ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Actuarial Modelling of Claim Counts and Losses in Motor Third Party Liability Insurance

By

George J. Tzougas

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfillment of the requirements for

the degree of Doctor of Philosophy in Statistics

Athens, Greece

July 2013

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Αναλογιστική Μοντελοποίηση του Αριθμού και του Κόστους των Απαιτήσεων στην Ασφάλιση Αστικής Ευθύνης Έναντι Τρίτων στον Κλάδο των Αυτοκινήτων

Γεώργιος Ι. Τζουγάς

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών

ως μέρος των απαιτήσεων για την απόκτηση

Διδακτορικού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα

Ιούλιος 2013

DEDICATION

To my parents Ioannis and Constantina

and my beloved family

ACKNOWLEDGEMENTS

First, I would like to express my profound gratitude to my supervisor Dr. Nicholas Frangos for his insight, excellent research guidance, constant encouragement, and for his support in so many aspects. He is a wonderful advisor, not only knowledgeable but also truly cares for his students. I am also very grateful to Dr. Spyridon Vrontos for his assistance and valuable suggestions for my dissertation and also for all the help he has offered me over the past 4 years. I would never have finished my dissertation without them. I also want to take this opportunity to thank my committee members Dr. Nicholas Frangos, Dr. Spyridon Vrontos and Dr. Michael Zazanis, the professors, students, classmates and friends in the Department of Statistics of the Athens University of Economics and Business. Last but not least, I express my special gratitude and heartfelt thanks to my parents for their continuous support and encouragement to go on with my education.

VITA

I was born in Athens on April 19, 1983. During my early school years, I found great interest in mathematics, participating in several mathematical competitions and gaining awards. Also, at that time I took music lessons and played basketball for a local team. In July 2001, I graduated from Doukas High School with an average grade 17.6 (out of 20) and in September of the same year I enrolled, as a first year student, in the Department of Mathematics of the University of Athens, where I met most of my current friends. During my graduate studies, I gained my first working experience between 2004 and 2005 when I had full time active participation in several stages of the project "A risk analysis information system", supported by the General Secretary of Development and Investment, of the Ministry of National Economy and Finance. In 2006, I completed my bachelor degree with an average grade 7.4 (out of 10). During my postgraduate studies in the Department of Statistics of the Athens University of Economics and Business, I successfully attained and finished with an average grade 7.2 the following courses: Theoretical Statistics I, Introduction to Statistical Packages and Data Analysis, Sampling Theory, Computational Statistics I, Time series, Generalized Linear Models, Stochastic Process with Emphasis in Finance I, Theoretical Statistics II, Computational Statistics II, Stochastic Process with Emphasis in Finance II. Furthermore, I participated in several projects of the above courses and in one project in Multivariate Analysis. In 2008, I was awarded my Master of Science degree, which was on the topic "A Study of Optimal Bonus-Malus Systems in Automobile Insurance Using Different Underlying Approaches ". I began work on my PhD studies in 2009. While undertaking this work I assisted in teaching the following courses: Calculus I, Linear Algebra and Applications, Generalized Linear Models. I further wrote course notes on the latter of these, which are available in the electronic class of the department. In 2010, I participated in the 6th Conference in Actuarial Science & Finance on Samos where I submitted a work entitled "On the design of Some Optimal Bonus-Malus Systems Using Frequency and Severity Components ". In 2013 my work entitled "The Design of an Optimal Bonus-Malus System Using the Sichel Distribution as a Model of Claim Counts" was presented at the International Cramér Symposium on Insurance Mathematics in Stockholm, June 11-14, 2013. An extended version of the study presented in aforementioned conference entitled "The Design of an Optimal Bonus-Malus System Based on the Sichel Distribution" was accepted for publication in the collective book "Modern Problems in Insurance Mathematics" that will be edited by Springer Verlag**.** Furthermore, during this year, my paper entitled "Optimal Bonus-Malus Systems Using Finite Mixture Models" was accepted for publication by ASTIN bulletin and my paper entitled "Risk Classification for Claim Counts and Losses Using Generalized Additive Models for Location, Scale and Shape" was submitted for publication to a leading actuarial journal. I also refereed two papers for ASTIN Bulletin. Finally, apart from Greek, I speak English well (First Certificate in English, Certificate in Advanced English) and French reasonably and have a good knowledge of the following software packages: MS Office, Matlab, R Language, S-Plus, E-Views, SPSS, Mathematica and C^{++}, LaTeX, Scientific Workplace.

ABSTRACT

Actuarial science is the discipline that deals with uncertain events where clearly the concepts of probability and statistics provide for an indispensable instrument in the measurement and management of risks in insurance and finance. An important aspect of the business of insurance is the determination of the price, typically called premium, to pay in exchange for the transfer of risks. It is the duty of the actuary to evaluate a fair price given the nature of the risk. Actuarial literature research covers a wide range of actuarial subjects among which is risk classification and experience rating in motor third-party liability insurance, which are the driving forces of the research presented in this thesis. This is an area of applied statistics that has been borrowing tools from various kits of theoretical statistics, notably empirical Bayes, regression, and generalized linear models, GLM, (Nelder and Wedderburn, 1972). However, the complexity of the typical application, featuring unobservable risk heterogeneity, imbalanced design, and nonparametric distributions, inspired independent fundamental research under the label `credibility theory', now a cornerstone in contemporary insurance mathematics. Our purpose in this thesis is to make a contribution to the connection between risk classification and experience rating with generalized additive models for location scale and shape, GAMLSS, (Rigby and Stasinopoulos, 2005) and finite mixture models (Mclachlan and Peel, 2000). In Chapter 1, we present a literature review of statistical techniques that can be practically implemented for pricing risks through ratemaking based on a priori risk classification and experience rated or Bonus-Malus Systems. The idea behind a priori risk classification is to divide an insurance portfolio into different classes that consist of risks with a similar profile and to design a fair tariff for each of them. Recent actuarial literature research assumes that the risks can be rated a priori using generalized linear models GLM, (see, for example, Denuit et al., 2007 & Boucher et al., 2007, 2008). Typical response variables involved in this process are the number of claims (or the claim frequency) and its corresponding severity (i.e. the amount the insurer paid out, given that a claim occurred). In Chapter 2, we extend this setup following the GAMLSS approach of Rigby and Stasinopoulos (2005). The GAMLSS models extend GLM framework allowing joint modeling of location and shape parameters. Therefore both mean and variance may be assessed by choosing a

marginal distribution and building a predictive model using ratemaking factors as independent variables. In the setup we consider, risk heterogeneity is modeled as the distribution of frequency and cost of claims changes between clusters by a function of the level of ratemaking factors underlying the analyzed clusters. GAMLSS modeling is performed on all frequency and severity models. Specifically, we model the claim frequency using the Poisson, Negative Binomial Type II, Delaporte, Sichel and Zero-Inflated Poisson GAMLSS and the claim severity using the Gamma, Weibull, Weibull Type III, Generalized Gamma and Generalized Pareto GAMLSS as these models have not been studied in risk classification literature. The difference between these models is analyzed through the mean and the variance of the annual number of claims and the costs of claims of the insureds, who belong to different risk classes. The resulting a priori premiums rates are calculated via the expected value and standard deviation principles with independence between the claim frequency and severity components assumed. However, in risk classification many important factors cannot be taken into account a priori. Thus, despite the a priori rating system, tariff cells will not be completely homogeneous and may generate a ratemaking structure that is unfair to the policyholders. In order to reduce the gap between the individual's premium and risk and to increase incentives for road safety, the individual's past record must taken into consideration under an a posteriori model. Bonus-Malus Systems (BMSs) are a posteriori rating systems that penalize insureds responsible for one or more accidents by premium surcharges or maluses and reward claim-free policyholders by awarding them discounts or bonuses. A basic interest of the actuarial literature is the construction of an optimal or `ideal' BMS defined as a system obtained through Bayesian analysis. A BMS is called optimal if it is financially balanced for the insurer: the total amount of bonuses must be equal to the total amount of maluses and if it is fair for the policyholder: the premium paid by each policyholder is proportional to the risk that they impose on the pool. The study of such systems based on different statistical models will be the main objective of this thesis. In Chapter 3, we extend the current BMS literature using the Sichel distribution to model the claim frequency distribution. This system is proposed as an alternative to the optimal BMS obtained by the Negative Binomial model (see, Lemaire, 1995). We also consider the optimal BMS provided by the Poisson-Inverse Gaussian distribution, which is a special case of the Sichel distribution. Furthermore, we introduce a generalized BMS that takes into account both the a priori and a posteriori characteristics of each policyholder, extending the framework developed by Dionne and Vanasse (1989, 1992). This is achieved by employing GAMLSS modeling on all the frequency models considered in this chapter, i.e. the Negative Binomial, Sichel and Poisson-Inverse Gaussian models. In the above setup optimality is achieved by minimizing the insurer's risk. The majority of optimal BMSs in force assign to each policyholder a premium based on their number of claims disregarding their aggregate amount. In this way, a policyholder who underwent an accident with a small size of loss will be unfairly penalized in comparison to a policyholder who had an accident with a big size of loss. Motivated by this, the first objective of Chapter 4 is the integration of claim severity into the optimal BMSs based on the a posteriori criteria of Chapter 3. For this purpose we consider that the losses are distributed according to a Pareto distribution, following the setup used by Frangos and Vrontos (2001). The second objective of Chapter 4 is the development of a generalized BMS with a frequency and a severity component when both the a priori and the a posteriori rating variables are used. For the frequency component we assume that the number of claims is distributed according to the Negative Binomial Type I, Poisson Inverse Gaussian and Sichel GAMLSS. For the severity component we consider that the losses are distributed according to a Pareto GAMLSS. This system is derived as a function of the years that the policyholder is in the portfolio, their number of accidents, the size of loss of each of these accidents and of the statistically significant a priori rating variables for the number of accidents and for the size of loss that each of these claims incurred. Furthermore, we present a generalized form of the one obtained in Frangos and Vrontos (2001). Finally, in Chapter 5 we give emphasis on both the analysis of the claim frequency and severity components of an optimal BMS using finite mixtures of distributions and regression models (see Mclachlan and Peel, 2000 & Rigby and Stasinopoulos, 2009) as these methods, with the exception of Lemaire(1995), have not been studied in the BMS literature. Specifically, for the frequency component we employ a finite Poisson, Delaporte and Negative Binomial mixture, while for the severity component we employ a finite Exponential, Gamma, Weibull and Generalized Beta Type II (GB2) mixture, updating the posterior probability. We also consider the case of a finite Negative Binomial mixture and a finite Pareto mixture updating the posterior mean. The generalized BMSs we propose adequately integrate risk classification and experience rating by taking into account both the a priori and a posteriori characteristics of each policyholder.

ΠΕΡΙΛΗΨΗ

Η αναλογιστική επιστήμη ασχολείται με αβέβαια γεγονότα, στα οποία με σαφήνεια οι έννοιες των πιθανοτήτων και της στατιστικής παρέχουν ένα απαραίτητο εργαλείο για τη μέτρηση και τη διαχείριση των κινδύνων στον τομέα της ασφάλισης και της οικονομίας. Μια σημαντική πτυχή του επαγγέλματος της ασφάλισης είναι ο καθορισμός της τιμής, που συνήθως ονομάζεται ασφάλιστρο, η οποία πρέπει να πληρωθεί ως αντάλλαγμα για τη μεταβίβαση των κινδύνων. Είναι καθήκον του αναλογιστή να αξιολογήσει μια δίκαιη τιμή, δεδομένης της φύσης του κινδύνου. Η βιβλιογραφική έρευνα της αναλογιστικής επιστήμης καλύπτει ένα ευρύ φάσμα θεμάτων, μεταξύ των οποίων είναι η ταξινόμηση των κινδύνων σε κλάσεις (risk classification) και η εμπειρική τιμολόγηση (experience rating) στην ασφάλιση αστικής ευθύνης έναντι τρίτων στον κλάδο των αυτοκινήτων, όπου αποτελούν τις κινητήριες δυνάμεις της έρευνας που παρουσιάζεται σε αυτή τη διατριβή. Τα ανωτέρω συνιστούν μια περιοχή της εφαρμοσμένης στατιστικής η οποία έχει δανειστεί εργαλεία από διαφορετικές εργαλειοθήκες της θεωρητικής στατιστικής, κυρίως εμπειρικές μεθόδους Bayes, παλινδρόμηση και γενικευμένα γραμμικά μοντέλα, ΓΓΜ (Nelder και Wedderburn, 1972). Ωστόσο, η πολυπλοκότητα της τυπικής εφαρμογής, συμπεριλαμβανομένης της μη παρατηρήσιμης ετερογένειας κινδύνου (risk heterogeneity), της μη ισορροπημένης σχεδίασης και των μη παραμετρικών κατανομών, ενέπνευσε μια ανεξάρτητη θεμελιώδη έρευνα γνωστή ως ‹‹θεωρία αξιοπιστίας›› (credibility theory), η οποία θεωρείται ο ακρογωνιαίος λίθος των σύγχρονων ασφαλιστικών μαθηματικών. Στόχος της παρούσας διδακτορικής διατριβής είναι η συνεισφορά στη σύνδεση μεταξύ της a priori ταξινόμησης των κινδύνων σε κλάσεις (a priori risk classification) και της εμπειρικής τιμολόγησης με τα γενικευμένα προσθετικά μοντέλα για τη θέση, την κλίμακα και το σχήμα, γνωστά ως GAMLSS μοντέλα (βλέπε Rigby και Stasinopoulos, 2005), και με τα πεπερασμένα μοντέλα μείξης, (finite mixture models, βλέπε Mclachlan και Peel, 2000). Στο Κεφάλαιο 1, παρουσιάζουμε τη βιβλιογραφική αναφορά των στατιστικών τεχνικών που μπορούν να εφαρμοστούν στην πράξη για την τιμολόγηση των κινδύνων με βάση την a priori ταξινόμηση τους σε κλάσεις καθώς και τα συστήματα εμπειρικής τιμολόγησης ή συστήματα εκπτώσεων-επιβαρύνσεων (Bonus-Malus systems, BMSs). Η ιδέα πίσω από την a priori ταξινόμηση των κινδύνων σε κλάσεις είναι ο

διαχωρισμός ενός ασφαλιστικού χαρτοφυλακίου σε διαφορετικές τάξεις, οι οποίες απαρτίζονται από κινδύνους με παρόμοιο προφίλ, καθώς και ο σχεδιασμός μια δίκαιης τιμολόγησης για κάθε μία από αυτές. Η σύγχρονη αναλογιστική βιβλιογραφική έρευνα θεωρεί ότι οι κίνδυνοι είναι δυνατό να ταξινομηθούν a priori με την βοήθεια των γενικευμένων γραμμικών μοντέλων (βλέπε, για παράδειγμα, Denuit et al., 2007 & Boucher et al., 2007, 2008). Οι απαντητικές μεταβλητές που εμπλέκονται σε αυτή τη διαδικασία είναι ο αριθμός (ή συχνότητα) των απαιτήσεων (claims) του ασφαλιζομένου προς την ασφαλιστική εταιρεία καθώς και η αντίστοιχη σφοδρότητά τους (δηλαδή το ποσό που ο ασφαλιστής κατέβαλε, δεδομένου ότι μια απαίτηση έχει καταγραφεί). Στο Κεφάλαιο 2, επεκτείνουμε αυτή τη θεώρηση ακολουθώντας την προσέγγιση των GAMLSS μοντέλων των Rigby και Stasinopoulos (2005). Τα GAMLSS μοντέλα επεκτείνουν το πλαίσιο των ΓΓΜ επιτρέποντας την κοινή μοντελοποίηση των παραμέτρων της θέσης και του σχήματος μιας κατανομής. Συνεπώς τόσο η μέση τιμή όσο και η διακύμανση μπορούν να εκτιμηθούν μέσω της επιλογής μιας περιθώριας κατανομής και της οικοδόμησης ενός προβλεπτικού μοντέλου χρησιμοποιώντας τους παράγοντες τιμολόγησης (ratemaking factors) ως ανεξάρτητες μεταβλητές. Στο ανωτέρω πλαίσιο, η ετερογένεια του κινδύνου μοντελοποιείται ως η κατανομή της αλλαγής των συχνοτήτων και/ή του κόστους των απαιτήσεων προς την ασφαλιστική εταιρεία μεταξύ των διαφορετικών ομάδων ασφαλιζομένων, σε συνάρτηση με το επίπεδο των παραγόντων τιμολόγησης στους οποίους στηρίζονται οι προαναφερθείσες ομάδες. Η GAMLSS μοντελοποίηση πραγματοποιείται σε όλα τα μοντέλα που αναπαριστούν τη συχνότητα και την σφοδρότητα των απαιτήσεων του ασφαλιζομένου προς την ασφαλιστική εταιρεία (claim frequency and severity models). Συγκεκριμένα, μοντελοποιούμε την συχνότητα των απαιτήσεων με βάση το Poisson, Αρνητικό Διωνυμικό Τύπου II, Delaporte, Sichel και Zero-Inflated Poisson GAMLSS και τη σφοδρότητα των απαιτήσεων με βάση το Γάμμα, Weibull, Weibull Τύπου III, Γενικευμένο Γάμμα και Γενικευμένο Pareto GAMLS καθώς τα μοντέλα αυτά δεν έχουν μελετηθεί στη βιβλιογραφία της a priori ταξινόμησης των κινδύνων σε κλάσεις. Η διαφορά μεταξύ αυτών των μοντέλων αναλύεται μέσω της μέσης τιμής και τη διακύμανσης του ετήσιου αριθμού καθώς και του κόστους των απαιτήσεων των ασφαλισμένων, οι οποίοι ανήκουν σε διαφορετικές κλάσεις κινδύνου. Οι προκύπτουσες τιμές των a priori ασφαλίστρων υπολογίζονται μέσω των αρχών της αναμενόμενης τιμής και της τυπική απόκλισης (expected value and standard deviation principles) με βάση την υπόθεση της ανεξαρτησίας μεταξύ της συχνότητας και της σφοδρότητας των απαιτήσεων προς την ασφαλιστική εταιρεία. Ωστόσο, κατά την ταξινόμηση των κινδύνων σε κλάσεις πολλοί σημαντικοί παράγοντες δεν μπορούν να ληφθούν υπόψη εκ των προτέρων. Συνεπώς, παρά το a priori σύστημα, οι κλάσεις τιμολόγησης, δεν θα είναι απολύτως ομοιογενείς και μπορεί να δημιουργηθεί μια δομή τιμολόγησης που είναι άδικη για τους ασφαλισμένους. Προκειμένου να μειωθεί το χάσμα μεταξύ του ατομικού ασφαλίστρου και του υποβόσκοντος κινδύνου και να αυξηθούν τα κίνητρα για την οδική ασφάλεια, το ατομικό ιστορικό ζημιών πρέπει να ληφθεί υπόψη στο πλαίσιο ενός a posteriori μοντέλου. Τα συστήματα εκπτώσεων-επιβαρύνσεων (BMSs) είναι συστήματα a posteriori τιμολόγησης τα οποία επιβάλλουν ποινές στους ασφαλισμένους που είναι υπαίτιοι για ένα ή περισσότερα ατυχήματα μέσω της επιβολής ενός επασφάλιστρου (malus) και αντίστοιχα επιβραβεύουν τους ασφαλισμένους χωρίς ατυχήματα με έκπτωση (bonus). Βασικό ενδιαφέρον της βιβλιογραφίας της αναλογιστικής επιστήμης αποτελεί η κατασκευή ενός βέλτιστου ή «ιδανικού» συστήματος BMS το οποίο ορίζεται ως ένα σύστημα το οποίο αποκτήθηκε μέσω της Bayesian analysis. Ο βασικός στόχος της παρούσας διατριβής είναι η μελέτη αυτών των συστημάτων μέσω της χρήσης διαφόρων στατιστικών μοντέλων. Στο Κεφάλαιο 3, επεκτείνουμε την τρέχουσα βιβλιογραφία για BMSs χρησιμοποιώντας την κατανομή Sichel για την μοντελοποίηση της κατανομής της συχνότητας των απαιτήσεων του ασφαλιζόμενου προς την ασφαλιστική εταιρεία. Το σύστημα αυτό προτείνεται ως εναλλακτικό του βέλτιστου BMS που λαμβάνεται από το Αρνητικό Διωνυμικό μοντέλο (βλέπε, Lemaire, 1995). Παρουσιάζουμε επίσης και το βέλτιστο BMS που προέρχεται από την Poisson-Inverse Gaussian κατανομή, η οποία μπορεί να θεωρηθεί ως μια ειδική περίπτωση της κατανομής Sichel. Επιπροσθέτως, προτείνουμε ένα γενικευμένο BMS που λαμβάνει υπόψη τα a priori και a posteriori χαρακτηριστικά του κάθε ασφαλισμένου, επεκτείνοντας το πλαίσιο που αναπτύχθηκε από τους Dionne και Vanasse (1989, 1992). Αυτό επιτυγχάνεται με την πραγματοποίηση GAMLSS μοντελοποίησης σε όλα τα μοντέλα του τρέχοντος κεφαλαίου που αναπαριστούν τη συχνότητα των απαιτήσεων του ασφαλιζόμενου προς την ασφαλιστική εταιρεία, δηλαδή τα Αρνητικό Διωνυμικό, Sichel και Poisson-Inverse Gaussian μοντέλα. Στην ανωτέρω θεώρηση η βελτιστοποίηση καθίσταται εφικτή μέσω της ελαχιστοποίησης του κινδύνου του ασφαλιστή. Η πλειονότητα των BMSs σε ισχύ βασίζεται στην συχνότητα των απαιτήσεων του ασφαλιζομένου προς την ασφαλιστική εταιρεία για την ανάθεση του ασφάλιστρου και αγνοεί το συνολικό τους κόστος. Συνεπώς, η επιβαλλόμενη ποινή σε ένα ασφαλισμένο ο οποίος υπέστη ατύχημα μικρού κόστους θα είναι αδίκως όμοια με την ποινή που επιβλήθηκε σε ένα ασφαλισμένο ο οποίος υπέστη ατύχημα μεγάλου κόστους. Παρακινούμενοι από αυτό το γεγονός, θέσαμε ως πρώτο στόχο του Κεφαλαίου 4 την ενσωμάτωση της σφοδρότητας των απαιτήσεων του ασφαλιζομένου προς την ασφαλιστική εταιρεία στα βέλτιστα BMSs του Κεφαλαίου 3, τα οποία βασίζονται στα a posteriori κριτήρια ταξινόμησης των ασφαλισμένων. Για αυτό τον σκοπό, υποθέτουμε ότι το ύψος των ζημιών κατανέμεται με βάση την κατανομή Pareto, ακολουθώντας την θεώρηση των Frangos and Vrontos (2001). Ο δεύτερος στόχος του Κεφαλαίου 4 είναι η δημιουργία ενός βέλτιστου BMS το οποίο περιλαμβάνει μια συνιστώσα για τη συχνότητα και μια συνιστώσα για τη σφοδρότητα των ατυχημάτων και βασίζεται στα a priori και στα a posteriori κριτήρια κατηγοριοποίησης των ασφαλιζομένων, καθώς ενσωματώνει στο αρχικό σύστημα και τις a priori επεξηγηματικές μεταβλητές για κάθε ασφαλισμένο. Η συνιστώσα της συχνότητας μοντελοποιείται υποθέτοντας ότι ο αριθμός των απαιτήσεων του ασφαλιζόμενου προς την ασφαλιστική εταιρεία κατανέμεται με βάση τα Αρνητικό Διωνυμικό Τυπου I, Poisson-Inverse Gaussian και Sichel GAMLSS. Η συνιστώσα της σφοδρότητας μοντελοποιείται υποθέτοντας ότι το ύψος των ζημιών κατανέμεται με βάση το Pareto GAMLSS. Το σύστημα αυτό προκύπτει ως μια συνάρτηση των ετών κατά τα οποία ο ασφαλισμένος βρίσκεται στο χαρτοφυλάκιο της εταιρείας, του αριθμού των ατυχημάτων του, του μεγέθους της ζημίας καθενός από αυτά τα ατυχήματα και των στατιστικά σημαντικών επεξηγηματικών μεταβλητών για τον αριθμό των ατυχημάτων και για το μέγεθος της ζημίας καθενός από αυτά τα ατυχήματα. Επιπλέον, παρουσιάζουμε μια πιο γενικευμένη μορφή από εκείνη που προτάθηκε στους Frangos and Vrontos (2001). Εν κατακλείδι, στο Κεφάλαιο 5 δίνουμε έμφαση στην ανάλυση της συνιστώσας της συχνότητας και της συνιστώσας της σφοδρότητας ενός βέλτιστου BMS χρησιμοποιώντας πεπερασμένες μείξεις κατανομών και μοντέλων παλινδρόμησης (βλέπε Mclachlan και Peel, 2000 και Rigby και Stasinopoulos, 2009), καθώς οι μέθοδοι αυτές, με εξαίρεση τον Lemaire (1995), δεν έχουν μελετηθεί στην βιβλιογραφία για BMSs. Η συνιστώσα της συχνότητας μοντελοποιείται με τη χρήση μιας πεπερασμένης μείξης Poisson, Αρνητικών Διωνυμικών και Delaporte κατανομών, ενώ η συνιστώσα της σφοδρότητας μοντελοποιείται με τη χρήση μιας πεπερασμένης μείξης Εκθετικών, Γάμμα, Weibull και Γενικευμένων Βήτα Τύπου ΙΙ (GB2) κατανομών. Παρουσιάζουμε επίσης, την περίπτωση μιας πεπερασμένης μείξης Αρνητικών Διωνυμικών κατανομών και μιας πεπερασμένης μείξης κατανομών Pareto, ενημερώνοντας τον posterior μέσο. Τα γενικευμένα BMS που προτείνουμε ενσωματώνουν επαρκώς την ταξινόμηση των κινδύνων σε κλάσεις, καθώς και την εμπειρική τιμολόγηση, λαμβάνοντας υπόψη τόσο τα a priori όσο και τα a posteriori χαρακτηριστικά του κάθε ασφαλισμένου.

Table of Contents

List of Tables

Chapter 1

Introduction

1.1 Motor Insurance

The history of the automobile dates back to 1769, with the creation of the first self-propelled steam engined road vehicle capable of human transport. In 1806, the first cars powered by an internal combustion engine running on fuel gas appeared, which led to the introduction in 1885 of the ubiquitous modern gasoline (or petrol) fueled internal combustion engine. Since then, the number of motor vehicles has grown constantly. In 2010 the number of cars on the world's roads was estimated to exceed 1.015 billion as compared with a few thousand at the end of the 19^{th}century. Unfortunately, a result of this has been a great increase in the rates of accidents and casualties, with hundreds of thousands killed and many times this number injured each year. As a result, automobile third-party liability insurance is required by law in most countries for a vehicle to be allowed on the public road network. Compulsory third party coverage provides protection if the vehicle's owner causes harm to another party, who recovers their cost from the policyholder. In developed countries third party coverage represents a considerable share of the yearly nonlife premium collection. This share becomes more prominent when first party coverages are considered (i.e. medical benefits, uninsured or underinsured motorist coverage, and collision and other than collision insurance). Moreover, insurance companies maintain large data bases, recording policyholders' characteristics as well as claim histories. The economic importance and availability of detailed information have to do with the fact that a large body of actuarial literature is devoted to this line of business.

Most actuaries worldwide are required to design tariff structures to spread the claim burden among policyholders in a fair way. This is the objective of a ratemaking process. A prime actuarial ratemaking principle is cost-based pricing of individual risks. An estimate of the future costs related to the insurance coverage is the amount policyholders are required to pay. The price of an insurance policy is defined by the pure premium approach as the ratio of estimated costs of all future claims against the coverage that the insurance policy provides while it is in effect to the risk exposure, plus expenses. The ratemaking of the property/casualty is based both on a claim frequency distribution and a loss distribution. The claim frequency is defined as the number of incurred claims per unit of earned exposure. The exposure is measured in car-years for motor third party liability insurance, the rate manual lists rates per car-year. The average payment per incurred claim is the average loss severity. Under mild conditions, the pure premium equals the product of the average claim frequency multiplied by the average loss severity. In liability insurance, several years are usually required for larger claims to be settled. Therefore, much of the data available for recent accident years will be incomplete, as the final claim cost will be unknown. In this case, a final cost estimate can be obtained by loss development factors and the average loss severity is based on incurred loss data. In contrast to paid loss data, which are purely objective, representing the company's actual payments, incurred loss data include subjective reserve estimates. Large claims have to be analyzed carefully by the actuary due to the fact they represent a considerable share of the insurer's yearly expenses.

1.2 A Priori Risk Classification

Insurance companies must maintain cross subsidies between different risk categories in order to remain competitive in the current market environment. Consequently, a constant challenge within the actuarial profession is that of the construction of a fair tariff structure. In light of the heterogeneity within a motor insurance portfolio, an insurance company should not apply the same premium to all insured risks. Otherwise the solvability of the company may be undermined by the concept known as adverse selection. 'Good' risks, i.e. those with low risk profiles, will pay too much and leave the company, whereas 'bad' risks will be encouraged to be insured there due to the favorable tariff. The idea behind risk classification is to partition all policies into homogenous classes with all policyholders belonging to the same class paying the same premium. Every time an additional rating factor is used by a competitor, the partition must be adjusted so that the best drivers with respect to this factor will not be lost. Because of this, we can understand that insurance companies use so many factors, not because it is required by actuarial theory, but due to competition among insurers.

Moreover, in a free market insurance companies have to use a rating structure that matches the premiums for the risks as closely as possible as the rating structures used by competitors. This entails that virtually every available classification variable correlated to the risk must be used. Failing to do so would mean sacrificing the chance to select against competitors and incurring the risk of suffering adverse selection by them. Thus the competition between insurers leads to more and more partitioned portfolios and not actuarial science. Social disasters are also frequently caused by this trend towards more risk classification. For instance, drivers categorized as 'bad' are tempted to drive without insurance as they do not find coverage for a reasonable price. At this point we should mention that even if a correlation exists between the rating factor and the risk covered by the insurer, there may be no causal relationship between that factor and risk. Requiring that insurance companies establish such a causal relationship to be allowed to use a rating factor is subject to debate.

Property and liability motor vehicle insurers use classification plans for the creation of risk classes. The classification variables introduced to partition risks into cells are called a-priori variables (as their values can be determined before the policyholder starts to drive). In motor third-party liability insurance, they commonly include the age, gender, and occupation of the policyholders the type and use of their car, and other personal data. Thus premiums for motor liability coverage often vary according to these individual characteristics. If any of these classification variables are misrepresented by the policyholders in their declaration, they can lose the coverage when they are involved in an accident so a strong incentive for accurate reporting of risk characteristics exists. Recent actuarial literature research assumes that the risks can then be rated a priori using generalized linear models (GLM). Typical response variables involved in this process are the claim frequency and its corresponding severity. The method can be summarized as follows: As the base cell we choose one risk classification cell. Normally it has the largest amount of exposure. The rate for the base cell is referred to as the base rate. A variety of risk classification variables define other rate cells. For each risk classification variable, there is a vector of differentials; with the base cell characteristics always assigned one hundred per cent. References for a priori risk classification include, for example, Dionne and Vanasse (1989, 1992), Dean, Lawless, and Willmot (1989), Denuit and Lang (2004), Gourieroux and Jasiak (2004), Yip and Yau (2005), Denuit et al. (2007) and Boucher et al. (2007, 2008).

1.3 Bonus-Malus Systems

The trend towards more classification factors has led the supervising authorities to exclude from the tariff structure certain risk factors, even though they were significantly correlated to losses. Classifications based on items that are beyond the control of the insured such as gender or age, were banned by many states. Moreover, many important factors cannot be taken into account a priori when pricing motor third party liability insurance products. For instance, reaction times, aggressive driving behavior or theoretical and practical driving experience are difficult to integrate into a priori risk classification. As a result, heterogeneity is still observed in tariff cells despite the use of many classification variables. Therefore, the idea to use the past number of claims in order to correct the inadequacies resulting from an a priori rating system. Experience rated or Bonus-Malus Systems, BMSs in short, are systems that impose penalties on policyholders responsible for one or more accidents by premium surcharges or maluses and reward policyholders with no claims by giving discounts or bonuses. Their prime objective, apart from promoting careful driving amongst policyholders, is to assess individual risks more accurately so that on a long term basis everyone will pay a premium corresponding to their own claim frequency. The mathematical definition of a Bonus-Malus system was introduced by Loimaranta (1972) and assumes that it can be modelled using conditional Markov chains, provided they possess a certain "memory-less" property: the knowledge of the present class and of the number of claims of the present year suffices to determine the class for the next year. More precisely, the Markov property is satisfied by the Bonus-Malus systems as follows: the future level of year $t+1$, depends on the present level of year t and the number of accidents reported during that year and does not depend on the past i.e. the claim frequency history and the levels occupied during years $1, 2, ..., t-1$. Thus, we can determine the optimal relativities using an asymptotic criterion based on the stationary distribution or using transient distributions. More details on these Bonus-Malus systems can be found in Norberg (1976) and comprehensively in Lemaire (1995). Furthermore, these systems are fair since a priori ratemaking penalizes policyholders who may be considered as bad risks, even if they are actually very good drivers and they will never cause an accident, whereas BMSs adjust the amount of premium using the individual claim record. A balance between the likelihood of being a good but unlucky driver who suffered a claim and the likelihood of being a truly bad driver, to whom the insurance company should make an increase in the premium payable, is made by the use of actuarial credibility models. Also, BMSs may be more acceptable to policyholders than seemingly arbitrary a priori classifications as it is fair to correct the inadequacies of the a priori system by using a more adequate system.

In the United Kingdom, discounts for claim-free driving were awarded much earlier, in 1910, but their initial intention was merely to encourage the renewal of a policy with the same company rather than reward careful driving. Grenander (1957 a, b) through his pioneering works was the first to provide theoretical treatments of Bonus-Malus systems. The first ASTIN colloquium was held in France in 1959 and was exclusively devoted to no-claim discounts in insurance, with particular reference to the motor business.

Many countries around the world use various Bonus-Malus systems. A typical form of no-claim bonus in the United Kingdom is defined as follows: An extra year of bonus is earned by drivers for each year they remain without claims at fault up to a maximum of four years, but two years bonus is lost each time they report a claim at fault. In such a system, maximum bonus is achieved in only a few years and the majority of mature drivers have maximum bonus. In Continental Europe the Bonus-Malus systems that are used tend to be more elaborate. Bonus-Malus scales consist of a finite number of levels, each with its own relativity or relative premium. Then the amount of premium paid by a policyholder is the product of a base premium with the relativity corresponding to the level occupied in the scale. New policyholders enter at a specific level and the policy moves up or down according to transition rules of the Bonus-Malus system at the beginning of the next year. If a Bonus-Malus system is in force, all policies in the same tariff class are partitioned according to the level they occupy in the Bonus-Malus scale. Hence, the Bonus-Malus systems can be considered as a refinement of a priori risk evaluation as according to individual past claims histories, they split each risk class into a number of subcategories.

During the 20th century, a uniform Bonus-Malus system was imposed on all the companies in most European countries. In 1994, the European Union directed that the mandatory Bonus-Malus systems must be dropped by its entire member countries because they were in contradiction to the total rating freedom implemented by the Third Directive and so competition between insurers was reduced. Since then, Belgium for instance, has dropped its mandatory system but the former uniform system is still operated by many companies there, with minor modifications for the policyholders who occupy the lowest levels in the scale. However, in other European countries, like Spain and Portugal, insurers compete on the basis of Bonus-Malus systems. Nevertheless, the mandatory systems in France and Grand Duchy of Luxembourg are still in force as in 2004 the European Court of Justice decided that both these mandatory systems were not in violation of the rating freedom imposed by European legislation. Thus, French law still imposes on insurers operating in France a unique Bonus-Malus system which is not based on a scale but instead uses the concept of an increase-decrease coefficient: a malus of 25 % per claim and a bonus of 5 % per claim-free year are implied, so a base premium is assigned to each policyholder and it is adapted according to the claims number reported to the insurer, multiplying by 1.25 the premium each time a claim is reported, and by 0.95 per claim-free year.

Another important issue is the actuarial and economic justifications for BMSs. As we have already mentioned, BMSs match individual premium to risk and increase incentives for road safety by taking into consideration the claims history record of each policyholder. They can be justified by asymmetrical information between the insurance company and the policyholders. Indeed, they encourage policyholders to drive carefully (i.e., they counteract moral hazard) and respond to adverse selection in automobile insurance. In the context of compulsory motor third party liability insurance, adverse selection occurs when the policyholders take advantage of the information they have about their claim behavior, known to them but unknown to the insurer. However, the problem of adverse selection is not as important as the problem of moral hazard when insurance companies charge the same premium amount to every policyholder. In a deregulated environment with companies using many risk classification factors the situation deteriorates and adverse selection becomes unavoidable as very heterogeneous driving behaviors are observed among policyholders who share the same a priori variables. Experience rating is a response to both adverse selection and moral hazard, penalizing the more numerous claims of those with more dangerous driving patterns. It is interesting to confront the approaches of economists and actuaries to experience rating. In economic literature, discounts and penalties are introduced mainly to counteract the inefficiency which arises from moral hazard. In actuarial literature, the main purpose is to better assess individual risk so that everyone will pay, in the long run, a premium corresponding to their own claim frequency. Since the penalty induced by the Bonus-Malus system is in general independent of the claim amount, a crucial issue for the policyholder is therefore to decide whether it is profitable or not to report small claims. Low cost claims are likely to be defrayed by the policyholders themselves, and not reported to the company. This phenomenon, known as the hunger for bonus, limits claim handling costs since small claims are not reported to the insurer.

1.3.1 Optimal Bonus-Malus Systems

In the previous section the BMSs were defined as systems that are usually modeled in the framework of homogeneous Markov chains. Under the aggressiveness and competitiveness of the insurance markets, this assumption is not realistic. A basic interest of actuarial literature is the construction of an optimal or 'ideal' BMS defined as a system obtained through Bayesian analysis. A BMS is called optimal if it is:

1. Financially balanced for the insurer. That is the total amount of bonuses is equal to the total amount of maluses.

2. Fair for the policyholder. That is each policyholder pays a premium proportional to the risk that he imposes to the pool.

Furthermore, optimal BMSs can be divided into those based only on the a posteriori classification criteria and those based both on the a priori and the a posteriori classification criteria.

Typically, a posteriori classification criteria include the number and severity of individual claims, while a priori classification criteria include variables such as characteristics of the policyholder and automobile.

Lemaire (1995) developed the design of an optimal BMS based on the number of claims of each policyholder, following the game-theoretic framework introduced by Bichsel (1964) and Buhlmann (1964). Given that the premium is proportional to the unknown claim frequency and an estimate has to be employed instead, the insurer faces a loss. Minimizing this loss gives the optimal estimate of the policyholder's claim frequency. In his system each policyholder must pay a premium proportional to his unknown claim frequency. Tremblay (1992) designed an optimal BMS using the quadratic error loss function, the zero-utility premium calculation principle and the Poisson-Inverse Gaussian distribution to approximate the number of claims. Coene and Doray (1996) developed a method of obtaining a financially balanced BMS by minimizing a quadratic function of the difference between the premium for an optimal BMS with an infinite number of classes, weighted by the stationary probability of being in a certain class and by imposing various constraints on the system. Walhin and Paris (1997) obtained an optimal BMS using as the claim frequency distribution the Hofmann's distribution, which encompasses the Negative Binomial and the Poisson-Inverse Gaussian, and also using as a claim frequency distribution a finite Poisson mixture.

The models discussed above are functions of time and of past claim frequency and do not take into consideration the characteristics of each individual. In this way the premiums do not vary simultaneously with other variables that affect the claim frequency distribution. For instance, let us suppose that the age variable has a negative effect on the expected number of claims. This would imply that insurance premiums should decrease with age, even though premium tables derived from BMS based only on the a posteriori criteria, do not allow for a variation of the statistically significant variable of age. Dionne and Vanasse (1989, 1992) developed a BMS that integrates a priori and a posteriori information on an individual basis. For this purpose they used the Negative Binomial regression model for assessing claim frequency. The resulting generalized system was derived as a function of the years that the policyholder is in the portfolio, the number of accidents and the individual characteristics which are significant for the number of accidents. Picech (1994) and Sigalotti (1994) constructed a BMS that incorporates the a posteriori and the a priori classification variables, 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 (1997) designed 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 (1998) developed the design of optimal BMS from different types of claims, such as claims at fault and claims not at fault.

The Bonus-Malus systems mentioned above, assign to each policyholder a premium based on the number of their accidents but the size of loss that each accident incurred is not considered. This is unfair, because a policyholder who underwent an accident with a small size of loss is penalized in the same way with a policyholder who had an accident with a big size of loss.

Among the BMSs that take severity into consideration are those designed from Picard (1976), Lemaire (1995), Pinquet (1997), Frangos and Vrontos (2001), Pitrebois et al. (2006) and Mahmoudvand and Hassani (2009).

The Construction of an Optimal BMS in the Form of a Statistical Game

The Poisson distribution was discovered by Simeon-Denis Poisson (1781–1840). Typically, a Poisson random variable is a count of the number of events that occur in a certain time interval or spatial area. In motor third party liability insurance, the accident pattern of drivers conforms to a Poisson distribution and we adopt the Poisson to model the claim frequency of individual policyholders (see Lemaire, 1995).

Following the setup of Lemaire (1995), we present the construction of an optimal BMS, based only on claim frequency, as a series of statistical games between the actuary and the nature. Consider a policyholder who is observed for t years and denote by k_j, $j = 1, ..., t$, the number of claims in which they were at fault. So their claim frequency history will be in a form of a vector $(k_1, ..., k_t)$. We assume that the claim frequency doesn't change over time and that k_j are the realizations of independent and identically (i.i.d) random variables K_j distributed according to a Poisson(λ) distribution with probability density function (pdf) given by

$$P\left(K_j = k\right) = \tfrac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, ..$$

With each group of observations $k_1, ..., k_t$, we must designate a number $\lambda_{t+1}\left(k_1, ..., k_t\right)$, which is the best estimate of the Poisson density λ at $t + 1$. The decision problem can be presented as follows: Given a series of i.i.d random variables $K_1, ..., K_t, ...$, determine a set of functions $\lambda_{t+1} = \lambda_{t+1}\left(k_1, ..., k_t\right), \, t = 0, 1, ...,$ that estimate λ optimally and sequentially.

