	Hypothesis	χ^2 statistic $(\chi^2_{0.05;1} \approx 3.84)$	Number of policies
2013	NegBinom(0.895;0.987)	0.27	58821
2014	NegBinom(0.897;0.809)	3.8	69725
2015	NegBinom(0.933;0.543)	0.32	49548
All	NegBinom(0.897;0.504)	3.26	430091

Table 4.4

For the aggregate claim amount variable 42 different types of distributions were tested. Some of them were rejected with the 10% of significance. The distributions accepted with 90% of confidence are presented in Table 4.5. Some of the continuous probability distributions checked by goodness of fit test are presented in Appendix (Table A.2). Here we present the goodness of fit testing for the Log-Pearson Type 3 distribution function.

Let's denote η the random variable representing the aggregate claim amount given to a policyholder in a year by the insurance company. It will be a sum of random variables with the random count:

$$\eta = \sum_{i=1}^{\xi} X_i$$

where X_i is the amount of the policyholder's i-th claim in a year covered by company and ξ is the number of claims of the same policyholder in the same year. We are suggesting that η comes from the Log-Pearson Type 3 distribution (LP3), that is its probability density function is:

$$f(x) = \frac{1}{x|\beta|\Gamma(\alpha)}\left(\frac{\ln(x) - \gamma}{\beta}\right)^{\alpha-1} \exp\left(-\frac{\ln(x) - \gamma}{\beta}\right) \tag{4.2.7}$$

Here the statistical hypothesis is formulated as follows:

$$H:\ \eta \sim F_{LP3}(x;\Theta) \tag{4.2.8}$$

where $\Theta = (\alpha, \beta, \gamma)$ is the group of parameters of the distribution (4.2.7).

Calculation processes for the continuous distributions are much complicated and time-consuming, so the statistical package EasyFit has been used as a helping hand.

There are a number of well-known methods which can be used to estimate distribution parameters based on available sample data. For every supported distribution, EasyFit implements one of the following parameter estimation methods:

- method of moments (MOM)
- maximum likelihood estimates (MLE)
- least squares estimates (LSE)
- method of L-moments

EasyFit uses the least computationally intensive methods. Thus, it employs the method of moments for those distributions whose moment estimates are available for all possible parameter values, and do not involve the use of iterative numerical methods. For many distributions, EasyFit uses the MLE method involving the maximization of the log-likelihood function. For some distributions, such as the 2-parameter Exponential and the 2-parameter Weibull, a closed form solution of this problem exists. For other distributions, EasyFit implements the numerical method for multi-dimensional function minimization. Given the initial parameter estimates vector, this method tries to improve it on each step of subsequent iteration. The algorithm terminates when the stopping criteria is satisfied (the specified accuracy of the estimation is reached, or the number of iterations reaches the specified maximum). More information on this package and its possibilities can be found in [133].

#	Distributions	2013	2014	2015	All
1	Burr	k=0.11811 α=0.79622 β=1.4161 γ=1.0000E+5	N/A	N/A	N/A
2	Burr (4P)	N/A	k=0.11811 α=0.79622 β=1.4161 γ=1.0000E+5	k=0.11811 α=0.79622 β=1.4161 γ=1.0000E+5	k=0.11811 α=0.79622 β=1.4161 γ=1.0000E+5
3	Fatigue Life	α=0.78958 β=2.0154E+5	α=0.8066 β=2.0489E+5	α=0.74986 β=1.9550E+5	α=0.79033 β=2.0198E+5
4	Frechet	α=1.7745 β=1.2302E+5	α=1.7194 β=1.1859E+5	α=1.9671 β=1.2941E+5	N/A
5	Gamma (3P)	α=0.35406 β=4.4379E+5 γ=1.0000E+5	N/A	N/A	N/A
6	Gen. Gamma	N/A	N/A	k=1.2386 α=0.94755 β=3.1252E+5	N/A
7	Log-Gamma	α=282.96 β=0.04285	α=276.58 β=0.04388	α=304.21 β=0.03982	α=283.01 β=0.04286
8	Lognormal	σ=0.72086 μ=12.126	σ=0.72972 μ=12.136	σ=0.69454 μ=12.114	σ=0.72097 μ=12.129
9	Lognormal (3P)	N/A	σ=10.095 μ=7.7681 γ=1.0000E+5	N/A	N/A
10	Log-Pearson 3	α=2.5756 β=0.44918 γ=10.969	α=2.561 β=0.45598 γ=10.968	α=3.2099 β=0.38766 γ=10.87	α=2.6656 β=0.44159 γ=10.952
11	Pearson 5	α=2.7055 β=4.1055E+5	α=2.6462 β=4.0357E+5	N/A	N/A
12	Pearson 6	α_1=3755.1 α_2=2.6988 β=109.03	α_1=1281.3 α_2=2.6473 β=315.26	α_1=1370.0 α_2=2.8299 β=312.17	N/A
13	Weibull	α=1.1142 β=2.7455E+5	α=1.0796 β=2.7821E+5	α=1.1854 β=2.6639E+5	N/A
14	Weibull (3P)	α=0.52299 β=1.0842E+5 γ=1.0000E+5	N/A	N/A	N/A

Table 4.5

So, for the 2013-2015 aggregate data we get (4.2.9) estimates for the Log-Pearson 3 distribution, using EasyFit.

$$\hat{\alpha} = 2.6656 \ \ \hat{\beta} = 0.44159 \ \ \hat{\gamma} = 10.952 \tag{4.2.9}$$

The hypothesis on continuous distributions like (4.2.8) can be checked with the help of Kolmogorov-Smirnov's non-parametric test. This test is based on the following statistics:

$$D = \max_{1 \le i \le n} \left(F_{LP3}(x_i) - \frac{i-1}{n}, \frac{i}{n} - F_{LP3}(x_i) \right) \tag{4.2.10}$$

where n is the volume of the sample.

The hypothesis is rejected with the confidence level of $1 - \varepsilon$, if the calculated value of the statistics (4.2.10) is greater than the table-value of the corresponding level of confidence for the statistics D.

EasyFit calculates the value of statistics (4.2.10) and gets $D = 0.26213$. The table-value for this statistics is $D_{0.05} = 0.2749$. EasyFit gives the corresponding P-Value as well, which in this case take the value 0.0697. This means that the hypothesis is not rejected on the levels smaller than P-Value.

4.3. Comparative Analysis of the Alternative and Current BMS

Here we present a comparative analysis of the BMS currently working in RA and the Alternative BMS presented by (2.2.1) model in Chapter 2 of this dissertation. The comparison was made on the example of data described in section 4.1.

We can use the quantile method and find values of α and β with the help of distribution of the aggregate claim amount. For different distributions we find different values. Thus, for example in the case of Log-Pearson 3 distribution we get $\alpha = 0.47\%$ and $\beta = 0.89\%$ but in the case of Lognormal distribution they are $\alpha = 0.62\%$ and $\beta = 1.18\%$. As mentioned in *Remark 2.1*, the ratio of bonus and malus factors can be interpreted as loss ratio of the system, so we have loss ratio 52. 68%, this is about the value of the 2015 year. But the loss ratio was fluctuated between 60%-75.4% during 2011-2014 years, so it would be better to state the values of α and β by fixing some medium level of loss ratio. We can use only Lemma2.1 and fix loss ratio for example at the level 66.67% with $\alpha = 1\%$ and according to formula $(2.2.2)$ for malus factor get the value $\beta = 1.5\%$. Although the model parameters are not invariant to distributions, it gives opportunity to developers constructing flexible systems by fixing the loss ratio beforehand.

In Table 4.6 the next premium and coefficient of bonus-malus for a policyholder with base premium of 60,000 AMD in case of different aggregate claim amounts are presented. It is clear from Table 4.6 that in case of small aggregate claims (up to 50,000 AMD) a policyholder has no malus; moreover, in some cases he may have some small bonus.

In the current BMS a policyholder pays 116% of the base premium if he had one claim during the previous year. In case of the Alternative BMS the policyholder pays the same 116% of the base premium if he has about 700,000 AMD aggregate claim for the previous year. Analyzed data shows that only 1.5% of aggregate claims are about 700,000 AMD, but the policyholders with one claim are 91%. So, the policyholders with aggregate claim amount lower than 700,000 AMD will prefer the Alternative BMS.

A policyholder with two claims will pay 144% of the base premium for next year in the current BMS. In the analyzed data they form 8% of all policyholders.

In the Alternative BMS the policyholder is “punished” with the same amount if he had aggregate claim with 1,800,000 AMD. The policyholders with that risk are only 0.68%. In addition to this, note that the policyholders with aggregate claim amount lower than 1,800,000 AMD, which are about 99,5% of all policyholders, will pay less than 144% of the base premium in the Alternative BMS. So, they will prefer the Alternative BMS as well.

If the policyholder makes 3 and more claims in the current BMS, his next premium forms 200% of the base premium. The same amount pays a policyholder in the Alternative BMS in case of a 4,000,000 AMD aggregate claim.

Base Premium	α	β
60,000	0.01	0.015
Aggregate Claim Amount for Current year	Next year premium	Bonus-Malus
-	59,400	99%
10,000	59,550	99%
20,000	59,700	100%
30,000	59,850	100%
40,000	60,000	100%
50,000	60,150	100%
60,000	60,300	101%
700,000	**69,900**	**117%**
1,800,000	**86,400**	**144%**
4,000,000	**119,400**	**199%**
10,000,000	209,400	349%
11,800,000	236,400	394%

Table 4.6

It should be noted that Alternative BMS is not preferred for the policyholders having one claim with more than 700,000 AMD claim amount and for those having two claims with more than 1,800,000 AMD aggregate claim amount. These two groups of policyholders form only 4.7%.

From Table 4.6 it can be concluded that the Alternative BMS had very hard terms for the policyholders with more than 4,000,000 AMD aggregate claim amount, but they are 0.04% of all.

It would be very interesting to compare some extreme cases presented in Table 4.7, where the aggregate claim amounts, provided maluses and base premiums for some policyholders are included.

Case	Aggregate Claim amount (1 claim case)	Penalty by current BMS	Penalty by Alternative BMS	Base Premium
1	6,448,940	5,740	96,332	41,000
2	15,000	5,740	-177	41,000
3	5,352,000	20,720	79,521	74,000
4	300,000	20,720	4,081	37,000
5	45,000	21,560	-114	77,000
6	42,000	26,880	102	48,000
	Aggregate Claim amount (3 and more claims case)			
7	3,045,200	48,000	45,198	48,000
8	1,537,000	48,000	22,575	48,000
9	621,000	48,000	8,835	48,000

Table 4.7

Let compare the first and second cases, where the policyholders are in the same risk group (they pay the same base premium). They both make one claim during the year, but the claim amount of the first one is more than 6,000,000 AMD and the other's is only 15,000 AMD. According to the current BMS they will pay with 5,740 AMD more premium for next year. In reality it does not correspond to the loss incurred by them to the company. The Alternative BMS suggests taking with 96,332 AMD more his current premium from the first policyholder and from the other one not to take additional premium (moreover, give him some small bonus). The same analysis can be done for other policyholders presented in Table 4.7.

The next table shows the comparative analysis of the collected premiums according to current BMS and Alternative BMS suggestions. Presented data include all policyholders with one year insurance policy disregarding the fact of claiming. It is obvious from Table 4.8 that the aggregate premium collected by the insurance company according to the current BMS is less than the Base premium with 1.2%-2.3%. This fact means that the introduction of BMS leads to some additional losses to the company. The results got from the Alternative BMS are better, than from the current BMS as the gap between next premium and base premium is less than 0.54%. This fact comes to confirm that the Alternative BMS is constructed according to the “optimal” BMS statements.

So, we can conclude that for the most part of the policyholders the Alternative BMS is preferable than the current BMS. In addition to that, the Alternative BMS is not incurring additional losses to the company and the company can state its loss ratio at the start of the insurance year.

	2013		2014		2015	
	Current BMS	Alternative BMS	Current BMS	Alternative BMS	Current BMS	Alternative BMS
Year premium/Base premium	98.81%	99.68%	98.79%	99.73%	100.04%	100.03%
Next premium/Base premium	98.22%	99.56%	97.68%	99.46%	97.87%	99.33%
Year premium/ Previous year premium	99.41%	99.88%	98.88%	99.73%	97.83%	99.30%

Table 4.8

4.4. A Comparative Analysis of the Extended and Current BMS

The BMS presented in the Chapter 3 of this dissertation differs from the current one by the additional term, which is the aggregate claim amount. So, the next level of BMS for a policyholder is determined by the number of claims and aggregate claim amount incurred to the company during the current year. To discuss a more flexible system we take transition rules between BMS classes presented in the Table 4.9, where BMS classes are taken from current BMS presented in Table 4.1.

Number of claims	Aggregate Claim Interval	Level (+/-)
0		-1
1	(0-100,000]	+1
1	(100,000-500,000]	**+4**
1	(500,000-1,500,000]	+7
1	(1,500,000+)	+10
2	(0-100,000]	+2
2	(100,000-500,000]	+5
2	(500,000-1,500,000]	**+8**
2	(1,500,000+)	+11
3+	(0-100,000]	+3
3+	(100,000-500,000]	+6
3+	(500,000-1,500,000]	+9
3+	(1,500,000+)	**+12**

Table 4.9

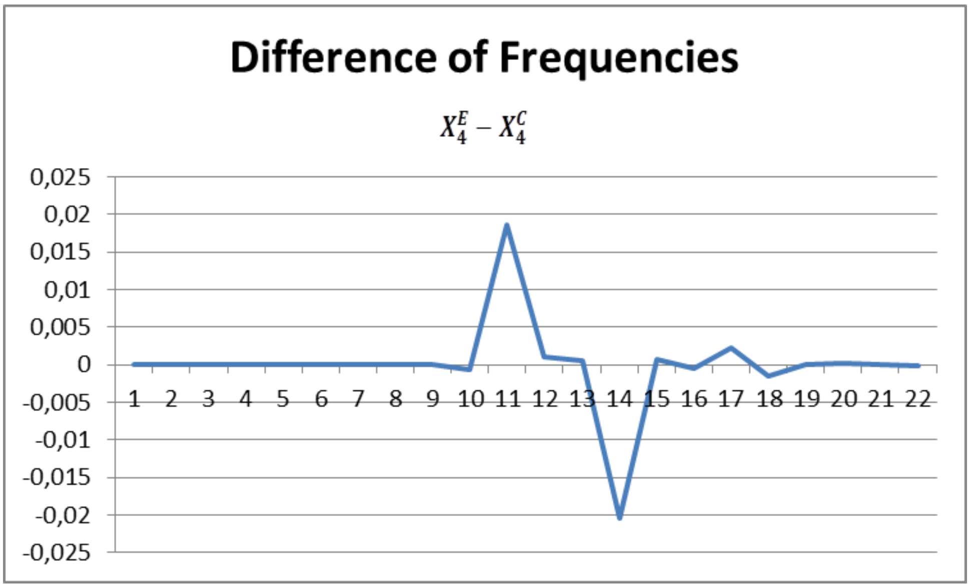
Bounds 100,000; 500,000 and 1,500,000 for aggregate claim amounts are taken from the fact that they are breakage points for the number of policyholders having aggregate claim amount under and above the mentioned points.

Claim history data for 2013, 2014 and 2015 as well as for the whole period shows that the policyholders with aggregate claim amount less than 100,000 AMD are 46% of those making claims. The policyholders with aggregate claim amount in the range (100,000AMD-500,000AMD] are form 45%, in the range (500,000AMD-1,500,000AMD] are 7% and more than 1,500,000AMD are 2%.