The design of an optimal BMS can thus be presented as a series of statistical games between the actuary and the nature, where the definition of each game is the following

$$G_{t+1} = \left(\Lambda_0, S_{t+1}, R_{t+1}\right),$$

where

- Λ_0, the space of strategies of nature, is the interval $[0, \infty)$, in which the unknown parameter λ belongs,
- S_{t+1}, the space of strategies of the actuary at time $t + 1$, is a class of decision functions $\lambda_{t+1}\left(k_1, ..., k_t\right)$, which associates a point $\lambda_{t+1}\epsilon\Lambda_0$ with each vector $(k_1, ..., k_t)$,
- $R = R_{t+1}\left(\lambda_{t+1}, \lambda\right)$, the risk function of the actuary at $t + 1$, is the expectation of the loss $L_{t+1}\left(\lambda_{t+1}, \lambda\right)$ that the actuary incurs when they take a decision $\lambda_{t+1}\left(k_1, ..., k_t\right)$ while nature is in state λ. The loss function $L_{t+1}\left(\lambda_{t+1}, \lambda\right)$ is a non-negative function of the difference between λ_{t+1} and λ so we have

$$R_{t+1}\left(\lambda_{t+1}, \lambda\right) = E\left[L_{t+1}\left(\lambda_{t+1}, \lambda\right)\right] = \sum L_{t+1}\left(\lambda_{t+1}, \lambda\right) P\left(k_1, ..., k_t|\lambda\right),$$

defining $\sum$ as the sum over all claim histories $(k_1, ..., k_t)$ and $P(k_1, ..., k_t|\lambda)$ as the the $t-$dimensional distribution of the number of claims for a policyholder characterized by their claim frequency λ.

Thus the set of the G_t, $t = 1, 2, 3, ...$ forms the statistical game

$$G = (\Lambda_0, S, R),$$

where $S = S_1 \times S_2 \times ... \times S_t \times ... \times$ is the Cartesian product of the S_t, and where

$$R = R(\lambda_1, ..., \lambda_t, ...; \lambda) = \sum_{t=1}^{\infty} R_t(\lambda_t, \lambda) = \sum_{t=1}^{\infty} E[L_t(\lambda_t, \lambda)],$$

is the total expected loss of the actuary.

A series $(\lambda_1^*, ..., \lambda_t^*, ...)$ is called *uniformly optimal* if

$$R(\lambda_1^*, ..., \lambda_t^*, ...; \lambda) \leq R(\lambda_1, ..., \lambda_t, ...; \lambda),$$

for each value of λ and for all $(\lambda_1, ..., \lambda_t)$.

However, in general, a uniformly optimal series does not exist since in a given heterogeneous portfolio each driver is characterized by the value of their parameter λ. For instance, the optimal BMS for a good risk policyholder, who has low claim frequency λ, is very different from the best system for a bad risk insured, who has high claim frequency λ.

In light of unobserved heterogeneity in the portfolio we assume that λ is the observed value of a random variable Λ with density function $u(\Lambda)$ called the structure function. The resulting distribution of the number of claims k

$$P(K_j = k) = \int_0^{\infty} \tfrac{e^{-\lambda}\lambda^k}{k!} u(\lambda)\, d\lambda, k = 0, 1, 2, ...$$

is called a mixed Poisson distribution[1]. Based on this assumption, an alternative is to minimize the average risk of the actuary

$$R(\lambda_1, ..., \lambda_t, ...) = \int_0^{\infty} R(\lambda_1, ..., \lambda_t, ...; \lambda)\, u(\lambda) d\lambda.$$

A series $(\lambda_1^*, ..., \lambda_t^*, ...)$ is defined as optimal if

$$R(\lambda_1^*, ..., \lambda_t^*, ...; \lambda) = \inf_{(\lambda_1, ..., \lambda_t, ...) \in S} R(\lambda_1, ..., \lambda_t, ...).$$

[1]Poisson mixtures are well-known counterparts to the simple Poisson distribution for the description of heterogeneous populations.

Based on a theorem of Wald and Wolfowitz (1951), it can be shown that an optimal solution of the above equality always exists.

The conditional distribution of $\lambda|k_1, ..., k_t$ is called the posterior structure function of λ, and is denoted as $u(\lambda|k_1, ..., k_t)$. It tells the insurer what the next year claim frequency might be given the information contained in past claim frequency history and it is the relevant distribution for risk analysis, management and decision making. Applying the Bayes theorem, one can find that $u(\lambda|k_1, ..., k_t)$ is equal to

$$u(\lambda|k_1, ..., k_t) = \frac{P(k_1, ..., k_t|\lambda)\, u(\lambda)}{\bar{P}(k_1, ..., k_t)} = \frac{P(k_1, ..., k_t|\lambda)\, u(\lambda)}{\int\limits_0^\infty P(k_1, ..., k_t|\lambda)\, u(\lambda)\, d\lambda},$$

where $\bar{P}(k_1, ..., k_t)$ is the distribution of claims during the t years of observation in the portfolio.

Thus, we must minimize

$$R(\lambda_1, ..., \lambda_t, ...) = \sum_{t=0}^{\infty} \int\limits_0^\infty \sum L_{t+1}(\lambda_{t+1}, \lambda)\, P(k_1, ..., k_t|\lambda)\, u(\lambda) d\lambda$$
$$= \sum_{t=0}^{\infty} \sum \int\limits_0^\infty L_{t+1}(\lambda_{t+1}, \lambda)\, P(k_1, ..., k_t)\, u(\lambda|k_1, ..., k_t) d\lambda.$$

Since the loss function $L_{t+1}(\lambda_{t+1}, \lambda) \geq 0$, the above expression has to be be minimized for each t and for each $(k_1, ..., k_t)$,

$$\int\limits_0^\infty L_{t+1}(\lambda_{t+1}, \lambda)\, u(\lambda|k_1, ..., k_t) d\lambda,$$

which is the a posteriori risk of Λ.

From the previous it is obvious that the Bayesian approach to this minimization problem is to find a loss function $L_{t+1}(\lambda_{t+1}, \lambda)$ to measure the loss incurred by estimating the value of the parameter λ as λ_{t+1}.

The loss is an increasing function of the size of the actuary's error. When $\lambda_{t+1} < \lambda$, the insured is undercharged, and the insurer will make a loss, while when $\lambda_{t+1} > \lambda$, the policyholder is overcharged, and the insurer risks losing the policyholder from his portfolio. With the aim of penalizing large mistakes more, it is assumed that the loss function is a non-negative convex function of the error. The actuary estimates correctly the insured's claim frequency and no error is made when $\lambda_{t+1} = \lambda$, i.e. the loss is zero and strictly positive in every other case.

Various loss functions are to be found in the statistical literature. The most classical choice is the quadratic error loss function

$$L_{t+1}\left(\lambda_{t+1},\lambda\right)=\left(\lambda_{t+1}-\lambda\right)^{2}.$$

In this case we must find the minimum of

$$\int_{0}^{\infty}\left(\lambda_{t+1}-\lambda\right)^{2}u(\lambda|k_{1},...,k_{t})d\lambda.$$

Thus, the optimal choice of λ at time $t+1$, $\hat{\lambda}_{t+1}\left(k_{1},...,k_{t}\right)$, for a policyholder who had a claim frequency history $k_{1},...,k_{t}$ is given by

$$\hat{\lambda}_{t+1}\left(k_{1},...,k_{t}\right)=\int_{0}^{\infty}\lambda u(\lambda|k_{1},...,k_{t})d\lambda.$$

This is the mean of the posterior structure function of λ, $E(\lambda|k_{1},...,k_{t})$. Thus a policyholder or a group of policyholders who underwent claims history $(k_{1},...,k_{t})$ will have to pay a net Bayesian credibility premium equal to their a posteriori claim frequency. By definition, a BMS designed using Bayesian analysis is called an optimal BMS.

1.4 Generalized Linear Models

Generalized linear models (GLMs) were introduced by Nelder and Weddeburn (1972). The history of GLMs in actuarial statistics goes back to the actuarial illustrations in the standard text presented by McCullagh and Nelder (1989). For an overview of the use of GLMs in typical problems in actuarial statistics see, for example, Haberman and Renshaw (1996). References for the use of GLM in a priori risk classification and experience rating include, for instance, Dionne and Vanasse (1989, 1992), Frangos and Vrontos (2001), Pitrebois et al. (2006), Denuit et al. (2007) and Boucher et al. (2008). Furthermore, with the exception of Jørgensen and Paes de Souza (1994), all actuarial analyses of the pure premium have examined claim frequencies and costs separately.

In what follows we present a short summary of the main characteristics of GLMs. For a broad introduction to GLMs we refer to McCullagh and Nelder (1989). The GLMs extend the framework of linear models to the class of distributions from the exponential family. Firstly, the normal distribution for the response variable Y is replaced by an exponential family distribution (denoted EF in general). Secondly, they provide a way around the transformation of data. Instead of a transformed data vector, a transformation of the mean is modelled as a linear function of explanatory variables through a (possibly non-linear) link function. The EF is very flexible and a considerable variety of possible outcome measures (such as continuous, count, binary and skew data) can be modelled within this framework.

A GLM consists of the following components:

1. The response variable Y has a distribution in the EF, with density function taking the form
$$f(y;\theta,\phi)=\exp\left\{\int\frac{[y-\mu(\theta)]}{\phi V(\mu)}d\mu(\theta)+c(y,\phi)\right\}, \tag{1.1}$$
where θ is called the natural parameter, ϕ is a dispersion parameter, $E(Y)=\mu=\mu(\theta)$ and $V(Y)=\phi V(\mu)$, for a given variance function V and a known bivariate function c. The form of (1.1) includes many important distributions such as the Normal, Poisson, Gamma, Inverse Gaussian and Negative distributions having variance functions $V(\mu)=1,\mu,\mu^2,\mu^3$ and $\mu+\frac{\mu}{\phi}$ respectively.

2. For a random sample $(Y_1,...,Y_n)$, the linear predictor is defined as
$$\eta_i=\mathbf{X}_i^T\boldsymbol{\beta}, i=1,...,n, \tag{1.2}$$
for some vector of parameters $\boldsymbol{\beta}=\left(\beta_1,...,\beta_p\right)^T$ and covariate $\mathbf{X}_i=(x_{i1},...,x_{ip})^T$ associated to Y_i.

3. A monotonic differentiable link function g describes how the expected response $\mu_i=E(Y_i)$ is related to the linear predictor $\boldsymbol{\eta}_i$.
$$g(\mu_i)=\eta_i, i=1,...,n. \tag{1.3}$$

For an observed independent random sample $(y_1,...,y_n)$ the log-likelihood function l of the vector of parameters $\boldsymbol{\beta}$ is given by

$$l(\boldsymbol{\beta})=\log(L(\boldsymbol{\beta}))=\sum_{i=1}^{n}\left\{\int\frac{[y_i-\mu_i(\theta)]}{\phi V(\mu_i)}d\mu_i(\theta)+c(y_i,\phi)\right\}, \tag{1.4}$$

where L denotes the likelihood function.
The derivative of l is equal to

$$\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}}=\sum_{i=1}^{n}\frac{dl(\boldsymbol{\beta})}{d\mu_i}\frac{d\mu_i}{d\boldsymbol{\beta}}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{\phi V(\mu_i)}\frac{d\mu_i}{d\mathbf{X}_i^T\boldsymbol{\beta}}\frac{\mathbf{X}_i^T\boldsymbol{\beta}}{d\boldsymbol{\beta}}, \tag{1.5}$$

where

$$\frac{d\mu_i}{d\mathbf{X}_i^T\boldsymbol{\beta}}=\frac{dg^{-1}\left(\mathbf{X}_i^T\boldsymbol{\beta}\right)}{d\mathbf{X}_i^T\boldsymbol{\beta}}=\frac{1}{g'(\mu_i)}. \tag{1.6}$$

Hence

$$\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{\phi V(\mu_i)}\frac{1}{g'(\mu_i)}\mathbf{X}_i^T. \tag{1.7}$$

Note that if Y_i has a Normal distribution, then $g'(\mu_i)=1$ and $V(\mu_i)=1$ for all i. Setting $\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}}=0$ yields $\sum_{i=1}^{n}\mathbf{X}_i\left(y_i-\mathbf{X}_i^T\boldsymbol{\beta}\right)=0$. However, in other EF cases a closed form solution to this system of p equations does not exist. Instead, the maximum likelihood estimator is obtained numerically, using iterative algorithms such as the Newton-Raphson or Fisher scoring methods (for more details, see McCullagh and Nelder, 1989).

1.5 Outline of the Thesis

The research projects we present in this dissertation deal with the important actuarial subjects of risk classification and experience rating in motor third-party liability insurance. Due to the quantitative nature of this area of applied statistics, the successful application of up-to-date statistical techniques in the analysis of motor insurance data is indispensable in the palette of actuarial skills. In this thesis, we apply concepts from both a priori risk classification and Bonus-Malus systems and link our work with contributions in this area.

In Chapter 2 we extend recent actuarial literature research which uses generalized linear models for pricing risks through ratemaking based on a priori risk classification (see, for example, Denuit et al. 2007 & Boucher et al., 2007, 2008). For this purpose we consider the generalized additive models for location scale and shape (GAMLSS). The GAMLSS models were introduced by Rigby and Stasinopoulos (2005) and Akantziliotou, Rigby, and Stasinopoulos (2002) as an alternative to the GLM framework. In the GAMLSS the systematic part of the model is expanded to allow modeling of location and shape parameters. Therefore, both mean and variance may be assessed by choosing a marginal distribution and building a predictive model using ratemaking factors as independent variables. In light of a priori ratemaking the GAMLSS are used to model the frequency and the severity of claims. Specifically, we assume that the number of claims is distributed according to the Poisson, Negative Binomial Type II, the Delaporte, Sichel and Zero-Inflated Poisson GAMLSS and that the losses are distributed according to the Gamma, Weibull, Weibull Type III, Generalized Gamma and Generalized Pareto GAMLSS as these models have not been studied in risk classification literature. Specification tests to select the optimal classification model for each case are considered. Differences between these models are analyzed through the mean and the variance of the annual number of claims and the costs of claims for the policyholders, who belong to different risk classes. The resulting a priori premiums are calculated via the expected value and standard deviation principles, assuming that the claim frequency and severity components are independent.

With Chapters 3, 4 and 5 we switch from a priori ratemaking techniques to experience rated or Bonus-Malus Systems (BMSs). These systems are much in line with the concept of fairness: a priori ratemaking penalizes policyholders who are characterized as bad drivers, even if they are actually very good drivers and will never cause any accident, whereas experience rating adjusts the amount of premium using the individual claim record. A basic interest of actuarial literature is the construction of an optimal or 'ideal' BMS, defined in 1.3.2 as a system obtained through Bayesian analysis. The study of such systems using different statistical models will be the main objective of these chapters.

In Chapter 3 our first contribution is the development of an optimal BMS using the Sichel distribution for assessing claim frequency. This system is proposed as an alternative to the optimal BMS obtained by the Negative Binomial model (Lemaire, 1995). The Sichel distribution differs from the Negative Binomial one by using a Generalized Inverse Gaussian (GIG) mixing distribution for the parameter of the Poisson density, i.e. the expected claim frequency, instead of the Gamma one, which the derivation of the Negative Binomial distribution is based on. An additional advantage of the Sichel model is that it can be considered as a candidate model for highly dispersed count data. We also consider the optimal BMS provided by the

Poisson-Inverse Gaussian (PIG) distribution, which is a special case of the Sichel distribution. Our second contribution is the development of a generalized BMS that integrates the a priori and the a posteriori information on an individual basis, extending the framework developed by Dionne and Vanasse (1989, 1992). This is achieved by using the Sichel GAMLSS for assessing claim frequency as an alternative to the Negative Binomial regression model of Dionne and Vanasse (1989, 1992). Furthermore, we consider the PIG GAMLSS for assessing claim frequency. With the aim of constructing an optimal BMS by updating the posterior mean claim frequency, we adopt the parametric linear formulation of these models and we allow only their mean parameter to be modelled as a function of the significant a priori rating variables for the number of claims. In the resulting generalized system, the premium is a function of the years that the policyholder is in the portfolio, the number of accidents and the significant a priori rating variables for the number of accidents.

In Chapter 4 our first objective is the integration of claim severity into the optimal BMSs based on the a posteriori criteria of Chapter 3. For this purpose we consider that the losses are distributed according to a Pareto distribution, following the framework proposed by Frangos and Vrontos (2001). The BMS resulting from the Sichel and Pareto models and that derived form the PIG and Pareto models are compared to the system provided by the Negative Binomial and Pareto models (Frangos and Vrontos, 2001). Our second objective is the development of a generalized BMS with a frequency and a severity component when both the a priori and the a posteriori rating variables are used. For the frequency component we assume that the number of claims is distributed according to the Negative Binomial Type I, Poisson Inverse Gaussian and Sichel GAMLSS. For the severity component we consider that the losses are distributed according to a Pareto GAMLSS. This system is derived as a function of the years that the policyholder was in the portfolio, their number of accidents, the size of loss of each of these accidents and of the statistically significant a priori rating variables for the number of accidents and for the size of loss that each of these claims incurred. Furthermore, we present a generalized form of the system obtained in Frangos and Vrontos (2001).

In Chapter 5 we put focus on both the analysis of the claim frequency and severity components of an optimal BMS using finite mixtures of distributions and regression models (see Mclachlan and Peel, 2000, and Rigby and Stasinopoulos, 2009). Finite mixture models are a popular statistical modelling technique given that they constitute a flexible and easily extensible model class for approximating general distribution functions in a semi-parametric way and accounting for unobserved heterogeneity. Finite mixture models have been widely applied in many areas, such as biology, biometrics, genetics, medicine and marketing. However, with the exception of Lemaire(1995), they have not been extensively studied in BMS literature. Our first contribution is the development of an optimal BMS based on the a posteriori frequency and severity component using various finite mixtures of distributions. For the frequency component we assume that the number of claims is distributed according to a finite Poisson, Delaporte and Negative Binomial mixture, and for the severity component we consider that the losses are distributed according to a finite Exponential, Gamma, Weibull and GB2 mixture. In this way we expand the setup of Lemaire (1995), who designed an optimal BMS based on the two component Poisson mixture distribution. Applying Bayes theorem we derive the posterior probability of the policyholder's classes of risk. Furthermore, we extend the setup of Frangos and Vrontos

(2001) for Negative Binomial and Pareto mixtures and derive the posterior distribution of both the mean claim frequency and the mean claim size, given the information we have about the claim frequency history and the claim size history for each policyholder for the time period they are in the portfolio. Our third contribution is the development of a generalized BMS that integrates the a priori and the a posteriori information on a individual basis extending the framework developed by Dionne and Vanasse (1989, 1992) and Frangos and Vrontos (2001). This is achieved by using finite mixtures of regression models. In the setup we consider, the heterogeneity in the data is accounted for in two ways. Firstly, the population heterogeneity is explained by choosing a finite number of unobserved latent components, each of which may be regarded as a sub-population. This is a discrete representation of heterogeneity in the data since the mean claim frequency and severity are approximated by a finite number of support points. Secondly, depending on the choice of the component distribution, heterogeneity can also accommodated within each component by including the explanatory variables in the mean rate function.

Concluding remarks and ideas for future research can be found in Chapter 6.

Chapter 2

A Priori Risk Classification for Claim Counts and Losses Using Generalized Additive Models for Location, Scale and Shape

2.1 Introduction

As we mentioned in Chapter 1, within the actuarial profession a major challenge is to design a tariff structure that will fairly distribute the burden of claims among policyholders. In light of the heterogeneity within a car insurance portfolio, an insurance company should not apply the same premium to all insured risks. Otherwise the phenomenon of adverse selection will undermine the solvability of the company. 'Good' risks, with low risk profiles, will pay too much and leave the company, whereas 'bad' risks are attracted by the favorable tariff. Every time an additional rating factor is used by a competitor, the actuary must adjust the partition in order to avoid losing the best drivers with respect to this factor. Because of this, we can understand why insurance companies use so many factors even though this is not required by actuarial theory, but instead is required by competition among insurers. The idea behind a priori risk classification is to split an insurance portfolio into classes that consist of risks with all policyholders belonging to the same class paying the same premium. In view of the economic importance of motor third party liability insurance in developed countries, actuaries have made many attempts to find a probabilistic model for the distribution of the number and costs of claims reported by policyholders.

Recent actuarial literature research assumes that the risks can be rated a priori using generalized linear models, GLM (see Nelder and Wedderburn, 1972) and generalized additive models, GAM (see Hastie and Tibshirani, 1990). For motor insurance, typical response variables in these regression models are the number of claims (or claim frequency) and its corresponding severity. References for a priori risk classification include, for example, Dionne and Vanasse (1989, 1992), Dean, Lawless and Willmot (1989), Denuit and Lang (2004), Yip and Yau (2005), and Boucher et al. (2007). Specifically, Dionne and Vanasse used a Negative Binomial Type I regression model, Dean, Lawless and Willmot used a Poisson-Inverse Gaussian regression model, Denuit and Lang used generalized additive models, Yip and Yau presented several parametric Zero-Inflated count distributions, and Boucher et al. presented a comparison of various Zero-Inflated Mixed Poisson and Hurdle Models. Also, a review of actuarial models for risk classification and insurance ratemaking can be found in Denuit et al. (2007).

In this Chapter, we extend this setup following the generalized additive models for location scale and shape (GAMLSS) approach of Rigby and Stasinopoulos (2005) as these models have not been studied in risk classification literature. In light of a priori ratemaking there is a substantial benefit in the GAMLSS approach since the GLM and GAM frameworks are extended by allowing all the parameters of the distribution of the claim frequency and severity response variables to be modelled as linear/non-linear or smooth functions of the explanatory variables. Therefore, both mean and variance may be assessed by choosing a marginal distribution and building a predictive model using all the available ratemaking factors as independent variables. A comprehensive actuarial application of GAMLSS can be found in Heller et al. (2007), where GAMLLS have been used for the statistical analysis of the total amount of insurance paid out on a policy. In the setup we consider, we adopt the parametric linear formulation of these models and we model risk heterogeneity as the distribution of frequency and/or severity of claims changes between clusters by a function of the level of ratemaking factors underlying the analyzed clusters. Specifically, we model the claim frequency using the Poisson, Negative Binomial Type II, Delaporte, Sichel and Zero-Inflated Poisson GAMLSS and the claim severity using the Gamma, Weibull, Weibull Type III, Generalized Gamma and Generalized Pareto GAMLSS. Our contribution puts focus on the comparison of these models through their variance values. To the best of our knowledge it is the first time that the variance of the claim frequency and severity is modelled in an actuarial context. Our analysis reveals that the differences in the variance values alter significantly the premiums calculated through the standard deviation principle. Furthermore, the variance of the claim frequency and severity is an important risk measure. Thus, GAMLSS modelling is justified because it enables us to use all the available information in the estimation of these values through the use of the important explanatory variables for the claim frequency and severity respectively.

The rest of this chapter proceeds as follows. Section 2.2 briefly discusses the basic concepts of the GAMLSS models and introduces the various claim frequency and severity distributions we use within this family of models. Section 2.3 contains an application to a data set concerning car-insurance claims at fault. Specifically, these classification models are compared on the basis of a sample of the automobile portfolio of a major company operating in Greece employing the Generalized Akaike Information Criterion (GAIC) which is valid for both nested or non-nested model comparisons (as suggested by Rigby and Stasinopoulos, 2005 & 2009). Furthermore, differences between these models are analyzed through the mean and the variance of the annual number of claims and the costs of claims of the policyholders who belong to different risk classes, which are formed by dividing the portfolio into clusters defined by the relevant ratemaking factors. Finally, the resulting premiums rates are calculated via the expected value and standard deviation principles with independence between the claim frequency and severity components assumed.

2.2 Generalized Additive Models for Location, Scale and Shape

In what follows we provide a short summary of the main characteristics of the Generalized additive models for location, scale and shape (GAMLSS). For a broad introduction to GAMLSS models the readers can refer to Rigby and Stasinopoulos (2005) and Akantziliotou, Rigby, and Stasinopoulos (2002).

The GAMLSS are semi-parametric regression type models. They are parametric, in the sense that they require a parametric distribution assumption for the response variable, and "semi" in the sense that the modelling of the parameters of the distribution, as functions of explanatory variables, may involve using non-parametric smoothing functions. These models were introduced as a flexible alternative to the popular generalized linear models, GLM (see Chapter 1), and generalized additive models, GAM (see Hastie and Tibshirani, 1990). In the GAMLSS the exponential family distribution assumption for the response variable Y is relaxed and replaced by a general distribution family, including distributions based on Box-Cox transformations (such as the Box-Cox t-distribution, Rigby and Stasinopoulos, 2004, or the Box-Cox power exponential distribution, Rigby and Stasinopoulos, 2006) and zero adjusted-distributions (such as the zero adjusted Inverse Gaussian distribution, which is useful for insurance data, see Heller et al., 2007). Another key feature of the GAMLSS is that the systematic part of the model is expanded to allow modelling not only of the mean (or location) but other parameters of Y as, linear and/or non-linear, parametric and/or additive non-parametric functions of explanatory variables and/or random effects. Common distribution parameters are location, scale, skewness and kurtosis but degrees of freedom (of a t-distribution) and zero inflation probabilities can be modelled as well. Thus, in the GAMLSS approach, the full conditional distribution of a multi-parameter model is related to a set of predictor variables of interest.

A GAMLSS model assumes independent observations y_i, for $i = 1, ..., n$ with probability density function $f\left(y_i|\boldsymbol{\theta}^i\right)$ conditional on $\boldsymbol{\theta}^i = (\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i}) = (\mu_i, \sigma_i, \nu_i, \tau_i)$ a vector of four distribution parameters, each of which can be a function to the explanatory variables. The $\mu_i, \sigma_i, \nu_i, \tau_i$ are referred to as the distribution parameters. The first two parameters μ_i and σ_i defined as location and scale parameters, while the parameters ν_i and τ_i are defined as shape parameters, although the model may be applied more generally to the parameters of any population distribution. Let $\mathbf{y}^T = (y_1, ..., y_n)$ be the vector of the response variable observations. Also, for $k = 1, 2, 3, 4$ let $g_k(.)$ be known monotonic link functions relating the k^{th} parameter θ_k to explanatory variables and random effects through an additive model given by

$$g_k\left(\boldsymbol{\theta}_k\right) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk}\boldsymbol{\gamma}_{jk}, \tag{2.1}$$

i.e.

$$g_k\left(\boldsymbol{\mu}\right) = \boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1}\boldsymbol{\gamma}_{j1}$$

$$g_k(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2}\boldsymbol{\gamma}_{j2}$$

$$g_k(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3}\boldsymbol{\gamma}_{j3}$$

$$g_k(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \mathbf{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4}\boldsymbol{\gamma}_{j4}$$

where $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}$ and $\boldsymbol{\eta}_k$ are vectors of length n, $\boldsymbol{\beta}_k^T$ is a parameter vector of length $J_k^{'}$, $\mathbf{X}_k$ is a fixed known design matrix of order $n \times J_k^{'}$, $\mathbf{Z}_{jk}$ is a fixed known $n \times q_{jk}$ design matrix and $\boldsymbol{\gamma}_{jk}$ is a q_{jk} dimensional random variable which is assumed to be distributed as $\boldsymbol{\gamma}_{jk} \sim N_{q_{jk}}\left(0, \mathbf{G}_{jk}^{-1}\right)$, where $\mathbf{G}_{jk}^{-1}$ is the (generalized) inverse of a $q_{jk} \times q_{jk}$ symmetric matrix $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\boldsymbol{\lambda}_{jk})$ which may depend on a vector of hyperparameters $\boldsymbol{\lambda}_{jk}$, and where if $\mathbf{G}_{jk}$ is singular then γ_{jk} is understood to have an improper prior density function proportional to $\exp\left(-\frac{1}{2}\boldsymbol{\gamma}_{jk}^T\mathbf{G}_{jk}\boldsymbol{\gamma}_{jk}\right)$.

The model in (2.1) allows us to model each distribution parameter as a linear function of explanatory variables and/or as linear functions of random effects.

There are several important sub-models of the GAMLSS. First, let $\mathbf{Z}_{jk} = \mathbf{I}_n$, where $\mathbf{I}_n$ is an $n\times\ n$ identity matrix, and $\boldsymbol{\gamma}_{jk} \equiv \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ for all combinations of j and k in (2.1), then the semi-parametric additive GAMLSS model is given by

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \tag{2.2}$$

where $\boldsymbol{\theta}_k$ for $k = 1, 2, 3, 4$ is used to represent the distribution parameter vectors $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}$ and $\boldsymbol{\tau}$. The function h_{jk} is an unknown function of the explanatory variable X_{jk} and $\mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$, $1, 2, ..., J_k$, where $\mathbf{x}_{jk}$ are also vectors of length n.

If there are no additive terms in any of the distribution parameters, we can have the parametric linear GAMLSS model,

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k, \tag{2.3}$$

In Eq. (2.2) and Eq. (2.3), replace $\mathbf{X}_k\boldsymbol{\beta}_k$ with $h_k(\mathbf{X}_k, \boldsymbol{\beta}_k)$, where h_k for $k = 1, 2, 3, 4$ are non-linear functions and $\mathbf{X}_k$ is a known design matrix of order $n \times J_k^{''}$, then the (2.4) and (2.5) are the non-linear semi-parametric additive and non-linear parametric GAMLSS models, respectively.

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = h_k(\mathbf{X}_k, \boldsymbol{\beta}_k) + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \tag{2.4}$$

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = h_k(\mathbf{X}_k, \boldsymbol{\beta}_k) \tag{2.5}$$

Once the GAMLSS model is determined, the parametric vectors $\boldsymbol{\beta}_k$ and the random effects parameters $\boldsymbol{\gamma}_{jk}$, for $j = 1, 2, ..., J_k$ and $k = 1, 2, 3, 4$ are estimated within the GAMLSS framework (for fixed values of the smoothing hyper-parameters λ_{jk}'s) by maximizing a penalized likelihood function l_p given by

$$l_p = l - \frac{1}{2} \sum_{k=1}^{p} \sum_{j=1}^{J_k} \lambda_{jk} \boldsymbol{\gamma}_{jk}^{'} \mathbf{G}_{jk} \boldsymbol{\gamma}_{jk}, \tag{2.6}$$

where $l = \sum_{i=1}^{n} \log f\left(y_i | \boldsymbol{\theta}^i\right)$ is the log-likelihood function of the data given $\boldsymbol{\theta}^i$ for $i = 1, ..., n$.

For parametric GAMLSS model (2.3) or (2.5), l_p reduces to l, and the unknown parameters $\boldsymbol{\beta}_k$ for $k = 1, 2, 3, 4$ are estimated by maximizing l. Two algorithms can be used to maximize the likelihood function l_p given by Eq. (2.6) . The first, the CG algorithm, is a generalization of the Cole and Green (1992) algorithm. The second, the RS algorithm, is a generalization of the algorithm used by Rigby and Stasinopoulos (1996 a,b) for fitting mean and dispersion additive models. The algorithms require the first (and optionally exact or approximated expected second) derivatives of the log-likelihood with respect to the parameters μ, σ, ν and τ to be computable. The "CG" algorithm also needs the exact or approximated expected cross derivatives. The algorithms are Fisher Scoring or quasi-Newton, depending on whether the expected second (and cross) derivatives, or the negative squares (and cross products) of the first derivatives are used. For more details about these two algorithms see Stasinopoulos and Rigby (2005). In this thesis we select the parametric linear GAMLSS (Eq. (2.3)) and thus the objective is to maximize the likelihood function l.

2.2.1 Claim Frequency Models

This section summarizes the characteristics of the various count models used in this chapter. Specifically, we consider the Poisson, Negative Binomial Type II, Delaporte, Sichel and Zero-Inflated Poisson GAMLSS for assessing claim frequency.

The Poisson Model

In insurance practice, the Poisson distribution has been widely used for modelling the number of claims reported to an insurer by an insured driver in a given period. In Poisson regression, we have a collection of independent Poisson counts whose means are modelled as non-negative functions of covariates. Specifically, consider a policyholder i whose number of claims, denoted as K_i, are independent, for $i = 1, .., n$. All the observable individual characteristics are summarized into the $1 \times J_1'$ vector $c_{1i}\left(c_{1i,1}, ..., c_{1i,J_1'}\right)$. We assume that K_i follows the Poisson distribution with parameter μ, the expected number of claims. We can allow the μ parameter to vary from one individual to another. Let $\mu_i = e_i \exp\left(c_{1i}\beta_1\right)$, where e_i denotes the exposure[1] of policy i, $0 < e_i \leq 1$, and $\beta_1^T \left(\beta_{1,1}, ..., \beta_{1,J_j'}\right)$ is the $1 \times J_1'$ vector of the coefficients. The

[1]Exposure is the proportion of the period of observation for which the policy has been in force.

exponential form ensures the non-negativity of μ_i. Then, the probability that the policyholder i has reported k claims to the insurer is given by

$$P(K_i = k) = \frac{e^{-\mu_i}\mu_i^k}{k!}, \tag{2.7}$$

for $k = 0, 1, 2, ...,$ where

$$E(K_i) = Var(K_i) = \mu_i = e_i \exp(c_{1i}\beta_1). \tag{2.8}$$

Equidispersion is a typical characteristic associated with the Poisson distribution, i.e. the variance of the Poisson distribution is equal to its mean. The Poisson regression model has been widely used by insurance practitioners for modelling claim count data (for example see Renshaw, 1994).

Multiplicative Poisson Random Effect Models

The Poisson distribution has a descriptive adequacy as a model when only randomness is present and the underlying population is homogeneous. Unfortunately, homogeneity is not a realistic assumption to make in modelling many real insurance count (or frequency) response data which often exhibit overdispersion, i.e., a situation where the variance of the response variable exceeds the mean. Overdispersion is defined as the extra variation occurring in count data modelling which is not explained by the Poisson model alone. The problem of unobserved heterogeneity arises because of the differences in driving behavior among policyholders and it is due to the fact that many important a priori rating variables are unknown to the insurance company or cannot be incorporated in the regression relationship (for legal, moral or economic reasons). Inappropriate imposition of the Poisson model may underestimate the standard errors and overstate the Chi-square statistics, and consequently, give misleading inference about the regression parameters. Equidispersion implied by the Poisson distribution is usually corrected by the introduction of a random variable ε_i into the regression component. The random term ε_i may reflect the omission of unobserved exogenous variables or purely random effects. According to Gourieroux, Montfort and Trognon (1984 a), (1984 b) we can write

$$\mu_i = e_i \exp(c_{1i}\beta_1 + \varepsilon_i) = e_i \exp(c_{1i}\beta_1) u_i, \tag{2.9}$$

where e_i is the corresponding risk exposure, $c_{1i}\left(c_{1i,1}, ..., c_{1i,J_1'}\right)$ is the $1 \times J_1'$ vector of the a priori rating variables and $\beta_1^T\left(\beta_{1,1}, ..., \beta_{1,J_1'}\right)$is the $1 \times J_1'$ vectors of the coefficients and where $u_i = \exp(\varepsilon_i)$, yielding a random μ_i. Therefore, heterogeneity is taken into account by assuming that the number of claims K_i conforms to the Poisson distribution with mean $\mu_i u_i$, where $\mu_i = e_i \exp(c_{1i}\beta_1)$. At the portfolio level, the u_i's are assumed to be independent and identically distributed continuous random effect variables with probability density function $\upsilon(u_i)$ defined on $\mathcal{R}^+$. The model can be considered as a multiplicative Poisson random effect model. Depending on the chosen parametric form of u_i, this general compound Poisson distribution will lead to different models. The following are well-known results applied to the above situation (see Boyer et al., 1992, Lemaire, 1995 & Boucher et al., 2007, 2008). The mean of K_i is given by $E(K_i) = \mu E(u_i)$ and the variance of K_i is given by $Var(K_i) = \mu_i E(u_i) + \mu_i^2 Var(u_i)$. If we let u_i have a unit mean, then the mean and variance of K_i are given by $E(K_i) = \mu_i$ and $Var(K_i) = \mu_i + \mu_i^2 Var(u_i)$ respectively, so that any mixed Poisson model allows for overdispersion. Note that in general $Var(u_i) = h(\sigma, \nu, \tau)$ is a function of the parameters σ, ν and τ of the mixing distribution $\upsilon(u_i)$. Recent actuarial literature research assumes that the μ_i parameter of the Poisson mixtures is modelled as a function of the explanatory variables while the parameters σ, ν and τ are constant. In the above setup the heterogeneity term u_i has a constant variance α and so the variance of K_i is equal to $Var(K_i) = \mu_i + \mu_i^2 \alpha$ and exceeds the mean μ_i. As we have already mentioned, in this chapter we extend this setup, following the general GAMLSS approach of Rigby and Stasinopoulos (2005), where all the parameters of the Poisson mixtures can be modelled as functions of explanatory variables with parametric linear functional forms[2]. In this way we are able to use all the available information in the estimation of the claim frequency distribution in order to group risks with similar risk characteristics and to establish "fair" premium rates. For this purpose we assume that the individual number of claims K_i , $i = 1, .., n$, are i.i.d mixed Poisson random variables and we specify the following models on the *ith* element of the vector parameters $\mu^T(\mu_1, ..., \mu_n), \sigma^T(\sigma_1, ..., \sigma_n), \nu^T(\nu_1, ..., \nu_n)$ and $\tau^T(\tau_1, ..., \tau_n)$:

$$\mu_i = e_i \exp(c_{1i}\beta_1), \tag{2.10}$$
$$\sigma_i = g_1^{-1}(c_{2i}\beta_2), \tag{2.11}$$
$$\nu_i = g_2^{-1}(c_{3i}\beta_3), \tag{2.12}$$
$$\tau_i = g_3^{-1}(c_{4i}\beta_4), \tag{2.13}$$

where g_z, $z = 1, 2, 3$ are known monotonic link functions chosen to ensure a valid range for the distribution parameters σ_i, ν_i and τ_i, where $c_{ji}\left(c_{ji,1}, ..., c_{ji,J_j'}\right)$are the $1 \times J_j'$ vectors of the a priori rating variables and where $\beta_j^T\left(\beta_{j,1}, ..., \beta_{j,J_j'}\right)$are the are the $1 \times J_j'$ vectors of the coefficients, for $j = 1, 2, 3, 4$. All models considered in this section belong to this category.

The Negative Binomial Type II Model

Let u have a Gamma distribution with probability density function given by

$$\upsilon(u) = \frac{u^{\sigma-1}\sigma^{\sigma}\exp(-\sigma u)}{\Gamma(\sigma)}, \tag{2.14}$$

$u > 0, \sigma > 0$. Parameterization (2.14) ensures that $E(u) = 1$. Note also that $Var(u) = \sigma$. Then it can be shown that the marginal distribution of K_i is a Negative Binomial Type II (NBII)

[2]Note that the relationship between claim frequency and explanatory variables may not be limited to a parametric linear functional form. If we want to explore more flexible functional forms, the semi-parametric additive model, the non-linear semi-parametric additive model and the non-linear parametric model can also be examined.

distribution with probability density function given by

$$P(K_i = k) = \frac{\Gamma\left(k + \frac{\mu}{\sigma}\right)\sigma^k}{\Gamma\left(\frac{\mu}{\sigma}\right)\Gamma(k+1)\left[1+\sigma\right]^{k+\frac{\mu}{\sigma}}}, \tag{2.15}$$

for $\mu > 0$ and $\sigma > 0$. This parameterization was used by Evans (1953) as pointed out by Johnson et al. (1994). The mean and variance of K_i are given by $E(K_i) = \mu$ and $Var(K_i) = \mu(1+\sigma)$ respectively. Clearly the variance exceeds the mean and the distribution allows for overdispersion. Note that alternative variance mean relationships can be obtained by reparameterization. For example, if σ is reparameterized to $\sigma_1\mu$ giving a Negative Binomial Type I distribution[3] then $Var(K_i) = \mu + \sigma_1\mu^2$. A priori ratemaking using the NBI distribution for K_i with a linear model in c_{1i} for the log of mean parameter and a constant for the scale parameter has been recommended by, for example, Boucher et al. (2007, 2008). More generally, a family of reparameterizations of the Negative Binomial Type I distribution can be obtained by reparameterizing σ to $\sigma_1\mu^{\nu-2}$ giving $Var(K_i) = \mu + \sigma_1\mu^{\nu}$. This gives a three parameter model with parameters μ, σ_1 and ν. The model can be fitted by maximum likelihood estimation. Note that a family of reparameterizations can be applied to any Poisson mixture model as defined in the previous subsection.

Following the general GAMLSS approach, we assume that the *ith* element of the vector parameters $\mu^T(\mu_1, ..., \mu_n)$ and $\sigma^T(\sigma_1, ..., \sigma_n)$ is given by

$$\mu_i = e_i \exp(c_{1i}\beta_1) \tag{2.16}$$

and

$$\sigma_i = \exp(c_{2i}\beta_2), \tag{2.17}$$

where e_i is the corresponding risk exposure. Then the mean and the variance of K_i are given by

$$E(K_i) = e_i \exp(c_{1i}\beta_1) \tag{2.18}$$

and

$$Var(K_i) = e_i \exp(c_{1i}\beta_1)\left[1 + \exp(c_{2i}\beta_2)\right]. \tag{2.19}$$

The Delaporte Model

Let u have a shifted Gamma distribution with probability density function given by

$$\upsilon(u) = \frac{(u-\nu)^{\frac{1}{\sigma}-1}\exp\left[-\frac{(u-\nu)}{\sigma(1-\nu)}\right]}{\sigma^{\frac{1}{\sigma}}(1-\nu)^{\frac{1}{\sigma}}\Gamma\left(\frac{1}{\sigma}\right)}, \tag{2.20}$$

for $u > \nu$, where $\sigma_i > 0$ and $0 \leq \nu < 1$. This parameterization ensures that $E(u) = 1$. Note also that $Var(u) = \sigma(1-\nu)^2$. Then K_i follows a Delaporte distribution with probability density function given by

[3] A more general proof of Eq. (2.15) can be found in Chapter 5 where we consider the case of the n-component mixture of Negative Binomial Type I (NBI) regression models, derived by updating the posterior mean claim frequency. Specifically, the proof of Eq. (2.15) can be obtained by this reparameterization and by letting $n = 1$.

$$P\left(K_i = k\right) = \frac{e^{-\mu\nu}}{\Gamma\left(\frac{1}{\sigma}\right)}\left[1 + \mu\sigma\left(1 - \nu\right)\right]^{-\frac{1}{\sigma}} S, \tag{2.21}$$

where

$$S = \sum_{m=0}^{k} \binom{k}{m} \frac{\mu^k \nu^{k-m}}{k!} \left[\mu + \frac{1}{\sigma\left(1 - k\right)}\right]^{-m} \Gamma\left(\frac{1}{\sigma} + m\right). \tag{2.22}$$

The mean and variance of K_i are given by $E(K_i) = \mu$ and $Var(K_i) = \mu + \mu^2\sigma\left(1 - \nu\right)^2$ respectively. This parameterization of Delaporte was given by Rigby and Stasinopoulos (2008)[4]. The special case $\sigma = 1$ in (2.21) gives the parameterization of a Poisson-Shifted Exponential distribution since the mixing distribution reduces to a Shifted Exponential distribution.