Considering the 22 levels of BMS mentioned in table 4.1, 4 intervals for aggregate claims and 4 possibility of claims numbers mentioned in table 4.9 and taking for example 50 policyholders, we get the cardinality of the processes X, H, N of the Chapter 3 equals: $|S_X| = 22^{50} \approx 1.32 * 10^{67}$, $|S_H| = |S_N| = 4^{50} \approx 1.27 * 10^{30}$. It is obvious that it would be too hard to work with matrixes of these sizes in practice. When a policyholder changes his/her level of BMS, the processes X, H and N change their state, so instead of transition matrixes B, Q and A it is enough to analyze the one step transition matrix of policyholders between BMS classes, groups of aggregate claims and groups of claims number. In addition to this we do not have the same group of policyholders each year in practice, so it would be convenient to consider the proportions of the policyholders among BMS classes instead of numbers of policyholders. At the start of BMS all policyholders were at the base level which is the 10th level. So, the initial value of the process X is: $X_0^T = (0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0)$. Aggregate claim intervals are (0-100,000]; (100,000-500,000]; (500,000-1,500,000] and (1,500,000+) for which we have $H_0 = (1,0,0,0,)$. Claims number can take the values 0,1,2, and 3+, so $N_0 = (1,0,0,0,)$. By this approach matrix B will be a 22×352 matrix instead of $1.32 * 10^{67} \times 2.13 * 10^{127}$. For 2016 we have $H_4 = (0.981,0.016,0.003,0.0004)$ and $N_4 = (0.963,0.035,0.002,0.0001)$ by statistical data. The frequencies of the policyholders by BMS classes take the following values for the current and Extended BMSs accordingly

X_4^C	X_4^E
0	0
0	0
0	0
0	0
0	0
0.006	0.006
0.005	0.005
0.005	0.005
0.938	0.938
7E-04	2E-05
0.004	0.022
0.002	0.003
0.001	0.002
0.035	0.014
6E-04	0.001
6E-04	2E-04
3E-04	0.003
0.002	3E-04
8E-05	6E-05
2E-04	3E-04
1E-04	1E-04
2E-04	4E-05

Graph 4.1 shows the smoothness of the Extended BMS in comparison to the current one. The detailed analysis of the models is presented below.



Graph 4.1

In the case of current BMS a policyholder is “punished” with 4 classes of BMS for one claim in the year, but in the Extended BMS he can be “punished” up to 10 classes for one claim, if the claim amount is more than 1,500,000AMD and in contrast to this, he is “punished” with one class if the claim amount is small (less than 100,000AMD).

Table 4.10 helps to identify the percentage of policyholders to which the Extended BMS has harder or smoothed terms compared with the current BMS.

Number of Claims	Aggregate Claim Amount (AMD)			
	(0-100000]	(100000-500000]	(500000-1500000]	1500000+
1	49%	**43%**	7%	1%
2	14%	69%	**14%**	3%
3+	2%	66%	28%	**4%**

Table 4.10

Focusing our attention on transition rules of current BMS and on tables 4.9 and 4.10, we conclude that: for the policyholders making one claim the terms of Extended BMS are smoother for 49%, are the same as current BMS for 43% and are harder for 8%. For the policyholders with two claims the Extended BMS is smoother for 83%, and the current BMS for 3%.

From the last row of the Table 4.10 it is obvious that the Extended BMS is the same for only 4% of the policyholders with three and more claims, but for the others it is smoother than the current BMS.

So, enlarging this analysis to the group of all policyholders with any claim we conclude that two models have the same malus for 39.5% of them. The Extended BMS has harder terms than the current one for 7.5% of policyholders having any claim. On the other hand, the current BMS is harder than suggested Extended BMS for 53% of claimed policyholders.

Although, Extended BMS is smoother for most of policyholders, forecast amount of premiums to be collected for 2016 is 100.11% of the base premium in opposite to the current BMS's 97.87%.

The Table 4.11 includes one step transition matrix for current BMS according to the distribution $\mathrm{NegBinom}(0.897;\ 0.504)$ mentioned in Table 4.4.

For the construction of one step transition matrix of Extended BMS the same $\mathrm{NegBinom}\ (0.897;\ 0.504)$ distribution for claims number was used and for the aggregate claim amount the $\mathrm{Log-Pearson3}\ (\alpha = 2.6656\ \ \beta = 0.44159\ \ \gamma = 10.952)$ distribution have been chosen. The transition matrix of the Extended BMS is presented in Table 4.12. Although, all distributions in Table 4.4 are accepted with the same level of accuracy, visual analysis for distribution Log-Pearson 3 gives better results, i.e. the distribution, density, P-P and Q-Q plots lay on according sample data more accurately than for other distributions.

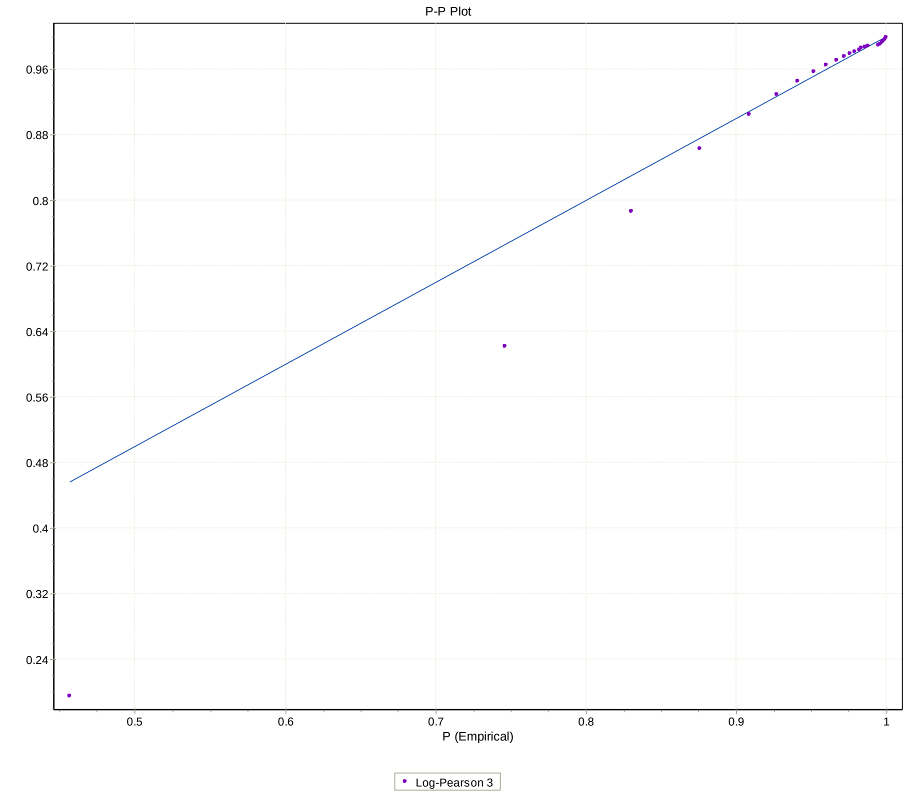
Transition Matrix of Current BMS																						
Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0.923	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-	-	-	-	-	-	-	-
2	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-	-	-	-	-	-	-
3	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-	-	-	-	-	-
4	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-	-	-	-	-
5	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-	-	-	-
6	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-	-	-
7	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-	-
8	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-	-
9	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001	-
10	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	-	0.001
11	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	-	0.001
12	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	-	0.001
13	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007	0.001
14	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	-	0.007
15	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	-	0.007
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	-	0.007
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.070	0.007
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	-	0.077
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	-	0.077
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	-	0.077
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.077
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	0.077