Following the general GAMLSS approach, we assume that the ith element of the vector parameters $\mu^T\left(\mu_1, ..., \mu_n\right), \sigma^T\left(\sigma_1, ..., \sigma_n\right)$ and $\nu^T\left(\nu_1, ..., \nu_n\right)$ is given by

$$\mu_i = e_i \exp\left(c_{1i}\beta_1\right), \tag{2.23}$$

$$\sigma_i = \exp\left(c_{2i}\beta_2\right) \tag{2.24}$$

and

$$\nu_i = \frac{\exp\left(c_{3i}\beta_3\right)}{1 + \exp\left(c_{3i}\beta_3\right)}, \tag{2.25}$$

where e_i is the corresponding risk exposure. Then the mean and variance of K_i are given by

$$E(K_i) = e_i \exp\left(c_{1i}\beta_1\right) \tag{2.26}$$

and

$$Var(K_i) = e_i \exp\left(c_{1i}\beta_1\right) + \left[e_i \exp\left(c_{1i}\beta_1\right)\right]^2 \exp\left(c_{2i}\beta_2\right)\left[1 - \frac{\exp(c_{3i}\beta_3)}{1 + \exp(c_{3i}\beta_3)}\right]^2. \tag{2.27}$$

.

The Sichel Model

Let u have a Generalized Inverse Gaussian (GIG) distribution with probability density function given by

$$\upsilon\left(u\right) = \frac{c^\nu u^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(cu + \frac{1}{cu}\right)\right]}{2K_\nu\left(\frac{1}{\sigma}\right)}, \tag{2.28}$$

for $u > 0$, where $\sigma > 0$ and $-\infty < \nu < \infty$ and where $c = \frac{K_{\nu+1}\left(\frac{1}{\sigma}\right)}{K_\nu\left(\frac{1}{\sigma}\right)}$ where

$$K_\nu\left(z\right) = \frac{1}{2}\int_0^\infty x^{\nu-1} \exp\left[-\frac{1}{2}z\left(x + \frac{1}{x}\right)\right] dx, \tag{2.29}$$

[4]For information about the proof of Eq. (2.21) refer to Rigby and Stasinopoulos (2008) and the references therein.

is the modified Bessel function of the third kind of order ν with argument z. Parameterization (2.28) was given by Rigby and Stasinopoulos (2008). This parameterization ensures that $E(u) = 1$. Note also that $Var(u) = \frac{2\sigma(\nu+1)}{c} + \frac{1}{c^2} - 1$. The gamma and reciprocal gamma are limiting distributions of (2.28), obtained by letting $\sigma \to \infty$ for $\nu > 0$ and $\nu < -1$ respectively. Then K_i follows a Sichel distribution with probability density function[5] given by

$$P(K_i = k) = \frac{\left(\frac{\mu}{c}\right)^k K_{k+\nu}(a)}{k!\,(a\sigma)^{k+\nu} K_\nu\left(\frac{1}{\sigma}\right)}, \tag{2.30}$$

where $a^2 = \sigma^{-2} + 2\mu(c\sigma)^{-1}$. The mean and variance of K_i are given by $E(K_i) = \mu$ and $Var(K_i) = \mu + \mu^2\left[\frac{2\sigma(\nu+1)}{c} + \frac{1}{c^2} - 1\right]$ respectively.

Following the general GAMLSS approach, we assume that the ith element of the vector parameters $\mu^T(\mu_1, ..., \mu_n)$, $\sigma^T(\sigma_1, ..., \sigma_n)$ and $\nu^T(\nu_1, ..., \nu_n)$ is given by

$$\mu_i = e_i \exp(c_{1i}\beta_1), \tag{2.31}$$

$$\sigma_i = \exp(c_{2i}\beta_2) \tag{2.32}$$

and

$$\nu_i = c_{3i}\beta_3, \tag{2.33}$$

where e_i is the corresponding risk exposure. Then the mean and variance of K_i are given by

$$E(K_i) = e_i \exp(c_{1i}\beta_1) \tag{2.34}$$

and

$$Var(K_i) = e_i \exp(c_{1i}\beta_1) + [e_i \exp(c_{1i}\beta_1)]^2 \left\{\frac{2\exp(c_{2i}\beta_2)[\exp(c_{3i}\beta_3) + 1]}{c_i} + \frac{1}{c_i^2} - 1\right\}, \tag{2.35}$$

where $c_i = \frac{K_{\nu_i+1}\left(\frac{1}{\sigma_i}\right)}{K_{\nu_i}\left(\frac{1}{\sigma_i}\right)}$ and where σ_i and ν_i are given by Eq. (2.32) and Eq. (2.33) respectively.

The Zero-Inflated Poisson Model

Zero-inflated count models provide a parsimonious yet powerful way to handle data sets that contain a large number of zeros. Such models assume that the data are a mixture of two distributions: a degenerate distribution for the zero case and a standard count distribution (see Lambert, 1992 and Greene, 1994). The Zero-Inflated Poisson (ZIP) distribution arises if we let $K_i = 0$ with probability π and $K_i \sim Po(\mu)$ with probability $(1-\pi)$. The probability density function of the ZIP distribution is defined as

$$P(K_i = k) = \begin{cases} \pi + (1-\pi)e^{-\mu}, \text{ if } k = 0 \\ (1-\pi)\frac{e^{-\mu}\mu^k}{k!}, \text{ if } k = 1, 2, 3, ... \end{cases} \tag{2.36}$$

[5]The proof of Eq. (2.30) can be found in subsection 3.3.2 of Chapter 3.

This parameterization was used by Johnson et al. (1994) and Lambert (1992). The mean and variance of K_i are given by $E(K_i) = \mu(1-\pi)$ and $Var(K_i) = E(K_i) + E(K_i)(\mu - E(K_i))$ respectively. The ZIP distribution thus allows for overdispersion as well. Note that the ZIP model is a special case of a mixed Poisson distribution obtained with u equal to 0 or μ (with respective probabilities π and $(1-\pi)$). Note also that if overdispersion in the Poisson part is still present then all the distributions seen in the previous subsection can be used since a heterogeneity term may be incorporated in the model. For instance, see Yip and Yau (2005) for an application to insurance claim count data.

Following Rigby and Stasinopoulos (2005), we assume that the ith element of the vector parameters $\mu^T(\mu_1, ..., \mu_n)$ and $\pi^T(\sigma_1, ..., \sigma_n)$ is given by:

$$\mu_i = e_i \exp(c_{1i}\beta_1) \tag{2.37}$$

and

$$\pi_i = \frac{\exp(c_{2i}\beta_2)}{1+\exp(c_{2i}\beta_2)}, \tag{2.38}$$

where e_i is the corresponding risk exposure, and where $c_{ji}\left(c_{ji,1}, ..., c_{ji,J_j^/}\right)$ are covariate vectors of length $1 \times J_j'$, $j = 1, 2$ for μ_i and π_i, which may be different, the same, or may have some but all not elements in common, and $\beta_j^T\left(\beta_{j,1}, ..., \beta_{j,J_j^/}\right)$ are the corresponding parameter vectors of length $1 \times J_j'$, $j = 1, 2$. Then the mean and the variance of K_i are given by

$$E(K_i) = e_i \exp(c_{1i}\beta_1)\left[1 - \exp(c_{2i}\beta_2)\right] \tag{2.39}$$

and

$$Var(K_i) = e_i \exp(c_{1i}\beta_1)\left[1 - \exp(c_{2i}\beta_2)\right]\left[1 + e_i \exp(c_{1i}\beta_1)\exp(c_{2i}\beta_2)\right]. \tag{2.40}$$

2.2.2 Claim Severity Models

In this section, we need to consider the claim severities. Different models are used to describe the behavior of the costs of claims as a function of the explanatory variables; including Gamma, Weibull, Weibull Type III, Generalized Gamma, and Generalized Pareto GAMLSS. In this thesis, we adopt the parametric linear formulation of these models[6].

The Gamma Model

Here we use the parameterization of the two parameter Gamma distribution given by Rigby and Stasinopoulos (2009), defined so that the mean m, i.e. the expected claim severity, is an explicit parameter of the distribution and the parameter s is related to the variation coefficient. This allows easier interpretation of regression type models for m and provides a more orthogonal

[6] Note that similarly to the case of the claim frequency models, the relationship between claim severity and explanatory variables may not be limited to a parametric linear functional form. If we want to explore more flexible functional forms, the semi-parametric additive model, the non-linear semi-parametric additive model and the non-linear parametric model can also be examined.

parameterization. Let $X_{i,k}$ be the cost of the kth claim reported by policyholder i, $i = 1, ..., n$ and assume that the individual claim costs $X_{i,1,} X_{i,2}, ..., X_{i,n}$ are i.i.d Gamma random variables. The probability density function of the Gamma distribution is given by

$$f(x) = \frac{1}{(s^2 m)^{\frac{1}{s^2}}} \frac{x^{\frac{1}{s^2}-1} \exp\left(-\frac{x}{s^2 m}\right)}{\Gamma\left(\frac{1}{s^2}\right)}, \tag{2.41}$$

for $x > 0$, where $m > 0$ and $s > 0$. If $X_{i,k}$ has probability density function (2.41), then the first two moments are given by $E(X_{i,k}) = m$ and $Var(X_{i,k}) = s^2 m^2$, so that the variance is proportional to the square of the mean.

Following Rigby and Stasinopoulos (2009), we assume that the *ith* element of the vector parameters $m^T\,(m_1, ..., m_n)$ and $s^T\,(s_1, ..., s_n)$ is given by

$$m_i = \exp\left(d_{1i}\gamma_1\right) \tag{2.42}$$

and

$$s_i = \exp\left(d_{2i}\gamma_2\right), \tag{2.43}$$

where $d_{ji}\left(d_{ji,1}, ..., d_{ji,J_j^{/}}\right)$ are vectors of exogenous variables of length $1 \times J_j'$, $j = 1, 2$ for m_i and s_i, which may be different, the same, or may have some but all not elements in common, and $\gamma_j^T\left(\gamma_{j,1}, ..., \gamma_{j,J_j^{/}}\right)$ are the corresponding vectors of the coefficients of length $1 \times J_j'$, $j = 1, 2$. Then the mean and the variance of $X_{i,k}$ are given by

$$E(X_{i,k}) = \exp\left(d_{1i}\gamma_1\right) \tag{2.44}$$

and

$$Var(X_{i,k}) = \left[\exp\left(d_{2i}\gamma_2\right)\right]^2 \left[\exp\left(d_{1i}\gamma_1\right)\right]^2. \tag{2.45}$$

Note also that a priori ratemaking using the Gamma distribution for $X_{i,k}$ with a linear model only in d_{1i} for the log of m and a constant for s can be found in, for example, Denuit et al. (2007).

The Weibull Model

Next we fit a Weibull distribution to model the individual claim sizes $X_{i,k}$. The specific parameterization of the two parameter Weibull distribution used here was that used by Johnson et al. (1994) p 629 and is defined as

$$f(x) = \frac{s x^{s-1}}{m^s} \exp\left[-\left(\frac{x}{m}\right)^s\right], \tag{2.46}$$

for $X_{i,k} > 0$, where $m > 0$ and $s > 0$. The mean and variance of $X_{i,k}$ are given by $E(X_{i,k}) = m\Gamma\left(\frac{1}{s} + 1\right)$ and $Var(X_{i,k}) = m^2\left\{\Gamma\left(\frac{2}{s} + 1\right) - \left[\Gamma\left(\frac{1}{s} + 1\right)\right]^2\right\}$ respectively. Although the parameter m is a scale parameter, it also affects the mean of $X_{i,k}$. Note that the Weibull distribution interpolates between the Exponential distribution ($s = 1$) and the Rayleigh distribution ($s = 2$).

We parallel here the treatment of the Gamma model. Following Rigby and Stasinopoulos (2009), we assume that the ith element of the vector parameters $m^T(m_1, ..., m_n)$ and $s^T(s_1, ..., s_n)$ is given by

$$m_i = \exp(d_{1i}\gamma_1) \tag{2.47}$$

and

$$s_i = \exp(d_{2i}\gamma_2), \tag{2.48}$$

where $d_{ji}\left(d_{ji,1}, ..., d_{ji,J_j^{\prime}}\right)$ are covariate vectors of length $1 \times J_j^{\prime}$, $j = 1, 2$ for m_i and s_i, which may be different, the same, or may have some but all not elements in common, and $\gamma_j^T\left(\gamma_{j,1}, ..., \gamma_{j,J_j^{\prime}}\right)$ are the corresponding parameter vectors of length $1 \times J_j^{\prime}$, $j = 1, 2$. Then the mean and the variance of $X_{i,k}$ are given by

$$E(X_{i,k}) = \exp(d_{1i}\gamma_1)\Gamma\left(\frac{1}{\exp(d_{2i}\gamma_2)} + 1\right) \tag{2.49}$$

and

$$Var(X_{i,k}) = [\exp(d_{1i}\gamma_1)]^2\left\{\Gamma\left(\frac{2}{\exp(d_{2i}\gamma_2)} + 1\right) - \left[\Gamma\left(\frac{1}{\exp(d_{2i}\gamma_2)} + 1\right)\right]^2\right\}. \tag{2.50}$$

The Weibull Type III Model

This is a parameterization of the Weibull distribution where m is the mean of the distribution. This probability density function of the Weibull Type III (WEI3) distribution is given by

$$f(x) = \frac{s}{m}\Gamma\left(\frac{1}{s} + 1\right)\left[\frac{x}{m}\Gamma\left(\frac{1}{s} + 1\right)\right]^{s-1}\exp\left\{-\left[\frac{x}{m}\Gamma\left(\frac{1}{s} + 1\right)\right]^{s}\right\}, \tag{2.51}$$

for $X_{i,k} > 0$, where $m > 0$ and $s > 0$. The mean and variance of $X_{i,k}$ are given by $E(X_{i,k}) = m$ and $Var(X_{i,k}) = m^2\left\{\Gamma\left(\frac{2}{s} + 1\right)\left[\Gamma\left(\frac{1}{s} + 1\right)\right]^{-2} - 1\right\}$ respectively.

Following Rigby and Stasinopoulos (2009), we assume that the ith element of the vector parameters $m^T(m_1, ..., m_n)$ and $s^T(s_1, ..., s_n)$ is given by

$$m_i = \exp(d_{1i}\gamma_1) \tag{2.52}$$

and

$$s_i = \exp(d_{2i}\gamma_2), \tag{2.53}$$

where $d_{ji}\left(d_{ji,1}, ..., d_{ji,J_j^{\prime}}\right)$ are vectors of the a priori rating variables of length $1 \times J_j^{\prime}$, $j = 1, 2$ for m_i and s_i, which may be different, the same, or may have some but all not elements in common, and $\gamma_j^T\left(\gamma_{j,1}, ..., \gamma_{j,J_j^{\prime}}\right)$ are the vectors of coefficients of length $1 \times J_j^{\prime}$, $j = 1, 2$. Then the mean and the variance of $X_{i,k}$ are given by

$$E(X_{i,k}) = \exp(d_{1i}\gamma_1) \tag{2.54}$$

and

$$Var(X_{i,k}) = [\exp(d_{1i}\gamma_1)]^2 \left\{ \Gamma\left(\frac{2}{\exp(d_{2i}\gamma_2)} + 1\right) \left[\Gamma\left(\frac{1}{\exp(d_{2i}\gamma_2)} + 1\right)\right]^{-2} - 1 \right\}. \tag{2.55}$$

The Generalized Gamma Model

Let us now fit the three-parameter Generalized Gamma (GG) distribution on the individual claim sizes $X_{i,k}$. The parameterization of the Generalized Gamma distribution we use was that used by Lopatatzidis and Green (2000), and is defined as

$$f(x) = \frac{|n|\theta^{\theta}\left(\frac{x}{m}\right)^{n\theta} \exp\left[-\theta\left(\frac{x}{m}\right)^{n}\right]}{\Gamma(\theta)\, x}, \tag{2.56}$$

for $X_{i,k} > 0$, where $m > 0$ and $s > 0$, where $-\infty < n < \infty$ and where $\theta = \frac{1}{s^2 n^2}$. The mean and the variance of $X_{i,k}$ are given by

$$E(X_{i,k}) = \frac{m\Gamma\left(\theta + \frac{1}{n}\right)}{\theta^{\frac{1}{n}}\Gamma(\theta)}$$

and

$$Var(X_{i,k}) = \frac{m^2\left\{\Gamma(\theta)\Gamma\left(\theta + \frac{2}{n}\right) - \left[\Gamma\left(\theta + \frac{1}{n}\right)\right]^2\right\}}{\theta^{\frac{2}{n}}\left[\Gamma(\theta)\right]^2}.$$

Note that if we let $s = \frac{1}{n}$ in Eq. (2.56), for $n > 0$, the GG distribution reduces to the Weibull distribution, with pdf given by Eq. (2.46). Note also that if we let $n = 1$ in Eq. (2.56), the GG distribution reduces to the Gamma distribution, with pdf given by Eq. (2.41).

Following Rigby and Stasinopoulos (2008), we assume that the ith element of the vector parameters $m^T(m_1, ..., m_n)$, $s^T(s_1, ..., s_n)$ and $n^T(n_1, ..., n_n)$ is given by

$$m_i = \exp(d_{1i}\gamma_1), \tag{2.57}$$

$$s_i = \exp(d_{2i}\gamma_2) \tag{2.58}$$

and

$$n_i = d_{3i}\gamma_3, \tag{2.59}$$

where $d_{ji}\left(d_{ji,1}, ..., d_{ji,J_1'}\right)$ are covariate vectors of length $1 \times J_j'$, $j = 1, 2, 3$ for m_i, s_i and n_i, which may be different, the same, or may have some but all not elements in common, and $\gamma_j^T\left(\gamma_{j,1}, ..., \gamma_{j,J_1'}\right)$ are the corresponding vectors of the coefficients of length $1 \times J_j'$, $j = 1, 2, 3$. Then the mean and the variance of $X_{i,k}$ are given by

$$E(X_{i,k}) = \frac{\exp(d_{1i}\gamma_1)\Gamma\left(\theta_i + \frac{1}{d_{3i}\gamma_3}\right)}{\theta_i^{\frac{1}{d_{3i}\gamma_3}}\Gamma(\theta_i)} \tag{2.60}$$

and

$$Var(X_{i,k}) = \frac{\left[\exp\left(d_{1i}\gamma_1\right)\right]^2 \left\{ \Gamma\left(\theta_i\right) \Gamma\left(\theta_i + \frac{2}{d_{3i}\gamma_3}\right) - \left[\Gamma\left(\theta_i + \frac{1}{d_{3i}\gamma_3}\right)\right]^2 \right\}}{\theta_i^{\frac{2}{d_{3i}\gamma_3}} \left[\Gamma\left(\theta_i\right)\right]^2}, \tag{2.61}$$

where $\theta_i = \frac{1}{s_i^2 n_i^2}$ and where s_i and n_i are given by Eq. (2.58) and Eq. (2.59) respectively.

The Generalized Pareto Model

Finally, we consider the case of the three parameter Generalized Pareto distribution for assessing claim sizes $X_{i,k}$. The probability density function of the Generalized Pareto distribution is given by

$$f\left(x\right) = \frac{\Gamma\left(n+t\right)}{\Gamma\left(n\right)\Gamma\left(t\right)} \frac{m^t x^{n-1}}{\left(x+m\right)^{n+t}}, \tag{2.62}$$

for $X_{i,k} > 0$, where $m > 0$, $n > 0$ and $t > 0$. The above parameterization of the Generalized Pareto distribution can be found, for example, in Klugman et al. (2004). If $X_{i,k}$ has probability density function (2.62), then the two first moments are given by $E(X_{i,k}) = \frac{mn}{t-1}$ and $Var(X_{i,k}) = \frac{m^2 n}{(t-1)}\left[\frac{n+t-1}{(t-1)(t-2)}\right]$ respectively. Note that if we let $n = 1$ in Eq. (2.62), the Generalized Pareto distribution reduces to the Pareto distribution.

Following Rigby and Stasinopoulos (2008), we assume that the ith element of the vector parameters $m^T\left(m_1, ..., m_n\right)$, $n^T\left(n_1, ..., n_n\right)$ and $t^T\left(t_1, ..., t_n\right)$ is given by

$$m_i = \exp\left(d_{1i}\gamma_1\right), \tag{2.63}$$

$$n_i = \exp\left(d_{2i}\gamma_2\right) \tag{2.64}$$

and

$$t_i = \exp\left(d_{3i}\gamma_3\right) \tag{2.65}$$

where $d_{ji}\left(d_{ji,1}, ..., d_{ji,J_j'}\right)$are vectors of exogenous variables of length $1 \times J_j'$, $j = 1, 2, 3$ for m_i, n_i and t_i which may be different ,the same, or may have some but all not elements in common and $\gamma_j^T\left(\gamma_{j,1}, ..., \gamma_{j,J_j'}\right)$ are the corresponding vectors of the coefficients of length $1 \times J_j'$, $j = 1, 2, 3$. Then the mean and the variance of $X_{i,k}$ are given by

$$E(X_{i,k}) = \frac{\exp\left(d_{1i}\gamma_1\right)\exp\left(d_{2i}\gamma_2\right)}{\exp\left(d_{3i}\gamma_3\right) - 1} \tag{2.66}$$

and

$$Var(X_{i,k}) = \frac{\left[\exp\left(d_{1i}\gamma_1\right)\right]^2 \exp\left(d_{2i}\gamma_2\right)}{\exp\left(d_{3i}\gamma_3\right) - 1} \left\{ \frac{\exp\left(d_{2i}\gamma_2\right) + \exp\left(d_{3i}\gamma_3\right) - 1}{\left[\exp\left(d_{3i}\gamma_3\right) - 1\right]\left[\exp\left(d_{3i}\gamma_3\right) - 2\right]} \right\}. \tag{2.67}$$

2.3 Application

Let us briefly present the data used to illustrate the techniques described in this chapter. The data were kindly provided by a Greek insurance company and concern a motor third party liability insurance portfolio observed during 3.5 years (from Aril 2008 up to October 2011). The data set comprises 15641 policies. Both private cars and fleet vehicles have been considered in this sample. Our interest lies in identifying the factors that affect the frequency and severity of claims at fault, and specifically the factors that correspond to each policyholder and their characteristics, including the characteristics of the car. For this purpose, the data consist of the available exogenous variables for every policy as the Bonus-Malus category, the gender of the driver, and the horsepower of the car, as well as the total number and the costs of claims at fault that were reported within the 3.5 year period. Nevertheless, it is important to note that gender has recently been ruled out by the European Court as a rating factor. This Bonus-Malus System has 20 classes and the transition rules are described as follows: Each claim free year is rewarded by one class discount and each claim in given year is penalized by one class. Only policyholders with complete records, i.e. with availability of all the variables under consideration were considered[7].

Claim counts are modelled for all 15641 policies that have been in force for the entire sampling period. The expected frequency of claims at fault is 0.4848 and the variance is 0.73086. Response variable is the total number of claims registered for each insured vehicle in the data set and the following explanatory variables were available:

- Bonus-Malus category: The categories of neighboring classes of the current Greek BMS (variable BM category; five categories: C1 = "drivers who belong to BM classes 1 and 2", C2 = "drivers who belong to BM classes 3-5", C3 = "drivers who belong to BM classes 6-9", C4 = "drivers who belong to BM class 10" and C5 = "drivers who belong to BM classes 11-20").

- Horsepower: The horsepower of the car (variable HP; four categories: C1 = "drivers who had a car with a hp between 0-33", C2 = "drivers who had a car with a hp between 34-66, C3 = "drivers who had a car with a hp between 67-99" and C4 = "drivers who had a car with a hp between 100-132").

- Gender: Policyholders gender (variable Gender; two categories: M= "male", F = "female").

Descriptive statistics for claim counts are presented in Table 2.1.

[7] All the characteristics we consider are observable.

Table 2.1: Descriptive Statistics of Claim Counts

statistic	value	BMCAT		HPCAT		Gender	
# observations	15641	C1:	11332	C1:	135	M:	9731
Minimum	0	C2:	1964	C2:	2185	F:	5910
1st Quantile	0.0000	C3:	1060	C3:	10098		-
Median	0.0000	C4:	1266	C4:	3223		-
Mean	0.4848	C5:	19		-		-
Standard Deviation	0.85490		-		-		-
3rd Quantile	1.0000		-		-		-
Maximum	9		-		-		-

Regarding the amount paid for each claim (this was the response variable), there were 5590 observations that met our criteria of selection, i.e. those with complete records. This sample contains only policies that have been in force for the entire sampling period. Furthermore, the claims that ended up at zero have been removed from the sample and the last reserve was used for the claim amount. The expected claim severity is 328 euros and the variance is 41231.59. The following explanatory variables were available:

- Bonus-Malus category: As previously, the categories of neighboring classes of the current Greek BMS (variable BM category; five categories: C1 = "drivers who belong to BM classes 1 and 2", C2 = "drivers who belong to BM classes 3-5", C3 = "drivers who belong to BM classes 6-9", C4 = "drivers who belong to BM class 10" and C5 = "drivers who belong to BM classes 11-20").

- Horsepower: In this case there were several subcategories of this variable in relation to those for the claim counts (variable HP; eleven categories: C1 = "drivers who had a car with a hp between 0-33", C2 = "drivers who had a car with a hp between 34-44", C3 = "drivers who had a car with a hp between 45-55", C4 = "drivers who had a car with a hp between 56-66", C5 = "drivers who had a car with a hp between 67-74", C6 = "drivers who had a car with a hp between 75-82", C7 = "drivers who had a car with a hp between 83-90", C8 = "drivers who had a car with a hp between 91-99", C9 ="drivers who had a car with a hp between 100-110", C10 = "drivers who had a car with a hp between 111-121" and C11 = "drivers who had a car with a hp between 122-132").

- Gender: In this case, data for fleet vehicles used by both male or female drivers in turn were also available (variable Gender; three categories: M = "male", F = "female" and B = "both")[8].

Descriptive statistics for claim costs are depicted in Table 2.2.

[8] Note that records for fleet data were not availiable for the case of the claim counts.

Table 2.2: Descriptive Statistics of Claim Costs

statistic	value	BMCAT		HPCAT		Gender	
# observations	5590	C1:	3566	C1:	59	M:	3035
Minimum	7	C2:	1087	C2:	278	F:	2165
1st Quantile	227	C3:	623	C3:	451	B:	390
Median	269	C4:	295	C4:	793		-
Mean	328	C5:	19	C5:	1692		-
Standard Deviation	203.06		-	C6:	1204		-
3rd Quantile	386		-	C7:	209		-
Maximum	4614		-	C8:	291		-
-	-		-	C9:	360		-
-	-		-	C10:	191		-
-	-		-	C11:	62		-

In our application we regrouped the levels of each explanatory variable presented in Tables 2.1 and 2.2 with respect to risk profiles with similar number and costs of claims reported to the company over the 3.5 years of observation. This was done in this chapter only in order to obtain a reasonable number of risk classes[9].

Firstly, regarding the claim frequency modelled for all 15641 policies, the new levels of explanatory variables we employ are:

- Bonus-Malus category: Four categories A, B, C and D, where: A = C1, i.e. drivers who belong to BM classes 1 and 2, B = C2, i.e. drivers who belong to BM classes 3-5, C = C3 & C5, i.e. drivers who belong to BM classes 6-9 & 11-20 and D = C4, i.e. drivers who belong to BM class 10.

- Horsepower: Three categories A, B and C, where: A = C1 & C4, i.e. drivers who had a car with a hp between 0-33 & 100-132"), B = C2, i.e. drivers who had a car with a hp between 34-66 and C = C3, i.e. drivers who had a car with a hp between 67-99

- Gender: The same two categories: M= "male", F = "female".

Secondly, regarding the claim severity modelled for 5590 policies with complete that met our criteria of selection, the new levels of the explanatory variables we employ are:

- Bonus-Malus category: Three categories A, B and C, where: A = C1, i.e. drivers who belong to BM classes 1 and 2, B = C2 & C3 & C5, i.e. drivers who belong to BM classes 3-5 & 6-9 & 11-20 and C = C4, i.e. drivers who belong to BM class 10.

- Horsepower: Four categories A, B, C and D, where: A = C9 & C10 & C11, i.e. drivers who had a car with a hp between 100-110 & 111-121 & 122-132, B = C1 & C2 & C3 & C4, i.e. drivers who had a car with a hp between 0-33 & 34-44 & 45-55 & 56-66, C = C5,

[9] In the other chapters this was not necessary.

i.e. drivers who had a car with a hp between 67-74 and D= C6 & C7 & C8, i.e. drivers who had a car with a hp between 75-82 & 83-90 & 91-99.

- Gender: The same three categories: M = "male", F = "female" and B = "both" for fleet vehicles used by both male or female drivers.

2.3.1 Modelling Results

This subsection is divided into two parts. The first part describes the modelling results of the GAMLSS models that have been applied to model claim frequency and the second part provides those of the GAMLSS models that have been applied to claim severity analysis. As mentioned previously, the GAMLSS are a general framework for univariate regression analysis that allows joint modelling of the mean (or location), the scale and the shape parameters of the distribution of the response variable as, linear and/or non-linear, parametric and/or additive non-parametric functions of explanatory variables and/or random effects. In this thesis, we adopted the parametric linear formulation for both the claim frequency and claim severity models. An important aspect of insurance ratemaking with GAMLSS is the selection of explanatory variables for the location, scale and shape parameter of each model. For this purpose we used the function step.GAIC within the GAMLSS package in software R, which performs the stepwise model selection using a Generalized Akaike information criterion (GAIC). The final claim frequency and severity models we selected are those that yield the lowest Global deviance (DEV), Akaike information criterion (AIC), and Bayesian information criterion (BIC) values constrained to the statistical significance, at a 5% threshold, of every explanatory variable they contain. In what follows, we present the estimated parameter values, the standard error of the parameter estimates, and the t-values for the hypothesis that the associated coefficient is zero together with the p-value of this test based on asymptotic normality.

Claim Frequency Models

We describe first the modelling results of the Poisson, Negative Binomial Type II (NBII), Delaporte, Sichel and Zero-Inflated Poisson (ZIP) GAMLSS for assessing claim frequency. Based on the variable selection technique we described above, the following final models were selected.

- Poisson GAMLSS:

$\log(\mu)$ =BM category + horsepower category + gender

- NBII GAMLSS:

$\log(\mu)$ = Bonus-Malus category + horsepower category + gender
$\log(\sigma)$ = horsepower category + gender

- Delaporte GAMLSS:

$$\log(\mu) = \text{Bonus-Malus category} + \text{horsepower category} + \text{gender}$$
$$\log(\sigma) = \text{horsepower category}$$
$$\text{logit}(\nu) = \text{constant}$$

- Sichel GAMLSS:

$$\log(\mu) = \text{Bonus-Malus category} + \text{horsepower category} + \text{gender}$$
$$\log(\sigma) = \text{horsepower category}$$
$$\text{logit}(\nu) = \text{constant}$$

- ZIP GAMLSS:

$$\log(\mu) = \text{Bonus-Malus category} + \text{horsepower category} + \text{gender}$$
$$\log(\sigma) = \text{Bonus-Malus category} + \text{gender}$$

The results obtained from the Poisson, NBII, Delaporte, Sichel and ZIP models are presented in Tables 2.3, 2.4, 2.5, 2.6 and 2.7 respectively.

Table 2.3: Results of the Fitted Poisson GAMLSS

Variable μ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.8150	0.0268	-30.437	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	0.6078	0.0296	20.565	0.0000
Category C	0.8834	0.0336	26.284	0.0000
Category D	-0.9423	0.0715	-13.184	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.2371	0.0419	-5.666	0.0000
Category C	-0.0725	0.0288	-2.513	0.0120
Gender				
Male	0	0	-	-
Female	0.0683	0.0240	2.845	0.0044

Table 2.4: Results of the Fitted Negative Binomial Type II GAMLSS

Variable μ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.8131	0.0322	-25.278	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	0.6328	0.0349	18.124	0.0000
Category C	0.8388	0.0368	22.784	0.0000
Category D	-0.9736	0.0810	-12.023	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.2351	0.0477	-4.934	0.0000
Category C	-0.0730	0.0340	-2.147	0.0318
Gender				
Male	0	0	-	-
Female	0.0687	0.0269	2.551	0.0107
Variable σ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.3728	0.0724	-5.150	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.7777	0.1913	-4.065	0.0000
Category C	-0.6716	0.1022	-6.573	0.0000
Gender				
Male	0	0	-	-
Female	-0.4313	0.1247	-3.458	0.0005

Table 2.5: Results of the Fitted Delaporte GAMLSS

Variable μ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.8221	0.0341	-24.115	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	0.6429	0.0389	16.549	0.0000
Category C	0.8679	0.0427	20.349	0.0000
Category D	-0.9561	0.0650	-14.712	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.2434	0.0499	-4.879	0.0000
Category C	-0.0742	0.0362	-2.051	0.0403
Gender				
Male	0	0	-	-
Female	0.0880	0.0284	3.100	0.0010
Variable σ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	1.5821	0.1034	15.301	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.9700	0.2151	-4.510	0.0000
Category C	-0.8971	0.1350	-6.644	0.0000
Variable ν	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.2013	0.0653	-3.083	0.0021

Table 2.6: Results of the Fitted Sichel GAMLSS

Variable μ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.8201	0.0340	-24.097	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	0.6387	0.0388	16.478	0.0000
Category C	0.8694	0.0425	20.443	0.0000
Category D	-0.9804	0.0696	-14.090	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.2458	0.0500	-4.915	0.0000
Category C	-0.0759	0.0361	-2.100	0.0357
Gender				
Male	0	0	-	-
Female	0.0908	0.0284	3.201	0.0013
Variable σ	**Estimate**	**Std Error**	**t-value/Wald 95%**	**P-value**
Intercept	1.210	0.5015	2.413	0.0158
Horsepower				
Category A	0	0	-	-
Category B	-1.664	0.5468	-3.043	0.0024
Category C	-1.598	0.5130	-3.116	0.0018
Variable ν	**Estimate**	**Std Error**	**t-value/Wald 95%**	**P-value**
Intercept	-2.104	0.1401	-15.02	0.0000

Table 2.7: Results of the Fitted Zero-Inflated Poisson GAMLSS

Variable μ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.2210	0.0250	-8.850	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	0.1571	0.0290	5.409	0.0000
Category C	0.7160	0.0321	22.316	0.0000
Category D	-0.2085	0.0677	-3.078	0.0021
Horsepower				
Category A	0	0	-	-
Category B	-0.2492	0.0420	-5.929	0.0000
Category C	-0.0939	0.0271	-3.467	0.0005
Gender				
Male	0	0	-	-
Female	-0.1010	0.0250	-4.045	0.0000
Variable σ	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.2036	0.0408	-4.989	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-2.8671	0.3860	-7.427	0.0000
Category C	-0.4926	0.1055	-4.671	0.0000
Category D	1.2694	0.1108	11.452	0.0000
Gender				
Male	0	0	-	-
Female	-0.5648	0.0725	-7.788	0.0000

From Tables 2.3, 2.4, 2.5, 2.6 and 2.7 we observe, for all models, that Bonus-Malus (BM) categories B and C have a positive effect on μ while BM category D has a negative effect on μ. Furthermore, we see that Horsepower (HP) categories B and C both have a negative effect on μ. Female drivers have a positive effect on μ in the case of the Poisson, NBII, Delaporte and Sichel models, while they have a negative effect in the case of the ZIP model. BM category A, HP category A and male drivers are the reference categories of μ. The positive values of the coefficients indicate higher risk compared to the reference class, whereas negative values demonstrate lower risk than the reference class. HP category appears in model equations for both μ and σ in the case of the NBII, Delaporte and Sichel models (Tables 2.4, 2.5 and 2.6), gender appears in model equations for both μ and σ in the the case of the NBII and ZIP models (Tables 2.4 and 2.7), and BM category appears in the models equation for both μ and σ in the case of the ZIP model (Table 2.7). Of interest is whether these a priori rating variables have a similar effect on μ and σ. From Tables 2.4, 2.5 and 2.6 we observe that HP categories B and C also have a negative effect on σ in the case of the NBII, Sichel and Delaporte models. However, from Tables 2.4 and 2.7 we see that female drivers have a negative effect on σ in the case of the NBII and ZIP models. From Table 2.7 we observe that BM category has the exact opposite effect on σ, since BM categories B and C have a negative effect on σ, while BM category D has a positive effect on σ in the case of the ZIP model. HP category A and male drivers are the reference categories for σ in the case of the NBII model, HP category A is the reference category for σ in the case of the Delaporte and Sichel models, and BM category A and male drivers are the reference categories for σ in the case of the ZIP model. Furthermore, it is important to note that even if some of the estimated coefficients between the models have the same sign, their estimated values may differ significantly.

Claim Severity Models

Let us now describe the modelling results of the Gamma, Weibull, Weibull Type III, Generalized Gamma and Generalized Pareto GAMLSS for assessing claim severity. Based on the variable selection technique we described in 2.3.1, the following final models were selected.

- Gamma GAMLSS:

$\log(m)$ = Bonus-Malus category + horsepower category + gender
$\log(s)$ = Bonus-Malus category + horsepower category + gender

- Weibull GAMLSS:

$\log(m)$ = Bonus-Malus category + horsepower category + gender
$\log(s)$ = Bonus-Malus category + horsepower category + gender

- Weibull Type III GAMLSS:

$\log(m)$ = Bonus-Malus category + horsepower category + gender
$\log(s)$ = Bonus-Malus category + horsepower category + gender

- Generalized Gamma GAMLSS:

$\log(m)$ = Bonus-Malus category + horsepower category + gender
$\log(s)$ = Bonus-Malus category + horsepower category + gender
n = Bonus-Malus category + gender

- Generalized Pareto GAMLSS:

$\log(m)$ = Bonus-Malus category + horsepower category + gender
$\log(n)$ = Bonus-Malus category + horsepower category + gender
$\log(t)$ = Bonus-Malus category + horsepower category

The results obtained from the Gamma, Weibull, Weibull Type III, Generalized Gamma, and Generalized Pareto models are displayed in Tables 2.8, 2.9, 2.10, 2.11 and 2.12 respectively.

Table 2.8: Results of the Fitted Gamma GAMLSS

Variable m	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	6.3699	0.0351	181.76	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-0.6786	0.0483	-14.037	0.0000
Category C	0.0294	0.0114	2.5670	0.0103
Horsepower				
Category A	0	0	-	-
Category B	-0.6833	0.0250	-27.295	0.0000
Category C	-0.5807	0.0245	-23.676	0.0000
Category D	-0.4082	0.0249	-16.404	0.0000
Gender				
Both	0	0	-	-
Male	-0.1127	0.0304	-3.712	0.0002
Female	-0.0711	0.0307	-2.315	0.0206
Variable s	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.4621	0.0415	-11.143	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	0.5946	0.0403	14.751	0.0000
Category C	-0.0443	0.0205	-2.161	0.0308
Horsepower				
Category A	0	0	-	-
Category B	-0.3130	0.0329	-9.526	0.0000
Category C	-0.3797	0.0322	-11.778	0.0000
Category D	-0.2535	0.0322	-7.883	0.0000
Gender				
Both	0	0	-	-
Male	-0.1589	0.0370	-4.295	0.0000
Female	-0.1788	0.0383	-4.666	0.0000

Table 2.9: Results of the Fitted Weibull GAMLSS

Variable m	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	6.4939	0.0407	159.505	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-0.7118	0.0527	-13.516	0.0000
Category C	0.0307	0.0132	2.322	0.0203
Horsepower				
Category A	0	0	-	-
Category B	-0.6838	0.0298	-22.941	0.0000
Category C	-0.5851	0.0295	-19.858	0.0000
Category D	-0.4066	0.0299	-13.580	0.0000
Gender				
Both	0	0	-	-
Male	-0.1166	0.0335	-3.478	0.0005
Female	-0.0790	0.0340	-2.324	0.0202
Variable s	**Estimate**	**Std Error**	**t-value/Wald 95%**	**P-value**
Intercept	0.3899	0.0422	9.231	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-0.5492	0.0466	-11.778	0.0000
Category C	0.0455	0.0198	2.297	0.0216
Horsepower				
Category A	0	0	-	-
Category B	0.4145	0.0322	12.868	0.0000
Category C	0.4199	0.0315	13.353	0.0000
Category D	0.2806	0.0313	8.976	0.0000
Gender				
Both	0	0	-	-
Male	0.0962	0.0390	2.470	0.0135
Female	0.0967	0.0399	2.427	0.0153

Table 2.10: Results of the Fitted Weibull Type III GAMLSS

Variable m	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	6.3880	0.0397	161.059	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-0.6649	0.0502	-13.242	0.0000
Category C	0.0312	0.0133	2.343	0.0192
Horsepower				
Category A	0	0	-	-
Category B	-0.6968	0.0292	-23.849	0.0000
Category C	-0.5978	0.0289	-20.702	0.0000
Category D	-0.4208	0.0293	-14.359	0.0000
Gender				
Both	0	0	-	-
Male	-0.1184	0.0330	-3.586	0.0003
Female	-0.0798	0.0335	-2.379	0.0174
Variable s	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	0.3883	0.0421	9.233	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-0.5498	0.0462	-11.902	0.0000
Category C	0.0442	0.0199	2.226	0.0261
Horsepower				
Category A	0	0	-	-
Category B	0.4139	0.0322	12.865	0.0000
Category C	0.4197	0.0314	13.354	0.0000
Category D	0.2799	0.0312	8.963	0.0000
Gender				
Both	0	0	-	-
Male	0.0975	0.0389	2.504	0.0123
Female	0.1016	0.0399	2.547	0.0109

Table 2.11: Results of the Fitted Generalized Gamma GAMLSS

Variable m	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	6.3277	0.0502	126.120	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-1.2020	0.1038	-11.578	0.0000
Category C	0.0548	0.0138	3.982	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.6223	0.0247	-25.216	0.0000
Category C	-0.5142	0.0242	-21.284	0.0000
Category D	-0.3608	0.0244	-14.808	0.0000
Gender				
Both	0	0	-	-
Male	-0.1839	0.0465	-3.952	0.0000
Female	-0.1602	0.0469	-3.417	0.0006
Variable s	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.4366	0.0435	-10.026	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	0.5872	0.0548	10.722	0.0000
Category C	-0.0520	0.0211	-2.4645	0.0224
Horsepower				
Category A	0	0	-	-
Category B	-0.2622	0.0346	-7.573	0.0000
Category C	-0.3410	0.0337	-10.125	0.0000
Category D	-0.2311	0.0336	-6.881	0.0000
Gender				
Both	0	0	-	-
Male	-0.2133	0.0396	-5.382	0.0000
Female	-0.2423	0.0408	-5.932	0.0000
Variable n	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	0.7189	0.1904	3.776	0.0001
Bonus-Malus				
Category A	0	0	-	-
Category B	-0.9809	0.3079	-3.186	0.0014
Category C	0.2763	0.0998	2.770	0.0056
Gender				
Both	0	0	-	-
Male	-0.3272	0.1456	-2.247	0.0246
Female	-0.3516	0.1532	-2.295	0.0321

Table 2.12: Results of the Fitted Generalized Pareto GAMLSS

Variable m	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	7.2849	0.0387	188.413	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-1.8305	0.0521	-35.1209	0.0000
Category C	0.0734	0.0140	5.236	0.0000
Horsepower				
Category A	0	0	-	-
Category B	-0.3370	0.0271	-12.459	0.0000
Category C	-0.2263	0.0265	-8.544	0.0000
Category D	-0.1463	0.0266	-5.493	0.0000
Gender				
Both	0	0	-	-
Male	-0.4307	0.0343	-12.572	0.0000
Female	-0.4227	0.0348	-12.137	0.0006
Variable n	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	1.3215	0.0346	38.170	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-0.7347	0.0431	-17.039	0.0000
Category C	0.0445	0.0130	3. 4231	0.0024
Horsepower				
Category A	0	0	-	-
Category B	0.2362	0.0244	9.684	0.0000
Category C	0.2984	0.0239	12.486	0.0000
Category D	0.2250	0.0240	9.372	0.0000
Gender				
Both	0	0	-	-
Male	0.3062	0.0307	9.978	0.0000
Female	0.3400	0.0313	10.881	0.0000
Variable t	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	2.3395	0.0222	105.455	0.0000
Bonus-Malus				
Category A	0	0	-	-
Category B	-1.5622	0.0426	-36.6542	0.0000
Category C	0.0537	0.0132	4.066	0.0000
Horsepower				
Category A	0	0	-	-
Category B	0.5190	0.0246	21.113	0.0000
Category C	0.5859	0.0242	24.191	0.0000
Category D	0.4332	0.0244	17.735	0.0000

The results summarized in Tables 2.8, 2.9, 2.10 and 2.11 show that BM category, HP category and gender appear in the model equations for both m and s in the case of the Gamma, Weibull and Weibull Type III and Generalized Gamma models. BM category B has a negative effect on m and a positive effect on s and BM category C has the exact opposite effect on m and s in the case of the Gamma and Generalized Gamma models while BM category B has a negative effect on both m and s and BM category C has a positive effect on m and s in the case of the Weibull and Weibull Type III models. HP categories B, C and D and male and female drivers have a negative effect on both m and s in the case of the Gamma and Generalized Gamma models while HP categories B, C and D and male and female drivers have a negative effect on m and a positive effect on s in the case of the Weibull and Weibull Type III models. BM category A, HP category A and fleet vehicles used by both male or female drivers are the reference categories for m and s in the case of Gamma, Weibull, Weibull Type III and Generalized Gamma models. Furthermore, in the case of the Generalized Gamma model, BM category and gender are also in the model equations for n. From Table 2.11 we see that BM category B has a negative effect on n while BM category C has a positive effect on n, male and female drivers have a negative effect on n and BM category A and fleet vehicles are the reference categories for n. Finally, in the case of the Generalized Pareto model we observe that BM category, HP category and gender appear in the model equations for both m and n, and BM category and HP category are in the model equations for t. Table 2.12 shows that BM category B has a negative effect on m, n and t, while BM category C has a positive effect on m, n and t. HP categories B, C and D have a negative effect on m and a positive effect on both n and t. Male and female drivers have a negative effect on m and a positive effect on n. Note also that BM category A, HP category A and fleet vehicles are the reference categories for m and n, and BM category A and HP category A are the reference categories for n.

2.3.2 Models Comparison

So far, we have several competing models for the claim frequency and claim severity components. The differences between models produce different premiums. Consequently, to distinguish between these models, this section will purpose to compare them in order to select the optimal one for each case. As suggested by Rigby and Stasinopoulos (2005, 2009) the models have been calibrated with respect to Generalized Akaike Information Criterion (GAIC) which is valid for both nested or non-nested model comparisons. In what follows we present the definition of GAIC.

Let us present first the definition of the Global Deviance (DEV), Akaike Information Criterion (AIC) and Schwartz Bayesian Criterion (SBC), which sometimes is referred to as Bayesian Information Criterion, (BIC), goodness of fit indices. The Global Deviance is given by

$$\hat{D} = -2\log\left(\hat{L}\right), \tag{2.68}$$

where L is the Likelihood which is defined as the probability of observing the sample. Also, the AIC and the SBC are given by

$$AIC = \hat{D} + 2 \times df \tag{2.69}$$

and

$$SBC = BIC = \hat{D} + \log(n) \times df \tag{2.70}$$

respectively, where n is the number of the independent observations assumed by a GAMLSS model, df are the degrees of freedom, that is, the number of fitted parameters in the model and $\hat{D}$ is the fitted global deviance defined above. Both criteria penalized the deviance by a quantity # multiplied by the number of degrees of freedom used in the model (df). Despite their apparent simplicity, the AIC and SBC criteria are based on explicit theoretical considerations, and their aims are not the same. As shown by Kuha (2004), SBC proposes to identify the model with the highest probability to be true, giving that one model under investigation is true while AIC denies the existence of an identifiable true model and, for instance, minimizes the distance or discrepancy between densities. Moreover, in model selection, it has been argued that SBC penalizes large models too heavily.

Both criteria are a special case of the Generalized Akaike Information Criterion (GAIC), defined as

$$GAIC = \hat{D} + \# \times df \tag{2.71}$$

If we let $\# = 2$ we have the AIC, while if we let $\# = \log(n)$ we have the SBC. Note that the penalty is lot more severe for SBC, which means that model selection which is done using SBC will result to a much simpler model[10] than model selection done by AIC. The resulting Global Deviance, AIC and SBC are given in Table 2.13 for the different claim frequency (Panel A) and claim severity (Panel B) fitted models.

Table 2.13: Comparison of Models for the Greek Data Set

Panel A: Claim Frequency Models				
Model	df	Global Deviance	AIC	SBC
Poisson	7	29115.29	29129.29	29182.90
NBII	11	28323.32	28345.32	28429.55
Delaporte	11	28357.99	28379.99	28464.23
Sichel	11	28348.97	28370.97	28455.20
ZIP	12	28503.22	28527.22	28619.11
Panel B: Claim Severity Models				
Model	df	Global Deviance	AIC	SBC
Gamma	16	69665.05	69697.05	69803.11
WEI	16	70794.96	70826.96	70933.02
WEI3	16	70793.02	70825.02	70931.08
GG	21	69427.16	69469.16	69608.37
GP	22	69582.12	69526.12	69771.96

[10] By a simpler model here we mean a model with less degrees of freedom (less parameters).

Overall, with respect to the Global Deviance, AIC and BIC indices, from Panel A we observe the best fitted claim frequency model is the Negative Binomial Type II model, followed closely by the Sichel and Delaporte models, while from the claim severity models in Panel B we see that the best fitting performances are provided by the Generalized Gamma model followed by the Generalized Pareto and Gamma models.

2.3.3 A Priori Risk Classification for the Greek Data Set

In this subsection differences between the claim frequency and severity models are analyzed through the mean and the variance of the number and costs of claims of the policyholders who belong to different risk classes, which are determined by the availability of the relevant a priori characteristics.

The final a priori ratemaking for the claim frequency models contains 24 classes (see Table 2.14). In Table 2.14 a 'YES' indicates the presence of the characteristic corresponding to the column.

Table 2.14: Risk Classes-Claim Frequency Component

Risk Class	BM Category				HP Category			Gender	
	A	B	C	D	A	B	C	Male	Female
1	YES	NO	NO	NO	YES	NO	NO	YES	NO
2	YES	NO	NO	NO	YES	NO	NO	NO	YES
3	YES	NO	NO	NO	NO	YES	NO	YES	NO
4	YES	NO	NO	NO	NO	YES	NO	NO	YES
5	YES	NO	NO	NO	NO	NO	YES	YES	NO
6	YES	NO	NO	NO	NO	NO	YES	NO	YES
7	NO	YES	NO	NO	YES	NO	NO	YES	NO
8	NO	YES	NO	NO	YES	NO	NO	NO	YES
9	NO	YES	NO	NO	NO	YES	NO	YES	NO
10	NO	YES	NO	NO	NO	YES	NO	NO	YES
11	NO	YES	NO	NO	NO	NO	YES	YES	NO
12	NO	YES	NO	NO	NO	NO	YES	NO	YES
13	NO	NO	YES	NO	YES	NO	NO	YES	NO
14	NO	NO	YES	NO	YES	NO	NO	NO	YES
15	NO	NO	YES	NO	NO	YES	NO	YES	NO
16	NO	NO	YES	NO	NO	YES	NO	NO	YES
17	NO	NO	YES	NO	NO	NO	YES	YES	NO
18	NO	NO	YES	NO	NO	NO	YES	NO	YES
19	NO	NO	NO	YES	YES	NO	NO	YES	NO
20	NO	NO	NO	YES	YES	NO	NO	NO	YES
21	NO	NO	NO	YES	NO	YES	NO	YES	NO
22	NO	NO	NO	YES	NO	YES	NO	NO	YES
23	NO	NO	NO	YES	NO	NO	YES	YES	NO
24	NO	NO	NO	YES	NO	NO	YES	NO	YES

As we have mentioned, all the policies were observed for 3.5 years. Thus the estimated expected annual claim frequency and the variance for each risk class are obtained if we let $e_i = e = \frac{1}{3.5}$ in the Eqs (2.8, 2.18, 2.26, 2.34 and 2.39) and the Eqs (2.8, 2.19, 2.27, 2.35 and 2.40) for the case of the Poisson, Negative Binomial Type II (NBII), Delaporte (DEL), Sichel and Zero-Inflated Poisson (ZIP) GAMLSS respectively. The results are summarized in Table

2.15. As expected, the variance of the NBII, Delaporte, Sichel and ZIP GAMLSS exceeds the mean and these models allow for overdispersion. Furthermore, we observe that the biggest differences lie in the variance values of these models. For example, the variance of the expected number of claims for a man who belongs to BM category A and has a car that belongs to HP category A, i.e. for the reference class, is equal to 0.1264, 0.2140, 0.1868, 0.1884 and 0.1391 while the variance of the expected number of claims for a woman who shares common characteristics is equal to 0.1354, 0.1964, 0.2100, 0.2128 and 0.1507 in the case of the Poisson, NBII, Delaporte, Sichel and ZIP GAMLSS respectively.

Table 2.15: A Priori Risk Classification for the Greek Dataset, Claim Frequency Models

Risk	Poisson		NBII		DEL		Sichel		ZIP	
Class	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1	0.1264	0.1264	0.1267	0.2140	0.1255	0.1868	0.1258	0.1884	0.1261	0.1391
2	0.1354	0.1354	0.1357	0.1964	0.1371	0.2100	0.1377	0.2128	0.1414	0.1507
3	0.0997	0.0997	0.1001	0.1318	0.0984	0.1127	0.0984	0.1046	0.0983	0.1062
4	0.1068	0.1068	0.1072	0.1293	0.1075	0.1245	0.1078	0.1152	0.1102	0.1158
5	0.1176	0.1176	0.1178	0.1592	0.1165	0.1381	0.1166	0.1260	0.1148	0.1256
6	0.1259	0.1259	0.1261	0.1550	0.1273	0.1529	0.1277	0.1390	0.1288	0.1365
7	0.2323	0.2323	0.2385	0.4029	0.2388	0.4602	0.2383	0.4629	0.2742	0.2777
8	0.2486	0.2486	0.2555	0.3699	0.2608	0.5247	0.2610	0.5302	0.2527	0.2543
9	0.1832	0.1832	0.1885	0.2483	0.1872	0.2388	0.1863	0.2089	0.2136	0.2158
10	0.1961	0.1961	0.2020	0.2435	0.2044	0.2659	0.2040	0.2311	0.1969	0.1980
11	0.2160	0.2160	0.2217	0.2998	0.2217	0.2995	0.2208	0.2548	0.2496	0.2524
12	0.2312	0.2312	0.2375	0.2918	0.2422	0.3349	0.2418	0.2825	0.2300	0.2314
13	0.3059	0.3059	0.2931	0.4950	0.2991	0.6462	0.3001	0.6564	0.3127	0.3616
14	0.3276	0.3276	0.3140	0.4545	0.3266	0.7406	0.3286	0.7559	0.3301	0.3610
15	0.2413	0.2413	0.2317	0.3050	0.2344	0.3153	0.2347	0.2705	0.2438	0.2734
16	0.2584	0.2584	0.2482	0.2992	0.2560	0.3525	0.2571	0.2999	0.2573	0.2761
17	0.2845	0.2845	0.2725	0.3684	0.2777	0.3997	0.2782	0.3320	0.2847	0.3252
18	0.3047	0.3047	0.2919	0.3586	0.3032	0.4487	0.3047	0.3692	0.3005	0.3261
19	0.0493	0.0493	0.0478	0.0808	0.0482	0.0573	0.0486	0.0579	0.0476	0.0542
20	0.0527	0.0527	0.0512	0.0742	0.0527	0.0634	0.0532	0.0645	0.0634	0.0701
21	0.0388	0.0388	0.0378	0.0498	0.0378	0.0399	0.0380	0.0389	0.0371	0.0411
22	0.0416	0.0416	0.0405	0.0489	0.0413	0.0438	0.0417	0.0427	0.0494	0.0534
23	0.0458	0.0458	0.0444	0.0601	0.0448	0.0480	0.0450	0.0465	0.0433	0.0488
24	0.0490	0.0490	0.0476	0.0585	0.0489	0.0527	0.0493	0.0510	0.0577	0.0632

The final a priori ratemaking for the claim severity models contains 36 classes (see Table 2.16).

Table 2.16: Risk Classes-Claim Severity Component

Risk Class	BM Category			HP Category				Gender		
	A	B	C	A	B	C	D	Both	Male	Female
1	YES	NO	NO	YES	NO	NO	NO	YES	NO	NO
2	YES	NO	NO	YES	NO	NO	NO	NO	YES	NO
3	YES	NO	NO	YES	NO	NO	NO	NO	NO	YES
4	YES	NO	NO	NO	YES	NO	NO	YES	NO	NO
5	YES	NO	NO	NO	YES	NO	NO	NO	YES	NO
6	YES	NO	NO	NO	YES	NO	NO	NO	NO	YES
7	YES	NO	NO	NO	NO	YES	NO	YES	NO	NO
8	YES	NO	NO	NO	NO	YES	NO	NO	YES	NO
9	YES	NO	NO	NO	NO	YES	NO	NO	NO	YES
10	YES	NO	NO	NO	NO	NO	YES	YES	NO	NO
11	YES	NO	NO	NO	NO	NO	YES	NO	YES	NO
12	YES	NO	NO	NO	NO	NO	YES	NO	NO	YES
13	NO	YES	NO	YES	NO	NO	NO	YES	NO	NO
14	NO	YES	NO	YES	NO	NO	NO	NO	YES	NO
15	NO	YES	NO	YES	NO	NO	NO	NO	NO	YES
16	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO
17	NO	YES	NO	NO	YES	NO	NO	NO	YES	NO
18	NO	YES	NO	NO	YES	NO	NO	NO	NO	YES
19	NO	YES	NO	NO	NO	YES	NO	YES	NO	NO
20	NO	YES	NO	NO	NO	YES	NO	NO	YES	NO
21	NO	YES	NO	NO	NO	YES	NO	NO	NO	YES
22	NO	YES	NO	NO	NO	NO	YES	YES	NO	NO
23	NO	YES	NO	NO	NO	NO	YES	NO	YES	NO
24	NO	YES	NO	NO	NO	NO	YES	NO	NO	YES
25	NO	NO	YES	YES	NO	NO	NO	YES	NO	NO
26	NO	NO	YES	YES	NO	NO	NO	NO	YES	NO
27	NO	NO	YES	YES	NO	NO	NO	NO	NO	YES
28	NO	NO	YES	NO	YES	NO	NO	YES	NO	NO
29	NO	NO	YES	NO	YES	NO	NO	NO	YES	NO
30	NO	NO	YES	NO	YES	NO	NO	NO	NO	YES
31	NO	NO	YES	NO	NO	YES	NO	YES	NO	NO
32	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO
33	NO	NO	YES	NO	NO	YES	NO	NO	NO	YES
34	NO	NO	YES	NO	NO	NO	YES	YES	NO	NO
35	NO	NO	YES	NO	NO	NO	YES	NO	YES	NO
36	NO	NO	YES	NO	NO	NO	YES	NO	NO	YES

Table 2.17 gives the estimated expected claim severity and the variance for each risk class obtained from the Gamma (GA), Weibull (WEI), Weibull Type III (WEI3), Generalized Gamma (GG) and Generalized Pareto (GP) GAMLSS according to the Eqs (2.44, 2.49, 2.54, 2.60 and 2.66) and the Eqs (2.45, 2.50, 2.55, 2.61 and 2.67) respectively. As expected, similarly to the case of the claim frequency models, we see that the biggest differences between the claim severity models lie in their variance values. For instance, the variance of the expected claim costs for a fleet vehicle that belongs to HP category A, used by both a man and a woman, and belongs to BM category A, i.e. for the reference class, is equal to 135347.30, 169637.36, 168267.90, 148196.45 and 142078.20, while the variance of the expected claim costs for a private car that belongs to HP category A and is used by a man who belongs to BM category A is equal to 78621.46, 110315.30, 111018.27, 72875.39 and 89891.64 in the case of the Gamma, WEI, WEI3, Generalized Gamma and Generalized Pareto GAMLSS.

Table 2.17: A Priori Risk Classification for the Greek Dataset, Claim Severity Models

Risk	GA		WEI		WEI3		GG		GP	
Class	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1	584.00	135347.30	597.96	169637.36	594.66	168267.90	591.62	148196.45	583.03	142078.20
2	521.75	78621.46	526.73	110315.30	528.26	111018.27	504.93	72875.39	514.78	89891.64
3	543.92	82108.76	546.89	118812.19	549.06	119033.67	516.38	72022.76	536.75	95624.76
4	294.89	18453.33	295.51	19539.26	296.25	19714.32	310.72	24073.97	300.72	26138.91
5	263.46	10719.29	263.36	13061.64	263.17	13063.90	262.37	11431.24	265.51	16207.29
6	274.65	11194.75	273.45	14069.47	273.53	14009.16	268.44	11300.70	276.84	17199.88
7	326.75	19827.00	326.18	23575.68	327.07	23782.38	344.55	25257.37	333.03	29934.69
8	291.93	11517.24	290.72	15762.85	290.55	15759.88	290.30	11905.58	294.05	18551.62
9	304.32	12028.09	301.85	16979.11	301.99	16900.22	297.05	11770.71	306.59	19686.62
10	388.27	36033.34	390.33	43363.58	390.41	43561.39	404.41	43421.71	394.23	46566.35
11	346.88	20931.28	346.96	28820.08	346.82	28847.75	341.83	20685.10	348.08	29009.46
12	361.62	21859.70	360.26	31043.01	360.47	30934.51	349.72	20448.12	362.94	30803.37
13	296.28	114416.43	352.27	172055.65	305.85	130297.57	265.02	129671.66	250.44	178704.02
14	264.70	66462.96	297.20	100325.75	271.70	84002.18	164.63	25281.89	221.13	121573.35
15	275.95	69410.96	308.51	107997.62	282.39	89988.87	165.62	23924.98	230.56	130384.50
16	149.60	15599.59	151.45	13989.85	152.36	14234.31	119.36	13878.38	108.56	11957.62
17	133.66	9061.59	132.51	8946.20	135.36	9359.92	83.06	3737.71	95.85	7832.20
18	139.34	9463.52	137.58	9634.51	140.68	10034.46	84.12	3595.64	99.94	8364.86
19	165.77	16760.83	166.98	16833.12	168.23	17162.40	127.92	13265.26	118.52	12850.63
20	148.10	9736.14	146.14	10772.70	149.44	11287.22	91.28	3837.95	104.64	8402.63
21	154.39	10167.99	151.73	11601.59	155.32	12100.73	92.58	3705.93	109.11	8972.35
22	196.98	30460.93	206.75	33670.27	200.79	31936.98	157.66	26065.04	145.27	23671.24
23	175.98	17694.34	179.31	21059.67	178.37	20903.82	108.54	6804.49	128.26	15622.57
24	183.46	18479.18	186.15	22677.63	185.39	22406.46	109.84	6535.28	133.74	16699.22
25	601.42	131373.54	613.31	164111.60	613.51	165097.24	591.91	131860.30	618.24	151126.30
26	537.32	76313.11	541.27	107216.06	545.01	109018.66	511.81	65142.30	545.87	95523.00
27	560.14	79698.02	561.99	115476.70	566.45	116893.25	524.41	64612.06	569.17	101603.68
28	303.69	17911.53	304.87	19167.52	305.64	19385.37	317.57	22467.66	319.63	28068.18
29	271.32	10404.57	271.88	12831.92	271.51	12847.07	270.40	10712.80	282.22	17391.14
30	282.84	10866.07	282.31	13822.11	282.20	13776.66	276.98	10614.65	294.27	18454.66
31	336.50	19244.87	336.52	23129.37	337.44	23385.76	353.80	23820.14	354.06	32168.91
32	300.64	11179.09	300.14	15486.68	299.76	15498.33	300.25	11270.97	312.61	19922.38
33	313.40	11674.94	311.64	16681.73	311.56	16619.75	307.55	11165.35	325.96	21139.48
34	399.85	34975.39	402.16	42412.94	402.78	42819.65	412.50	40339.87	418.90	49941.43
35	357.23	20316.73	357.83	28251.56	357.81	28364.50	351.74	19299.08	369.87	31088.48
36	372.41	21217.89	371.54	30430.94	371.89	30416.57	360.31	19124.05	385.65	33007.98

Overall, from Tables 2.15 and 2.17 it can be seen that both the claim frequency and severity models are better compared through their variance values, and thus GAMLSS modelling is justified because it enables us to use all the available information in the estimation of these values through the use of the important a priori rating variables for the number and the costs of claims respectively.

2.3.4 Calculation of the Premiums According to the Expected Value and Standard Deviation Principles

Consider a policyholder i who belongs to a group of policyholders, whose number of claims, denoted as K_i, are independent, for $i = 1, .., n$. Let $X_{i,k}$ be the cost of the kth claim reported by the policyholder i and assume that the individual claim costs $X_{i,1}, X_{i,2}, ..., X_{i,n}$ are independent. It is assumed that the number of claims of each policyholder that belongs to a certain group is independent of the severity of each claim in order to deal with the frequency and the severity components separately.

A premium principle is a rule for assigning a premium to an insurance risk. In this section the premiums rates will be calculated via two well-known premium principles, the expected value and the standard deviation premium principles.

- The premium rates calculated according to the expected value principle are given by

$$P_1 = (1 + w_1)\, E(K_i)\, (1 + w_2)\, E(X_{i,k}), \tag{2.72}$$

where $w_1 > 0$ and $w_2 > 0$ are risk loads.

- The premium rates calculated according to the standard deviation principle are given by

$$P_2 = \left[E(K_i) + \omega_1 \sqrt{Var(K_i)}\right] \left[E(X_{i,k}) + \omega_2 \sqrt{Var(X_{i,k})}\right], \tag{2.73}$$

where $\omega_1 > 0$ and $\omega_2 > 0$ are risk loads.

In the following example (Table 2.18), six different groups of policyholders have been considered. In Table 2.18 a 'YES' indicates the presence of the characteristic corresponding to the column.

Table 2.18: The Six Different Groups of Policyholders to Be Compared

Group	BM Category A	HP NO-33	HP 34-66	HP 100-132	Male	Female
1	YES	YES	NO	NO	YES	NO
2	YES	YES	NO	NO	NO	YES
3	YES	NO	YES	NO	YES	NO
4	YES	NO	YES	NO	NO	YES
5	YES	NO	NO	YES	YES	NO
6	YES	NO	NO	YES	NO	YES

We will calculate the premiums P_1 and P_2 that must be paid by a specific group of policyholders based on the alternative models for assessing claim frequency and the various claim severity models. We assume that $w_1 = w_2 = \omega_1 = \omega_2 = \frac{1}{10}$. The premiums P_1 and P_2 are obtained in Table 2.19 by substituting into Eqs (2.72 and 2.73) the corresponding $E(K_i)$ and $Var(K_i)$, and $E(X_{i,k})$ and $Var(X_{i,k})$ values to these six different groups of policyholders, which were displayed in Tables 2.15 and 2.17 for the case of the Poisson, NBII, Delaporte, Sichel and ZIP GAMLSS, and the Gamma, Weibull, Weibull Type III, Generalized Gamma and Generalized Pareto GAMLSS respectively. From Table 2.19 we observe, for all models, that the premiums P_1 and P_2 do not differ much. For instance, consider a man who belongs to BM category A and has a car with a HP between 34-66. In the case of the Poisson model and the corresponding severity models, P_1 is equal to 31.78, 31.77, 31.75, 31.65 and 32.03 euros, while P_2 equals 35.95, 36.07, 36.05, 35.85 and 36.52 euros. In the case of the NBII model and the corresponding severity models, P_1 is equal to 31.91, 31.90, 31.88, 31.78 and 32.16 euros, while P_2 equals 37.35, 37.48, 37.46, 37.25 and 37.95 euros. In the case of the Delaporte model and the corresponding severity models, P_1 is equal to 31.37, 31.36, 31.33, 31.24 and 31.61 euros, while P_2 equals 36.14, 36.27, 36.24, 36.04 and 36.72 euros. In the case of the Sichel model and the corresponding severity models, P_1 is equal to 31.37, 31.36, 31.33, 31.24 and 31.61 euros, while P_2 equals 35.80, 35.93, 35.90, 35.70 and 36.38 euros. In the case of the ZIP model and the corresponding severity models, P_1 is equal to 36.62, 36.46, 36.47, 35.80 and 36.91 euros, while P_2 equals 41.14, 41.15, 41.16, 40.25 and 41.82 euros.

Table 2.19: Premium Rates Calculated Via the Expected Value and Standard Deviation Principles

Group	PO-GA		PO-WEI		PO-WEI3		PO-GG		PO-GP	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
1	40.2946	44.3448	40.2793	44.5028	40.2503	44.4722	40.1279	44.2231	40.6082	45.0619
2	44.9970	49.1158	44.8004	49.1297	44.8135	49.1391	43.9796	48.0550	45.3558	49.9293
3	31.7830	35.9450	31.7710	36.0730	31.7480	36.0482	31.6515	35.8463	32.0303	36.5261
4	35.4925	39.7840	35.3374	39.7953	35.3477	39.8030	34.6900	38.9248	35.7755	40.4430
5	79.7985	89.0400	80.5602	90.6845	80.7942	90.9493	77.2260	86.1468	78.7325	88.2257
6	89.1126	98.5955	89.5992	100.1082	89.9547	100.4874	84.6006	93.5402	87.9379	97.7515
Group	NBII-GA		NBII-WEI		NBII-WEI3		NBII-GG		NBII-GP	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
1	40.3903	47.3588	40.3750	47.5275	40.3458	47.4948	40.2232	47.2290	40.7045	48.1246
2	45.0967	51.3464	44.9000	51.3610	44.9128	51.3708	44.0770	50.2374	45.4563	52.1968
3	31.9105	37.3493	31.8984	37.4824	31.8754	37.4566	31.7785	37.2468	32.1588	37.9532
4	35.6254	40.8331	35.4700	40.8447	35.4801	40.8525	34.8200	39.9513	35.9095	41.5094
5	79.9880	95.0917	80.7514	96.8480	80.9860	97.1309	77.4093	92.0020	78.9194	94.2221
6	89.3100	103.0732	89.7977	104.6550	90.1540	105.0510	84.7881	97.7883	88.1327	102.1909
Group	DEL-GA		DEL-WEI		DEL-WEI3		DEL-GG		DEL-GP	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
1	40.0077	46.1980	39.9925	46.3625	39.9637	46.3306	39.8422	46.0711	40.3190	46.9449
2	45.5620	52.1760	45.3630	52.1908	45.3762	52.2008	44.5318	51.0492	45.9253	53.0402
3	31.3686	36.1354	31.3567	36.2641	31.3341	36.2392	31.2388	36.0362	31.6127	36.7197
4	35.7251	40.7265	35.5690	40.7381	35.5794	40.7459	34.9173	39.8470	36.0100	41.4011
5	79.2304	92.7607	79.9866	94.4740	80.2190	94.7500	76.6762	89.7467	78.1720	91.9124
6	90.2314	104.7387	90.7241	106.3456	91.0841	106.7484	85.6628	99.3684	89.0420	103.8421
Group	SI-GA		SI-WEI		SI-WEI3		SI-GG		SI-GP	
	P_1	$P2$	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
1	40.1034	46.3306	40.0881	46.4957	40.0592	46.4637	39.9374	46.2034	40.4154	47.0800
2	45.7614	52.4340	45.5614	52.4489	45.5748	52.4590	44.7267	51.3016	46.1263	53.3025
3	31.3686	35.7989	31.3567	35.9264	31.3341	35.9017	31.2388	35.7006	31.6127	36.3777
4	35.8248	40.4289	35.6683	40.4404	35.6787	40.4481	35.0148	39.5558	36.1105	41.0985
5	79.4197	93.0272	80.1778	94.7453	80.4107	95.0221	76.8594	90.0045	78.3588	92.1765
6	90.6263	105.2565	91.1212	106.8714	91.4827	107.2762	86.0377	99.8600	89.4317	104.3555
Group	ZIP-GA		ZIP-WEI		ZIP-WEI3		ZIP-GG		ZIP-GP	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
1	40.1990	44.7401	40.1837	44.8994	40.1547	44.8685	40.0327	44.6172	40.5118	45.4635
2	46.9910	51.4043	46.7857	51.4189	46.7993	51.4287	45.9285	50.2941	47.3657	52.2557
3	31.3367	35.8390	31.3248	35.9666	31.3022	35.9420	31.2071	35.7406	31.5806	36.4185
4	36.6224	41.1386	36.4624	41.1503	36.4730	41.1582	35.7943	40.2502	36.9144	41.8200
5	79.6091	89.8335	80.3690	91.4926	80.6024	91.7600	77.0427	86.9146	78.5457	89.0120
6	93.0615	103.1895	93.5696	104.7726	93.9409	105.1700	88.3495	97.8986	91.8347	102.3061

A framework has been provided for modelling count and loss data. Note that different count and loss models can be fitted and the parameters of the distributions can be modelled generally as linear, nonlinear and/or smoothing functions of explanatory variables. In this case the problem of the choice between models becomes more acute and is an interesting topic of further research.

Chapter 3

The Design of Optimal Bonus-Malus Systems Using Alternative Mixed Poisson Distributions as Models of Claim Counts

3.1 Introduction

In a priori risk classification, many important factors cannot be taken into account a priori when pricing motor third party liability insurance products. For instance, reaction times, aggressive driving behavior or theoretical and practical driving experience are difficult to integrate into a priori risk classification. As a result, heterogeneity is still observed in tariff cells despite the use of many classification variables. In order to reduce the gap between the individual's premium and risk and to increase incentives for road safety, the individual's past record must taken into consideration through the use of an a posteriori model. In motor insurance, credibility techniques can be used to re-evaluate the annual expected claim frequency given the past claim record. Bayesian statistics offer an intellectually acceptable approach to credibility theory. One of the commercial simplifications of credibility theory is known as Bonus-Malus Systems (BMSs).

As we mentioned in Chapter 1, a BMS penalizes policyholders responsible for one or more claims by a premium surcharge (malus) and rewards the policyholders who had a claim-free year by awarding discount of the premium (bonus). A basic interest of the actuarial literature is the construction of an optimal or 'ideal' BMS defined as a system obtained through Bayesian analysis. A BMS is called optimal if it is financially balanced for the insurer: the total amount of bonuses must be equal to the total amount of maluses and if it is fair for the policyholder: the premium paid for each policyholder is proportional to the risk that they impose on the pool. Optimal BMSs can be broadly derived in two ways; based only on the a posteriori classification criteria and based on both the a priori and the a posteriori classification criteria. Typically, classification criteria such as the number of accidents of the policyholder and the severity of each accident are considered as a posteriori, while variables such as the characteristics of the driver and the automobile are considered as a priori classification criteria. These systems, besides encouraging policyholders to drive carefully (i.e. counteracting moral hazard), aim to better assess individual risks. The amount of premium is adjusted each year on the basis of the individual claims experience using techniques from credibility theory. Contributions to the literature of BMS include, among others, Lemaire (1995), Dionne and Vanasse (1989, 1992), Frangos and Vrontos (2001), Pinquet et al. (2001), Brouhns et al. (2003), Pitrebois et al. (2006) and Mahmoudvand and Hassani (2009) and the references therein. A review of BMS, and actuarial models for risk classification and insurance ratemaking can be found in Denuit et al. (2007). The literature more closely related to ours is Lemaire (1995) and Dionne and Vanasse (1989, 1992). Lemaire (1995) considered, among other BMS, the optimal BMS obtained using the quadratic error loss function and the expected value premium calculation principle approximating the claim frequency distribution by the Negative Binomial. Dionne and Vanasse (1989, 1992) developed a BMS that integrates a priori and a posteriori information on an individual basis. For this purpose they used the Negative Binomial regression model for assessing claim frequency.

Our first contribution is the development of an optimal BMS using the Sichel distribution for assessing claim frequency. This system is proposed as an alternative to the optimal BMS provided by the Negative Binomial distribution (Lemaire,1995). In fact the Sichel distribution (Sichel, 1985) differs from the standard Negative Binomial one by using an Generalized Inverse Gaussian (GIG) mixing distribution for the parameter of the Poisson density, i.e. the expected claim frequency, instead of the Gamma one, which the derivation of the Negative Binomial distribution is based on. It is important to note that different parameterizations of the Generalized Inverse Gaussian distribution may lead to other models. An additional advantage of the Sichel model is that it can be considered as a candidate model for highly dispersed count data. We also consider the optimal BMS obtained by the Poisson-Inverse Gaussian distribution (PIG), which is a special case of the Sichel distribution. Our second contribution is the development of a generalized BMS that integrates the a priori and the a posteriori information on an individual basis, extending the framework developed by Dionne and Vanasse (1989, 1992). This is achieved by using the generalized additive models for location, scale and shape (GAMLSS). As mentioned in Chapter 2, in the GAMLSS, the exponential family distribution assumption for the response variable is relaxed and replaced by a general distribution family, including highly skewed continuous and discrete distributions. Thus, the GAMLSS are suited to model highly dispersed count data. Within the framework of the GAMLSS we present the Sichel GAMLSS for assessing claim frequency as an alternative to the Negative Binomial regression model of Dionne and Vanasse (1989, 1992). Furthermore we consider the PIG GAMLSS for assessing claim frequency. With the aim of constructing an optimal BMS by updating the posterior mean claim frequency, we adopt the parametric linear formulation of these models and we allow only their mean parameter to be modelled as a function of the significant a priori rating variables for the number of claims. In the resulting generalized system, the premium is a function of the years that the policyholder is in the portfolio, the number of accidents and the significant a priori rating variables for the number of accidents.

This chapter is laid out as follows. In Section 3.2 we consider the design of an optimal BMS based on the a posteriori criteria. The design presented in Section 3.3 is based on both the a posteriori and the a priori classification criteria. Finally, an application to the data set presented in Chapter 2 can be found in Section 3.4.

3.2 The Design of an Optimal BMS Based on the a Posteriori Criteria

This Section presents the development of an optimal BMS using the Sichel distribution for assessing claim frequency. This system is proposed as an alternative to the optimal BMS provided by the Negative Binomial distribution (see Lemaire, 1995). In fact the Sichel distribution works very well when the data is highly dispersed. In other situations, it works similar to the Negative Binomial distribution. Furthermore, we consider the optimal BMS obtained by the Poisson-Inverse Gaussian distribution, which is a special case of the Sichel distribution. In the above setup optimality is achieved by minimizing the insurer's risk, following the current methodology as presented in Chapter 1.

3.2.1 The Negative Binomial Model

We consider first the design of an optimal BMS using the Negative Binomial distribution for assessing claim frequency[1]. The portfolio is considered to be heterogeneous and all policyholders have constant but unequal underlying risks of having an accident. We assume that the number of claims k given the parameter λ is distributed as a Poisson(λ),

$$P(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, \tag{3.1}$$

for $k = 0, 1, 2, 3, ...$ and $\lambda > 0$, where the parameter λ is the mean claim frequency which varies from individual to individual, denoting the different underlying risk of each policyholder having an accident. Following the setup of Lemaire (1995), we consider that the structure function follows a Gamma distribution which has a probability density function of the form

$$u(\lambda) = \frac{\lambda^{\alpha-1}\tau^{\alpha}\exp(-\tau\lambda)}{\Gamma(\alpha)}, \tag{3.2}$$

for $\lambda > 0, \alpha > 0, \tau > 0$, with mean $E(\lambda) = \frac{\alpha}{\tau}$ and variance $Var(\lambda) = \frac{\alpha}{\tau^2}$. Then it can be proved that the unconditional distribution of the number of claims k is a Negative Binomial (α, τ) distribution with probability density function

$$P(k) = \binom{k+\alpha-1}{k} p^{\alpha} q^{k},\ p = \left(\frac{\tau}{1+\tau}\right), q = \left(\frac{1}{1+\tau}\right), \tag{3.3}$$

for $k = 0, 1, 2, 3, ...$, where $\lambda > 0, \alpha > 0, \tau > 0$. The mean and the variance of k are given by $E(k) = \mu = \frac{\alpha}{\tau}$ and $Var(k) = \frac{\alpha}{\tau}\left(1 + \frac{1}{\tau}\right)$ respectively.

[1] We use the same notation as in Frangos and Vrontos (2001).

Posterior Structure Function

Consider a policyholder with claim history $k_1, ..., k_t$ where k_j is the number of claims that the policyholder had in year $j, j = 1, ..., t$. Let us denote with $K = \sum_{j=1}^{t} k_j$ the total number of claims that the policyholder had in t years. Applying Bayes theorem we obtain the posterior structure function of λ for a policyholder or a group of policyholders with claim history $k_1, ..., k_t$, denoted as $u(\lambda|k_1, ..., k_t)$ and given by

$$u(\lambda|k_1, ..., k_t) = \frac{(\tau + t)^{K+\alpha} \lambda^{K+\alpha-1} e^{-(\tau+t)\lambda}}{\Gamma(\alpha + K)}, \tag{3.4}$$

which is the probability density function of a gamma $(\alpha + K, \tau + t)$. For information about the proof of Eqs (3.3 and 3.4) refer to Lemaire (1995). Also, a more general proof of Eqs (3.3 and 3.4) can be found in Chapter 5 where we consider the case of the n-component mixture of Negative Binomial distributions derived by assuming that the number of claims $k|\lambda$ is distributed according to a Poisson(λ) and that the structure function follows an n-component mixture of Gamma distributions.

Optimal Choice of $\hat{\lambda}_i^{t+1}$

Consequently, by using the quadratic error loss function, the optimal choice of λ at time $t+1$, $\hat{\lambda}_{t+1}$, for a policyholder with claim history $k_1, ..., k_t$, is the mean of the posterior structure function given by Eq. (3.4), that is

$$\hat{\lambda}_{t+1}(k_1, ..., k_t) = \frac{K + \alpha}{\tau + t}. \tag{3.5}$$

3.2.2 The Sichel Model

Let us consider now the construction of an optimal BMS using the Sichel distribution to model the claim frequency distribution. The Sichel is a compound Poisson distribution and it can be derived by assuming that the mixing distribution of the Poisson rate λ is a Generalized Inverse Gaussian distribution.

As previously, the portfolio is considered to be heterogeneous and all policyholders have constant but unequal underlying risks of having an accident and $k|\lambda$ is distributed according to a Poisson(λ). Let us now assume that the mean claim frequency λ follows a Generalized Inverse Gaussian distribution, denoted as GIG(μ, σ, ν), with probability density function given by

$$u(\lambda) = \frac{\left(\frac{c}{\mu}\right)^{\nu} \lambda^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2K_{\nu}\left[\frac{1}{\sigma}\right]}, \tag{3.6}$$

for $\lambda > 0$, where $\mu > 0, \sigma > 0$ and $-\infty < \nu < \infty$, and where $c = \frac{K_{\nu+1}\left(\frac{1}{\sigma}\right)}{K_{\nu}\left(\frac{1}{\sigma}\right)}$, where

$$K_\nu(z) = \frac{1}{2}\int_0^\infty x^{\nu-1}\exp\left[-\frac{1}{2}z\left(x+\frac{1}{x}\right)\right]dx,$$

is the modified Bessel function of the third kind of order ν with argument z. Eq. (3.6) is obtained from a reparameterization of equation (2.2) of the GIG distribution of Jørgensen (1982) or equation (15.74) from Johnson et al. (1994) p 284. The mean and the variance of λ are given by $E(\lambda) = \mu$ and $Var(\lambda) = \mu^2\left[\frac{2\sigma(\nu+1)}{c} + \frac{1}{c^2} - 1\right]$ respectively. Note that the gamma is a limiting distribution of Eq. (3.6) obtained by letting $\sigma \to \infty$ for $\nu > 0$.

Proposition 1 *If we let $\nu = -0.5$ in (3.6) the Generalized Inverse Gaussian distribution can be reduced to an Inverse Gaussian distribution with pdf given by*

$$u(\lambda) = \frac{\sqrt{\mu}}{\sqrt{2\pi\sigma\lambda^3}}\exp\left[-\frac{1}{2\sigma\mu\lambda}(\lambda-\mu)^2\right], \tag{3.7}$$

for $\lambda > 0$ and $\sigma > 0$, where $E(\lambda) = \mu$ and where $Var(\lambda) = \mu^2\sigma$.