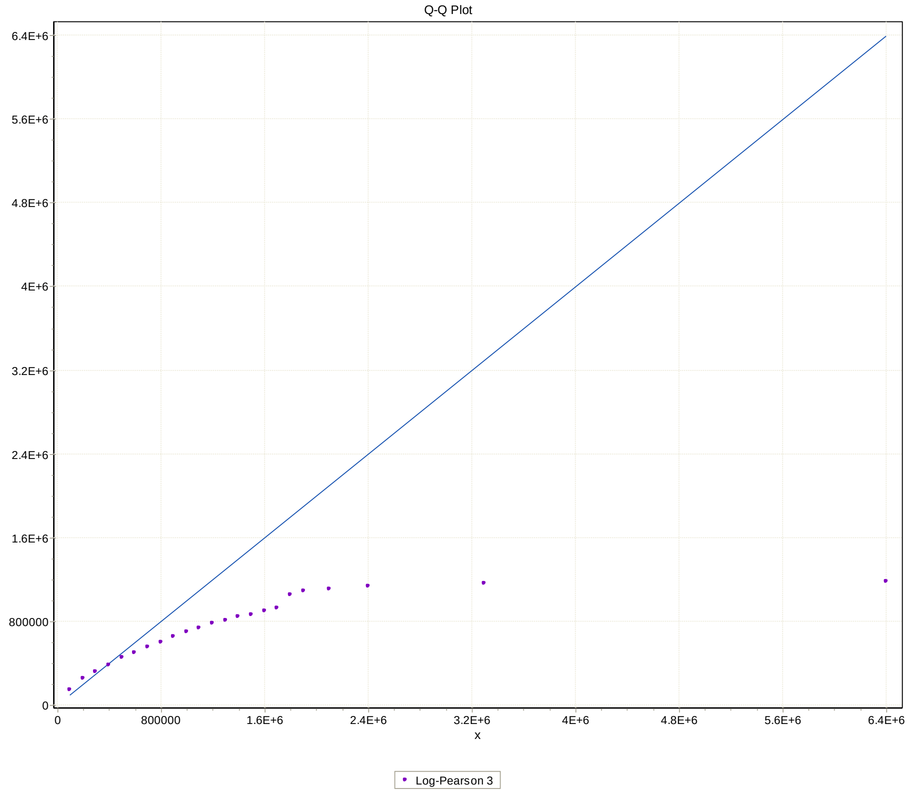
Table 4.11

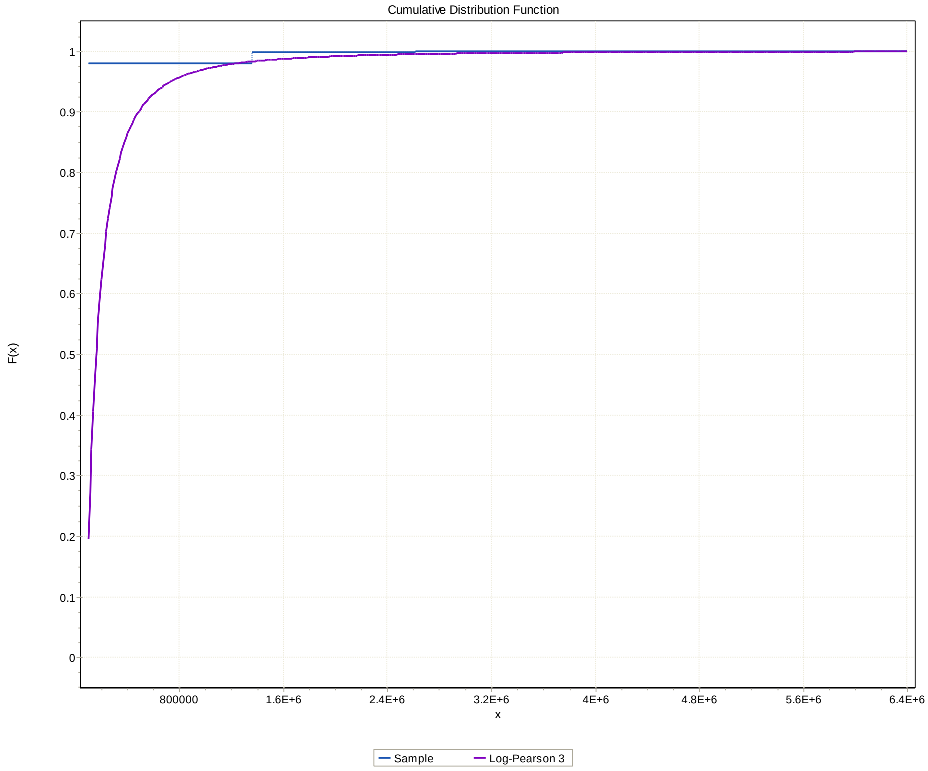
Transition Matrix of Extended BMS																						
Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0.923	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-	-	-	-	-	-	-	-
2	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-	-	-	-	-	-	-
3	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-	-	-	-	-	-
4	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-	-	-	-	-
5	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-	-	-	-
6	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-	-	-
7	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-	-
8	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-	-
9	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05	-
10	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	9E-05	1E-05
11	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03	1E-04
12	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	6E-05	1E-03
13	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	5E-04	1E-03
14	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.006	2E-03
15	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	5E-04	0.007
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.005	0.008
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.050	0.013
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	1E-04	0.062
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.001	0.062
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.014	0.064
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	-	0.077
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.923	0.077

Table 4.12

Here the above mentioned plots are presented:







Cumulative Distribution Function
F(x)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
800000
1.6E+6
2.4E+6
3.2E+6
4E+6
4.8E+6
5.6E+6
6.4E+6
x
Sample
Log-Pearson 3

Conclusion

In the first chapter of the dissertation a literature review of current BMS is presented. Analyzing the distributions applied for the claim number processes it can be concluded that claim history for each portfolio is unique and there is not a universal probability distribution that fits to all claim number processes good. Some problem-solving methods and approaches have been discussed to eliminate BMS's current issues which are consequences of the transition rules between BMS classes including only claims number component. So, it is suggested to construct new BMS's, which will include the claim severity as a premium calculation component.

In the second chapter a new BMS model is presented, which was constructed on the basis of "optimal" BMS principle. Presented model eliminates the discussed disadvantages of many BMS's. In the scope of the model

- ✓ The policyholder has an opportunity to pay the same premium and even have a small bonus in the case of small size of claim.
- ✓ Each policyholder will be penalized proportionally to his claim.
- ✓ Policyholders will cover only a part of the claim occurred, so there will be no bonus hunger behavior.
- ✓ Policyholders will report about each claim so the insurers and regulators will have an opportunity for the best evaluation of the system (risk frequency, claim amount, distribution of future losses, premium for new drivers etc.).

The Extended BMS model, presented in the third chapter of this dissertation, can be a tool for constructing more flexible systems to eliminate some of the disadvantages of the current ones. The model development is based on the Markov property. With the help of measure change techniques and HMMs, the recursive formulas for transition matrices and m step predictions for claim numbers and aggregate claim amounts were derived.

The last chapter was devoted to the testing of new models presented in this dissertation by a comparative analysis with the BMS used in Armenian CMTPL insurance. The analysis showed that the Alternative BMS is preferable for most of policyholders as well as for the insurers. It can be noted that in the case of Extended BMS there is no concentration of policyholders in some classes. Starting from the first insurance year policyholders are spread out on the most of BMS classes. This means that the risks of the system are distributed among policyholders accordingly to their own part of risk. And due to this property, the Extended BMS being harder only for a small part of the policyholders (7.5%) provides good results in comparison with the current system.

Abbreviations

ASTIN- **A**ctuarial **St**udies **i**n **N**on-life Insurance

BMS-**B**onus-**M**alus **S**ystem

CASCO- **Cas**ualty and **Co**llision (voluntary vehicle insurance)

CMPTL-**C**ompulsory **M**otor **T**hird **P**arty **L**iability

EM-**E**xpectation **M**aximization

HMM-**H**idden **M**arkov **M**odel

m.g.f. – **M**oment **G**enerating **F**unction

MTPL-**M**otor **T**hird **P**arty **L**iability

NPMLE- **N**on-**P**arametric **M**aximum **L**ikelihood **E**stimation

TPL-**T**hird **P**arty **L**iability

Bibliography

[1]. **Bischel, F.** "Erfahrung-Tarifieung in der Motorfahrzeug-haftplichtversiherung". *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.* 1964, pp. 119-129.

[2]. **Bühlmann, H.** "Optimale Pramienstufensysteme". *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.* 1964, pp. 193-213.

[3]. **Derron, M.** "A theoretical study of the no-claim bonus problem". *ASTIN Bulletin.* 1965, 3, pp. 62-74.

[4]. **Gurtler, M.** "Das subjective Risico in der Motorfahrzeugversicherung". *Zeitschrift fur die Gesamte Versicherungs-Wissenschaft.* 1960, 49, pp. 203-223.

[5]. **Gurtler., M.** ″Der Bonus als Mittel zur Erfassung des subjectiven Risicos″. *Zeitschrift fur die Gesamte Versicherungs-Wissenschaft.* 1961, 1, pp. 1-47.