Proof. If we let $\nu = -0.5$ in (3.6), then we have

$$u(\lambda) = \frac{\left(\frac{c}{\mu}\right)^{-\frac{1}{2}}\lambda^{-\frac{3}{2}}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda+\frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2K_{-\frac{1}{2}}\left[\frac{1}{\sigma}\right]}.$$

Making use of the relationships between the modified Bessel functions of different orders (see Abramowitz and Stegun, 1974):

$$K_{1/2}(z) = K_{-1/2}(z) = \left(\frac{\pi}{2z}\right)^{\frac{1}{2}}\exp(-z),\ K_\nu(z) = K_{-\nu}(z),$$
$$K_{\nu+1}(z) = \frac{2\nu}{z}K_\nu(z) + K_{\nu-1}(z) \text{ and } K_{3/2}(z) = K_{1/2}(z)\left(z+\frac{1}{z}\right)$$

one can obtain that

$$c = \frac{K_{1/2}\left(\frac{1}{\sigma}\right)}{K_{-1/2}\left(\frac{1}{\sigma}\right)} = 1 \text{ and } K_{-\frac{1}{2}}\left(\frac{1}{\sigma}\right) = \left(\frac{\pi\sigma}{2}\right)^{\frac{1}{2}}e^{-\frac{1}{\sigma}}$$

and hence

$$u(\lambda) = \frac{\left(\frac{1}{\mu}\right)^{-\frac{1}{2}}\lambda^{-\frac{3}{2}}\exp\left[-\frac{1}{2\sigma}\left(\frac{1}{\mu}\lambda+\frac{1}{c}\frac{1}{\lambda}\right)\right]}{2\left(\frac{\pi\sigma}{2}\right)^{\frac{1}{2}}e^{-\frac{1}{\sigma}}} = \frac{\sqrt{\mu}}{\sqrt{2\pi\sigma\lambda^3}}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda+\frac{\mu}{c}\frac{1}{\lambda}+2\right)\right]$$

$$= \frac{\sqrt{\mu}}{\sqrt{2\pi\sigma\lambda^3}}\exp\left[-\frac{1}{2\sigma\mu\lambda}(\lambda-\mu)^2\right].$$

Proposition 2 *Let $k|\lambda$ be distributed according to a Poisson (λ) and let the distribution of the parameter λ be the GIG(μ, σ, ν) given by Eq. (3.6). The unconditional distribution of the number of claims k is given by a Sichel (μ, σ, ν) distribution, which has a probability density function of the form*

$$P(k) = \frac{\left(\frac{\mu}{c}\right)^k K_{k+\nu}(a)}{k!\,(a\sigma)^{k+\nu} K_\nu\left(\frac{1}{\sigma}\right)}, \tag{3.8}$$

for $k = 0, 1, 2, 3, ...,$ where $a^2 = \sigma^{-2} + 2\mu\,(c\sigma)^{-1}$.

Note that the mean of k is equal to $E(k) = \mu$ and the variance of k is equal to $Var(k) = \mu + \mu^2\left[\frac{2\sigma(\nu+1)}{c} + \frac{1}{c^2} - 1\right]$. Like the Negative Binomial the variance of the Sichel exceeds its mean, a desirable property which is common for all mixtures of Poisson distributions and allows us to deal with data that present overdispersion.

Proof. Considering the assumptions of the model the unconditional distribution of the number of claims k can be derived by solving the convolution integral

$$P(k) = \int_0^\infty P(k|\lambda)\,u(\lambda)\,d\lambda$$

$$= \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!} \frac{\left(\frac{c}{\mu}\right)^\nu \lambda^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2K_\nu\left(\frac{1}{\sigma}\right)} d\lambda$$

$$= \frac{\left(\frac{\mu}{c}\right)^k}{k!K_\nu\left(\frac{1}{\sigma}\right)} \frac{1}{2} \int_0^\infty \left(\frac{c}{\mu}\right)^{k+\nu} \lambda^{k+\nu-1} \exp\left[-\frac{1}{2}\left(\left(\frac{c}{\mu\sigma}+2\right)\lambda + \frac{\mu}{c\sigma}\frac{1}{\lambda}\right)\right] d\lambda.$$

If we let $a = \sqrt{\left(\frac{c}{\mu\sigma}+2\right)\frac{\mu}{c\sigma}} = \sqrt{\sigma^{-2} + 2\mu\,(c\sigma)^{-1}}$, then we have

$$P(k) = \frac{\left(\frac{\mu}{c}\right)^k}{k!K_\nu\left(\frac{1}{\sigma}\right)(a\sigma)^{k+\nu}} \frac{1}{2} \int_0^\infty \left(\frac{c}{\mu}\right)^{k+\nu} (a\sigma)^{k+\nu} \lambda^{k+\nu-1} \exp\left[-\frac{1}{2}\left(\left(\frac{c}{\mu\sigma}+2\right)\lambda + \frac{\mu}{c\sigma}\frac{1}{\lambda}\right)\right] d\lambda$$

$$= \frac{\left(\frac{\mu}{c}\right)^k}{k!K_\nu\left(\frac{1}{\sigma}\right)(a\sigma)^{k+\nu}} \frac{1}{2} \int_0^\infty \left(\frac{c}{\mu}a\sigma\lambda\right)^{k+\nu-1} \exp\left[-\frac{1}{2}\left(a^2\frac{c\sigma}{\mu}\lambda + \frac{1}{\frac{c\sigma}{\mu}\lambda}\right)\right] d\left(\frac{c}{\mu}a\sigma\lambda\right) =$$

$$\frac{\left(\frac{\mu}{c}\right)^k}{k!K_\nu\left(\frac{1}{\sigma}\right)(a\sigma)^{k+\nu}} \frac{1}{2} \int_0^\infty \left(\frac{c}{\mu}a\sigma\lambda\right)^{k+\nu-1} \exp\left[-\frac{1}{2}a\left(\frac{c}{\mu}a\sigma\lambda + \frac{1}{\frac{c}{\mu}a\sigma\lambda}\right)\right] d\left(\frac{c}{\mu}a\sigma\lambda\right).$$

The integrand of the above expression is equal to $2K_{k+\nu}(a)$, where $K_{k+\nu}(a)$ is the modified Bessel function of the third kind of order $k+\nu$ with argument a.

Thus we have

$$P\left(k\right)=\frac{\left(\frac{\mu}{c}\right)^{k}K_{k+\nu}(a)}{k!(a\sigma)^{k+\nu}K_{\nu}\left(\frac{1}{\sigma}\right)}.$$

Proposition 3 *The special case $\nu=-0.5$ in (3.8) gives the parameterization of the Poisson-Inverse Gaussian (PIG) distribution used by Dean et al. (1989). The probability density function of the PIG is given by*

$$P\left(k\right)=\left(\frac{2a}{\pi}\right)^{\frac{1}{2}}\frac{\mu^{k}e^{\frac{1}{\sigma}}K_{k-\frac{1}{2}}\left(a\right)}{\left(a\sigma\right)^{k}k!},\tag{3.9}$$

for $k=0,1,2,3,...,$ where $a^{2}=\sigma^{-2}+2\frac{\mu}{\sigma}$.

The PIG(μ,σ) can arise if we assume that the mixing distribution of the Poisson rate λ is an Inverse Gaussian distribution with probability density function given by Eq. (3.7). Note that the mean and the variance of the PIG distribution are given by $E(k)=\mu$ and $Var(k)=\mu+\mu^{2}\sigma$ respectively. Note also that the Poisson-gamma, i.e. Negative Binomial Type I is a limiting case of (3.8) obtained by letting $\sigma\rightarrow\infty$ for $\nu>0$.

Proof. If we let $\nu=-0.5$ in (3.8), then we have that

$$P\left(k\right)=\frac{\left(\frac{\mu}{c}\right)^{k}K_{k-\frac{1}{2}}(a)}{K_{-\frac{1}{2}}\left(\frac{1}{\sigma}\right)k!(a\sigma)^{k-\frac{1}{2}}},$$

for $k=0,1,2,3,...,$ where $c=\frac{K_{1/2}\left(\frac{1}{\sigma}\right)}{K_{-1/2}\left(\frac{1}{\sigma}\right)}$ and where $a^{2}=\sigma^{-2}+2\frac{\mu}{\sigma}$. As we mentioned in the proof of (3.7), using the relationships between the modified Bessel functions of different order one can find that

$$c=1\text{ and }K_{-\frac{1}{2}}\left(\frac{1}{\sigma}\right)=\left(\frac{\pi\sigma}{2}\right)^{\frac{1}{2}}e^{-\frac{1}{\sigma}}.$$

and hence

$$P\left(k\right)=\frac{\mu^{k}K_{k-\frac{1}{2}}(a)}{\left(\frac{\pi\sigma}{2}\right)^{\frac{1}{2}}e^{-\frac{1}{\sigma}}k!(a\sigma)^{k-\frac{1}{2}}}=\frac{\left(\frac{2}{\pi\sigma}\right)^{\frac{1}{2}}\mu^{k}e^{\frac{1}{\sigma}}K_{k-\frac{1}{2}}(a)}{k!(a\sigma)^{k-\frac{1}{2}}}$$

$$=\left(\frac{2a}{\pi}\right)^{\frac{1}{2}}\frac{\mu^{k}e^{\frac{1}{\sigma}}K_{k-\frac{1}{2}}(a)}{(a\sigma)^{k}k!}.$$

Another proof of Eq. (3.9) can be obtained in a similar way to the proof of Eq. (3.8) if we let $\nu=-0.5$ and solve the convolution integral

$$P\left(k\right)=\int_{0}^{\infty}P\left(k|\lambda\right)u\left(\lambda\right)d\lambda,$$

where $k|\lambda$ follows a Poisson (λ) and λ follows the Inverse Gaussian with pdf given by (3.7).

Posterior Structure Function

In what follows we present the design of an optimal BMS using the Sichel distribution for assessing claim frequency. We also present the system provided by the PIG distribution. Consider a policyholder observed during t years and denote by k_j the number of accidents in which they were at fault in year $j = 1, ..., t$, so their claim frequency history will be in a form of a vector $(k_1, ..., k_t)$. Let us denote by $K = \sum_{j=1}^{t} k_j$ the total number of claims that this insured had in t years. Our goal is to calculate the posterior structure function of λ for a policyholder or a group of policyholders with claim history $k_1, ..., k_t$ for the case of the Sichel and PIG models respectively.

Proposition 4 *Let $k_j|\lambda$, for $j = 1, ..., t$, be distributed according to a Poisson (λ) and let the prior structure function of the parameter λ be the GIG(μ, σ, ν) given by Eq. (3.6). The posterior structure function of λ for a policyholder or a group of policyholders with claim history $k_1, ..., k_t$, denoted as $u(\lambda|k_1, ..., k_t)$, is a GIG $(w_1, w_2, K + \nu)$ distribution with probability density function of the form*

$$u(\lambda|k_1, ..., k_t) = \frac{\left(\frac{w_1}{w_2}\right)^{\frac{K+\nu}{2}} \lambda^{K+\nu-1}}{2K_{K+\nu}\left(\sqrt{w_1 w_2}\right)} \exp\left[-\frac{1}{2}\left[w_1\lambda + w_2\frac{1}{\lambda}\right]\right], \tag{3.10}$$

for $\lambda > 0$, where $w_1 = \frac{c}{\sigma\mu} + 2t$ and $w_2 = \frac{\mu}{\sigma c}$, with $\sigma > 0, -\infty < \nu < \infty$ and $c = \frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_\nu\left[\frac{1}{\sigma}\right]}$ and where $K_{K+\nu}(z)$ is the modified Bessel function of the third kind of order $K + \nu$ with argument z.

Proof. Considering the assumptions of the model, we have

$$P(k_1, ..., k_t|\lambda) = P(k_1|\lambda) \cdot ... \cdot P(k_t|\lambda) = \frac{e^{-\lambda}\lambda^{k_1}}{k_1!} \cdot ... \cdot \frac{e^{-\lambda}\lambda^{k_t}}{k_t!} = \frac{e^{-t\lambda}\lambda^K}{\prod_{j=1}^{t} k_j!}.$$

By Bayes theorem and Eq. (3.6),

$$u(\lambda|k_1, ..., k_t) = \frac{P(k_1,...,k_t|\lambda)u(\lambda)}{P(k_1,...,k_t)} = \frac{P(k_1,...,k_t|\lambda)u(\lambda)}{\int_0^\infty P(k_1,...,k_t|\lambda)u(\lambda)d\lambda}$$

$$= \frac{\frac{e^{-t\lambda}\lambda^K}{\prod_{j=1}^{t} k_j!} \frac{\left(\frac{c}{\mu}\right)^\nu \lambda^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2K_\nu\left[\frac{1}{\sigma}\right]}}{\int_0^\infty \frac{e^{-t\lambda}\lambda^K}{\prod_{j=1}^{t} k_j!} \frac{\left(\frac{c}{\mu}\right)^\nu \lambda^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\mu}\lambda + \frac{\mu}{c}\frac{1}{\lambda}\right)\right]}{2K_\nu\left[\frac{1}{\sigma}\right]} d\lambda}$$

$$= \frac{\lambda^{K+\nu-1} e^{-t\lambda} \exp\left[-\frac{1}{2}\left(\frac{c}{\sigma\mu}\lambda+\frac{\mu}{c\sigma}\frac{1}{\lambda}\right)\right]}{\int\limits_0^\infty \lambda^{K+\nu-1} e^{-t\lambda} \exp\left[-\frac{1}{2}\left(\frac{c}{\sigma\mu}\lambda+\frac{\mu}{\sigma c}\frac{1}{\lambda}\right)\right] d\lambda}$$

$$= \frac{\lambda^{K+\nu-1} \exp\left[-\frac{1}{2}\left(\left(\frac{c}{\sigma\mu}+2t\right)\lambda+\frac{\mu}{c\sigma}\frac{1}{\lambda}\right)\right]}{\int\limits_0^\infty \lambda^{K+\nu-1} \exp\left[-\frac{1}{2}\left(\left(\frac{c}{\sigma\mu}+2t\right)\lambda+\frac{\mu}{c\sigma}\frac{1}{\lambda}\right)\right] d\lambda}.$$

If we let $w_1 = \left(\frac{c}{\sigma\mu} + 2t\right)$ and $w_2 = \frac{\mu}{\sigma c}$ then we have

$$u\left(\lambda|k_1, ..., k_t\right) = \frac{\lambda^{K+\nu-1} \exp\left[-\frac{1}{2}\left(w_1\lambda+w_2\frac{1}{\lambda}\right)\right]}{\int\limits_0^\infty \lambda^{K+\nu-1} \exp\left[-\frac{1}{2}\left(w_1\lambda+w_2\frac{1}{\lambda}\right)\right] d\lambda}$$

$$= \frac{\left(\frac{w_1}{w_2}\right)^{\frac{K+\nu}{2}} \lambda^{K+\nu-1} \exp\left[-\frac{1}{2}\left(w_1\lambda+w_2\frac{1}{\lambda}\right)\right]}{\int\limits_0^\infty \left(\frac{\sqrt{w_1}}{\sqrt{w_2}}\right)^{K+\nu} \lambda^{K+\nu-1} \exp\left[-\frac{1}{2}\sqrt{w_1 w_2}\left(\frac{\sqrt{w_1}}{\sqrt{w_2}}\lambda+\frac{\sqrt{w_1}}{\sqrt{w_2}}\frac{1}{\lambda}\right)\right] d\lambda}$$

$$= \frac{\left(\frac{w_1}{w_2}\right)^{\frac{K+\nu}{2}} \lambda^{K+\nu-1} \exp\left[-\frac{1}{2}\left(w_1\lambda+w_2\frac{1}{\lambda}\right)\right]}{\int\limits_0^\infty \left(\frac{\sqrt{w_1}}{\sqrt{w_2}}\lambda\right)^{K+\nu-1} \exp\left[-\frac{1}{2}\sqrt{w_1 w_2}\left(\frac{\sqrt{w_1}}{\sqrt{w_2}}\lambda+\frac{1}{\frac{\sqrt{w_1}}{\sqrt{w_2}}\lambda}\right)\right] d\left(\frac{\sqrt{w_1}}{\sqrt{w_2}}\lambda\right)}.$$

The integrand of the above expression is equal to $2K_{K+\nu}\left(\sqrt{w_1 w_2}\right)$, i.e. the modified Bessel function of the third kind of order $K+\nu$ with argument $\sqrt{w_1 w_2}$.

Thus we have that

$$u\left(\lambda|k_1, ..., k_t\right) = \frac{\left(\frac{w_1}{w_2}\right)^{\frac{K+\nu}{2}} \lambda^{K+\nu-1}}{2K_{K+\nu}(\sqrt{w_1 w_2})} \exp\left[-\frac{1}{2}\left[w_1\lambda + w_2\frac{1}{\lambda}\right]\right].$$

Proposition 5 *In the special case when $\nu = -0.5$, i.e. when the simple Inverse Gaussian, given by Eq. (3.7), is the structure function of λ, the posterior structure function of λ for a policyholder or a group of policyholders with claim history $k_1, ..., k_t$ is a GIG $\left(h_1, h_2, K - \frac{1}{2}\right)$ with pdf given by*

$$u\left(\lambda|k_1, ..., k_t\right) = \frac{\left(\frac{h_1}{h_2}\right)^{\frac{K-\frac{1}{2}}{2}} \lambda^{K-\frac{3}{2}}}{2K_{K-\frac{1}{2}}\left(\sqrt{h_1 h_2}\right)} \exp\left[-\frac{1}{2}\left[h_1\lambda + h_2\frac{1}{\lambda}\right]\right], \tag{3.11}$$

for $\lambda > 0, \sigma > 0$, *where* $h_1 = \frac{1}{\sigma\mu} + 2t$ *and* $h_2 = \frac{\mu}{\sigma}$ *and where* $K_{K-\frac{1}{2}}(z)$ *is the modified Bessel function of the third kind of order* $K - \frac{1}{2}$ *with argument* z.

Proof. If let $\nu = -0.5$ in (3.10), then the proof can be obtained in a similar way to that of Proposition 1.

Optimal Choice of $\hat{\lambda}_i^{t+1}$

Subsequently, by using the quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}$ for a policyholder with claim history $k_1, ..., k_t$ is the mean of the GIG $(w_1, w_2, K+\nu)$, i.e. the posterior structure function given by Eq. (3.10), that is

$$\begin{aligned} \hat{\lambda}_{t+1}(k_1, ..., k_t) &= \int_0^\infty \lambda u(\lambda | k_1, ..., k_t)\, d\lambda \\ &= \left(\sqrt{\frac{w_2}{w_1}}\right) \frac{K_{K+\nu+1}(w_1 w_2)}{K_{K+\nu}(w_1 w_2)}, \end{aligned} \tag{3.12}$$

where $w_1 = \frac{c}{\sigma\mu} + 2t$ and $w_2 = \frac{\mu}{\sigma c}$, with $\sigma > 0, -\infty < \nu < \infty$ and $c = \frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_\nu\left[\frac{1}{\sigma}\right]}$ and where $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z.

In the special case when $\nu = -0.5$, using again the quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}$ for a policyholder with claim history $k_1, ..., k_t$ is the mean of the GIG $\left(h_1, h_2, K - \frac{1}{2}\right)$, i.e. the posterior structure function given by Eq. (3.11), that is

$$\begin{aligned} \hat{\lambda}_{t+1}(k_1, ..., k_t) &= \int_0^\infty \lambda u(\lambda | k_1, ..., k_t)\, d\lambda \\ &= \left(\sqrt{\frac{h_2}{h_1}}\right) \frac{K_{K+\frac{1}{2}}(h_1 h_2)}{K_{K-\frac{1}{2}}(h_1 h_2)}, \end{aligned} \tag{3.13}$$

where $h_1 = \frac{1}{\sigma\mu} + 2t$ and $h_2 = \frac{\mu}{\sigma}$, where $\sigma > 0$ and where $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z.

3.2.3 Calculation of the Premiums According to the Net Premium Principle

We calculate the premium based on the net premium principle for the set of distributions considered in the previous sections. Consider a policyholder or a group of policyholders who in t years have produced K claims. The net premium that should be paid from that specific group of policyholders is equal to their expected number of claims for period $t+1$, $\lambda_{t+1}(k_1, ..., k_t)$,i.e. is equal to

$$Premium = e\hat{\lambda}_{t+1}(k_1, ..., k_t), \tag{3.14}$$

where $e = \frac{1}{3.5}$ denotes the exposure to risk[2], since, as mentioned in Chapter 2, all 15641 policyholders were observed for 3.5 years and where $\hat{\lambda}_{t+1}(k_1, ..., k_t)$ in (3.14) is given by the Eqs (3.5, 3.13 and 3.12) for the case of Negative Binomial, Poisson-Inverse Gaussian and Sichel distributions respectively.

In order to find the premium that must be paid we have to know:

1. the maximum likelihood estimates of the parameters α and τ of the Negative Binomial distribution with pdf given by Eq. (3.3),
2. the maximum likelihood estimates of the parameters μ and σ of the Poisson-Inverse Gaussian distribution with pdf given by Eq. (3.9),
3. the maximum likelihood estimates of the parameters μ, σ and ν of the Sichel distribution with pdf given by Eq. (3.8),
4. the number of years t that the policyholder is under observation,
5. their total number of claims $K = \sum_{j=1}^{t} k_j$, where k_j the number of accidents in which they were at fault in year $j = 1, ..., t$.

3.2.4 Properties of the Optimal BMS Based on the a Posteriori Criteria

1. The system is fair, in a Bayesian sense. Every insured has to pay a premium proportional to the estimate of their claim frequency taking into account, through the Bayes theorem, all the information available gathered in the past.
2. The system is financially balanced. Every single year, the average premium per policyholder remains constant at the initial level thus

$$P = e\frac{\alpha}{\tau}, \tag{3.15}$$

for the case of the Negative Binomial distribution and

$$P = e\mu, \tag{3.16}$$

for the case of the Sichel and the Poisson-Inverse Gaussian distributions respectively. The above are a direct consequence of the property of conditional expectation

$$E_{\Lambda}\left[\Lambda\right] = E\left[E\left[\lambda | k_1, ...k_t\right]\right].$$

[2]Exposure is the proportion of the period of observation for which the policy has been in force.

3. In the beginning all the policyholders are paying the same premium, which is equal to (3.15), when we consider the Negative Binomial distribution, and equal to (3.16), when we consider the PIG or the Sichel distributions.

4. The Sichel distribution has a thicker tail than the Negative Binomial and the Poisson-Inverse Gaussian distributions and offers the advantage of being able to model count data with high dispersion.

3.3 The Design of an Optimal BMS Based Both on the a Priori and the a Posteriori Criteria

In this section we develop a generalized BMS that integrates the a priori and the a posteriori information on an individual basis. For this purpose we consider the generalized additive models for location, scale and shape (GAMLSS) in order to use all available information in the estimation of the claim frequency distribution. As we have seen in Chapter 2, the GAMLSS basically consist of four different formulations: the semi-parametric additive model, the parametric linear model, the non-linear semi-parametric additive model and the non-linear parametric model. Within the framework of the GAMLSS we propose the Sichel GAMLSS for assessing claim frequency as an alternative to the Negative Binomial regression model of Dionne and Vanasse (1989, 1992). Furthermore, we consider the PIG GAMLSS for approximating the number of claims. With the aim of constructing an optimal BMS by updating the posterior mean claim frequency, we adopt the parametric linear formulation of these models and we allow only their mean parameter to be modelled as a function of the significant a priori rating variables for the number of claims. In this generalized BMS, the premium is a function of the years that the policyholder is in the portfolio, the number of accidents and the explanatory variables for the number of accidents.

3.3.1 The Negative Binomial Model

This generalized optimal BMS is developed according to the design of Dionne and Vanasse (1989, 1992) and Frangos and Vrontos (2001). Consider a policyholder i with an experience of t periods whose number of claims for period j, denoted as K_i^j are independent. If we assume that K_i^j follows the Poisson distribution with parameter λ^j, the expected number of claims for period j then the probability of having k accidents is

$$P\left(K_i^j = k\right) = \frac{e^{-\lambda^j}\left(\lambda^j\right)^k}{k!},$$

for $k = 0, 1, 2, 3, ...$ and $\lambda^j > 0$, where $E(K_i^j) = \lambda^j$ and $Var(K_i^j) = \lambda^j$. We can allow the λ^j parameter to vary from one individual to another. Let $\lambda_i^j = \exp\left(c_i^j \beta^j\right)$, where $c_i^j\left(c_{i,1}^j, ..., c_{i,h}^j\right)$ is the $1 \times h$ vector of h individual characteristics[3], which represent different a priori rating variables

[3] All the characteristics we consider are observable.

and β^j is the vector of the coefficients. The exponential form ensures the non-negativity of λ_i^j. The conditional to c_i^j probability that policyholder i will be involved in k accidents during the period j will become

$$P\left(K_i^j = k|c_i^j\right) = \frac{e^{-\exp\left(c_i^j\beta^j\right)}\left[\exp\left(c_i^j\beta^j\right)\right]^k}{k!}, \tag{3.17}$$

for $k = 0, 1, 2, 3, ...$ and $\lambda_i^j > 0$, where $E(K_i^j|c_i^j) = Var(K_i^j|c_i^j) = \lambda_i^j = \exp\left(c_i^j\beta^j\right)$. For the determination of the expected number of claims in this model we assume that the h individual characteristics provide enough information. However, if one assumes that the a priori rating variables do not contain all the significant information for the expected number of claims then a random variable ε_i has to be introduced into the regression component. According to Gourieroux, Montfort and Trognon (1984 a), (1984 b) we can write

$$\lambda_i^j = \exp\left(c_i^j\beta^j + \varepsilon_i\right) = \exp\left(c_i^j\beta^j\right)u_i,$$

where $u_i = \exp\left(\varepsilon_i\right)$, yielding a random λ_i^j. Assume that u_i follows a Gamma distribution with probability density function

$$\upsilon\left(u_i\right) = \frac{u_i^{\frac{1}{\alpha}-1}\frac{1}{\alpha}^{\frac{1}{\alpha}}\exp\left(-\frac{1}{\alpha}u_i\right)}{\Gamma\left(\frac{1}{\alpha}\right)}, \tag{3.18}$$

$u_i > 0, \alpha > 0$, with mean $E(u_i) = 1$ and variance $Var(u_i) = \alpha$. Under this assumption the conditional distribution of $K_i^j|c_i^j$ becomes

$$P\left(K_i^j = k|c_i^j\right) = \binom{k+\frac{1}{\alpha}-1}{k}\frac{\left[\alpha\exp\left(c_i^j\beta^j\right)\right]^k}{\left[1+\alpha\exp\left(c_i^j\beta^j\right)\right]^{k+\frac{1}{\alpha}}}, \tag{3.19}$$

which is a Negative Binomial Type I (NBI) distribution with parameters α and $\exp\left(c_i^j\beta^j\right)$. It can be shown that the above parameterization does not affect the results if there is a constant term in the regression. We choose $E(u_i) = 1$ in order to have $E(\varepsilon_i) = 0$. Note that $E(K_i^j|c_i^j) = \mu_i^j = \exp\left(c_i^j\beta^j\right)$ and $Var(K_i^j|c_i^j) = \exp\left(c_i^j\beta^j\right)\left[1+\alpha\exp\left(c_i^j\beta^j\right)\right]$. More details about the Negative Binomial regression can be found in Lawless (1987) and Hilbe (2011). Note also that Eq. (3.19) gives the parametric linear GAMLSS where only the mean parameter of the NBI response distribution is modelled as a function of the explanatory variables.

Posterior Structure Function

We are going to build an optimal BMS based on the number of past claims and on individual's characteristics in order to adjust that individual's premiums over time. The problem is to determine, at the renewal of the policy, the expected claim frequency of the policyholder i for the period $t+1$ given the observation of the reported accidents in the preceding t periods and observable characteristics in the preceding $t+1$ periods and the current period. Consider a policyholder i with $K_i^1, ..., K_i^t$ claim history and $c_i^1, ..., c_i^{t+1}$ characteristics and denote as

$K = \sum_{j=1}^{t} K_i^j$ the total number of claims that they had. The mean claim frequency of the individual i for period $t+1$ is $\lambda_i^{t+1}\left(c_i^{t+1}, u_i\right)$, a function of both the vector of the individual's characteristics and a random factor u_i with probability density function given by Eq. (3.18). Based on the assumptions of the model one can find that the probability density function of $\lambda_i^{t+1}\left(c_i^{t+1}, u_i\right)$, denoted as $f\left(\lambda_i^{t+1}\right)$, is given by

$$f\left(\lambda_i^{t+1}\right) = \frac{\left(\frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)\alpha}\right)^{\frac{1}{\alpha}} \left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha}-1} e^{-\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)\alpha}}}{\Gamma\left(\frac{1}{\alpha}\right)}, \tag{3.20}$$

for $\lambda_i^{t+1} > 0$ and $\alpha > 0$, which is a Gamma$\left(\alpha, \exp\left(c_i^{t+1}\beta^{t+1}\right)\right)$ distribution. The posterior distribution of the mean claim frequency λ_i^{t+1} for an individual i observed over $t+1$ periods with $K_i^1, ..., K_i^t$ claim history and $c_i^1, ..., c_i^{t+1}$ characteristics is obtained using Bayes theorem and is given by a Gamma with updated parameters $\frac{1}{\alpha} + K$ and S_i^j, with pdf

$$f\left(\lambda_i^{t+1} | K_i^1, ..., K_i^t; c_i^1, ..., c_i^t\right) = \frac{\left(S_i^j\right)^{K+\frac{1}{\alpha}} \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha}-1} \exp\left[-S_i^j \lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha}+K\right)}, \tag{3.21}$$

where $S_i^j = \frac{\frac{1}{\alpha}+\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}$ with $\lambda_i^{t+1} > 0$ and $\alpha > 0$. Let us consider, as a special case, the situation in which the vector of the individual characteristics remains the same from one year to the next, i.e. $c_i^1 = c_i^2 = ... = c_i^{t+1} = c_i$ and $\beta_z^1 = \beta_z^2 = ... = \beta_z^t = \beta$. Then the posterior distribution of the mean claim frequency λ_i^{t+1} is simplified to

$$f\left(\lambda_i^{t+1} | K_i^1, ..., K_i^t; c_i^1, ..., c_i^t\right) = \frac{\left(Z_i^j\right)^{K+\frac{1}{\alpha}} \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha}-1} \exp\left[-Z_i^j \lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha}+K\right)}, \tag{3.22}$$

where $Z_i = \left[\frac{1}{\exp(c_i\beta)\alpha} + t\right]^{K+\frac{1}{\alpha}}$. For more information about the proof of Eqs (3.19 and 3.21) refer to Dionne and Vanasse (1989, 1992). Also, a more general proof can be found in Chapter 5 where we consider the case of the n-component Negative Binomial mixture regression model derived by updating the posterior mean.

Optimal Choice of $\hat{\lambda}_i^{t+1}$

Using the quadratic loss function, in the general case, one can find that the optimal estimator of $\hat{\lambda}_{t+1}$ is the mean of the posterior structure function, given by

$$\begin{aligned}
\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_i^1,...,c_i^{t+1}\right) &= \int_0^\infty \lambda_i^{t+1}(c_i^{t+1},u_i) f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right) d\lambda_i^{t+1} \\
&= \exp\left(c_i^{t+1}\beta^{t+1}\right)\left[\frac{\frac{1}{\alpha}+\sum_{j=1}^{t}K_i^j}{\frac{1}{\alpha}+\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}\right] \qquad (3.23)
\end{aligned}$$

This estimator defines the premium and corresponds to the multiplicative tariff formula where the base premium is the a priori frequency $\exp\left(c_i^{t+1}\beta^{t+1}\right)$ and where the Bonus-Malus factor is represented by the expression in brackets. When the vector of the individual characteristics remains the same from one year to the next $\hat{\lambda}_i^{t+1}$ is simplified to

$$\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_i^1,...,c_i^{t+1}\right) = \exp\left(c_i\beta\right)\left[\frac{\frac{1}{\alpha}+\sum_{j=1}^{t}K_i^j}{\frac{1}{\alpha}+t\exp\left(c_i\beta\right)}\right]$$

When $t = 0, \hat{\lambda}_i^1\left(c_i^1\right) = \exp\left(c_i^1\beta\right)$, which implies that only a priori rating is used in the first period. Moreover, when the regression component is limited to a constant β_0 one obtains

$$\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t\right) = \exp\left(\beta_0\right)\left[\frac{\frac{1}{\alpha}+\sum_{j=1}^{t}K_i^j}{\frac{1}{\alpha}+t\exp\left(\beta_0\right)}\right],$$

which corresponds to the 'univariate',without regression component, model.

3.3.2 The Sichel Model

Let us now consider the generalized BMS obtained by using the Sichel parametric linear GAMLSS for assessing claim frequency. The Sichel GAMLSS can be considered as a candidate model for highly dispersed claim count data when the observed high dispersion cannot be efficiently handled by the Negative Binomial regression model.

Consider a policyholder i with an experience of t periods whose number of claims for period j, denoted as K_i^j, are independent. We assume again that K_i^j follows Poisson distribution with parameter $\lambda_i^j = \exp\left(c_i^j \beta^j\right)$, where $c_i^j \left(c_{i,1}^j, ..., c_{i,h}^j\right)$ is the vector of h individual characteristics and β^j is the vector of the coefficients. The conditional to c_i^j probability that policyholder i will be involved in k accidents during the period j is given by Eq. (3.17). For the determination of the expected number of claims in this model we assume that the h individual characteristics provide enough information. Nevertheless, if one assumes that the a priori rating variables do not contain all the significant information for the expected number of claims then a random variable ε_i has to be introduced into the regression component, and for $u_i = \exp(\varepsilon_i)$ we have

$$\lambda_i^j = \exp\left(c_i^j \beta^j + \varepsilon_i\right) = \exp\left(c_i^j \beta^j\right) u_i,$$

yielding a random λ_i^j. Let u_i have a Generalized Inverse Gaussian distribution GIG$(1, \sigma, \nu)$ with probability density function given by

$$\upsilon(u_i) = \frac{c^\nu u_i^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(c u_i + \frac{1}{c u_i}\right)\right]}{2K_\nu\left(\frac{1}{\sigma}\right)}, \tag{3.24}$$

for $u_i > 0$, where $\sigma > 0$ and $-\infty < \nu < \infty$ and where $c = \frac{K_{\nu+1}\left(\frac{1}{\sigma}\right)}{K_\nu\left(\frac{1}{\sigma}\right)}$, where

$$K_\nu(z) = \frac{1}{2}\int_0^\infty x^{\nu-1} \exp\left[-\frac{1}{2}z\left(x + \frac{1}{x}\right)\right] dx,$$

is the modified Bessel function of the third kind of order ν with argument z. Eq. (3.24) is obtained from Eq. (3.6) if we let $\mu = 1$. Parameterization (3.24) ensures that $E(u_i) = 1$. Note also that $Var(u_i) = \frac{2\sigma(\nu+1)}{c} + \frac{1}{c^2} - 1$.

Proposition 6 *If we let $\nu = -0.5$ in (3.24) the Generalized Inverse Gaussian distribution can be reduced to an Inverse Gaussian distribution with pdf given by*

$$\upsilon(u_i) = \frac{1}{\sqrt{2\pi\sigma u_i^3}} \exp\left[-\frac{1}{2\sigma u_i}(u_i - 1)^2\right], \tag{3.25}$$

for $u_i > 0$, where $\sigma > 0$ and where $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z. Parameterization (3.25) also ensures that $E(u_i) = 1$. Note also that $Var(u_i) = \sigma$.

Proof. The proof of (3.25) follows from the proof of (3.7) if we let $\mu = 1$.

Proposition 7 *Considering the assumptions of the model, i.e. Eq. (3.17) and Eq. (3.24), the conditional distribution of $K_i^j | c_i^j$ will be a Sichel distribution with parameters* $\exp\left(c_i^j \beta^j\right), \sigma, \nu$ *and probability density function of the following form*

$$P\left(K_i^j = k | c_i^j\right) = \frac{\left[\frac{\exp\left(c_i^j \beta^j\right)}{c}\right]^k K_{k+\nu}(a)}{k!\,(a\sigma)^{k+\nu} K_\nu\left(\frac{1}{\sigma}\right)}, \tag{3.26}$$

for $k = 0, 1, 2, 3, \ldots$ *and* $a^2 = \sigma^{-2} + 2\exp\left(c_i^j \beta^j\right)(c\sigma)^{-1}$.

The above parameterization of the Sichel distribution ensures that the location parameter is the mean of $K_i^j | c_i^j$, given by

$$E(K_i^j | c_i^j) = \mu_i^j = \exp\left(c_i^j \beta^j\right),$$

and the variance of $K_i^j | c_i^j$ is given by

$$Var(K_i^j | c_i^j) = \exp\left(c_i^j \beta^j\right) + \left(\exp\left(c_i^j \beta^j\right)\right)^2 \left[\frac{2\sigma(\nu+1)}{c} + \frac{1}{c^2} - 1\right].$$

Thus Eq. (3.26) gives the parametric linear GAMLSS where only the mean parameter of the Sichel response distribution is modelled as a function of the significant a priori rating variables for the number of claims.

Proof. The conditional distribution of $K_i^j | c_i^j$ can be derived by solving the convolution integral

$$P\left(K_i^j = k | c_i^j\right) = \int_0^\infty \frac{e^{-\exp\left(c_i^j \beta^j\right) u_i} \left[\exp\left(c_i^j \beta^j\right) u_i\right]^k}{k!} u(u_i)\, du_i$$

$$= \int_0^\infty \frac{e^{-\exp\left(c_i^j \beta^j\right) u_i} \left[\exp\left(c_i^j \beta^j\right) u_i\right]^k}{k!} \frac{c^\nu u_i^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(c u_i + \frac{1}{c u_i}\right)\right]}{2 K_\nu\left(\frac{1}{\sigma}\right)} du_i$$

$$= \frac{\left[\frac{\exp\left(c_i^j \beta^j\right)}{c}\right]^k}{k! K_\nu\left(\frac{1}{\sigma}\right)} \frac{1}{2} \int_0^\infty \left[\frac{c}{\exp\left(c_i^j \beta^j\right)}\right]^{k+\nu} u_i^{k+\nu-1} \exp\left[-\frac{1}{2}\left(\left(\frac{c}{\exp\left(c_i^j \beta^j\right)\sigma} + 2\right) u_i + \frac{\exp\left(c_i^j \beta^j\right)}{c\sigma} \frac{1}{u_i}\right)\right] du_i.$$

If we let $a = \sqrt{\left(\frac{c}{\exp\left(c_i^j \beta^j\right)\sigma} + 2\right) \frac{\exp\left(c_i^j \beta^j\right)}{c\sigma}} = \sqrt{\sigma^{-2} + 2\exp\left(c_i^j \beta^j\right)(c\sigma)^{-1}}$, then we have

$$P(k) = \frac{\left(\frac{\exp\left(c_i^j \beta^j\right)}{c}\right)^k}{k! K_\nu\left(\frac{1}{\sigma}\right)(a\sigma)^{k+\nu}} \frac{1}{2} \int_0^\infty \left(\frac{c}{\exp\left(c_i^j \beta^j\right)}\right)^{k+\nu} (a\sigma)^{k+\nu} u_i^{k+\nu-1} \cdot$$

$$\cdot \exp\left[-\frac{1}{2}\left(\left(\frac{c}{\exp\left(c_i^j \beta^j\right)\sigma} + 2\right) u_i + \frac{\exp\left(c_i^j \beta^j\right)}{c\sigma} \frac{1}{u_i}\right)\right] du_i$$

$$= \frac{\left(\frac{\exp\left(c_i^j \beta^j\right)}{c}\right)^k}{k! K_\nu\left(\frac{1}{\sigma}\right)(a\sigma)^{k+\nu}} \frac{1}{2} \int_0^\infty \left(\frac{c}{\exp\left(c_i^j \beta^j\right)} a\sigma u_i\right)^{k+\nu-1} \cdot$$

$$\cdot \exp\left[-\frac{1}{2}\left(a^2\frac{c\sigma}{\exp\left(c_i^j\beta^j\right)}u_i+\frac{1}{\frac{c\sigma}{\exp\left(c_i^j\beta^j\right)}u_i}\right)\right]d\left(\frac{c}{\exp\left(c_i^j\beta^j\right)}a\sigma u_i\right)$$

$$=\frac{\left(\frac{\exp\left(c_i^j\beta^j\right)}{c}\right)^k}{k!K_\nu\left(\frac{1}{\sigma}\right)(a\sigma)^{k+\nu}}\frac{1}{2}\int_0^\infty\left(\frac{c}{\exp\left(c_i^j\beta^j\right)}a\sigma u_i\right)^{k+\nu-1}.$$

$$\cdot \exp\left[-\frac{1}{2}a\left(\frac{c}{\exp\left(c_i^j\beta^j\right)}a\sigma u_i+\frac{1}{\frac{c}{\exp\left(c_i^j\beta^j\right)}a\sigma u_i}\right)\right]d\left(\frac{c}{\exp\left(c_i^j\beta^j\right)}a\sigma u_i\right).$$

The integrand of the above expression is equal to $2K_{k+\nu}(a)$, where $K_{k+\nu}(a)$ is the modified Bessel function of the third kind of order $k+\nu$ with argument a.

Thus we have

$$P(k)=\frac{\left[\frac{\exp\left(c_i^j\beta^j\right)}{c}\right]^k K_{k+\nu}(a)}{k!(a\sigma)^{k+\nu}K_\nu\left(\frac{1}{\sigma}\right)}.$$

Proposition 8 *If we let $\nu=-0.5$ in (3.26) the Sichel distribution reduces to a Poisson-Inverse Gaussian (PIG) distribution with probability density function given by*

$$P\left(K_i^j=k|c_i^j\right)=\left(\frac{2a}{\pi}\right)^{\frac{1}{2}}\frac{\left[\exp\left(c_i^j\beta^j\right)\right]^k e^{\frac{1}{\sigma}}K_{k-\frac{1}{2}}(a)}{(a\sigma)^k\, k!},\tag{3.27}$$

for $k=0,1,2,3,...,$ where $a^2=\sigma^{-2}+2\frac{\exp\left(c_i^j\beta^j\right)}{\sigma}$.

The mean and the variance of the PIG distribution are given by $E(k)=\exp\left(c_i^j\beta^j\right)$ and $Var(k)=\exp\left(c_i^j\beta^j\right)+\left[\exp\left(c_i^j\beta^j\right)\right]^2\sigma$ respectively. Thus Eq. (3.27) gives the parametric linear GAMLSS where only the mean parameter of the PIG response distribution is modelled as a function of the significant a priori rating variables for the number of claims. Note also that the Poisson-Inverse Gaussian distribution can arise if we let u_i have an Inverse Gaussian distribution with probability density function given by Eq. (3.25).

Proof. The proof of Eq. (3.27) follows from the proof of Eq. (3.9) if we let $\mu=\exp\left(c_i^j\beta^j\right)$.

Posterior Structure Function

Our goal is to build an optimal BMS which integrates a priori and a posteriori information on an individual basis, using the Sichel parametric linear GAMLSS for assessing claim frequency. We will also consider the system provided by the PIG parametric linear GAMLSS. These optimal BMSs are based on the number of past claims and on individual's characteristics in order to adjust that individual's premiums over time. Similarly to the case of the Negative Binomial model, the problem is to determine at the renewal of the policy the expected claim frequency of the policyholder i for the period $t+1$ given the observation of the reported accidents in the preceding t periods and observable characteristics in the preceding $t+1$ periods and the current period. Consider a policyholder i with claim history $K_i^1, ..., K_i^t$ and $c_i^1, ..., c_i^{t+1}$ characteristics and denote by $K = \sum_{j=1}^{t} K_i^j$ the total number of claims that they had. The mean claim frequency of the individual i for period $t+1$ is $\lambda_i^{t+1}\left(c_i^{t+1}, u_i\right)$, a function of both the vector of individual's characteristics and a random factor u_i with pdf given by Eq. (3.24).

Using the following theorem, we are going to calculate first $f(\lambda_i^{t+1})$, which represents the pdf of the mean claim frequency of the individual i for period $t+1$ called the structure function, and then applying the Bayes theorem we will find $f\left(\lambda_i^{t+1} | K_i^1, ..., K_i^t; c_i^1, ..., c_i^t\right)$, i.e. the posterior distribution of λ_i^{t+1} given the observation of $K_i^1, ..., K_i^t$ and $c_i^1, ..., c_i^t$ for the case of the Sichel and PIG GAMLSS respectively.

Theorem 9 *Let X have probability density function $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Let X and Y be defined by*

$$X = \{x : f_X(x) > 0\} \text{ and } Y = \{y : y = g(x),\ x \in X\}.$$

Suppose that $f_X(x)$ is continuous on X and that $g^{-1}(y)$ has a continuous derivative on Y. Then the probability density function of Y is given by

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|, y \in Y.$$

Lemma 10 *Based on the above theorem one can find that $\lambda_i^{t+1}\left(c_i^{t+1}, u_i\right)$ is distributed according to a GIG$\left(\exp\left(c_i^{t+1}\beta^{t+1}\right), \sigma, \nu\right)$ with probability density function given by*

$$f\left(\lambda_i^{t+1}\right) = \frac{\left(\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1} + \frac{1}{\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1}}\right)\right]}{2K_\nu\left(\frac{1}{\sigma}\right)}, \tag{3.28}$$

for $\lambda_i^{t+1} > 0, \sigma > 0$ and $-\infty < \nu < \infty$, where $c = \frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_\nu\left[\frac{1}{\sigma}\right]}$ and where $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z.

Proof. If we let $g(u_i) = \exp\left(c_i^{t+1}\beta^{t+1}\right)u_i$, then g is a strictly increasing function. Also, as we have already mentioned, u_i follows a continuous Generalized Inverse Gaussian distribution with pdf given by Eq.(3.24).

Note that here the support sets:

$$X = \{u_i : u(u_i) > 0\} \text{ and } Y = \left\{\lambda_i^{t+1} : \lambda_i^{t+1} = g(u_i),\ x \in X\right\}$$

are both the interval $(0, \infty)$. From Eq.(3.24) we can easily see that the pdf of u_i is continuous on X. If we let $\lambda_i^{t+1} = g(u_i)$, then:

$$g^{-1}\left(\lambda_i^{t+1}\right) = \frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)} \text{ and } \frac{d}{d\lambda_i^{t+1}} g^{-1}\left(\lambda_i^{t+1}\right) = \frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}$$

and g^{-1} is continuous on Y. Applying the above theorem, for $\lambda_i^{t+1} \in (0,\infty)$, we get:

$$f\left(\lambda_i^{t+1}\right) = \upsilon\left(g^{-1}\left(\lambda_i^{t+1}\right)\right)\left|\frac{d}{d\lambda_i^{t+1}} g^{-1}\left(\lambda_i^{t+1}\right)\right| =$$

$$= \frac{c^{\nu}\left(\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(c\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)} + \frac{1}{c\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}}\right)\right]}{2K_{\nu}\left(\frac{1}{\sigma}\right)} \frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}$$

$$= \frac{\left(\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1} + \frac{1}{\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1}}\right)\right]}{2K_{\nu}\left(\frac{1}{\sigma}\right)}.$$

Proposition 11 *For $\nu = -0.5$ in (3.28) the Generalized Inverse Gaussian distribution can be reduced to an Inverse Gaussian distribution with pdf given by*

$$f\left(\lambda_i^{t+1}\right) = \frac{\sqrt{\exp\left(c_i^{t+1}\beta^{t+1}\right)}}{\sqrt{2\pi\sigma\left(\lambda_i^{t+1}\right)^3}} \exp\left[-\frac{1}{2\sigma\exp\left(c_i^{t+1}\beta^{t+1}\right)\lambda_i^{t+1}}\left(\lambda_i^{t+1} - \exp\left(c_i^{t+1}\beta^{t+1}\right)\right)^2\right], \quad (3.29)$$

for $\lambda_i^{t+1} > 0$ and $\sigma > 0$.

Proof. If we let $\nu = -0.5$ in (3.28) we have that

$$f\left(\lambda_i^{t+1}\right) = \frac{\left(\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)^{-\frac{1}{2}}\left(\lambda_i^{t+1}\right)^{-\frac{3}{2}} \exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1} + \frac{1}{\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1}}\right)\right]}{2K_{-\frac{1}{2}}\left(\frac{1}{\sigma}\right)}.$$

Making use of the relationships between the modified Bessel functions of different orders one can obtain that

$$c = \frac{K_{1/2}\left(\frac{1}{\sigma}\right)}{K_{-1/2}\left(\frac{1}{\sigma}\right)} = 1 \text{ and } K_{-\frac{1}{2}}\left(\frac{1}{\sigma}\right) = \left(\frac{\pi\sigma}{2}\right)^{\frac{1}{2}} e^{-\frac{1}{\sigma}}$$

and hence

$$f\left(\lambda_i^{t+1}\right) = \frac{\left(\frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)^{-\frac{1}{2}}\left(\lambda_i^{t+1}\right)^{-\frac{3}{2}}\exp\left[-\frac{1}{2\sigma}\left(\frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1}+\frac{1}{c}\frac{1}{\lambda_i^{t+1}}\right)\right]}{2\left(\frac{\pi\sigma}{2}\right)^{\frac{1}{2}}e^{-\frac{1}{\sigma}}}$$

$$= \frac{\sqrt{\exp\left(c_i^{t+1}\beta^{t+1}\right)}}{\sqrt{2\pi\sigma\left(\lambda_i^{t+1}\right)^3}}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\lambda_i^{t+1}+\frac{\exp\left(c_i^{t+1}\beta^{t+1}\right)}{c}\frac{1}{\lambda_i^{t+1}}+2\right)\right]$$

$$= \frac{\sqrt{\exp\left(c_i^{t+1}\beta^{t+1}\right)}}{\sqrt{2\pi\sigma\left(\lambda_i^{t+1}\right)^3}}\exp\left[-\frac{1}{2\sigma\exp\left(c_i^{t+1}\beta^{t+1}\right)\lambda_i^{t+1}}\left(\lambda_i^{t+1}-\exp\left(c_i^{t+1}\beta^{t+1}\right)\right)^2\right].$$

Proposition 12 *Considering the assumptions of the model, i.e. Eq. (3.17) and Eq. (3.28),the posterior distribution of the mean claim frequency λ_i^{t+1} for an individual i observed over $t+1$ periods with $K_i^1, ..., K_i^t$ claim history and $c_i^1, ..., c_i^{t+1}$ characteristics is given by a GIG $(w_1, w_2, K+\nu)$ distribution with probability density function of the following form*

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^{t+1}\right) = \frac{\left(\frac{w_1}{w_2}\right)^{\frac{K+\nu}{2}}\left(\lambda_i^{t+1}\right)^{K+\nu-1}}{2K_{K+\nu}\left(\sqrt{w_1w_2}\right)}\exp\left[-\frac{1}{2}\left(w_1\lambda_i^{t+1}+w_2\frac{1}{\lambda_i^{t+1}}\right)\right], \tag{3.30}$$

for $\lambda_i^{t+1}>0$, where $w_1 = \frac{c+2\sigma\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}{\sigma\exp\left(c_i^{t+1}\beta^{t+1}\right)}$ and $w_2 = \frac{\exp\left(c_i^{t+1}\beta^{t+1}\right)}{\sigma c}$ with $\sigma>0$ and $-\infty<\nu<\infty$, where $c = \frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_\nu\left[\frac{1}{\sigma}\right]}$ and where $K_{K+\nu}(z)$ is the modified Bessel function of the third kind of order $K+\nu$ with argument z. When the vector of the individual characteristics remains the same from one year to the next, i.e. $c_i^1 = c_i^2 = ... = c_i^{t+1} = c_i$ and $\beta^1=\beta^2=...=\beta^t=\beta$, Eq. (3.30) is simplified to

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^{t}\right) = \frac{\left(\frac{\omega_1}{\omega_2}\right)^{\frac{K+\nu}{2}}\left(\lambda_i^{t+1}\right)^{K+\nu-1}}{2K_{K+\nu}\left(\sqrt{\omega_1\omega_2}\right)}\exp\left[-\frac{1}{2}\left(\omega_1\lambda_i^{t+1}+\omega_2\frac{1}{\lambda_i^{t+1}}\right)\right], \tag{3.31}$$

where $\omega_1 = \frac{c}{\sigma\exp(c_i\beta)}+2t$ and where $\omega_2 = \frac{\exp(c_i\beta)}{\sigma c}$.

Proof. In what follows we provide the proof of Eq. (3.30).
By Bayes rule

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right) = \frac{P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)}{\bar{P}\left(K_i^1,...,K_i^t|c_i^1,...,c_i^t\right)} \tag{3.32}$$

and where by definition

$$\bar{P}\left(\left(K_i^1,...,K_i^t\right)|c_i^1,...,c_i^t\right)=\int_0^\infty P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)d\lambda_i^{t+1}. \tag{3.33}$$

From (3.32) and (3.33) we have that

$$\begin{aligned} & f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right) \\ = & \frac{P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)}{\int_0^\infty P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)d\lambda_i^{t+1}}. \end{aligned} \tag{3.34}$$

The probability of the sequence $K_i^1,...,K_i^t$ given the frequency of accidents at $t+1$ and the individual's characteristics over the t periods $c_i^1,...,c_i^t$, will be a t-dimension Poisson distribution:

$$P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)=\frac{e^{-\sum_{j=1}^{t}\lambda_i^j}\prod_{j=1}^{t}\left(\lambda_i^j\right)^{K_i^j}}{\prod_{j=1}^{t}K_i^j!}. \tag{3.35}$$

If we let $\lambda_i^j=\exp\left(c_i^j\beta^j\right)u_i\equiv\dot{\lambda}_i^j u_i$, then from (3.28) we get:

$$f\left(\lambda_i^{t+1}\right)=\frac{\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{2K_\nu\left(\frac{1}{\sigma}\right)}, \tag{3.36}$$

for $u_i>0,\sigma>0$ and $-\infty<\nu<\infty$ and where $c=\frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_\nu\left[\frac{1}{\sigma}\right]}$ and $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z.

By substituting (3.35) and (3.36) into (3.34), we get:

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)=$$

$$
= \frac{\dfrac{e^{-\sum_{j=1}^{t}\lambda_i^j}\prod_{j=1}^{t}\left(\lambda_i^j\right)^{K_i^j}}{\prod_{j=1}^{t}K_i^j!}\;\dfrac{\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{2K_{\nu}\left(\frac{1}{\sigma}\right)}}{\displaystyle\int_0^{\infty}\dfrac{e^{-\sum_{j=1}^{t}\lambda_i^j}\prod_{j=1}^{t}\left(\lambda_i^j\right)^{K_i^j}}{\prod_{j=1}^{t}K_i^j!}\;\dfrac{\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{2K_{\nu}\left(\frac{1}{\sigma}\right)}d\lambda_i^{t+1}}
$$

$$
= \frac{e^{-\sum_{j=1}^{t}\lambda_i^j}\prod_{j=1}^{t}\left(\lambda_i^j\right)^{K_i^j}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{\displaystyle\int_0^{\infty}e^{-\sum_{j=1}^{t}\lambda_i^j}\prod_{j=1}^{t}\left(\lambda_i^j\right)^{K_i^j}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]d\lambda_i^{t+1}}
$$

[4]

$$
= \frac{e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j u_i}\prod_{j=1}^{t}\dot{\lambda}_i^j u_i^K\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{\displaystyle\int_0^{\infty}e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j u_i}\prod_{j=1}^{t}\dot{\lambda}_i^j u_i^K\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]d\lambda_i^{t+1}}
$$

[5]

$$
= \frac{e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}}\left(\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}\right)^{K}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{\displaystyle\int_0^{\infty}e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}}\left(\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}\right)^{K}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^{\nu}\left(\lambda_i^{t+1}\right)^{\nu-1}\exp\left[-\frac{1}{2\sigma}\left(\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]d\lambda_i^{t+1}}
$$

---

[4]If we let $\lambda_i^j \equiv \dot{\lambda}_i^j \cdot u_i$ and $K = \sum_{j=1}^{t} K_i^j$

[5]If we let $\lambda_i^{t+1} \equiv \dot{\lambda}_i^{t+1} \cdot u_i$

$$
= \frac{\left(\frac{1}{\dot{\lambda}_i^{t+1}}\right)^K \left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^\nu \left(\lambda_i^{t+1}\right)^{K+\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1} + \frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{\int_0^\infty \left(\frac{1}{\dot{\lambda}_i^{t+1}}\right)^K \left(\frac{c}{\dot{\lambda}_i^{t+1}}\right)^\nu \left(\lambda_i^{t+1}\right)^{K+\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1} + \frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right] d\lambda_i^{t+1}}
$$

$$
= \frac{\left(\lambda_i^{t+1}\right)^{K+\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1} + \frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right]}{\int_0^\infty \left(\lambda_i^{t+1}\right)^{K+\nu-1} \exp\left[-\frac{1}{2\sigma}\left(\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1} + \frac{1}{\frac{c}{\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}}\right)\right] d\lambda_i^{t+1}}
$$

$$
= \frac{\dfrac{\left(\frac{\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}}{\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}\right)^{\frac{K+\nu}{2}}}{2K_{K+\nu}\left(\sqrt{\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}\right)}\left(\lambda_i^{t+1}\right)^{K+\nu-1} \exp\left[-\frac{1}{2}\left(\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}\lambda_i^{t+1} + \frac{\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}{\lambda_i^{t+1}}\right)\right]}{\int_0^\infty \dfrac{\left(\frac{\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}}{\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}\right)^{\frac{K+\nu}{2}}}{2K_{K+\nu}\left(\sqrt{\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}\right)}\left(\lambda_i^{t+1}\right)^{K+\nu-1} \exp\left[-\frac{1}{2}\left(\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}\lambda_i^{t+1} + \frac{\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}{\lambda_i^{t+1}}\right)\right] d\lambda_i^{t+1}}.
$$

The integrand of the above expression is of the same form as a Generalized Inverse Gaussian with parameters $\frac{c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j}{\sigma\dot{\lambda}_i^{t+1}}$, $\frac{\dot{\lambda}_i^{t+1}}{\sigma c}$ and $K+\nu$, thus we have

$$\int_0^\infty \frac{\left(\frac{\frac{c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j}{\sigma\dot{\lambda}_i^{t+1}}}{\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}\right)^{\frac{K+\nu}{2}}\left(\lambda_i^{t+1}\right)^{K+\nu-1}}{2K_{K+\nu}\left(\sqrt{\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}\right)}\exp\left[-\frac{1}{2}\left(\frac{\left(c+2\sigma\sum_{j=1}^{t}\dot{\lambda}_i^j\right)}{\sigma\dot{\lambda}_i^{t+1}}\lambda_i^{t+1}+\frac{\frac{\dot{\lambda}_i^{t+1}}{\sigma c}}{\lambda_i^{t+1}}\right)\right]d\lambda_i^{t+1}=1.$$

For $\dot{\lambda}_i^j \equiv \exp\left(c_i^j\beta^j\right)$ we get

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)=$$

$$=\frac{\left(\frac{w_1}{w_2}\right)^{\frac{K+\nu}{2}}\left(\lambda_i^{t+1}\right)^{K+\nu-1}}{2K_{K+\nu}(\sqrt{w_1w_2})}\exp\left[-\frac{1}{2}\left(w_1\lambda_i^{t+1}+w_2\frac{1}{\lambda_i^{t+1}}\right)\right],$$

for $\lambda_i^{t+1}>0$, where $w_1=\frac{c+2\sigma\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}{\sigma\exp\left(c_i^{t+1}\beta^{t+1}\right)}$ and $w_2=\frac{\exp\left(c_i^{t+1}\beta^{t+1}\right)}{\sigma c}$.

When the vector of the individual characteristics remains the same from one year to the next, we have that $\exp\left(c_{z,i}^j\beta_z^j\right)\equiv\exp\left(c_{z,i}\beta_z\right)$ and it can be easily verified that w_1 and w_2 are simplified to $\omega_1=\frac{c}{\sigma\exp(c_i\beta)}+2t$ and $\omega_2=\frac{\exp(c_i\beta)}{\sigma c}$ respectively.

Proposition 13 *In the special case when $\nu=-0.5$, i.e. when the simple Inverse Gaussian, given by Eq. (3.29), is the prior structure function of λ_i^{t+1}, the posterior structure function of λ_i^{t+1} for a policyholder with $K_i^1,...,K_i^t$ claim history and $c_i^1,...,c_i^{t+1}$ characteristics is a GIG$\left(h_1,h_2,K-\frac{1}{2}\right)$ with pdf given by*

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^{t+1}\right)=\frac{\left(\frac{h_1}{h_2}\right)^{\frac{K-\frac{1}{2}}{2}}\left(\lambda_i^{t+1}\right)^{K-\frac{3}{2}}}{2K_{K-\frac{1}{2}}\left(\sqrt{h_1h_2}\right)}\exp\left[-\frac{1}{2}\left(h_1\lambda_i^{t+1}+h_2\frac{1}{\lambda_i^{t+1}}\right)\right],\quad(3.37)$$

for $\lambda_i^{t+1}>0$, where $h_1=\frac{1+2\sigma\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}{\sigma\exp\left(c_i^{t+1}\beta^{t+1}\right)}$ and $h_2=\frac{\exp\left(c_i^{t+1}\beta^{t+1}\right)}{\sigma}$, where $\sigma>0$ and where $K_{K-\frac{1}{2}}(z)$ is the modified Bessel function of the third kind of order $K-\frac{1}{2}$ with argument z.

When the vector of the individual characteristics remains the same from one year to the next, i.e. $c_i^1 = c_i^2 = ... = c_i^{t+1} = c_i$ *and* $\beta^1 = \beta^2 = ... = \beta^t = \beta$, *Eq. (3.37) is simplified to*

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^{t+1}\right) = \frac{\left(\frac{\eta_1}{\eta_2}\right)^{\frac{K-\frac{1}{2}}{2}}\left(\lambda_i^{t+1}\right)^{K-\frac{3}{2}}}{2K_{K-\frac{1}{2}}\left(\sqrt{\eta_1\eta_2}\right)}\exp\left[-\frac{1}{2}\left(\eta_1\lambda_i^{t+1}+\eta_2\frac{1}{\lambda_i^{t+1}}\right)\right], \quad (3.38)$$

for $\lambda_i^{t+1} > 0$, *where* $\eta_1 = \frac{1}{\sigma\exp(c_i\beta)} + 2t$ *and* $\eta_2 = \frac{\exp(c_i\beta)}{\sigma}$ *with* $\sigma > 0$.

Proof. As we have already mentioned, if we let $\nu = -0.5$ in (3.30) and (3.31) respectively, then using the relationships between the modified Bessel functions of different order one can obtain that $c = 1$. Thus (3.37) and (3.38) can be easily proved.

Optimal Choice of $\hat{\lambda}_i^{t+1}$

Using the quadratic loss function, in the case of the Sichel model, one can find that the optimal estimator of $\hat{\lambda}_i^{t+1}$ is the mean of the GIG $(w_1, w_2, K+\nu)$, i.e. the posterior structure function given by Eq. (3.30), that is

$$\begin{aligned}
&\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_i^1,...,c_i^{t+1}\right) \\
&= \int_0^\infty \lambda_i^{t+1}(c_i^{t+1},u_i) f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)d\lambda_i^{t+1} \\
&= \left(\sqrt{\frac{w_2}{w_1}}\right)\frac{K_{K+\nu+1}\left(w_1w_2\right)}{K_{K+\nu}\left(w_1w_2\right)}, \quad (3.39)
\end{aligned}$$

for $w_1 = \frac{c+2\sigma\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}{\sigma\exp\left(c_i^{t+1}\beta^{t+1}\right)}$ and $w_2 = \frac{\exp\left(c_i^{t+1}\beta^{t+1}\right)}{\sigma c}$, where $\sigma > 0$ and $-\infty < \nu < \infty$ and where $c = \frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_\nu\left[\frac{1}{\sigma}\right]}$ and $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z. When the vector of the individual characteristics remains the same from one year to the next $\hat{\lambda}_i^{t+1}$, given by Eq. (3.39), is simplified to

$$\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_i^1,...,c_i^{t+1}\right) = \left(\sqrt{\frac{\omega_2}{\omega_1}}\right)\frac{K_{K+\nu+1}\left(\omega_1\omega_2\right)}{K_{K+\nu}\left(\omega_1\omega_2\right)},$$

for $\omega_1 = \frac{c}{\sigma\exp(c_i\beta)} + 2t$ and $\omega_2 = \frac{\exp(c_i\beta)}{\sigma c}$. When $t = 0$, $\hat{\lambda}_i^1\left(c_i^1\right) = \exp\left(c_i^1\beta\right)$, which implies that only a priori rating is used in the first period. Moreover, when the regression component is limited to a constant β_0 one obtains

$$\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t\right) = \left(\sqrt{\frac{\omega_2}{\omega_1}}\right)\frac{K_{K+\nu+1}\left(\omega_1\omega_2\right)}{K_{K+\nu}\left(\omega_1\omega_2\right)},$$

for $\omega_1 = \frac{c}{\sigma \exp(\beta_0)} + 2t$ and $\omega_2 = \frac{\exp(\beta_0)}{\sigma c}$, which corresponds to the 'univariate', without regression component, model.

In the special case when $\nu = -0.5$, i.e. when the Sichel distribution reduces to the PIG distribution, using again the quadratic error loss function the optimal choice of $\hat{\lambda}_{t+1}$ is the mean of the GIG $\left(h_1, h_2, K - \frac{1}{2}\right)$, i.e. the posterior structure function given by Eq. (3.37), that is

$$\begin{aligned}
&\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right) \\
&= \int_0^\infty \lambda_i^{t+1}(c_i^{t+1}, u_i) f\left(\lambda_i^{t+1} | K_i^1, ..., K_i^t; c_i^1, ..., c_i^t\right) d\lambda_i^{t+1} \\
&= \left(\sqrt{\frac{h_2}{h_1}}\right) \frac{K_{K+\nu+1}(h_1 h_2)}{K_{K+\nu}(h_1 h_2)},
\end{aligned} \tag{3.40}$$

for $\lambda_i^{t+1} > 0$, where $h_1 = \frac{1 + 2\sigma \sum_{j=1}^{t} \exp\left(c_i^j \beta^j\right)}{\sigma \exp\left(c_i^{t+1} \beta^{t+1}\right)}$ and $h_2 = \frac{\exp\left(c_i^{t+1} \beta^{t+1}\right)}{\sigma}$, where $\sigma > 0$ and where $K_{K-\frac{1}{2}}(z)$ is the modified Bessel function of the third kind of order $K - \frac{1}{2}$ with argument z. When the vector of the individual characteristics remains the same from one year to the next $\hat{\lambda}_i^{t+1}$ is simplified to

$$\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right) = \left(\sqrt{\frac{\eta_2}{\eta_1}}\right) \frac{K_{K+\nu+1}(\eta_1 \eta_2)}{K_{K+\nu}(\eta_1 \eta_2)},$$

for $\eta_1 = \frac{1}{\sigma \exp(c_i \beta)} + 2t$ and $\eta_2 = \frac{\exp(c_i \beta)}{\sigma}$. When $t = 0$, $\hat{\lambda}_i^1(c_i^1) = \exp(c_i^1 \beta)$, which implies that only a priori rating is used in the first period. Moreover, when the regression component is limited to a constant β_0 one obtains

$$\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t\right) = \left(\sqrt{\frac{\eta_2}{\eta_1}}\right) \frac{K_{K+\nu+1}(\eta_1 \eta_2)}{K_{K+\nu}(\eta_1 \eta_2)},$$

for $\eta_1 = \frac{1}{\sigma \exp(\beta_0)} + 2t$ and $\eta_2 = \frac{\exp(\beta_0)}{\sigma}$, which corresponds to the 'univariate' ,without regression component, model.

3.3.3 Calculation of the Premiums of the Generalized BMS

Now we are able to compute the premiums of the optimal BMS based both on a priori and the a posteriori criteria. Consider a policyholder or a group of policyholders who in t years have produced K claims. The net premium that should be paid from that specific group of policyholders is equal to their expected number of claims for the period $t+1$, i.e. is equal to

$$Premium = e\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right), \tag{3.41}$$

where $e = \frac{1}{3.5}$ is the corresponding risk exposure and where $\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right)$ in (3.41) is given by the Eqs (3.23, 3.40 and 3.39) for the case of the NBI, PIG and Sichel models respectively.

In order to find the premiums that must be paid we have to know:

1. the estimates of the parameter α and the vector β^j of the significant a priori rating variables for the number of claims for the case of the NBI model given by Eq. (3.19),

2. the estimates of the parameter σ and the vector β^j for the case of the PIG model given by Eq. (3.27),

3. the estimates of the parameters σ and ν and the vector β^j for the case of the Sichel model given by Eq. (3.26),

4. the number of years t that the policyholder is under observation,

5. the total number of claims $K = \sum_{j=1}^{t} K_i^j$, where K_i^j is the number of accidents in which policyholder i was at fault in year $j = 1, ..., t$.

3.3.4 Properties of the Optimal BMS Based Both on the a Priori and the a Posteriori Criteria

1. It is fair since it takes into account the number of claims and the significant a priori rating variables for the number of claims.

2. It is financially balanced for the insurer. Each year the average premium will be equal to

$$P = e \exp\left(c_i^{t+1}\beta^{t+1}\right) \tag{3.42}$$

   In order to prove Eq. (3.42) it is sufficient to show that

$$E\left[\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right)\right] = e \exp\left(c_i^{t+1}\beta^{t+1}\right).$$

3. In the beginning, all the policyholders with the same characteristics are paying the same premium, which is equal to (3.42).

4. The premiums vary simultaneously with the variables that affect the distribution of the number of claims.

3.4 Application

We use the same data set we used in Chapter 2, the descriptive statistics of which can be found in Table 2.1 of Chapter 2. As mentioned in that chapter, claim counts are modelled for all 15641 policies. The a priori rating variables we employ are the Bonus-Malus category, the horsepower of the car and gender of the driver. Recall that this Bonus-Malus System has 20 classes and the transition rules are described as follows: Each claim free year is rewarded by one class discount and each claim in given year is penalized by one class. The Bonus-Malus category consists of five categories of neighboring BM classes: C1 = "drivers who belong to BM classes 1 and 2", C2 = "drivers who belong to BM classes 3-5", C3 = "drivers who belong to BM classes 6-9", C4 = "drivers who belong to BM class 10" and C5 = "drivers who belong to BM classes 11-20". The horsepower of the car consists of four categories: C1 = "drivers who had a car with a hp between 0-33", C2 = "drivers who had a car with a hp between 34-66", C3 = "drivers who had a car with a hp between 67-99", C4 ="drivers who had a car with a hp between 100-132". Finally, the gender consist of two categories: M = "male", F = "female". Nevertheless, as we mentioned in Chapter 2, gender has recently been ruled out by the European Court as a rating factor. Firstly, the Negative Binomial, Poisson-Inverse Gaussian (PIG) and Sichel distributions were fitted on the number of claims. Secondly, the NBI, PIG and Sichel GAMLSS were applied to model claim frequency. For the GAMLSS models we selected the parametric linear formulation considering a linear model in the explanatory variables only for the log of their mean parameter in order to derive optimal an optimal BMS by updating the posterior mean. The log link function ensured that the mean number of insurance claims predicted from the fitted models is positive. The distributions and the GAMLSS models were estimated using the GAMLSS package in the software R. The likelihood functions were maximized iteratively using the RS algorithm of Rigby and Stasinopoulos (2005). The ratio of Bessel functions of the third kind whose orders are different was calculated using the HyperbolicDist package in software R. Subsequently, we are able to compute the premiums determined by the optimal BMS based on the a posteriori criteria and the premiums determined by the optimal BMS based both on the a priori and the a posteriori criteria according to the current methodology as presented in Sections 3.2 and 3.3.

3.4.1 Modelling Results

This subsection describes the modelling results of the distributions and the GAMLSS models that have been applied to model claim frequency.

Firstly, the Negative Binomial, Poisson-Inverse Gaussian (PIG) and Sichel distributions have been incorporated into the GAMLSS package in R and the following results were obtained:

- In the case of the Negative Binomial[6] distribution, with pdf given by Eq. (3.3), the

[6] Note that the GAMLSS package allow us to find the maximum likelihood estimators of the parameters of the Negative Binomial Type I distribution (NBI). The pdf of the of the Negative Binomial(τ, α) distribution can be derived from a reparameterization of the pdf of the NBI(μ, σ) distribution if we let $\mu = \frac{\alpha}{\tau}$ and $\sigma = \frac{1}{\alpha}$. Thus $\hat{\tau} = \frac{1}{\hat{\mu}\hat{\sigma}}$ and $\hat{\alpha} = \frac{1}{\hat{\sigma}}$.

maximum likelihood estimators of the parameters are $\hat{\tau} = \frac{1}{\exp(-0.08603)\cdot\exp(-0.72410)} = 2.2482$ and $\hat{\alpha} = \frac{1}{\exp(-0.08603)} = 1.0898$. Note also that $\frac{\hat{\alpha}}{\hat{\tau}} = 0.4847$

- In the case of the Poisson-Inverse Gaussian (PIG) distribution, with pdf given by Eq. (3.9) the maximum likelihood estimators of the parameters are $\hat{\mu} = \exp(-0.72412) = 0.4848$ and $\hat{\sigma} = 0.9890$.

- In the case of the Sichel distribution, with pdf given by Eq. (3.8) the maximum likelihood estimators of the parameters are $\hat{\mu} = \exp(-0.72409) = 0.4848$, $\hat{\sigma} = 0.9905$ and $\hat{\nu} = -1.2440$.

Secondly, the Negative Binomial Type I, PIG and Sichel GAMLSS have been implemented in the GAMLSS package in R. We used the function step.GAIC, within the GAMLSS package, which performs the stepwise model selection using a Generalized Akaike information criterion in order to find the variables that are considered as better predictors. The models presented below are the best fitted models. For each parameter in these models we present the estimated parameter values, the standard error of the parameter estimates, and the t-values for the hypothesis that the associated coefficient is zero together with the p-value of this test based on asymptotic normality.

The results are summarized in Tables 3.1, 3.2 and 3.3 for the case of the NBI, PIG and Sichel models respectively.

Table 3.1: Results of the Fitted Negative Binomial Type I GAMLSS

Variable	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.8366	0.1381	-6.0570	0.0000
Bonus-Malus				
Category 1	0	0	-	-
Category 2	0.6084	0.0356	17.0903	0.0000
Category 3	0.8467	0.0431	19.6328	0.0000
Category 4	-0.9402	0.0755	-12.4524	0.0000
Category 5	1.9670	0.0229	8.5590	0.0000
Horsepower				
Category 1	0	0	-	-
Category 2	-0.2235	0.1419	-1.5748	0.1153
Category 3	-0.0537	0.1379	-0.3891	0.6972
Category 4	0.0151	0.1403	0.1075	0.9144
Gender				
Male	0	0	-	-
Female	0.0794	0.0281	2.8260	0.0037
α	-0.4222	0.0555	-7.5960	0.0000

Table 3.2: Results of the Fitted Poisson-Inverse Gaussian GAMLSS

Variable	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.8551	0.1561	-5.4793	0.0000
Bonus-Malus				
Category 1	0	0	-	-
Category 2	0.6304	0.0578	10.9046	0.0000
Category 3	0.8483	0.0426	19.9098	0.0000
Category 4	-0.9435	0.0797	-11.8441	0.0000
Category 5	2.0822	0.2119	9.8251	0.0000
Horsepower				
Category 1	0	0	-	-
Category 2	-0.2079	0.1492	-1.3937	0.1634
Category 3	-0.0372	0.1505	-0.2475	0.8045
Category 4	0.0178	0.1460	0.1219	0.9030
Gender				
Male	0	0	-	-
Female	0.0858	0.0463	1.8521	0.0007
σ	-0.3209	0.0189	-16.9300	0.0000

Table 3.3: Results of the Fitted Sichel GAMLSS

Variable	Estimate	Std Error	t-value/Wald 95%	P-value
Intercept	-0.8731	0.1308	-6.6743	0.0000
Bonus-Malus				
Category 1	0	0	-	-
Category 2	0.6487	0.0385	16.8401	0.0000
Category 3	0.8500	0.0429	19.8365	0.0000
Category 4	-0.9780	0.0697	-14.0254	0.0000
Category 5	2.0944	0.2602	8.0485	0.0000
Horsepower				
Category 1	0	0	-	-
Category 2	-0.1907	0.1352	-1.4105	0.1584
Category 3	-0.0179	0.1307	-0.1370	0.8910
Category 4	0.0146	0.1329	0.1101	0.9124
Gender				
Male	0	0	-	-
Female	0.0949	0.0280	3.3172	0.0008
σ	0.8890	0.4234	2.0990	0.0298
ν	-1.2650	0.1210	-10.4546	0.0000

From Tables 3.1, 3.2 and 3.3 we observe that model selection using GAIC results in the same set of significant explanatory variables (BM category, horsepower category and gender) for the location parameter μ_i^j of each model. Bonus-Malus categories 2, 3 and 5 have a positive effect on the mean claim frequency μ_i^j while the Bonus-Malus category 4 has a negative effect on μ_i^j. The horsepower categories are not so important for μ_i^j. Note also that Bonus-Malus category 1, horsepower category 1 and male drivers are the reference categories. Also, as expected, the coefficient values are almost equal in the case of the NBI, PIG and Sichel models respectively. The positive values of the coefficients indicate higher risk compared to the reference class, whereas negative values demonstrate lower risk than the reference class.

3.4.2 Models Comparison

In this subsection we compare the fit of the models for the observed claim frequencies in the portfolio of 15641 policyholders analyzed earlier. These models are all non-nested. In order to accept or reject some models, classical hypothesis/specification tests for non-nested models can be used (see, Boucher et al., 2007, 2008). Firstly, we compare the non-nested distributions presented in Section 2. In this case, information criteria like AIC or SBC are useful as well as the Vuong test (Vuong, 1989). Table 3.4 (Panels A and B) reports our results with respect to the aforementioned non-nested comparisons. Specifically, from Panel A and Panel B we observe the superiority of the Poisson-Inverse Gaussian distribution vs the Negative Binomial distribution. Overall, the best fit is given by the Sichel distribution.

Table 3.4: Comparison of Distributions for the Greek Data Set

Panel A: Based on AIC, BIC				
Model	df	AIC		SBC
Negative Binomial	2	29338.6		29353.9
PIG	2	29313.2		29328.5
Sichel	3	29311.9		29334.9
Panel B: Based on Vuong test				
Model 1	Model 2	Vuong Test Statistic	p-value	Decision
Negative Binomial	PIG	-2.38	0.00	PIG
PIG	Sichel	-0.71	0.00	Sichel

Secondly, we compare the non-nested GAMLSS[7] models presented in Section 3 employing Global Deviance, AIC, SBC (see, Rigby and Stasinopoulos, 2009) and the Vuong test. The results are displayed in Table 3.5. Specifically, when the Global Deviance, AIC and BIC are used (Table 3.5, Panel A)) our findings suggest that the PIG GAMLSS is superior to the NBI GAMLSS. However, when the Vuong test is used, (Table 3.5, Panel B) we observe the superiority of the NBI GAMLSS vs the PIG GAMLSS. Finally, with respect to the Global Deviance, AIC, BIC and the Vuong test results, the Sichel GAMLSS provided the best fitting performances.

[7]As we have already mentioned, we fitted the paramatric linear GAMLSS and we allowed only the mean parameter to be modelled as a function of the significant a priori rating variables for the number of claims.

Table 3.5: Comparison of GAMLSS Models for the Greek Data Set

Panel A: Based on Global Deviance, AIC, BIC				
Model	df	Global Deviance	AIC	SBC
NBI	10	28417.9	28437.9	28514.5
PIG	10	28380.1	28380.1	28476.6
Sichel	11	28347.28	28369.3	28453.5
Panel B: Based on Vuong test				
Model 1	Model 2	Vuong Test Statistic	p-value	Decision
NBI	PIG	26.45	0.00	NBI
NBI	Sichel	-2.46	0.00	Sichel

Note that the same conclusions may not necessarily apply to another observed portfolio.

3.4.3 Optimal BMS Based on the a Posteriori Criteria

In this subsection we consider the premiums determined by the optimal BMS based on the a posteriori classification criteria. As we have already mentioned, all the policies were in force for 3.5 years thus the expected claim frequencies must be multiplied by the exposure to risk $e = \frac{1}{3.5}$ in order to compute the premiums. In the following examples, the premiums will be divided by the premium when $t = 0$, since we are not so much interested in the absolute premium values as in the differences between various classes. We will present the results so that the premium for a new policyholder is 100.

Let us consider a policyholder observed for 7 years whose number of claims range from 1 to 6. In the following tables we compute this individual's scaled premiums for the case of the Negative Binomial, PIG and Sichel models respectively. We consider first the Negative Binomial model with pdf given by Eq. (3.3), following Lemaire (1995). The maximum likelihood estimators of the parameters are $\hat{\tau} = 7.868$ and $\hat{\alpha} = 1.089$. Note also that $\hat{\mu} = \frac{\hat{\alpha}}{\hat{\tau}} = 0.138$. The optimal BMS resulting from the Negative Binomial distribution will be defined by Eq. (3.5) and is presented in Table 3.6.

Table 3.6: Optimal BMS Based on the a Posteriori Classification Criteria, Negative Binomial Model

	Number of Claims						
Year				k			
t	0	1	2	3	4	5	6
0	100.00	0.00	0.00	0.00	0.00	0.00	0.00
1	88.72	170.14	251.55	332.95	414.37	495.77	577.19
2	79.73	152.89	226.05	299.21	372.40	445.54	518.70
3	72.40	138.82	205.25	271.68	338.11	404.55	471.00
4	66.29	127.13	187.96	248.79	309.63	370.46	431.30
5	61.14	117.25	173.35	229.46	285.56	341.67	397.80
6	56.73	108.79	160.85	212.91	265.00	317.04	369.09
7	52.92	101.48	150.03	198.60	247.15	295.71	344.27

Let us consider next the Poisson-Inverse Gaussian (PIG) distribution with pdf given by Eq. (3.9). The maximum likelihood estimators of the parameters are $\hat{\mu} = 0.138$ and $\hat{\sigma} = 0.989$. The BMS derived by the PIG distribution will be defined by Eq. (3.13) and is presented in Table 3.7.

Table 3.7: Optimal BMS Based on the a Posteriori Classification Criteria, Poisson-Inverse Gaussian Model

	Number of Claims						
Year				k			
t	0	1	2	3	4	5	6
0	100.00	0.00	0.00	0.00	0.00	0.00	0.00
1	88.60	156.62	254.20	371.02	497.35	628.04	760.81
2	80.37	131.16	201.63	286.00	378.14	474.22	572.34
3	74.08	113.86	167.54	231.65	302.14	376.17	452.15
4	69.08	101.31	143.81	194.37	250.22	309.22	370.06
5	64.95	91.77	126.42	167.44	212.90	261.14	311.11
6	61.50	84.25	113.16	147.20	185.00	225.25	267.10
7	58.54	78.17	102.72	131.50	163.45	197.61	233.24

Finally, we consider the Sichel distribution with pdf given by Eq. (3.8). The maximum likelihood estimators of the parameters are $\hat{\mu} = 0.138$, $\hat{\sigma} = 0.990$ and $\hat{\nu} = -1.244$. This system provided by this model will be defined by Eq. (3.12) and is presented in Table 3.8.

Table 3.8: Optimal BMS Based on the a Posteriori Classification Criteria, Sichel Model

Number of Claims							
Year				k			
t	0	1	2	3	4	5	6
0	100.00	0.00	0.00	0.00	0.00	0.00	0.00
1	94.32	158.79	262.10	400.55	561.44	733.96	912.19
2	88.83	134.68	201.93	289.09	390.60	500.82	615.94
3	83.96	118.78	166.88	227.55	298.02	375.08	456.26
4	79.71	107.34	143.89	188.97	241.06	298.24	358.87
5	76.00	98.62	127.58	162.66	202.94	247.21	294.36
6	72.72	91.71	115.37	143.60	175.81	211.21	249.02
7	69.82	86.05	105.86	129.17	155.60	184.63	215.69

It is interesting to compare the optimal BMS provided by the Sichel distribution with the systems obtained from the Poisson-Inverse Gaussian and Negative Binomial distributions respectively. From Table 3.6, Table 3.7 and Table 3.8 we observe that these three systems are fair since if the policyholder has a claim free year the premium is reduced, while if the policyholder has one or more claims the premium is increased. Furthermore, we notice that they can be considered generous with good risks and strict with bad risks. For example, the bonuses given for the first claim free year are 11.28%, 11.4% and 5.68% of the basic premium in the case of the Negative Binomial (Table 3.6), Poisson-Inverse Gaussian (Table 3.7) and Sichel (Table 3.8) models respectively. On the contrary, policyholders who had one claim over the first year of observation will have to pay a malus of 70.14%, 56.62% and 58.79% of the basic premium in the case of the Negative Binomial, Poisson-Inverse Gaussian and Sichel models respectively. Also, policyholders who had one claim over the second year of observation will have to pay a malus of 51.55%, 54.20% and 62.10% in the case of the Negative Binomial, Poisson-Inverse Gaussian model and Sichel models respectively.