[6]. **Gurtler, M.** "Der optimale Bonus". *Zeitschrift fur die Gesamte Versicherungs-Wissenschaft.* 1962, 3, pp. 4-51.

[7]. **Welten, C.P.** "The unearned no claim bonus". *The ASTIN Bulletin.* 1969, 5, pp. 25-32.

[8]. **Brouhns, N., Guillen,M., Denuit, M. and Pinquet,J.** "Bonus-malus scales in segmented tariffs with stochastic migration between segments". *Journal of Risk and Insurance.* 2003, Vol. 70, 4, pp. 577-599.

[9]. **Denuit, M. and Dhaene,J.** "Bonus-Malus Scales Using Exponential Loss Functions". August 2000, Belgium.

[10]. **Lemaire, J.** *"Bonus-Malus Systems in automobile insurance".* Boston : Kluwer Academic Publisher, 1995.

[11]. **Dionne, G. and Vanasse, C.** "A Generalization of Actuarial Automobile Insurance Rating Models: the Negative Binomial Diistribution with a Regression Component" . *ASTIN Bulletin.* 1989, 19, pp. 199-212.

[12]. **Teodorescu, S.** "The Bonus-Malus System Modelling Using The Transition Matrix". *Challenges Of The Knowledge Society. Economics.* 2012, pp. 1502-1507.

[13]. **Aggoun, L. and Benkherouf, L.** "An Insurance Model with Bonus-Malus System". *International Mathematical Forum, 1.* 2006, 26, pp. 1255 – 1272.

[14]. **Panjer, H.H., and Willmot, G.E.** *"Insurance Risk Models".* Schaunburg III. s.l. : Society or Actuaries, Schaunburg III, 1992.

[15]. **Mikosch, T.** *"Non-life Insurance Mathematics".* s.l. : Springer, 2009.

[16]. **Denuit, M., Marechal, X., Pitrebois, S., Walhin, J.F.** *"Actuarial Modelling of Claim Counts".* s.l. : John Wiley & Sons, 2007.

[17]. **Asmussen, S.** "Modeling and Performance of Bonus-Malus Systems: Stationarity versus Age-Correction". *Risks.* 2014, Vol. 2, pp. 49-73.

[18]. **Gosslaux, A. M., and Lemaire, J.** “Methodes d'ajustement de distributions de sinistres”. *MVSV.* 1981, pp. 87-95.

[19]. **Kestmont, R.-M., and Paris,J.** “Sur l'ajustement du nombre de sinistres”. *Bulletin of the Swiss Association of the Actuaries.* 1985, pp. 157-164.

[20]. **Willmot, G.E.** "The Poisson-lnverse Gaussian distribution as an alternative to the Negative Binomial". *Scandinavian Actuarial Journal.* 1987, pp. 113-127.

[21]. **Tremblay, L.** ″Using the Poisson-Inverse Gaussian Distribution in Bonus-Malus Systems″. *ASTIN Bulletin.* 1992, Vol. 22, 1, pp. 97–106.

[22]. **Besson, J.L., and Partrat, C.** "Loi de Poisson inverse gaussienne et systèmes de bonus-malus". *Proceedings of the Astin Colloquium.* 1990, Montreux 81, pp. 418–419.

[23]. **Ruohonen, M.** "On a model for the claim number process". *ASTIN Bulletin.* 1987, 18, pp. 57–68.

[24]. **Panjer, H.H.** " Models of claim frequency". In: I.B. Mac Neill and G.J. Umphrey Editors, Advances in the Statistical Sciences, 1987, Vol. VI, Actuarial Sciences, The University of Western Ontario Series in Philosophy of Science, Reidel, Dordrecht 39, pp. 115–122.

[25]. **Consul, P.C.** ″A model for distributions of injuries in auto-accidents″ . *Bulletin of the SWISS Association of Actuaries.* 1990, pp. 161–168.

[26]. **Elvers, E.** ″A note on the Generalized Poisson distribution″. *ASTIN Bulletin.* 1991, 21, p. 167.

[27]. **Ter Berg, P.** ″A loglinear Lagrangian Poisson Model″. *ASTIN Bulletin.* 1996, 26, pp. 123-129.

[28]. **Denuit, M.** ″A new distribution of Poisson-type for the number of claims″. *ASTIN Bulletin.* 1997, 27, pp. 229–242.

[29]. **Lefèvre, Cl., & Picard, Ph.** ″On the first crossing of a Poisson process in a lower boundary″. In: C.C. Heyde, Yu V. Prohorov, R. Pyke and S.T. Rachev Editors, Athens Conference on Applied Probability and Time Series, 1996, Vol. I, Applied Probability. Lecture Notes in Statistics, Springer, Berlin, 114, pp. 159-175.

[30]. **Albrecht, P.** “On Some Statistical methods Connected with the Mixed Poisson Process”. *Scandinavian Actuarial Journal.* 1982, pp. 1-14.

[31]. **Lemaire, J.** *″Automobile Insurance: Actuarial Models".* Netherlands : Kluwer-Nijhoff, 1985.

[32]. **Petauton, P.** *″Theorie de l’Assurance Dommages".* Paris : Dunod, 2000.

[33]. **Reinhart, J.M.** *″Assurance Non-Vie" (course notes).* Belgium : Universit´e Libre de Bruxelles, 2005.

[34]. **Lazar, D., Lazar, I., Parpucea, I.** "Theoretical And Practical Considerations Regarding Bonus-Malus System". *ACTA Universitatis Apulensis.* 2007, 13, pp. 57-66.

[35]. **Walhin, J. F. and Paris, J.** “Using Mixed Poisson Processes in Connection With Bonus-Malus Systems”. *ASTIN BULLETIN.* 1999, Vol. 29, 1, pp. 81-99.

[36]. **Simar, L.** “Maximum Likelihood Estimation of a Compound Poisson Process”. *The Annals of Statistics.* 1976, Vol. 4, pp. 1200-1209.

[37]. **Denuit, M. and Lambert P.** “Smoothed NPML Estimation of the Risk Distribution Underlying Bonus-Malus Systems”. *Proceedings.* November 2001, Vol. LXXXVIII, 169, Part 2, pp. 142-174.

[38]. **Carriere, J.** “Nonparametric Tests for Mixed Poisson Distributions”. *Insurance: Mathematics and Economics.* 1993, 13, pp. 75-81.

[39]. **Dufresne, F.** ″Distribution stationnaire d’un systeme bonus-malus et probabilite de ruine″. *ASTIN Bulletin.* 1988, 18, pp. 31-46.

[40]. **Dufresne., F.** ″The efficiency of the Swiss Bonus-Malus system″. *Bulletin of the Swiss Association of Actuaries.* 1995, 1, pp. 29-41.

[41]. **Norberg, R.** ″A credibility theory for automobile bonus system″. *Scandinavian Actuarial Journal.* 1976, pp. 92-107.

[42]. **Borgan, Ø., Hoem, J.M. and Norberg, R.** "A nonasymptotic criterion for the evaluation of automobile bonus systems". *Scandinavian Actuarial Journal.* 1981, pp. 165–178.

[43]. **Bühlmann, H.** "Mathematical Methods in Risk Theory". *Springer-Verlag.* 1970.

[44]. **Bühlmann, H. and Gisler, A.** "A Course in Credibility Theory and its Applications". *Springer-Verlag.* 2005.

[45]. **Niemiec, M.** "Bonus-Malus Systems as Markov Set-Chains". *ASTIN Bulletin.* 2007, Vol. 37, 1, pp. 53-65.

[46]. **Hartfiel, D.J.** "Markov Set-Chains". *Springer, New York.* 1998.

[47]. **Lemaire, J.** "Construction of the new Belgian motor third party tariff structure". *ASTIN Bulletin.* 1988, Vol. 18, 1, pp. 99–112.

[48]. **Centeno, M. and Silva, J.M.A.** ″Optimal Bonus Scales Under Path-Dependent Bonus Rules″. *Scandinavian Actuarial Journal.* 2002, 24, pp. 615-627.