3.4.4 Optimal BMS Based Both on the a Priori and the a Posteriori Criteria

In this subsection we consider the premiums determined by the generalized optimal BMS that integrates the a priori and the a posteriori information on an individual basis. The expected claim frequencies are multiplied again by the exposure to risk $e = \frac{1}{3.5}$ in order to derive the premiums. The premiums are divided again by the premium when $t = 0$, as it is interesting to see the percentage change in the premiums after one or more claims.

Let us see an example in order to understand better how this BMS works. Consider a group of policyholders who share the following common characteristics. The policyholder i is a woman who has a car with horsepower between 0-33 and her Bonus-Malus (BM) class varies over time, starting from BM class 1[8]. Implementing the NBI GAMLSS (Eq. (3.19)) we found

[8]Recall that BM class 1 corresponds to BM category 1 according to the grouping of the levels of the BM

that $\hat{\alpha} = 0.655$, implementing the PIG GAMLSS (Eq. (3.27)) we found that $\hat{\sigma} = 0.725$, and implementing the Sichel GAMLSS (Eq. (3.26)) we found that $\hat{\sigma} =$ 0.889 and $\hat{\nu} = -3.023$. As we have already mentioned, the mean (or location) parameter of these models is given by $E(K_i^j|c_i^j) = \mu_i^j = \exp\left(c_i^j \beta^j\right)$, where $c_i^j\left(c_{i,1}^j, ..., c_{i,h}^j\right)$ is the $1 \times h$ vector of h individual characteristics, which represent different a priori rating variables and β^j is the vector of the coefficients. The estimation of the vector β^j and therefore of the mean parameter for the NBI, PIG and Sichel distributions respectively led to the following results presented in Table 3.9.

Table 3.9: Women, Horse Power 0-33

Bonus-Malus Category	NBI $\hat{\mu}_i^j$	PIG $\hat{\mu}_i^j$	Sichel $\hat{\mu}_i^j$
1	0.1339	0.1323	0.1314
2	0.2459	0.2483	0.2514
3	0.3123	0.3088	0.3073
4	0.0523	0.0515	0.0490
5	0.9571	1.0610	1.0642

Based on the above estimates for this group of individuals we are now able to derive the generalized optimal BMSs resulting from the Eqs (3.23, 3.40 and 3.39) for the case of the NBI, PIG and Sichel models respectively. These BMSs are presented in Table 3.10. Note that Bonus-Malus class varies substantially depending on the number of claims of policyholder i for period j. For this reason in Table 3.10 we specify the exact order of the claims history in order to derive the scaled premiums that must be paid by this group of policyholders, based on the transition rules of this system (see Chapter 2) and assuming that the age of the policy is up to 2 years.

class explanatory variable.

Table 3.10: Women, Horse Power 0-33, Varying Bonus-Malus Class

Year	Number of Claims k_t	Optimal BMS NBI	Optimal BMS PIG	Optimal BMS Sichel
$t = 0$	$k_0 = 0$	100	100	100
	$k_1 = 0$	91.93	91.60	98.31
$t = 1$	$k_1 = 1$	279.57	247.92	297.21
	$k_1 = 2$	390.29	347.12	463.42
	$k_1 = 0, k_2 = 0$	85.06	85.01	92.77
$t = 2$	$k_1 = 0, k_2 = 1$	258.69	220.31	257.80
	$k_1 = 0, k_2 = 2$	361.15	297.74	371.50
	$k_1 = 1, k_2 = 0$	132.56	107.47	121.53
$t = 2$	$k_1 = 1, k_2 = 1$	339.90	265.83	318.49
	$k_1 = 1, k_2 = 2$	554.13	423.95	520.57
	$k_1 = 2, k_2 = 0$	339.90	265.83	318.49
$t = 2$	$k_1 = 2, k_2 = 1$	554.13	423.95	520.57
	$k_1 = 2, k_2 = 2$	676.59	527.64	672.74

Consider now another group of policyholders who share the following common characteristics. The policyholder i is now a man who has a car with horsepower between 0-33 and his BM class varies over time, starting from BM class 1. The estimation of the vector β^j and thus of the mean parameter of the NBI, PIG and Sichel distributions respectively led to the following results displayed in Table 3.11.

Table 3.11: Men, Horse Power 0-33

Bonus-Malus Category	NBI $\hat{\mu}_i^j$	PIG $\hat{\mu}_i^j$	Sichel $\hat{\mu}_i^j$
1	0.1237	0.1215	0.1194
2	0.2272	0.2282	0.2286
3	0.2886	0.2837	0.2795
4	0.0483	0.0472	0.0446
5	0.8844	0.9745	0.9678

Based on the above estimates for this new group of policyholders we can derive the generalized optimal BMSs provided by the Eqs (3.23, 3.40 and 3.39) for the case of the NBI, PIG and Sichel models respectively. In Table 3.12 we specify again the exact order of the claims history in order to compute the scaled premiums that must be paid by this new group of policyholders assuming again that the age of the policy is up to 2 years. For example, consider again a policyholder who at $t = 2$ has a total number of claims $K = 2$. From Table 3.12 we can see that if he has claim frequency history $k_1 = 0, k_2 = 2$ then his premium increases from 100 to 365.29, 304.69 and 384.82, in the case of the NBI, PIG and Sichel models respectively. On the contrary, if he has claim frequency history $k_1 = 1, k_2 = 1$ then his premium increases from 100 to 345.13, 273.55 and 331.80 in the case of the NBI, PIG and Sichel models respectively.

Table 3.12: Men, Horse Power 0-33, Varying Bonus-Malus Class

Year	Number of Claims k_t	Optimal BMS NBI	Optimal BMS PIG	Optimal BMS Sichel
$t = 0$	$k_0 = 0$	100	100	100
	$k_1 = 0$	92.49	92.21	98.85
$t = 1$	$k_1 = 1$	281.29	250.58	301.71
	$k_1 = 2$	392.70	351.99	474.61
	$k_1 = 0, k_2 = 0$	86.04	85.98	93.72
$t = 2$	$k_1 = 0, k_2 = 1$	261.66	224.27	263.82
	$k_1 = 0, k_2 = 2$	365.29	304.69	384.82
	$k_1 = 1, k_2 = 0$	134.60	109.89	124.95
$t = 2$	$k_1 = 1, k_2 = 1$	345.13	273.55	331.80
	$k_1 = 1, k_2 = 2$	562.66	438.37	548.06
	$k_1 = 2, k_2 = 0$	345.13	273.55	331.80
$t = 2$	$k_1 = 2, k_2 = 1$	562.66	438.37	548.06
	$k_1 = 2, k_2 = 2$	687.00	547.37	713.46

Note that from Table 3.10 and Table 3.12 we observe that the premiums that should be paid by a woman who has a car with horsepower between 0-33 and her BM class varies over time do not differ much from those that should be paid by a man who shares common characteristics. Note also that other combinations of a priori characteristics could be used and also different claim frequency histories.

It is interesting to compare these BMSs with those obtained when only the a posteriori classification criteria are used. Using these BMSs we saw from Table 3.6, Table 3.7 and Table 3.8 that a policyholder who at $t = 2$ has two claims faces a malus of 126.05%, 101.63% and 101.93% of the basic premium in the case of the Negative Binomial, Poisson-Inverse Gaussian and Sichel distributions respectively. Using the generalized optimal BMSs based both on the a priori and the a posteriori classification criteria we consider first a woman, who has a car with horsepower between 0-33 and her BM class varies over time. From Table 3.10 we saw that if at $t = 2$ she has claim frequency history $k_1 = 0, k_2 = 2$, she faces a malus of 261.15%, 197.74% and 271.50% of the basic premium in the case of the NBI, PIG and Sichel GAMLSS respectively, while if she has $k_1 = 1, k_2 = 1$ claim frequency history then she faces a malus of 239.90%, 165.83% and 218.49% of the basic premium in the case of the NBI, PIG and Sichel GAMLSS respectively. Consider also a man, who has a car with horsepower between 0-33 and his BM class varies over time. From Table 3.12 we saw that if at $t = 2$ he has claim frequency history $k_1 = 0, k_2 = 2$, he faces a malus 265.29%, 204.69% and 284.82% of the basic premium, in the case of the NBI, PIG and Sichel GAMLSS respectively, while if he has $k_1 = 1, k_2 = 1$ claim frequency history then he faces a malus of 245.13%, 173.55% and 231.80% of the basic premium in the case of the NBI, PIG and Sichel GAMLSS respectively. These systems are more fair since they consider all the important a priori and a posteriori information for the number of claims of each policyholder in order to estimate their risk of having an accident and thus they permit the differentiation of the premiums for various number of claims based on the expected claim frequency of each policyholder as this is estimated both from the a priori and the a posteriori classification criteria. The optimal BMSs obtained have all the attractive properties of the BMS developed by Dionne and Vanasse (1989, 1992).

A possible line of future research is the use of different claim frequency models within the framework of the GAMLSS.

Chapter 4

Modelling Claim Losses in Optimal Bonus-Malus Systems

4.1 Introduction

As noted in Chapters 1 and 3, a BMS is called optimal if it satisfies two conditions: firstly, it must be financially balanced for the insurer, that is, the total amount of bonuses is equal to the total amount of maluses, and secondly, it must be fair to the policyholders, that is, each policyholder pays a premium proportionate to the risk that he brings to the pool. However, if there is no difference in penalty between the policyholder having an accident with a small size of loss and a policyholder with a big size of loss, a BMS can be said to be unfair. Therefore, a system which takes both the frequency and the severity of claims into account must be used to set the premium an insured will pay.

Among the BMSs that take severity into consideration are those designed from Picard (1976), Lemaire (1995), Pinquet (1997), Frangos and Vrontos (2001), Pitrebois et al. (2006) and Mahmoudvand and Hassani (2009). Picard (1976) generalized the traditional Negative Binomial model in order to take into account the subdivision of claims into small and large losses. In order to separate large from small losses, two options were used. The first option was to consider that those losses under a threshold are regarded as small and the remainder as large. The second option was to subdivide the accidents into those that caused property damage and those that caused bodily injury, penalizing more severely the insureds who had a bodily injury accident. Lemaire (1995) first applied Picard's (1976) thought to Belgian data. Lemaire found that this classification would lead to serious practical problems, since it is time consuming to access exact amounts, and insureds who have claims just over the limit protest strongly. Hence, he proposed an improved categorization where only two categories were analyzed. This first category was the accidents with merely property damage and the second category was the accidents with bodily injuries. He further assumed that the frequency of accidents with bodily injuries conformed to a Beta distribution. It has been found that the malus that must be paid for one accident with bodily injuries is as high as those for four accidents with only property damage. Pinquet (1997) proposed the design of an optimal BMS which makes allowance for the severity of claims. Starting from a rating model and based on the analysis of claim frequency and claim severity he added two heterogeneity components to represent unobserved factors that explain the severity variables. The size of the claims was expected to follow gamma or lognormal distribution. The rating factors, as well as the heterogeneity components were included in the scale parameter of the distribution. Considering that the heterogeneity also follows a gamma or lognormal distribution, a credibility expression was obtained to provide a predictor for the average claim size for the following period. Frangos and Vrontos (2001) developed the design of an optimal BMS with a frequency and a severity component expanding the setup that Lemaire (1995) used to design an optimal BMS based on the number of claims. They assumed that the number of claims was distributed according to the Negative Binomial distribution and that the losses of the claims were distributed according to an Exponential-Inverse Gamma , the Pareto distribution. Applying Bayes theorem they obtained the posterior distribution of the mean claim size given the information they had about the claim size history for each policyholder for the time period they were in the portfolio. Furthermore, Frangos and Vrontos proposed a generalized BMS that integrates the a priori and the a posteriori information on a individual basis, expanding the framework developed by Dionne and Vanasse (1989, 1992). Pitrebois et al. (2006) extended the model proposed by Lemaire (1995) by making a more detailed classification. Four different types of claims were studied: accident with or without bodily injury and those with or without partial liability of the driver. Furthermore, they extended Lemaire's Beta distribution to a multivariate Beta distribution, the Dirichlet distribution. Finally, Mahmoudvand and Hassani (2009) also considered a generalized BMS with a frequency and a severity component following the setup proposed by Frangos and Vrontos.

The first objective of this chapter is the integration of claim severity into the optimal BMS based on the a posteriori criteria, which was presented in Chapter 3 for the case of the Negative Binomial, Poisson Inverse Gaussian and Sichel distributions respectively. For this purpose we consider that the losses are distributed according to a Pareto distribution. The optimal BMS resulting from the Sichel distribution for assessing claim frequency and the Pareto distribution for assessing claim severity is proposed as an alternative to the system provided by the Negative Binomial and Pareto models (see Frangos and Vrontos, 2001). Furthermore, we consider the system obtained by the Poisson-Inverse Gaussian distribution (PIG) for assessing claim frequency and the Pareto distribution for assessing claim severity, since the PIG is a special case of the Sichel distribution. The second objective of this chapter is the development of a generalized BMS with a frequency and a severity component when both the a priori and the a posteriori rating variables are used. For the frequency component we assume that the number of claims is distributed according to the Negative Binomial Type I, Poisson Inverse Gaussian and Sichel GAMLSS respectively, based on the methodology presented in Chapter 3. For the severity component we consider that the losses are distributed according to a Pareto GAMLSS, extending the framework developed by Frangos and Vrontos (2001). With the aim of constructing an optimal BMS by updating the posterior mean claim frequency and the posterior mean claim severity we adopt the parametric linear formulation of these GAMLSS models and we allow only their mean parameter to be modelled as a function of the explanatory variables. The generalized system we propose will be derived as a function of the years that the policyholder was in the portfolio, their number of accidents, the size of loss of each of these accidents and of the statistically significant a priori rating variables for the number of accidents and for the size of loss that each of these claims incurred. Furthermore, we present a generalized form of the BMS obtained by Frangos and Vrontos (2001).

Let us now detail the contents of this chapter. Section 4.2 describes the design of optimal BMS with a frequency and a severity component based on the a posteriori criteria. The system presented in Section 4.3 is based on both the a posteriori and the a priori classification criteria and Section 4.4 contains an application to the data set presented in Chapter 1.

4.2 The Design of an Optimal BMS with a Severity Component Based on the a Posteriori Criteria

The optimal BMS based on the a posteriori frequency component was presented in Chapter 3 for the case of Negative Binomial, Poisson-Inverse Gaussian (PIG) and Sichel distributions respectively. We assume that the number of claims of each policyholder is independent[1] of the severity of each claim in order to deal with the frequency and the severity component separately.

Similarly to the design of the optimal BMS based on the claim frequency component, the design of the optimal BMS based on the claim severity component will be developed again through Bayesian analysis. Each policyholder will have to pay a premium proportional to his unknown claim severity and the loss of the insurer will derive from the use of the estimated claim severity instead of the true unknown claim severity. The estimate of the policyholder's claim severity that minimizes the loss incurred will be the optimal one and the quadratic error loss function will be used again for the penalization of the actuary's errors. In this way the estimate of the revised claim severity will be the a posteriori expectation. The resulting optimal BMS will be constructed according to the expected value premium calculation principle and it is fair to the policyholders and financially balanced for the insurance companies. In what follows we present an optimal BMS obtained by the Pareto distribution for assessing claim severity. Next we use the net premium principle for the calculation of the premiums determined by the optimal BMS with a frequency and a severity component based on the a posteriori criteria for the case of the Negative Binomial-Pareto, PIG-Pareto and Sichel-pareto models respectively, and finally we discuss the properties of this system.

4.2.1 The Pareto Model

We consider a heterogeneous portfolio with respect to the mean claim size of each policyholder. Let x be the claim size of each insured and consider that their mean claim size is denoted as y. We assume that the conditional distribution of $x|y$ is a one parameter Exponential distribution with probability density function (pdf), given by

$$f(x|y) = \frac{e^{-\frac{x}{y}}}{y}, \tag{4.1}$$

[1]This is at best an approximation, since for example city drivers have more but cheaper accidents than the drivers in rural areas.

for $x > 0$, $y > 0$, with mean $E(x|y) = y$ and variance $Var(x|y) = y^2$. Following the setup of Frangos and Vrontos (2001), we consider that the structure function follows an Inverse Gamma distribution, which has a pdf of the form

$$g(y) = \frac{\frac{1}{m}\exp\left(-\frac{m}{y}\right)}{\left(\frac{y}{m}\right)^{s+1}\Gamma(s)}, \tag{4.2}$$

for $y > 0, s > 0, m > 0$, with mean $E(y) = \frac{m}{s-1}$ and variance $Var(y) = \frac{m^2}{(s-1)^2(s-2)}$, for $s > 2$. Then it can be proved that the unconditional distribution of the claims severity x will be a Pareto distribution, with pdf given by

$$f(x) = sm^s(x+m)^{-s-1}, \tag{4.3}$$

for $y > 0, s > 0, m > 0$ where $E(x) = \frac{m}{s-1}$ and where $Var(x) = \frac{m^2}{(s-1)^2(s-2)}$. In this way the relatively tame exponential distribution gets transformed into the heavy-tailed Pareto distribution which can be considered as a good candidate for modeling the claim severity. Also by taking y distributed according to an Inverse Gamma, the heterogeneity that characterizes the severity of the claims of different policyholders is incorporated in the model. Such a generation of the Pareto distribution can be found in Herzog (1996) and in other actuarial papers, but Frangos and Vrontos (2001) were the first who proposed it for the design of an optimal BMS.

Posterior Structure Function

Consider that a policyholder stays in the portfolio for t years, the number of claims they had in the year j is denoted by k_j , the total number of claims that they had in t years is denoted by $K = \sum_{j=1}^{t} k_j$ and by x_k is denoted the claim amount for the k claim. Then the information we have for their claim size history will be in the form of a vector $x_1, ..., x_k$ and the total claim amount for that specific policyholder over the t years that they are in the portfolio will be equal to $\sum_{k=1}^{K} x_k$. Applying the Bayes theorem, we find that the posterior structure function of the mean claim size y, given the policyholder's claim size history $x_1, ..., x_K$, denoted as $g(y|x_1, ..., x_K)$, is given by

$$g(y|x_1, ..., x_K) = \frac{\left(m + \sum_{k=1}^{K} x_k\right)^{K+s} e^{-\frac{\left(m+\sum_{k=1}^{K} x_k\right)}{y}}}{y^{K+s+1}\Gamma(K+s)}, \tag{4.4}$$

which is the pdf of a of Inverse Gamma$\left(K+s, m+\sum_{k=1}^{K} x_k\right)$. For more information about the derivation of Eqs (4.3 and 4.4) refer to Frangos and Vrontos (2001). Also a more general proof of Eqs (4.3 and 4.4) can be found in Chapter 5, where we consider the case of the n-component of Pareto mixture distribution derived by assuming that the severity of claims x is distributed according to an Exponential(y), and that the structure function follows an n-component Inverse Gamma distribution.

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$

Consequently, by using the quadratic error loss function the optimal choice of $\hat{y}_{t+1}$ for a policyholder with claim size history $x_1, ..., x_K$ is the mean of the posterior structure function given by Eq. (4.4), that is

$$\hat{y}_{t+1}\left(x_1, ..., x_K\right) = \frac{m + \sum_{k=1}^{K} x_k}{K + s - 1}. \tag{4.5}$$

4.2.2 Calculation of the Premiums According to the Net Premium Principle

Consider a policyholder or a group of policyholders who in t years have produced K claims with total claim amount equal to $\sum_{k=1}^{K} x_k$. As mentioned in Chapter 3, their expected number of claims for period $t+1$, $\lambda_{t+1}(k_1, ..., k_t)$, is given by the Eqs (3.5, 3.13 and 3.12) for the case of Negative Binomial, Poisson-Inverse Gaussian (PIG) and Sichel distributions respectively and as shown previously their expected claim severity, $y_{t+1}(x_{1,...,}x_K)$ is given by Eq. (4.5). The net premium that should be paid by that specific group of policyholders is given by the product of their expected number of claims $\lambda_{t+1}(k_1, ..., k_t)$, and their expected claim severity, $y_{t+1}(x_{1,...,}x_K)$, for period $t+1$, and is equal to

$$Premium = e\hat{\lambda}_{t+1}(k_1, ..., k_t)\hat{y}_{t+1}(x_1, ..., x_K), \tag{4.6}$$

where $e = \frac{1}{3.5}$ denotes the exposure to risk, since as mentioned in Chapters 2 and 3 all policyholders were observed for 3.5 years.

- In the case of the Negative Binomial-Pareto model Eq. (4.6) becomes

$$Premium = e\frac{K+\alpha}{\tau + t}\frac{m + \sum_{k=1}^{K} x_k}{K + s - 1}, \tag{4.7}$$

  where $\alpha > 0, \tau > 0$ and where $s > 0, m > 0$.

- In the case of the PIG-Pareto model Eq. (4.6) becomes

$$Premium = e\left(\sqrt{\frac{h_2}{h_1}}\right)\frac{K_{K+\frac{1}{2}}\left(h_1h_2\right)}{K_{K-\frac{1}{2}}\left(h_1h_2\right)}\frac{m+\sum_{k=1}^{K}x_k}{K+s-1}, \tag{4.8}$$

where $h_1 = \frac{1}{\sigma\mu} + 2t$ and $h_2 = \frac{\mu}{\sigma}$ where $\sigma > 0$ and where $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z and with $s > 0, m > 0$.

- In the case of the Sichel-Pareto model Eq. (4.6) becomes

$$Premium = e\left(\sqrt{\frac{w_2}{w_1}}\right)\frac{K_{K+\nu+1}\left(w_1w_2\right)}{K_{K+\nu}\left(w_1w_2\right)}\frac{m+\sum_{k=1}^{K}x_k}{K+s-1}, \tag{4.9}$$

where $w_1 = \frac{c}{\sigma\mu} + 2t$ and $w_2 = \frac{\mu}{\sigma c}$ with $\sigma > 0, -\infty < \nu < \infty$ and $c = \frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_\nu\left[\frac{1}{\sigma}\right]}$ and where $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z and with $s > 0, m > 0$.

In order to find the premium that must be paid we have to know:

1. the maximum likelihood estimates of the parameters α and τ of the Negative Binomial distribution with pdf given by Eq. (3.3) in Chapter 3,
2. the maximum likelihood estimates of the parameters μ and σ of the Poisson-Inverse Gaussian distribution with pdf given by Eq. (3.9) in Chapter 3,
3. the maximum likelihood estimates of the parameters μ, σ and ν of the Sichel distribution with pdf given by Eq. (3.8) in Chapter 3,
4. the maximum likelihood estimates of the parameters m and s of the Pareto distribution with pdf given by Eq. (4.3),
5. the number of years t that the policyholder is under observation,
6. their total number of claims $K = \sum_{j=1}^{t} k_j$, where k_j the number of accidents in which they were at fault in year $j = 1, ..., t$ and their total claim amount $\sum_{k=1}^{K} x_k$.

4.2.3 Properties of the Optimal BMS with a Frequency and a Severity Component Based on the a Priori Criteria

The optimal Bonus-Malus System with a frequency and a severity component based on the a posteriori criteria has several important properties (see Frangos and Vrontos, 2001).

1. The system is fair in a Bayesian sense. Each insured is informed of both the number and the size of their claims while they are in the portfolio, and the premium they have to pay at each renewal is proportional to the estimate of their claim frequency and claim severity, taking into account, through the Bayes theorem, all the information gathered in the past. We use the exact loss x_k that is incurred from each claim in order to have a differentiation in the premium for policyholders with the same number of claims, not just a scaling with the average claim severity of the portfolio.

2. The system is financially balanced. Each year,the average premium per policyholder remains constant at the initial level

$$P = e\frac{\alpha}{\tau}\frac{m}{s-1}, \tag{4.10}$$

for the case of the Negative Binomial-Pareto model and

$$P = e\mu\frac{m}{s-1}, \tag{4.11}$$

for the case of the PIG-Pareto and Sichel-Pareto models respectively. The financial stability of the BMS is proved considering that the claim frequency and the claim severity are independent components and that

$$E_\Lambda\left[\Lambda\right] = E\left[E\left[\lambda|k_1, ...k_t\right]\right],$$

$$E_Y\left[Y\right] = E\left[E\left[y|x_1, ..., x_K\right]\right].$$

3. In the beginning all policyholders are paying the same premium, which is equal to (4.10), when we consider the NB-Pareto model, and equal to (4.11), when we consider the PIG-Pareto and Sichel-Pareto models respectively.

4. The Bayesian credibility premium increases proportionally to the number and to the severity of the claims and always decreases when no accidents are caused.

5. The phenomenon of bonus hunger will decrease and the estimate for the actual claim frequency will be more accurate since the claims with small loss will be reported due to the fact that the policyholders who had them will know that claim severity will be taken into consideration.

6. The introduction of the severity component is more crucial than the number of claims for the insurer since it determines the expenses of the insurer incurred by accidents, and thus the premium that must be paid.

7. The estimator of the mean of severity is not always robust and can be affected by variation. So in practice a more robust estimator could be used. (i.e. cutting of the data, M-estimator).

4.3 The Design of an Optimal BMS with a Severity Component Based Both on the a Priori and the a Posteriori Criteria

The objective of this section is the development of a generalized BMS with a frequency and a severity component when both the a priori and the a posteriori rating variables are used. For the frequency component we assume that the number of claims is distributed according to the Negative Binomial Type I, Poisson Inverse Gaussian and Sichel GAMLSS respectively, based on the methodology presented in Chapter 3. For the severity component we consider that the losses are distributed according to a Pareto GAMLSS, extending the framework developed by Frangos and Vrontos (2001). As mentioned in the previous chapters, the GAMLSS basically consist of four different formulations: the semi-parametric additive model, the parametric linear model, the non-linear semi-parametric additive model and the non-linear parametric model. With the aim of constructing an optimal BMS by updating the posterior mean claim frequency and the posterior mean claim severity, we adopt the parametric linear formulation of these models and we allow only their mean parameter to be modelled as a function of the explanatory variables. In the resulting generalized systems, the premium is a function of the years that the policyholder was in the portfolio, their number of accidents, the size of loss of each of these accidents and of the statistically significant a priori rating variables for the number of accidents and for the size of loss that each of these claims incurred. Furthermore, we present a generalized form of the system obtained by Frangos and Vrontos (2001).

The premiums of the generalized BMS we present will be derived using the following multiplicative tariff formula:

$$Premium = GBM_F \ \cdot GBM_S, \tag{4.12}$$

where GBM_F denotes the generalized BMS obtained for the frequency component where we employ the NBI, PIG and Sichel GAMLSS respectively, and GBM_S denotes the generalized BMS obtained for the severity component where we employ the Pareto GAMLSS. The generalized premiums obtained by the Eq. (4.12) vary simultaneously with the variables that affect the distributions of the number of claims and the size of loss distribution.

4.3.1 The Pareto Model

The generalized Bonus-Malus factor for the severity component is derived according to the structure proposed by Frangos and Vrontos (2001). Consider a policyholder i with an experience of t periods. Assume that the number of claims of the individual i for period j are independent and is denoted as K_i^j and by $X_{i,k}^j$ is denoted the loss incurred from his claim k for the period j.

We consider that $X_{i,k}^{j}$ follows the Exponential distribution with mean claim severity for period j, y^{j}. Then the probability of the loss incurred from claim k for the period j is:

$$f\left(X_{i,k}^{j}=x\right)=\frac{e^{-\frac{x}{y^{j}}}}{y^{j}},$$

for $X_{i,k}^{j}>0$ and $y_{i}^{j}>0$.

Since policyholders have different mean claim severity, it is fair for each policyholder to pay a premium proportional to the risk that they impose on the pool. We can allow the y^{j} parameter to vary from one individual to another. Let $y_{i}^{j}=\exp\left(d_{i}^{j}\gamma^{j}\right)$, where $d_{i}^{j}\left(d_{i,1}^{j},...,d_{i,h}^{j}\right)$ is the $1\times h$ vector of h individual's characteristics, which represent different a priori rating variables and γ^{j} is the vector of the coefficients. The exponential form ensures the non-negativity of y_{i}^{j}. Then, the conditional to d_{i}^{j} pdf of the claim size $X_{i,k}^{j}$, for a claim k of a policyholder i in period j will become

$$f\left(X_{i,k}^{j}|d_{i}^{j}\right)=\frac{e^{-\frac{x}{\exp\left(d_{i}^{j}\gamma^{j}\right)}}}{\exp\left(d_{i}^{j}\gamma^{j}\right)}, \tag{4.13}$$

where $E(X_{i,k}^{j}|d_{i}^{j})=y_{i}^{j}=\exp\left(d_{i}^{j}\gamma^{j}\right)$and where $Var(X_{i,k}^{j}|d_{i}^{j})=\left(y_{i}^{j}\right)^{2}=\left(\exp\left(d_{i}^{j}\gamma^{j}\right)\right)^{2}$. For the determination of the expected claim severity in this model we assume that the h individual characteristics provide enough information. Nevertheless, if one assumes that the a priori rating variables do not contain all the significant information for the mean claim severity then a random variable ξ_{i} has to be introduced into the regression component. Thus we can write

$$y_{i}^{j}=\exp\left(d_{i}^{j}\gamma^{j}+\xi_{i}\right)=\exp\left(d_{i}^{j}\gamma^{j}\right)w_{i},$$

where $w_{i}=\exp\left(\xi_{i}\right)$, yielding a random y_{i}^{j}. We will assume that w_{i} follows an Inverse Gamma distribution with probability density function

$$w\left(w_{i}\right)=\frac{\frac{1}{(s-1)}\exp\left(-\frac{(s-1)}{w_{i}}\right)}{\left(\frac{w_{i}}{s-1}\right)^{s+1}\Gamma\left(s\right)}, \tag{4.14}$$

$w_{i}>0, s>0$ for $=1,...,n,=1$ with mean $E(w_{i})=1$ and variance $Var\left(w_{i}\right)=\frac{1}{s-2}$, for $s>2$. It can be shown that the above parameterization does not affect the results if there is a constant term in the regression. We chose $E(w_{i})=1$ in order to have $E(\xi_{i})=0$. Under this assumption the conditional distribution of $X_{i,k}^{j}|d_{i}^{j}$ becomes

$$f\left(X_{i,k}^{j}|d_{i}^{j}\right)=s\frac{\left((s-1)\exp\left(d_{i}^{j}\gamma^{j}\right)\right)^{s}}{\left(x+(s-1)\exp\left(d_{i}^{j}\gamma^{j}\right)\right)^{s+1}}, \tag{4.15}$$

which is a Pareto distribution with parameters s and $(s-1)\exp\left(d_{i}^{j}\gamma^{j}\right)$. Note that $E(X_{i,k}^{j}|d_{i}^{j})=m_{i}^{j}=\exp\left(d_{i}^{j}\gamma^{j}\right)$ and $Var\left(X_{i,k}^{j}|d_{i}^{j}\right)=\frac{\left[(s-1)\exp\left(d_{i}^{j}\gamma^{j}\right)\right]^{2}}{s-1}\left(\frac{2}{s-2}-\frac{1}{s-1}\right)$. Note also that Eq. (4.15) gives the parametric linear GAMLSS where only the mean parameter of the distribution of the response variable, i.e. the Pareto distribution, is modelled as a function of the significant explanatory variables for the size of claims.

Posterior Structure Function

Our goal is to construct a generalized optimal BMS based on the past claim size history and on an individual's characteristics in order to adjust that individual's premiums over time. Thus the problem is to determine, at the renewal of the policy, the expected claim severity of the policyholder i for the period $t+1$ given the observation of the reported claim sizes in the preceding t periods and observable characteristics in the preceding $t+1$ periods and the current period.

Consider a policyholder i with $X_{i,1}^{1}, X_{i,2}^{2}, ..., X_{i,K_i^t}^{t}$ claim size history in t periods and $d_i^1, ..., d_i^{t+1}$ characteristics. The total number of claims for this specific policyholder in the preceding t periods will be denoted as $K = \sum_{j=1}^{t} K_i^j$ and the total claim amount produced by the accidents where they were at fault will be equal to $\sum_{j=1}^{t}\sum_{k=1}^{K_i^j} X_{i,k}^{j} = \sum_{k=1}^{K} X_{i,k}$. The mean claim severity of the policyholder i for period $t+1$ is $y_i^{t+1}\left(d_i^{t+1}, w_i\right)$, a function of both the vector of an individual's characteristics and a random factor w_i with pdf given by Eq. (4.14). Based on the assumptions of the model, one can find that the probability density function of $y_i^{t+1}\left(d_i^{t+1}, w_i\right)$ denoted as $g\left(y_i^{t+1}\right)$, is given by

$$g\left(y_i^{t+1}\right) = \frac{\left(\frac{(s-1)\exp\left(d_i^{t+1}\gamma^{t+1}\right)}{y_i^{t+1}}\right)^{s} \exp\left(-\frac{(s-1)\exp\left(d_i^{t+1}\cdot\gamma^{t+1}\right)}{y_i^{t+1}}\right)}{y_i^{t+1}\Gamma(s)}, \tag{4.16}$$

for $y_i^{t+1} > 0$ and $s > 0$, which is an Inverse Gamma distribution with parameters s and $\exp\left(d_i^{t+1}\gamma^{t+1}\right)$.

The posterior pdf of the mean claim severity y_i^{t+1} for an individual i observed over $t+1$ periods, with $X_{i,1}^{1}, X_{i,2}^{2}, ..., X_{i,K_i^t}^{t}$ claim size history and $d_i^1, ..., d_i^{t+1}$ characteristics is obtained by applying Bayes theorem and is an Inverse Gamma with updated parameters $s+K$ and C_i^j, with pdf given by

$$g\left(y_i^{t+1} | X_{i,1}^{1}, X_{i,2}^{2}, ..., X_{i,K_i^t}^{t}; d_i^1, ..., d_i^{t+1}\right) = \frac{\frac{1}{C_i^j}\exp\left(-\frac{C_i^j}{y_i^j}\right)}{\left(\frac{y_i^j}{C_i^j}\right)^{K+s+1}\Gamma(s+K)}, \tag{4.17}$$

for $y_i^{t+1} > 0, s > 0$, where

$$C_i^j = \left[(s-1) + \sum_{j=1}^{t} \frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_i^j \gamma^j\right)} \right] \exp\left(d_i^{t+1}\gamma^{t+1}\right).$$

In the case that the vector of the individual characteristics remains constant, i.e. $d_i^1 = d_i^2 = ... = d_i^{t+1} = d_i$ and $\gamma^1 = \gamma^2 = ... = \gamma^t = \gamma$ the posterior pdf of the mean claim severity is simplified to

$$g\left(y_i^{t+1}|X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right) = \frac{\frac{1}{D_i}\exp\left(-\frac{D_i}{y_i^j}\right)}{\left(\frac{y_i^j}{D_i}\right)^{K+s+1}\Gamma(s+K)}, \tag{4.18}$$

where $D_i = (s-1)\exp(d_i\gamma) + \sum_{j=1}^{t}\sum_{k=1}^{K_i^j} X_{i,k}^j$. For more information about the derivation of Eqs (4.15, 4.17 and 4.18) refer to Mahmoudvand and Hassani (2009) and Frangos and Vrontos (2001). A more general proof of Eqs (4.15, 4.17 and 4.18) can be found in Chapter 5 where we consider the case of the n-component Pareto mixture regression model derived by updating the posterior mean.

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$

In the general case, using the quadratic error loss function, the optimal estimator of $\hat{y}_i^{t+1}$ will be the mean of the posterior structure function and is given by

$$\begin{aligned} &\hat{y}_i^{t+1}\left(X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right) \\ = &\int_0^\infty y_i^{t+1}(d_i^{t+1}, w_i) g\left(y_i^{t+1}|X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right) dy_i^{t+1} \\ = &\exp\left(d_i^{t+1}\gamma^{t+1}\right)\left[\frac{(s-1) + \sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_i^j\gamma^j\right)}}{s+K-1}\right]. \end{aligned} \tag{4.19}$$

This estimator defines the premium and corresponds to the multiplicative tariff formula where the base premium is the a priori severity $\exp\left(d_i^{t+1}\gamma^{t+1}\right)$ and where the Bonus-Malus factor is represented by the expression in brackets. When the vector of the individual characteristics remains the same for all years the optimal estimator is simplified to

$$\hat{y}_i^{t+1}\left(X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right) = \frac{(s-1)\exp\left(d_i\gamma\right) + \sum_{j=1}^{t}\sum_{k=1}^{K_i^j} X_{i,k}^j}{s+K-1},$$

which coincides with the one obtained by Frangos and Vrontos (2001). When $t = 0$, $\hat{y}_i^1\left(d_i^1\right) = \exp\left(d_i^1\gamma\right)$ which implies that only a priori rating is used in the first period. Moreover, when the regression component is limited to a constant γ_0 one obtains

$$\hat{y}_i^{t+1}\left(X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t\right) = \frac{(s-1)\exp\left(\gamma_0\right) + \sum_{j=1}^{t}\sum_{k=1}^{K_i^j} X_{i,k}^j}{K+s-1},$$

which corresponds to the ‘univariate’, without regression component, model.

4.3.2 Calculation of the Premiums of the Generalized BMS

Now we are able to compute the premiums of the generalized optimal BMS based both on the frequency and the severity component. As we said, the premiums of the generalized optimal BMS will be given by the product of the generalized BMS based on the frequency component and of the generalized BMS based on the severity component. Consider a policyholder or a group of policyholders who in t years have produced K claims with total claim amount equal to $\sum_{k=1}^{K} X_{i,k}$. As mentioned in Chapter 3, their expected number of claims for period $t+1$, $\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right)$ is given by the Eqs (3.23, 3.40 and 3.39) for the case of the NBI, PIG and Sichel GAMLSS respectively and as shown previously their expected claim severity, $\hat{y}_i^{t+1}\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_i^1, ..., d_i^{t+1}\right)$ is given by Eq.(4.19). The net premium that should be paid by that specific group of policyholders is equal to the product of their expected number of claims and their expected claim severity for the period $t+1$ and is given by

$$\begin{aligned} Premium &= GBM_F \cdot GBM_S \\ &= e\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right) \cdot \\ &\quad \cdot \hat{y}_i^{t+1}\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_i^1, ..., d_i^{t+1}\right), \end{aligned} \tag{4.20}$$

where $e = \frac{1}{3.5}$ is the corresponding exposure to risk.

- In the case of the Negative Binomial-Pareto model Eq. (4.20) becomes

$$Premium = e\exp\left(c_i^{t+1}\beta^{t+1}\right)\left[\frac{\frac{1}{\alpha}+\sum_{j=1}^{t}K_i^j}{\frac{1}{\alpha}+\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}\right]\exp\left(d_i^{t+1}\gamma^{t+1}\right)\left[\frac{(s-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j}X_{i,k}^j}{\exp\left(d_i^j\gamma^j\right)}}{s+K-1}\right]. \tag{4.21}$$

When the vectors of the individual characteristics remain the same from one year to the next Eq. (4.21) is simplified to

$$Premium = e\exp(c_i\beta)\left[\frac{\frac{1}{\alpha}+\sum_{j=1}^{t}K_i^j}{\frac{1}{\alpha}+t\exp(c_i\beta)}\right]\frac{(s-1)\exp(d_i\gamma)+\sum_{j=1}^{t}\sum_{k=1}^{K_i^j}X_{i,k}^j}{s+K-1},$$

which coincides with the one obtained by Frangos and Vrontos (2001).

- In the case of the PIG-Pareto model Eq. (4.20) becomes

$$Premium = e\left(\sqrt{\frac{h_2}{h_1}}\right)\frac{K_{K+\nu+1}(h_1h_2)}{K_{K+\nu}(h_1h_2)}\exp\left(d_i^{t+1}\gamma^{t+1}\right)\left[\frac{(s-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j}X_{i,k}^j}{\exp\left(d_i^j\gamma^j\right)}}{s+K-1}\right], \tag{4.22}$$

where $h_1 = \frac{1+2\sigma\sum_{j=1}^{t}\exp\left(c_i^j\beta^j\right)}{\sigma\exp\left(c_i^{t+1}\beta^{t+1}\right)}$ and $h_2 = \frac{\exp\left(c_i^{t+1}\beta^{t+1}\right)}{\sigma}$, where $\sigma > 0$ and where $K_{K-\frac{1}{2}}(z)$ is the modified Bessel function of the third kind of order $K-\frac{1}{2}$ with argument z. When the vector of the individual characteristics remains the same from one year to the next Eq. (4.22) is simplified to

$$Premium = e\left(\sqrt{\frac{\eta_2}{\eta_1}}\right)\frac{K_{K+\nu+1}(\eta_1\eta_2)}{K_{K+\nu}(\eta_1\eta_2)}\frac{(s-1)\exp(d_i\gamma)+\sum_{j=1}^{t}\sum_{k=1}^{K_i^j}X_{i,k}^j}{s+K-1},$$

for $\eta_1 = \frac{1}{\sigma \exp(c_i \beta)} + 2t$ and $\eta_2 = \frac{\exp(c_i \beta)}{\sigma}$.

- In the case of the Sichel-Pareto model Eq. (4.20) becomes

$$Premium = e \left(\sqrt{\frac{w_2}{w_1}} \right) \frac{K_{K+\nu+1}(w_1 w_2)}{K_{K+\nu}(w_1 w_2)} \exp\left(d_i^{t+1} \gamma^{t+1}\right) \left[\frac{(s-1) + \sum_{j=1}^{t} \frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_i^j \gamma^j\right)}}{s+K-1} \right], \qquad (4.23)$$

where $w_1 = \frac{c + 2\sigma \sum_{j=1}^{t} \exp\left(c_i^j \beta^j\right)}{\sigma \exp\left(c_i^{t+1} \beta^{t+1}\right)}$ and $w_2 = \frac{\exp\left(c_i^{t+1} \beta^{t+1}\right)}{\sigma c}$, where $\sigma > 0$ and $-\infty < \nu < \infty$ and where $c = \frac{K_{\nu+1}\left[\frac{1}{\sigma}\right]}{K_{\nu}\left[\frac{1}{\sigma}\right]}$ and $K_\nu(z)$ is the modified Bessel function of the third kind of order ν with argument z. When the vector of the individual characteristics remains the same from one year to the next Eq. (4.23) is simplified to

$$Premium = e \left(\sqrt{\frac{\omega_2}{\omega_1}} \right) \frac{K_{K+\nu+1}(\omega_1 \omega_2)}{K_{K+\nu}(\omega_1 \omega_2)} \frac{(s-1)\exp(d_i \gamma) + \sum_{j=1}^{t} \sum_{k=1}^{K_i^j} X_{i,k}^j}{s+K-1},,$$

for $\omega_1 = \frac{c}{\sigma \exp(c_i \beta)} + 2t$ and $\omega_2 = \frac{\exp(c_i \beta)}{\sigma c}$
In order to find the premiums that must be paid we have to know:

1. the estimates of the parameter α and the the vector β^j of the significant a priori rating variables for the number of claims for the case of the NBI GAMLSS given by Eq. (3.19) in Chapter 3,

2. the estimates of the parameter σ and the the vector β^j for the case of the PIG GAMLSS given by Eq. (3.27) in Chapter 3,

3. the estimates of the parameters σ and ν and the the vector β^j for the case of the Sichel GAMLSS given by Eq. (3.26) in Chapter 3,

4. the estimates of the parameter s and the the vector γ^j of the significant a priori rating variables for the severity of claims for the case of the Pareto GAMLSS given by Eq. (4.15),

5. the number of years t that the policyholder is under observation,

6. their total number of claims $K = \sum_{i=1}^{t} k_i$, where k_i the number of accidents in which they were at fault in year $i = 1, ..., t$ and their total claim amount $\sum_{k=1}^{K} X_{i,k}$.