[49]. **Pitrebois, S., Denuit, M. and Walhin, J.F.** "Fitting the Belgian Bonus-Malus System". *BELGIAN ACTUARIAL BULLETIN.* 2003, Vol. 3, 1, pp. 58-62.

[50]. **Renshaw, A.E.** "Modelling the Claim Process in the Presence of Covariates". *ASTIN Bulletin.* 1994, 24, pp. 265-285.

[51]. **Pinquet, J.** "Allowance for Costs of Claims in Bonus-Malus Systems". *ASTIN Bulletin.* 1997, Vols. 33-57, 27.

[52]. **Pinquet., J.** "Experience rating through heterogeneous models". *Handbook of Insurance Dionne, G., Boston Mass., Kluwer.* 1999.

[53]. **Dionne, G. and Vanasse, C.** "Automobile Insurance Ratemaking in the Rresence of Asymmetrical Information". *Journal of Applied Econometrics.* 1992, 7, pp. 149-165.

[54]. **Picech, L.** "The Merit-Rating Factor in a Multiplicating Rate-Making model". *ASTIN Colloquium, Cannes.* 1994.

[55]. **Sigalotti.** "Equilibrium Premiums in a Bonus-Malus System". *ASTIN Colloquium, Cannes.* 1994.

[56]. **Taylor, G.** "Setting A Bonus-Malus Scale in the Presence of Other Rating Factors". *ASTIN Bulletin.* 1997, 27, pp. 319-327.

[57]. **Pinquet, J.** "Designing Optimal Bonus-Malus Systems From Different Types of Claims". *Astin Bulletin.* 1998, Vol. 28, pp. 205-220.

[58]. **Lazar, D., Ciumas, C. and Pop, F.** "On using econometrics in pricing automobile insurance". *In Poloucek, S., Stavarek, D. (eds.) Future of Banking after the Year 2000 in the World and in the Czech Republic, Karvina: Silesian University.* 2005, pp. 1285-1298.

[59]. **Lemaire, J. and Zi, H.** ″A comparative analysis of 30 bonus-malus systems″. *ASTIN Bulletin.* 1994, Vol. 24, 2, pp. 287–309.

[60]. **Loimaranta, K.** "Some asymptotic properties of bonus systems". *The ASTIN Bulletin.* 1972, 6, pp. 233-245.

[61]. **Vepsalainen, S.** "Applications to a theory of bonus systems". *The ASTIN Bulletin.* 1972, 6, pp. 212-221.

[62]. **Lemaire, J.** ″Driver versus company, optimal behaviour of the policy holder". *Scandinavian Actuarial Journal.* 1976, 59, pp. 209-219.

[63]. **De Pril, N.** "The efficiency of a bonus-malus system". *The ASTIN Bulletin.* 1978, 10, pp. 59-72.

[64]. **Lemaire, J.** “Bonus-Malus Systems: The European and Asian Approach to Merit-Rating”. *North American Actuarial Journal.* 1998, Vol. 2, No.1, pp. 26-38.

[65]. **Holtan, J.** "Bonus Made Easy". *ASTIN Bulletin.* 1994, 24, pp. 61-74.

[66]. **Pitrebois, S., Walhin, J.F., Denuit, M.** "Bonus-Malus Systems with varying deductibles". *ASTIN Bulletin.* 2005, Vol. 35, 1, pp. 261-274.

[67]. **Philipson, C.** ″The Swedish system of bonus″. *AUSTIN Bulletin.* 1960, 1, pp. 134-141.

[68]. **Alting von Geusau, B.J.J.** ″The influence of a no-claim-discount-system on the behaviour of the insured clientele of an insurance company: A first approach″. *University of Amsterdam.* 1968.

[69]. **Dellaert, N.P., Frenk, J.B.G. and Kouwenhoven, A.** "Optimal claim behaviour for third-party liability insurances or To claim or not to claim: that is the question". *Insurance: Mathematics and Economics.* 1990, 9, pp. 59-76.

[70]. **Grenander, U.** ″Some remarks on bonus systems in automobile insurance″. *Skandinavisk Aktuarietidskrift.* 1957, 40, pp. 180-198.

[71]. **Haehling von Lanzenauer, C.** ″Entscheidungsregeln fur das optimale Verhalten in der Kraftfahrtversicherung″. *Unternemensforschung.* 1969, 13, pp. 193-204.

[72]. **Haehling von Lanzenauer., C.** "Decision problems in the Canadian automobile insurance systems″. *Journal of Risk and Insurance.* (1972a), 39, pp. 79-92.

[73]. **Haehling von Lanzenauer, C.** ″The effects of the insurees' decisions on the insurer's profit″. *Proceedings in Operations Research, Physica Verlag, Heidelberg.* (1972b), pp. 399-421.

[74]. **De Leve, G. and Weeda, P.J.** ″Driving with Markov-programming″. *The ASTIN Bulletin.* 1968, 5, pp. 62-86.

[75]. **Weeda, J.** ″Technical aspects of the iterative solution of the automobile insurance problem″. *Report BN 27/75 Mathematical Centre, Amsterdam.* 1975.

[76]. **Martin-Lof, A.** "A method for finding the optimal decision rule for a policy holder of an insurance with a bonus system". *Scandinavian Actuarial Journal.* 1973, 56, pp. 23-29.

[77]. **Haehling von Lanzenauer, C. and Lundberg, W.N.** ″The propensity to cause accidents″. *The ASTIN Bulletin.* 1973, 7, pp. 154-164.

[78]. **Haehling von Lanzenauer, C.** "Optimal claim decisions by policyholders in automobile insurance with merit-rating structures". *Operations Research.* 1974, 22, pp. 979-990.

[79]. **Lemaire, J.** "La soif du bonus". *The ASTIN Bulletin.* 1977, 9, pp. 181-190.

[80]. **Hastings, N.A.J.** "Optimal claiming on vehicle insurance. *Operational Research Quarterly.* 1976, 27, pp. 805-813.

[81]. **De Pril, N.** ″Optimal claim decisions for a bonus-malus system: A continuous approach". *The ASTIN Bulletin.* 1979, 10, pp. 215-222.

[82]. **Norman, J.M. and D.C.S. Shearn.** ″Optimal claiming on vehicle insurance revisited″ . *Journal of the Operational Research Society.* 1980, 31, pp. 181-186.

[83]. **Tijms, H.C.** ″Stochastic modelling and analysis: A computational approach″. *Wiley, New York.* 1986 .

[84]. **Kolderman, J. and Volgenant, A.** ″Optimal claiming in an automobile insurance system with bonus-malus structure″. *Journal of the Operational Research Society.* 1985, 36, pp. 239-247.

[85]. **Menist, M. and Volgenant, A.** ″Optimaal claimgedrag bij autoverzekeringen voor eindige periode,. *Het Verzekeringsarchief.* 1986, 63, pp. 61-76.

[86]. **Dellaert, N.P., Frenk, J.B.G. and Van Rijsoort, L.P.** "Optimal claim behaviour for vehicle damage insurances". *Insurance: Mathematics and Economics.* 1993, 12, pp. 225-244.

[87]. **Chiappori, P.-A. and Salanie, B.** "Asymmetric information in insurance markets: predictions and tests". *In Georges Dionne, editor, Handbook of Insurance.* 2013.

[88]. **Rothschild, M. and Stiglitz, J. E.** ″Equilibrium in competitive insurance markets: an essay on the economics of imperfect information″. *Quarterly Journal of Economics.* 1976, Vol. 90, pp. 629-649.

[89]. **Stiglitz, J. E.** ″Risk, incentives and insurance: the pure theory of moral hazard″. *The Geneva Papers on Risk and Insurance.* 1983, Vol. 26, pp. 4-33.

[90]. **Dionne, G. and Lasserre, P.** ″Adverse selection, repeated insurance contracts and announcement strategy″. *Review of Economic Studies.* 1985, Vol. 52, pp. 719-723.

[91]. **Cooper R. and Hayes, B.** ″Multi-period insurance contracts″. *International Journal of Industrial Organization.* 1987, Vol. 5, pp. 211-231.

[92]. **Kim, H., Kim, D., Im, S. and Hardin, J.W.** ″ Evidence of asymmetric information in the automobile insurance market: dichotomous versus multinomial measurement of insurance coverage″. *The Journal of Risk and Insurance.* 2009, Vol. 76, pp. 343-366.

[93]. **Cohen, A. and Siegelman, P.** ″Testing for adverse selection in insurance markets″. *The Journal of Risk and Insurance.* 2010, Vol. 77, pp. 39-84.

[94]. **Dahlby, G.** ″Adverse selection and statistical discrimination: An analysis of Canadian automobile insurance″. *The Journal of Public Economics.* 1983, Vol. 20, pp. 207–219.

[95]. **Dahlby., G.** ″Testing for asymmetric information in Canadian automobile insurance″. *In Georges Dionne, editor, Contributions to Insurance Economics. Kluwer Academic, Boston.* 1992.

[96]. **Puelz, R. and Snow, A.** ″ Evidence on adverse selection: equilibrium signaling and Cross-Subsidization in the insurance market″. *Journal of Political Economy.* 1994, Vol. 102, pp. 236-257.

[97]. **Cohen, A.** ″Asymmetric information and learning: Evidence from the automobile insurance market″. *The Review of Economics and Statistics.* 2005, Vol. 87, 2, pp. 197–207.

[98]. **Chiappori, P.-A. and Salanie, B.** ″Testing for asymmetric information in insurance markets″. *The Journal of Political Economy.* 2000, Vol. 108, 1, pp. 56–78.

[99]. **Dionne, G., Gourieroux, C. and Vanasse, C.** "Testing for adverse selection in the automobile insurance market: A comment". *Journal of Political Economy.* 2001, Vol. 109, pp. 444–453.

[100]. **Saito, K.** ″An empirical analysis of the roles of risk aversion in the automobile insurance market″. *Annual Meeting of American Risk and Insurance Association.* 2006.

[101]. **Donnelly, C., Englund, M. and Nielsen, J.P.** "The Importance of the Choice of Test for Finding Evidence of Asymmetric Information". *ASTIN BULLETIN .* May 2014, Vol. 44, 02, pp. 173-195.

[102]. **Picard, P.** ″Generalisation de l'etude sur la survenance des sinistres en assurance automobile″. *Bulletin Trimestriel de l'Institute des Actuaires Francais.* 1976, pp. 204-267.

[103]. **Frangos, N.E. and Vrontos, S.D.** "Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance". *ASTIN Bulletin.* 2001, Vol. 31, 1, pp. 1-22.

[104]. **Vrontos, S.** "Design of an Optimal Bonus-Malus System in Automobile Insurance" . *Msc. Thesis, Department of Statistics, Athens University of Economics and Business, ISBN: 960-.* 1998 .

[105]. **Pitrebois, S., Denuit, M. and Walhin, J.-F.** "Setting a Bonus-Malus Scale in the Presence of Other Rating Factors: Taylor's Work Revisited". *ASTIN Bulletin.* 2003, Vol. 33, 2, pp. 419-436.

[106]. **Bonsdorff, H.** ″On asymptotic properties of Bonus-Malus systems based on the number and on the size of the claims″. *Scandinavian Actuarial Journal.* 2005, Vol. 4, pp. 309-320.

[107]. **Sammatitini, G.** "A Bonus-Malus System with Conditioned Bonus". *Insurance Mathematics and Economics.* 1990, Vol. 9, pp. 163-169.

[108]. **Coene, G. and Doray L.G.** "A Financially Balanced Bonus-Malus System". *ASTIN Bulletin.* 1996, Vol. 26, 1, pp. 107-116.

[109]. **Ширяев, А.Н.** ″Вероятность″. *изд. МЦНМО, Москва.* 2004.

[110]. **Gulyan, A.G.** "An Alternative Model for Bonus-Malus System". *Proceedings of the Yerevan State University (Mathematics).* 2015, 1, pp. 15-19.

[111]. **Bowers, N.L. at al.** "Actuarial Mathematics". *The Society of Actuaries.* 1997.

[112]. **Schmidt, K. D.** "On inequalities for moments and the covariance of monotone functions". *Insurance: Mathematics and Economics.* 2014, Vol. 55, pp. 91-95.

[113]. **Egozcue, M., Garcia,L.F. and Wong,W.-K.** "On some covariance inequalities for monotonic and non-monotonic functions". *Journal of inequalities of in pure and applied mathematics.* 2009, Vol. 10, 3, pp. Art. 75, 7 pp.

[114]. **Kozubik, A.** "Bonus–Malus Systems as Markov Process and Hungry for Bonus". *Journal of Information, Control and Management Systems.* 2006, Vol. 4, 1, pp. 9-17.

[115]. **Tsoi, A. C., Zhang, S. and Hagenbuchner, M.** "Hierarchical Hidden Markov Models: An Application to Health Insurance Data". *Data Mining, Lecture Notes in Computer Science, Springer-Verlag.* 2006, Vol. 3755, pp. 244-259.

[116]. **Paroli, R., Redaelli, G. and Spezia, L.** "Poisson Hidden Markov Models for Time Series of Overdispersed Insurance Counts". *ASTIN Bulletin.* 2000, pp. 461-474.

[117]. **Kluppelberg, C. and Severin, M.** "Prediction of Outstanding Insurance Claims". *Center of Mathematical Sciences, Munich University of Technology.* 2002.

[118]. **Bulla, J.** "Application of Hidden Markov Models and Hidden Semi-Markov Models to Financial Time Series". *PhD Dissertation.* 2006.

[119]. **Dizaji, A.K.** "A Review on Bonus Malus Systems and the Introduction to Dynamic Bonus Malus System by the Use of Hidden Markov Models". *MS Thesis.* 2011.

[120]. **Ouyang, K.X., Vaidya, B. and Makrakis, D.** "A probabilistic-based approach towards trust evaluation using Poisson Hidden Markov Models and Bonus Malus Systems". *Privacy, Security and Trust (PST), 2011 Ninth Annual International Conference.* 2011, pp. 150-155.

[121]. **Chitchyan, R.N. and Gulyan, A.G.** "A Bonus-Malus System with Aggregate Claim Amount Component". *Journal of Contemporary Mathematical Analysis.* 2015, Vol. 50, 4, pp. 74-88.

[122]. **Elliott, R. J., Aggoun, L. and Moore, J. B.** “Hidden Markov models: Estimation and Control”. *Springer-Verlag New York.* 1995.

[123]. **Aggoun, L., and Elliott, R.** “Measure Theory and Filtering: Introduction with applications”. *Cambridge University Press.* 2004.

[124]. **Ghahramani, Z.** “An introduction to Hidden Markov Models and Bayesian Networks”. *International Journal of Pattern Recognition and Artificial Intelligence.* 2001, Vol. 15, 1, pp. 9-42.

[125]. **Gulyan, A.G. and Chitchyan, R.N.** "Evaluation Methods of the Coefficients of Transition Matrices in Extended Bonus-malus System". *Contemporary Science and Education in Americas, Africa and Eurasia. The IX International Academic Congress (Technology and engineering).* 2015, Vol. 3, pp. 215-221.

[126]. www.paap.am.

[127]. **Chitchyan, R.N. and Gulyan, A.G.** "Comparative Analysis of the Bonus-Malus Systems". *Proceedings of Engineering Academy of Armenia (Economics and management).* 2016, Vol. 13, 1, pp. 5-16.

[128]. http://paap.am/datas/users/1-regulation%20factes.pdf.

[129]. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/.

[130]. **Айвазян, С.А., Мхитарян, В.С.** "Теория Вероятностей и Прикладная Статистика". *Юнити, Москва.* 2001.

[131]. **Huber-Carol, C., Balakrishnan,N., Nikulin,M. S., Mesbah,M.** "Goodness-of-Fit Tests and Model Validity". *Statistics for Industry and Technology.* 2002.

[132]. **Cox, D.R., Donnelly, C.A.** "Principles of Applied Statistics". *Cambridge University Press.* 2011.

[133]. *http://www.mathwave.com/.*

[134]. **Вентцель, Е.С.** "Теория вероятностей". *Высшая Школа.* 1999.

Appendix

Discrete Distributions

Table A.1

Distribution Name	Probability Function	Mean	Variance
Poisson $\lambda > 0$	$\frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ
Negative Binomial $x = k, k+1, \ldots$	$C_{x-1}^{k-}(1-p)^{x-k}p^k$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Geometric	$(1-p)^k p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson-Goncharov $n \in N$	$\frac{\theta(\theta+n\lambda)^{n-1}}{n!}e^{-\theta-n\lambda}$	N/A	N/A
Poisson-Inverse Gaussian	$e^{\frac{\mu}{\beta}\left(1-\sqrt{1+2\beta(1-z)}\right)}$	μ	$\mu(1+\beta)$

Continuous Distributions

(for $0 \leq x < +\infty$)

Table A.2

Distribution Name	Probability Density Function $\mathbf{f(x)}$	Cumulative Distribution Function $\mathbf{F(x)}$	Mean	Variance
Burr(XII) (3-parameter) $\alpha, \beta, k > 0$	$\frac{\alpha k\left(\frac{x}{\beta}\right)^{\alpha-1}}{\beta\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{k+1}}$	$1-\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{-k}$	Moments $\mu_r = \beta^r kB\left(\frac{\alpha k - r}{\alpha}, \frac{\alpha + r}{\alpha}\right), for\ r < \alpha k$	
Chi-Squared $v \in N$	$\frac{x^{\frac{v}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)}$	$\frac{\Gamma_{\frac{x}{2}}\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}$	v	$2v$
Dagum (3-parameter) $\alpha, \beta, k > 0$	$\frac{\alpha k\left(\frac{x}{\beta}\right)^{k\alpha-1}}{\beta\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{k+1}}$	$1-\left(1+\left(\frac{x}{\beta}\right)^{-\alpha}\right)^{-k}$	$-\frac{\beta}{\alpha}\frac{\Gamma\left(-\frac{1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}+k\right)}{\Gamma(k)},$ $for\ \alpha > 1$	$-\frac{\beta^2}{\alpha^2}\left(2\alpha\frac{\Gamma\left(-\frac{2}{\alpha}\right)\Gamma\left(\frac{2}{\alpha}+k\right)}{\Gamma(k)}+\right.$ $for\ \alpha > 2$
Erlang (3-parameter) $m > 0;\ \gamma \leq x < +\infty$	$\frac{(x-\gamma)^{m-1}}{\beta^m\Gamma(m)}e^{-\frac{x-\gamma}{\beta}}$	$\frac{\Gamma_{\frac{(x-\gamma)}{\beta}}(m)}{\Gamma(m)}$	$\gamma + \beta m$	$\beta^2 m$
Erlang (2-parameter) $m > 0$	$\frac{x^{m-1}}{\beta^m\Gamma(m)}e^{-\frac{x}{\beta}}$	$\frac{\Gamma_{\frac{x}{\beta}}(m)}{\Gamma(m)}$	βm	$\beta^2 m$
Exponential (2-parameter) $\lambda > 0;\ \gamma \leq x < +\infty$	$\lambda e^{-\lambda(x-\gamma)}$	$1-e^{-\lambda(x-\gamma)}$	$\gamma + \frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Distribution Name	Probability Density Function $\mathbf{f(x)}$	Cumulative Distribution Function $\mathbf{F(x)}$	Mean	Variance
Exponential $\boldsymbol{\lambda} > 0$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
F (Fisher-Snedecor) or Pearson 5 $\mathbf{v_1, v_2} \in \mathbf{N}$	$\frac{\left(\frac{v_1}{v_2}x\right)^{\frac{v_1}{2}}\left(1+\frac{v_1}{v_2}x\right)^{-\frac{v_1+v_2}{2}}}{xB\left(\frac{v_1}{2},\frac{v_2}{2}\right)}$	$I_x(v_1,v_2)$	Moments $\mu_r=\left(\frac{v_1}{v_2}\right)^r\frac{\Gamma\left(\frac{v_1}{2}+r\right)}{\Gamma\left(\frac{v_1}{2}\right)}\frac{\Gamma\left(\frac{v_2}{2}-r\right)}{\Gamma\left(\frac{v_2}{2}\right)}, for\ \ 2r<v_2$	
Fatigue life[6] $\boldsymbol{\alpha}, \boldsymbol{\beta} > 0$	$\frac{\left(\frac{x}{\beta}\right)^{0.5}+\left(\frac{\beta}{x}\right)^{0.5}}{2\alpha x}\cdot\phi\left(\frac{1}{\alpha}\left(\left(\frac{x}{\beta}\right)^{0.5}-\left(\frac{\beta}{x}\right)^{0.5}\right)\right)$	$\Phi\left(\frac{1}{\alpha}\left(\left(\frac{x}{\beta}\right)^{0.5}-\left(\frac{\beta}{x}\right)^{0.5}\right)\right)$	$\beta\left(1+\frac{\alpha^2}{2}\right)$	$(\alpha\beta)^2\left(1+\frac{5\alpha^2}{4}\right)$
Fretchet (2-parameter) $\boldsymbol{\alpha}, \boldsymbol{\beta} > 0$	$\frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1}e^{-\left(\frac{\beta}{x}\right)^{\alpha}}$	$e^{-\left(\frac{\beta}{x}\right)^{\alpha}}$	Moments $\mu_r=\beta\Gamma\left(1-\frac{r}{\alpha}+r\right), for\ \ r<\alpha$	
Gamma (2-parameter) $\boldsymbol{\alpha}, \boldsymbol{\beta} > 0$	$\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$	$\frac{\Gamma_x(\beta x)}{\Gamma(\alpha)}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Generalized Gamma (or 3-parameter) $\mathbf{a, d, p} > 0$	$\frac{\frac{p}{a^d}}{\Gamma\left(\frac{d}{p}\right)}x^{d-1}e^{-\left(\frac{x}{a}\right)^p}$	$\frac{\Gamma_{\frac{d}{p}}\left(\left(\frac{x}{a}\right)^p\right)}{\Gamma\left(\frac{d}{p}\right)}$	$a\frac{\Gamma\left(\frac{d+1}{p}\right)}{\Gamma\left(\frac{d}{p}\right)}$	$a^2\left(\frac{\Gamma\left(\frac{d+2}{p}\right)}{\Gamma\left(\frac{d}{p}\right)}-\left(\frac{\Gamma\left(\frac{d+1}{p}\right)}{\Gamma\left(\frac{d}{p}\right)}\right)\right)$
Pearson 5 or Inverse Gamma $\boldsymbol{\alpha}, \boldsymbol{\beta} > 0$	$\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{-\alpha-1}e^{-\frac{\beta}{x}}$	$\frac{\Gamma_{\alpha}\left(\frac{\beta}{x}\right)}{\Gamma(\alpha)}$	$\frac{\beta}{\alpha-1}, for\ \ \alpha>1$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, for\ \alpha>2$
Weibull $\boldsymbol{\alpha}, \boldsymbol{\beta} > 0$	$\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	$1-e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	$\beta\Gamma\left(1+\frac{1}{\alpha}\right)$	$\beta^2\left(\Gamma\left(1+\frac{2}{\alpha}\right)-\left(\Gamma\left(1+\frac{1}{\alpha}\right)\right)\right)$

The following functions have been used

1. Beta function

$$B(\alpha_1,\alpha_2)=\int_0^1 t^{\alpha_1-1}(1-t)^{\alpha_2-1}dt,\ \ \alpha_1,\alpha_2>0$$

2. Incomplete Beta function

$$B(\alpha_1,\alpha_2)=\int_0^x t^{\alpha_1-1}(1-t)^{\alpha_2-1}dt,\ \ \alpha_1,\alpha_2>0;0\le x\le 1$$

3. Gamma function

[6] where $\Phi(x)$ and $\phi(x)$ are the cumulative distribution and probability density functions of the Standard Normal Distribution

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt\,, \quad \alpha > 0$$

4. Incomplete Gamma function

$$\Gamma_x(\alpha) = \int_0^x t^{\alpha-1} e^{-t} dt\,, \quad \alpha > 0$$

5. Regularized Incomplete Beta function

$$I_x(\alpha_1, \alpha_2) = \frac{B_x(\alpha_1, \alpha_2)}{B(\alpha_1, \alpha_2)}$$