4.3.3 Properties of the Optimal BMS with a Frequency and a Severity Component Based Both on the a Priori and the a Posteriori Criteria

In what follows we present the properties of the generalized optimal BMS with a frequency and a severity component based both on the a priori and the a posteriori criteria (see Frangos and Vrontos, 2001).

1. It is fair since it takes into account the claim frequency, the claim severity, the significant a priori rating variables for the claim frequency and the significant a priori rating variables for the claim severity.

2. It is financially balanced for the insurer. Each year the average premium will be equal to

$$P = e \exp\left(c_i^{t+1} \beta^{t+1}\right) \exp\left(d_i^{t+1} \gamma^{t+1}\right) \tag{4.24}$$

   In order to prove Eq. (4.24) it is sufficient to show that

$$E\left[\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_i^1, ..., c_i^{t+1}\right)\right] = e \exp\left(c_i^{t+1} \beta^{t+1}\right)$$

   and that

$$E\left[\hat{y}_i^{t+1}\left(X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right)\right] = \exp\left(d_i^{t+1} \gamma^{t+1}\right).$$

3. All the properties we mentioned for the optimal BMS without the a priori rating variables hold for this BMS as well. In the beginning all the policyholders with the same characteristics are paying the same premium which is equal to (4.24).

4. The generalized premium increases proportionally to the number of claims and to the size of loss that each claim incurred and always decreases when no accidents are caused.

5. This generalized BMS could lead to a decrease in the phenomenon of bonus hunger. The policyholders who had an accident with a small loss will have one more reason to report the claim as they will know that the size of the claim will be taken into consideration and they will not have to pay the same premium as somebody who had a claim with a big loss.

6. The claim severity, which is more crucial than the claim frequency for the insurer, is introduced into the design of the generalized BMS.

7. The premiums vary simultaneously with the variables that affect the distributions of the number of claims and the size of loss distribution.

4.4 Application

We use the same data set we used in Chapter 2, the descriptive statistics of which can be found in Table 2.2 of Chapter 2. As mentioned in that chapter, concerning the amount paid for each claim, there were 5590 observations that met our criteria. Also, both private cars and fleet vehicles have been considered in this sample and the available a priori rating variables we employ are the Bonus Malus category, the horsepower of the car and gender of the driver. Recall that this Bonus-Malus System has 20 classes and the transition rules are described as follows: Each claim free year is rewarded by one class discount and each claim in given year is penalized by one class. The Bonus-Malus category consists of five categories of neighboring BM classes : C1 = "drivers who belong to BM classes 1 and 2", C2 = "drivers who belong to BM classes 3-5", C3 = "drivers who belong to BM classes 6-9", C4 = "drivers who belong to BM class 10" and C5 = "drivers who belong to BM classes 11-20". The horsepower of the car consists of eleven categories: C1 = "drivers who had a car with a hp between 0-33", C2 = "drivers who had a car with a hp between 34-44", C3 = "drivers who had a car with a hp between 45-55", C4 = "drivers who had a car with a hp between 56-66", C5 = "drivers who had a car with a hp between 67-74", C6 = "drivers who had a car with a hp between 75-82", C7 = "drivers who had a car with a hp between 83-90", C8 = "drivers who had a car with a hp between 91-99", C9 ="drivers who had a car with a hp between 100-110", C10 = "drivers who had a car with a hp between 111-121" and C11 = "drivers who had a car with a hp between 122-132". Finally, the gender consists of three categories: M = "male", F = "female" and B = "both", since in this case, data for fleet vehicles used by either male or female drivers were also available, i.e. shared use. Nevertheless, as we mentioned previously, gender has recently been ruled out by the European Court as a rating factor. The expected claim severity is 328 euros and the variance is 41231.59. The Pareto distribution and GAMLSS were fitted in turn on the costs of claims. For the Pareto GAMLSS we selected the parametric linear formulation considering a linear model in the explanatory variables only for the log of their mean parameter in order to derive an optimal BMS by updating the posterior mean. For this purpose we use the RS algorithm in order to maximize the likelihood function employing the methods described in Rigby and Stasinopoulos (2001, 2005, 2009).

4.4.1 Modelling Results

The modelling results for the claim frequency models were presented in Chapter 3. This subsection describes the modelling results for the Pareto distribution and the Pareto GAMLSS. The heavy-tailed Pareto distribution is often a good candidate for modelling the claim severity, as has been shown to be the case in many studies (see Klugman, Panjer & Willmot, 2004).

Firstly, we incorporated the Pareto distribution, with pdf given by Eq. (4.3), into the GAMLSS package in R and we obtained the maximum likelihood estimates of the location and shape parameters m and s respectively. It is $\hat{m} = 28001.$ and $\hat{s} = 85.798$, thus $E(x) = \frac{\hat{m}}{\hat{s}-1} = \frac{28001}{85.798-1} = 330.21$ The Akaike information criterion (AIC) and the Schwartz Bayesian criterion (SBC) of the estimated model are equal to 75989.3 and 76002.5 respectively.

Secondly, we examined the applicability of the Pareto regression model, given by Eq. (4.15), to the claim severity data in order to use all available information in the estimation of claim severity distribution. Note that the GAMLSS package allows us to find the maximum likelihood estimators of the parameters of the regression model where the distribution of the response variable, i.e. the costs of claims, is the Pareto2o (m', s') distribution, with pdf given by $f(x) = s'm'^{s'}(x+m')^{-s'-1}$ (see Eq. (4.3)). The Pareto(m, s) response distribution with pdf given by Eq. (4.15) can be derived from a reparameterization of the pdf of the Pareto2o (m', s') distribution with $s' = s$ and $m' = (s' - 1)m$. Thus $\hat{s} = \hat{s}'$ and $\hat{m} = \frac{\hat{m}'}{\hat{s}'-1}$. Variable selection techniques were applied in order to find the explanatory variables that are considered as better predictors. For this purpose we used the function step.GAIC, within the GAMLSS package. The model recorded in Table 4.1 is the best fitted model with Global Deviance equal to 75746.86, AIC equal to 75782.86 and SBC equal to 75902.18.

Table 4.1: Results of the Fitted Pareto GAMLSS

Variable	Estimate	Std Error	t-value/Wald	P-value
Intercept	8.6786	0.1729	50.1922	0.0000
Bonus-Malus				
Category 1	0	0	-	-
Category 2	-0.0216	0.0359	-0.6032	0.5464
Category 3	0.1148	0.0494	2.5554	0.0106
Category 4	-0.7272	0.0679	-10.7184	0.0000
Category 5	0.4113	0.2370	1.7359	0.0826
Horsepower				
Category 1	0	0	-	-
Category 2	-0.2105	0.1480	-1.4225	0.1549
Category 3	-0.1975	0.1430	-1.3811	0.1673
Category 4	-0.0127	0.1392	-0.0909	0.9276
Category 5	0.0057	0.1367	0.0418	0.9667
Category 6	0.1392	0.1377	1.0116	0.3118
Category 7	0.1564	0.1522	1.0283	0.3039
Category 8	0.3354	0.1476	2.2728	0.0231
Category 9	0.4405	0.1449	3.0396	0.0024
Category 10	0.6463	0.1538	4.2013	0.0000
Category 11	1.0669	0.1885	5.6610	0.0000
Gender				
Both	0	0	-	-
Male	-0.0807	0.0602	-1.3421	0.1796
Female	-0.0259	0.0619	-0.4180	0.6760
s'	2.9620	0.0945	31.330	0.0000

From Table 4.1, we observe that the Bonus-Malus categories 2 and 4 have a negative effect on the expected claim severity, y_i^j, while the Bonus-Malus categories 3 and 5 have a positive effect on y_i^j. The horsepower categories 2 up to 4 have a negative effect on mean claim severity.

This could be because policyholders who own cars with a very low horsepower are more likely to have cheaper and slower cars which correspondingly cause less serious damage in an accident. However, horsepower categories 2 up to 7 are not so important for the mean claim severity. Finally, horsepower categories 8 up to 11 have a positive effect on the mean claim severity. The gender has a negative effect on the mean claim severity. Bonus-Malus category 1, horsepower category 1 and fleet vehicles used by both male and female drivers in turn are the reference categories. The positive values of the coefficients indicate higher risk compared to the reference class, whereas negative values demonstrate lower risk than the reference class.

4.4.2 Optimal BMS Based on the a Posteriori Criteria

The premiums determined by the optimal BMS based on the a posteriori frequency component were reported in Chapter 3. In this subsection we consider the premiums determined by the optimal BMS based on the a posteriori severity component and the premiums determined by the optimal BMS based both on the a posteriori frequency and severity component. In the following examples, the premiums will be divided by the premium when $t = 0$, since we are not so much interested in the absolute premium values as in the differences between various classes. We will present the results so that the premium for a new policyholder is 100.

We consider first the optimal BMS based on the a posteriori severity component. In what follows we calculate the premiums that must be paid by a policyholder who is observed for the first year of their presence in the portfolio, has one accident and the claim amount of their accident ranges from 150 to 7000 euros. The optimal BMS resulting from the Pareto distribution will be defined by Eq. (4.5) and is presented in Table 4.2. We observe that for claim sizes up to 325 euros the policyholder receives a bonus[2]. Also, as expected, the higher the claim size the higher the premium. For example, for one claim size of 350 in the first year the premium increases from 100 to 100.07 and for one claim of size 7000 in the first year the premium increases from 100 to 123.54

[2]Note that the mean claim size is 328 euros.

Table 4.2: Optimal BMS Based on the A Posteriori Severity Component, One Claim in the First Year of Observation

Claim Size	Optimal BMS Pareto	Claim Size	Optimal BMS Pareto	Claim Size	Optimal BMS Pareto
150	99.36393	500	100.59931	2000	105.89380
175	99.45217	600	100.95227	2500	107.65863
200	99.54041	700	101.30524	3000	109.42346
225	99.62865	800	101.65821	3500	111.18829
250	99.71689	900	102.01117	4000	112.95312
275	99.80513	1000	102.36414	4500	114.71795
300	99.89337	1100	102.71710	5000	116.48278
325	99.98162	1200	103.07007	5500	118.24761
350	100.06986	1300	103.42304	6000	120.01244
375	100.15810	1400	103.77600	6500	121.77727
400	100.24634	1500	104.12897	7000	123.54210

Let us now consider the optimal BMS based both on the a posteriori frequency and severity component. In Tables 3.6, 3.7 and 3.8 of Chapter 3 we reported the premiums resulting from the Negative Binomial, Poisson-Inverse Gaussian and Sichel distributions respectively. We assume that the number of claims of each policyholder is independent from the severity of each claim. Subsequently, we are able to compute the premiums determined by the optimal BMS based both on the frequency and the severity component. As we said, the premiums of this optimal BMS will be given from the product of the BMS based on the frequency component and of the BMS based on the severity component. Consider again that a policyholder is observed for the first year of their presence in the portfolio, has one accident and the claim amount of their accident ranges from 150 to 7000 euros. In Table 4.3 we report the scaled premiums for the case of the Negative Binomial-Pareto (NB-PA), Poisson Inverse Gaussian-Pareto (PIG-PA) and Sichel-Pareto (SI-PA) models respectively. We observe that these three systems do not differ much. For example, a policyholder who had one claim with claim size 250 will have to pay a malus of 69.65%, 56.17% and 58.34% of the basic premium in the case of the NB-PA, PIG-PA and SI-PA models respectively, while a policyholder who had one claim with claim size 5000 will have to pay a malus of 98.18%, 82.43% and 84.96% of the basic premium in the case of the NB-PA, PIG-PA and SI-PA models respectively. It is obvious that these optimal BMSs allow the discrimination of the premiums with respect to the frequency and the severity of the claims.

Table 4.3: Optimal BMS Based on the Alternative Distributions for Assessing Claim Frequency Presented in Chapter 3 and the Pareto Distribution for Assessing Claim Severity, One Claim in the First Year of Observation

Claim Size	Optimal BMS NB-PA	Optimal BMS PIG-PA	Optimal BMS SI-PA
150	169.0578	155.6238	157.7800
175	169.2079	155.7620	157.9201
200	169.3581	155.9002	158.0602
225	169.5082	156.0384	158.2003
250	169.6583	156.1766	158.3405
275	169.8085	156.3148	158.4806
300	169.9586	156.4530	158.6207
325	170.1087	156.5912	158.7608
350	170.2589	156.7294	158.9009
375	170.4090	156.8676	159.0410
400	170.5591	157.0058	159.1812
500	171.1597	157.5586	159.7416
600	171.7602	158.1115	160.3021
700	172.3607	158.6643	160.8626
800	172.9613	159.2171	161.4231
900	173.5618	159.7699	161.9835
1000	174.1623	160.3227	162.5440
1100	174.7629	160.8755	163.1045
1200	175.3634	161.4283	163.6650
1300	175.9640	161.9812	164.2254
1400	176.5645	162.5340	164.7859
1500	177.1650	163.0868	165.3464
2000	180.1677	165.8509	168.1488
2500	183.1704	168.6149	170.9511
3000	186.1731	171.3790	173.7535
3500	189.1758	174.1431	176.5559
4000	192.1784	176.9072	179.3583
4500	195.1811	179.6713	182.1606
5000	198.1838	182.4353	184.9630
5500	201.1865	185.1994	187.7654
6000	204.1892	187.9635	190.5678
6500	207.1919	190.7276	193.3701
7000	210.1945	193.4916	196.1725

4.4.3 Optimal BMS Based Both on the a Priori and the a Posteriori Criteria

The premiums determined by the generalized BMS with a frequency component were reported in Chapter 3. In this subsection we consider the premiums determined by a generalized optimal BMS with a severity component and the premiums determined by a generalized BMS based both on the frequency and severity component when both the a priori and the a posteriori rating variables are used. The premiums are divided again by the premium when $t = 0$, in order to observe the percentage change in the premiums after one or more claims.

We consider first the generalized optimal BMS based on the severity component. In the following example we consider a group of policyholders who share the following common characteristics. The policyholder i is a woman, who has a car with horsepower between 0-33 and her Bonus-Malus class varies over time, starting from BM class 1[3]. Implementing the Pareto GAMLSS (Eq. (4.15)) we found that $\hat{s} = \exp(2.9620) = 19.337$. As we have already mentioned, the mean (or location) parameter of this model is given by $E(X_{i,k}^{j}|d_i^j) = m_i^j = \exp\left(d_i^j \gamma^j\right)$, where $d^j\left(d_{i,1}^j, ..., d_{i,h}^j\right)$ is the $1 \times h$ vector of h individual's characteristics, which represent different a priori rating variables and γ^j is the vector of the coefficients. The estimation of the vector γ^j, and therefore of the mean parameter for the Pareto model, led to the following results presented in Table 4.4.

Table 4.4: Women, Horse Power 0-33

Bonus-Malus Category	Pareto $\hat{m}_i^j$
1	312.3426
2	305.8518
3	350.0586
4	150.9729
5	471.1142

Based on the above estimates for this group of individuals we are now able to derive the generalized optimal BMS for the severity component according to the Eq. (4.19). Consider that the policyholder i is observed for the first year of her presence in the portfolio, has one accident and the claim amount of her accident ranges from 150 to 7000 euros. The premiums resulting from this system are displayed in Table 4.5. From Table 4.5 we observe that for claim sizes up to 400 euros the policyholder i receives a bonus due to the fact that the cost of the claim that the insurance company has to pay is not significant. Also, we observe that the higher the claim size the higher the premium, revealing the appropriateness of the modelling technique. For example, for one claim size of 900 in the first year her premium increases from 100 to 107.45 while for one claim size of 7000 her premium increases from 100 to 206.35.

[3]Recall that BM class 1 corresponds to BM category 1 according to the grouping of the levels of the BM class explanatory variable.

Table 4.5: Women, Horse Power 0-33, Varying Bonus-Malus Class, One Claim in the First Year of Observation

Claim Size	Optimal BMS Pareto	Claim Size	Optimal BMS Pareto	Claim Size	Optimal BMS Pareto
150	95.28981	500	100.96442	2000	125.28420
175	95.69514	600	102.58574	2500	133.39079
200	96.10047	700	104.20706	3000	141.49739
225	96.50579	800	105.82838	3500	149.60398
250	96.91112	900	107.44970	4000	157.71057
275	97.31645	1000	109.07101	4500	165.81717
300	97.72178	1100	110.69233	5000	173.92376
325	98.12711	1200	112.31365	5500	182.03035
350	98.53244	1300	113.93497	6000	190.13695
375	98.93777	1400	115.55629	6500	198.24354
400	99.34310	1500	117.17761	7000	206.35013

Consider now another group of policyholders who share the following common characteristics. The policyholder i is now a man, who has a car with horsepower between 0-33 and his Bonus-Malus class varies over time, starting from BM class 1. The estimation of the vector γ^j and therefore of the mean parameter for the Pareto model led to the following results depicted in Table 4.6.

Table 4.6: Men, Horse Power 0-33

Bonus-Malus Category	Pareto $\hat{m}_i^j$
1	295.9234
2	289.7738
3	331.6568
4	143.0366
5	446.3487

Based on the above estimates for this new group of policyholders we can derive the generalized optimal BMS according to the Eq. (4.19). Consider again that the policyholder i is observed for the first year of his presence in the portfolio, has one accident and the claim amount of his accident ranges from 150 to 1000 euros. The premiums resulting from this system are reported in Table 4.7. We observe that for a claim with size up to 400 euros the policyholder i receives a bonus. However, for claim sizes higher than 400 euros we observe that the he has to pay a malus. For instance, if the policyholder has one claim size of 800 in the first year then he faces a malus of 6.54% of the basic premium and if he has one claim size of 7000 in the first year then he faces a malus of 112.64% of the basic premium.

Table 4.7: Men, Horse Power 0-33, Varying Bonus-Malus Class, One Claim in the First Year of Observation

Claim Size	Optimal BMS Pareto	Claim Size	Optimal BMS Pareto	Claim Size	Optimal BMS Pareto
150	95.42474	500	101.41421	2000	127.08336
175	95.85256	600	103.12549	2500	135.63975
200	96.28038	700	104.83677	3000	144.19613
225	96.70820	800	106.54804	3500	152.75252
250	97.13602	900	108.25932	4000	161.30890
275	97.56384	1000	109.97060	4500	169.86529
300	97.99166	1100	111.68187	5000	178.42167
325	98.41948	1200	113.39315	5500	186.97805
350	98.84730	1300	115.10443	6000	195.53444
375	99.27512	1400	116.81570	6500	204.09082
400	99.70294	1500	118.52698	7000	212.64721

Note that from Tables 4.5 and 4.7 we observe that the premiums that should be paid by a woman, who has a car with horsepower between 0-33 and whose Bonus-Malus class varies over time do not differ much from those that should be paid by a man who shares the same characteristics. Note also that other combinations of a priori characteristics could be used for multiple claims and different claim severity histories. It is interesting to compare this BMS with the one obtained when only the a posteriori classification criteria are used. Using this system we saw from Table 4.2 that a policyholder with one accident with claim size of 3000 euros in one year has to pay a malus of 9.42% of the basic premium. Using the generalized optimal BMS with a severity component based both on the a priori and the a posteriori classification criteria, a woman who has a car with horsepower between 0-33 and whose Bonus-Malus category varies over time for one accident of claim size 3000 euros in one year will have to pay a malus of 41.49% of the basic premium, while a man, who shares the same characteristics for one accident of claim size 3000 euros in one year will have to pay a malus of 44.19% of the basic premium. This system is more fair since it considers all the important a priori and a posteriori information for each policyholder for the severity component in order to estimate their risk of having an accident.

Let us now consider the premiums determined by the generalized BMS with a frequency and a severity component based both on the a priori and the a posteriori classification criteria. The generalized BMS with a frequency component was presented in Chapter 3. This system was derived by adopting the parametric linear formulation of the Negative Binomial Type I (NBI), Poisson-Inverse Gaussian (PIG) and Sichel GAMLSS and by allowing only the mean parameter of these models to be modelled as a function of the significant explanatory variables for the number of claims. We assume that the number of claims of each policyholder is independent of the severity of each claim. Subsequently, we are able to derive the generalized BMS with a frequency and a severity component by using the Pareto GAMLSS to integrate accident severity into the BMSs presented in Chapter 3. In this generalized system, the premiums will be given from the product of the generalized BMS based on the frequency component, and of the generalized BMS based on the severity component. Let us see an example in order to understand better how this BMS works. Consider a group of policyholders who share the following common characteristics. The policyholder i is a woman who has a car with horsepower between 0-33 and her Bonus-Malus class varies over time, starting from BM class 1. The generalized premiums for the frequency component were presented in Table 3.10 of Chapter 3 for the case of the NBI, PIG and Sichel GAMLSS respectively. We assume that the policyholder i is observed for the first year of her presence in the portfolio, has one accident and the claim amount of her accident ranges from 150 to 7000 euros. In Table 4.8 we report the generalized premiums based on the frequency and the severity component for the case of the Negative Binomial Type I-Pareto (NBI-PA), Poisson Inverse Gaussian-Pareto (PIG-PA) and Sichel-Pareto (SI-PA) models respectively. From Table 4.8 we observe that these three systems do not differ much. For example if the policyholder i had one claim with claim size 1500 then she will have to pay a malus of 227.59%, 190.51% and 248.26% of the basic premium in the case of the NBI-PA, PIG-PA and SI-PA models respectively, while if she had one claim with claim size 5000 then she will have to pay a malus of 386.24%, 331.19% and 416.92% of the basic premium in the case of the NBI-PA, PIG-PA and SI-PA models respectively. These optimal BMSs allow the discrimination of the premiums with respect to the a priori and the a posteriori frequency and severity component.

Table 4.8: Women, Horse Power 0-33, Varying Bonus-Malus Class, One Claim in the First Year of Observation

Claim Size	Optimal BMS NBI-PA	Optimal BMS PIG-PA	Optimal BMS SI-PA
150	266.4017	236.2425	283.2108
175	267.5349	237.2474	284.4155
200	268.6681	238.2523	285.6202
225	269.8013	239.2572	286.8249
250	270.9344	240.2621	288.0296
275	272.0676	241.2670	289.2342
300	273.2008	242.2718	290.4389
325	274.3340	243.2767	291.6436
350	275.4672	244.2816	292.8483
375	276.6003	245.2865	294.0530
400	277.7335	246.2914	295.2576
500	282.2662	250.3110	300.0764
600	286.7990	254.3306	304.8951
700	291.3317	258.3501	309.7138
800	295.8644	262.3697	314.5325
900	300.3971	266.3893	319.3512
1000	304.9298	270.4089	324.1700
1100	309.4626	274.4284	328.9887
1200	313.9953	278.4480	333.8074
1300	318.5280	282.4676	338.6261
1400	323.0607	286.4872	343.4448
1500	327.5934	290.5067	348.2636
2000	350.2570	310.6046	372.3572
2500	372.9206	330.7025	396.4508
3000	395.5842	350.8003	420.5444
3500	418.2478	370.8982	444.6380
4000	440.9114	390.9961	468.7316
4500	463.5751	411.0939	492.8252
5000	486.2387	431.1918	516.9188
5500	508.9023	451.2896	541.0124
6000	531.5659	471.3875	565.1060
6500	554.2295	491.4854	589.1996
7000	576.8931	511.5832	613.2932

Consider now another group of policyholders who share the following common characteristics. The policyholder i is now a man who has a car with horsepower between 0-33 and his Bonus-Malus class varies over time, starting from BM class 1. The generalized premiums are displayed in Table 4.9 for the case of the Negative Binomial Type I-Pareto, Poisson Inverse

Gaussian-Pareto and Sichel-Pareto models respectively.

Table 4.9: Men, Horse Power 0-33, Varying Bonus-Malus Class, One Claim in the First Year of Observation

Claim Size	Optimal BMS NBI-PA	Optimal BMS PIG-PA	Optimal BMS SI-PA
150	268.4203	239.1153	287.9060
175	269.6237	240.1874	289.1968
200	270.8271	241.2594	290.4875
225	272.0305	242.3314	291.7783
250	273.2339	243.4034	293.0691
275	274.4373	244.4755	294.3599
300	275.6407	245.5475	295.6506
325	276.8441	246.6195	296.9414
350	278.0476	247.6916	298.2322
375	279.2510	248.7636	299.5230
400	280.4544	249.8356	300.8137
500	285.2680	254.1237	305.9768
600	290.0817	258.4119	311.1399
700	294.8953	262.7000	316.3030
800	299.7090	266.9881	321.4661
900	304.5226	271.2762	326.6292
1000	309.3363	275.5643	331.7923
1100	314.1499	279.8524	336.9554
1200	318.9636	284.1406	342.1185
1300	323.7772	288.4287	347.2816
1400	328.5909	292.7168	352.4447
1500	333.4045	297.0049	357.6078
2000	357.4728	318.4455	383.4232
2500	381.5410	339.8861	409.2387
3000	405.6093	361.3267	435.0542
3500	429.6776	382.7673	460.8696
4000	453.7458	404.2078	486.6851
4500	477.8141	425.6484	512.5006
5000	501.8823	447.0890	538.3160
5500	525.9506	468.5296	564.1315
6000	550.0188	489.9702	589.9470
6500	574.0871	511.4108	615.7624
7000	598.1553	532.8514	641.5779

From Tables 4.8 and 4.9 we observe that the premiums that should be paid by a woman, who has a car with horsepower between 0-33 and her Bonus-Malus class varies over time do not differ much from those that should be paid by a man who shares common characteristics. For example, if the policyholder i is a woman who had one claim with claim size 4500 euros then she will have to pay a malus of 363.57%, 311.09% and 392.82% of the basic premium in the case of the NBI-PA, PIG-PA and SI-PA models respectively, while if the policyholder is a man who had one claim with claim size 4500 euros then he will have to pay a malus of 377.81%, 325.65% and 412.50% of the basic premium in the case of the NBI-PA, PIG-PA and SI-PA models respectively. Furthermore, it is interesting to compare these BMSs with those obtained when only the a posteriori classification criteria are used. Using these BMS we saw from Table 4.3 that a policyholder with one accident with claim size of 6500 euros in one year has to pay a malus of 107.19%, 90.73% and 93.37% of the basic premium in the case of the NB-PA, PIG-PA and SI-PA models respectively. Using the generalized optimal BMS with a frequency and a severity component, a woman who has a car with horsepower between 0-33 and her Bonus-Malus class varies over time for one accident of claim size 6500 euros in one year will have to pay a malus of 454.23%, 391.48% and 489.20% of the basic premium in the case of the NBI-PA, PIG-PA and SI-PA models respectively, while a man, who shares common characteristics for one accident of claim size 6500 euros in one year will have to pay a malus of 474.08%, 411.41% and 515.76% of the basic premium in the case of the NBI-PA, PIG-PA and SI-PA models respectively. As mentioned in Frangos and Vrontos (2001), these systems are more fair since they consider all the important a priori and a posteriori information for each policyholder, both for the frequency and the severity components, in order to estimate their risk of having an accident and thus they permit the differentiation of the premiums for various number of claims and for various claim sizes based on the expected claim frequency and expected claim severity of each policyholder as these are estimated both from the a priori and the a posteriori classification criteria.

In the following chapter we give more emphasis on the analysis of the claim severity component using finite mixtures of distributions and regression, as these methods have not been studied in the BMS literature, as opposed to the claim frequency component for which many alternative models have been proposed, such as zero-inflated models, Hurdle models, and other count data models.

Chapter 5

The Design of Optimal Bonus-Malus Systems Using Finite Mixture Models for Assessing Claim Counts and Losses

5.1 Introduction

In Chapter 5 we put focus on both the analysis of the claim frequency and severity components of an optimal BMS using finite mixtures of distributions and regression models. Finite mixture models are a popular statistical modelling technique, given that they constitute a flexible and easily extensible model class for approximating general distribution functions in a semi-parametric way and accounting for unobserved heterogeneity. Finite mixture models have been widely applied in many areas, such as biology, biometrics, genetics, medicine and marketing. For a comprehensive list of the applications and numerical derivations of finite mixture models, readers are referred to Mclachlan and Peel (2000). However, these models have not been extensively studied in BMS literature. Specifically, only Lemaire (1995) considered the good risk/bad risk model, employing a finite Poisson mixture distribution with two components.

Our first contribution is the development of an optimal BMS that takes into account the number of claims of each policyholder and the exact size of loss that these claims incurred, using various finite mixtures of distributions. For the frequency component we assume that the number of claims is distributed according to a finite Poisson, Negative Binomial and Delaporte mixture, and for the severity component we consider that the losses are distributed according to a finite Exponential, Gamma, Weibull and GB2 mixture. In this way we expand the setup that Lemaire (1995) used to design an optimal BMS based on the number of claims. Applying Bayes theorem we derive the posterior probability of the policyholder's classes of risk. Furthermore, we extend the setup of Frangos and Vrontos (2001) for Negative Binomial and Pareto mixtures and derive the posterior distribution of both the mean claim frequency and the mean claim size, given the information we have about the claim frequency history and the claim size history for each policyholder for the time period he is in the portfolio. Our third contribution is the development of a generalized BMS that integrates the a priori and the a posteriori information on a individual basis extending the framework developed by Frangos and Vrontos (2001). This is achieved by using finite mixtures of regression models. In this generalized BMS, the premium is a function of the years that the policyholder is in the portfolio, the number of accidents, the size of loss that each of these accidents incurred, and the significant a priori rating variables for the number of accidents and for their severity.

The layout of this chapter is as follows. Section 5.2 includes a short review of the basic concepts of finite mixture models. Section 5.3 describes the design of optimal BMS with a frequency and a severity component based on the a posteriori criteria. The design presented in Section 5.4 is based on both the a posteriori and the a priori criteria and Section 5.5 contains an application to the data set concerning car insurance claims which was presented in the previous chapters.

5.2 Finite Mixture Models

In what follows we present a short summary of the main characteristics of finite mixture models. Mclachlan and Peel (2000) provide a detailed introduction to finite mixture models. Assume that a random variable Y comes from component z, having probability density function $f_z(y)$, with probability π_z, for $z = 1, .., n$, then the marginal density of Y is given by

$$f_Y(y) = \sum_{z=1}^{n} \pi_z f_z(y), \tag{5.1}$$

where $0 \leq \pi_z \leq 1$ is the prior (or mixing) probability of component z and where $\sum_{z=1}^{n} \pi_z = 1$, $z = 1, .., n$.

The simplest finite mixture models are finite mixtures of distributions which are used for model-based clustering. In this case the model is given by a convex combination of a finite number of different distributions where each of the distributions is referred to as a component.

An extension is to estimate finite mixture models assuming that the n components $f_z(y)$ can be represented by GAMLSS models. In this setup the probability density function $f_z(y)$ for component z depends on parameters $\boldsymbol{\theta}_z$, each of which can be a function to the explanatory variables $\mathbf{x}_z$, i.e. $f_z(y) = f_z(y|\boldsymbol{\theta}_z, \mathbf{x}_z)$ (see Chapter 2)[1]. Similarly, $f_Y(y)$ depends on parameters $\boldsymbol{\psi} = (\boldsymbol{\theta}, \boldsymbol{\pi})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ and $\boldsymbol{\pi}^T = (\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_n)$ and explanatory variables $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, i.e. $f_Y(y) = f_Y(y|\boldsymbol{\psi}, \mathbf{x})$, and

$$f_Y(y|\boldsymbol{\psi}, \mathbf{x}) = \sum_{z=1}^{n} \psi_z f_z(y|\boldsymbol{\theta}_z, \mathbf{x}_z). \tag{5.2}$$

Note that the prior probabilities may also depend on explanatory variables $\mathbf{x}_0$ and parameters through a multinomial logistic model (for more information, refer to Rigby and Stasinopoulos, 2009). In this thesis, we assume that all the component distributions, $f_z(y)$, arise from the same parametric distribution family, the prior probabilities are included in the linear predictor

[1]Reacall that GAMLSS models have up to four distributional parameters μ, σ, ν and τ.

for only the mean parameters of $f_z(y)$. Using this formulation, the heterogeneity in the data can be accounted for in two ways. Firstly, the population heterogeneity is accounted for by choosing a finite number of unobserved latent components, each of which may be regarded as a sub-population. This is a discrete representation of heterogeneity since the mean (of claim frequency or severity) is approximated by a finite number of support points. Secondly, depending on the choice of the $f_z(y)$ distribution, heterogeneity can also accommodated within each component by including the explanatory variables in the mean function.

For an observed independent random sample $(y_1, ..., y_{\mathrm{n}})$ from finite mixture model (5.2), the log-likelihood function l is given by

$$l = l(\boldsymbol{\psi}, \mathbf{y}) = \log(L(\boldsymbol{\psi}, \mathbf{y})) = \sum_{i=1}^{\mathrm{n}} \log\left[\sum_{z=1}^{n} \pi_z f_z(y_i)\right], \tag{5.3}$$

where L denotes the likelihood function and where $f_z(y_i) = f_z(y_i|\boldsymbol{\theta}_z, \mathbf{x}_{z,i})$. The log-likelihood function l can be maximized with respect to $\boldsymbol{\psi}$, i.e. with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$, using the EM algorithm (for more details see Rigby and Stasinopoulos, 2009).

5.3 The Design of an Optimal BMS Based on the a Posteriori Criteria

We assume that the number of claims of each policyholder is independent of the severity of each claim in order to deal with the frequency and the severity component separately.

5.3.1 Frequency Component Updating the Posterior Probability

The framework we consider is a generalization of the good risk/bad risk model proposed by Lemaire (1995). Specifically, the portfolio is considered to be heterogeneous, consisting of n categories of policyholders classified according to their driving skills. In this respect, we have fractions of drivers π_z where the risk that each policyholder of category z is imposing on the pool, with respect to their claim frequency, is denoted by λ_z, $z = 1, \cdots, n$. The distribution of the number of claims k in each category is denoted by $P_z(k)$. Thus, the structure function is an n-point discrete distribution and the distribution of the unconditional number of claims, denoted by $P(k)$, is given by

$$P(k) = \sum_{z=1}^{n} \pi_z P_z(k),$$

$k = 0, 1, ..., \pi_z > 0$, for $z = 1, \cdots, n$ and $\sum\limits_{z=1}^{n} \pi_z = 1$. The expected value of the number of claims is equal to $E(k) = \sum\limits_{z=1}^{n} \pi_z \lambda_z$.

In what follows, we consider a policyholder with claims history $k_1, ..., k_t$,where k_j is the number of claims that the policyholder had in year j, $j = 1, ..., t$. Let us denote with $K = \sum_{j=1}^{t} k_j$ the total number of claims that the policyholder had in t years and with R_l the risk, imposed on the insurance company, associated with the lth category of policyholders. Moreover, the posterior probability of the policyholder belonging to the lth category is given by $\pi_l(k_1, \cdots, k_t)$. Applying the Bayes theorem, the posterior probability of the policyholder belonging to the lth category is given by the following equation

$$\pi_l(k_1, ..., k_t) = \frac{P(k_1, ..., k_t|R_l)\pi_l}{\sum_{z=1}^{n} P(k_1, ..., k_t|R_z)\pi_z}$$

In this way, we update the posterior probability of belonging in category l given the information we have for the claim frequency history of the policyholder. Under a quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}$ for a policyholder with claim history $k_1, ..., k_t$ is the mean of the posterior structure function given by

$$\hat{\lambda}_{t+1}(k_1, ..., k_t) = \sum_{z=1}^{n} \pi_z(k_1, \cdots, k_t)\lambda_z. \tag{5.4}$$

The setup described is applied to finite Poisson, Negative Binomial and Delaporte mixture distributions. Extensions to other finite mixture distributions can be obtained in a similar way and is straightforward.

Finite Poisson Mixture

Poisson mixtures are counterparts to the simple Poisson distribution for the description of heterogeneous populations. Of special interest are populations consisting of a finite number of homogeneous sub-populations. In these cases the probability distribution of the population can be regarded as a finite mixture of Poisson distributions. The finite Poisson mixture is a generalization of the good risk/bad risk model proposed by Lemaire (1995). We have n fractions of drivers π_z with Poisson parameter λ_z and the distribution of the unconditional number of claims, k, is the following

$$P(k) = \sum_{z=1}^{n} \pi_z \frac{e^{-\lambda_z}\lambda_z^k}{k!}, k = 0, 1, 2, 3, ...\pi_z, \lambda_z > 0 \text{ for } z = 1, \cdots, n, \ \sum_{z=1}^{n} \pi_z = 1 \tag{5.5}$$

Posterior Probability The posterior probability of the policyholder belonging to the lth category of drivers is given by

$$\pi_l(k_1, ..., k_t) = \frac{\lambda_l^K e^{-t\lambda_l}\pi_l}{\sum_{z=1}^{n} \lambda_z^K e^{-t\lambda_z}\pi_z} \tag{5.6}$$

Proof. By means of the Bayes theorem, we have

$$\pi_l\left(k_1,...,k_t\right) = P\left(R_l|k_1,...,k_t\right) = \frac{P(k_1,...,k_t|R_l)P(R_l)}{\sum_{z=1}^{n} P(k_1,...,k_t|R_z)P(R_z)}$$

$$= \frac{\frac{e^{-\lambda_l}\cdot\lambda_l^{k_1}}{k_1!}\cdot ...\cdot\frac{e^{-\lambda_l}\cdot\lambda_l^{k_t}}{k_t!}\cdot\pi_l}{\frac{e^{-\lambda_1}\cdot\lambda_1^{k_1}}{k_1!}\cdot ...\cdot\frac{e^{-\lambda_1}\cdot\lambda_1^{k_t}}{k_t!}\cdot\pi_1+...+\frac{e^{-\lambda_l}\cdot\lambda_l^{k_1}}{k_1!}\cdot ...\cdot\frac{e^{-\lambda_l}\cdot\lambda_l^{k_t}}{k_t!}\cdot\pi_l+...+\frac{e^{-\lambda_n}\cdot\lambda_n^{k_1}}{k_1!}\cdot ...\cdot\frac{e^{-\lambda_n}\cdot\lambda_n^{k_t}}{k_t!}\cdot\pi_n}$$

$$= \frac{\frac{\lambda_l^K\cdot e^{-t\cdot\lambda_l}}{\prod_{j=1}^{t}k_j!}\cdot\pi_l}{\frac{\lambda_1^K\cdot e^{-t\cdot\lambda_1}}{\prod_{j=1}^{t}k_j!}\cdot\pi_1+...+\frac{\lambda_{l-1}^K\cdot e^{-t\cdot\lambda_{l-1}}}{\prod_{j=1}^{t}k_j!}\cdot\pi_{l-1}+\frac{\lambda_l^K\cdot e^{-t\cdot\lambda_l}}{\prod_{j=1}^{t}k_j!}\cdot\pi_l+\frac{\lambda_{l+1}^K\cdot e^{-t\cdot\lambda_{l+1}}}{\prod_{j=1}^{t}k_j!}\cdot\pi_{l+1}+...+\frac{\lambda_n^K\cdot e^{-t\cdot\lambda_n}}{\prod_{j=1}^{t}k_j!}\cdot\pi_n}$$

$$= \frac{\lambda_l^K e^{-t\lambda_l}\pi_l}{\sum_{z=1}^{n}\lambda_z^K e^{-t\lambda_z}\pi_z}.$$

Optimal Choice of $\hat{\boldsymbol{\lambda}}_{\mathbf{i}}^{\mathbf{t+1}}$ Under a quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}\left(k_1,...,k_t\right)$ for a policyholder with claim history $k_1,\cdots,k_t$ is the mean of the posterior structure function and is given by substituting (5.6) into (5.4).

Finite Negative Binomial Type I Mixture

For the case of the finite Negative Binomial mixture, we assume that the portfolio consists of fractions of drivers π_z where the number of claims $k|R_z$ follows a Negative Binomial(α_z,λ_z)[2]. Thus, the structure function is an n-point discrete distribution and the number of claims k, given the parameters α_z,λ_z, is distributed according to

$$P\left(k\right) = \sum_{z=1}^{n}\pi_z\binom{k+\frac{1}{\alpha_z}-1}{k}\left(\frac{1}{1+\alpha_z\lambda_z}\right)^{\frac{1}{\alpha_z}}\left(\frac{\alpha_z\lambda_z}{1+\alpha_z\lambda_z}\right)^{k}, \tag{5.7}$$

for $k=0,1,...$ where $\alpha_z,\lambda_z>0$ for $z=1,...,n$ and $\sum_{z=1}^{n}\pi_z=1$.

[2]We use the parameterization of Negative Binomial Type I given by Rigby and Stasinopoulos (2009) and Johnson et al (2005).

Posterior Probability The lth category posterior probability is equal to

$$\pi_l(k_1,...,k_t) = \frac{\prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_l}-1}{k_j} \left(\frac{1}{1+\alpha_l\lambda_l}\right)^{\frac{t}{\alpha_l}} \left(\frac{\alpha_l\lambda_l}{1+\alpha_l\lambda_l}\right)^{K} \pi_l}{\sum_{z=1}^{n} \prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_z}-1}{k_j} \left(\frac{1}{1+\alpha_z\lambda_z}\right)^{\frac{t}{\alpha_z}} \left(\frac{\alpha_z\lambda_z}{1+\alpha_z\lambda_z}\right)^{K} \pi_z} \tag{5.8}$$

Proof. Applying the Bayes theorem, we have

$$\pi_l(k_1,...,k_t) = P(R_l|k_1,...,k_t) = \frac{P(k_1,...,k_t|R_l)P(R_l)}{\sum_{z=1}^{n} P(k_1,...,k_t|R_z)P(R_z)}$$

$$= \frac{\prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_l}-1}{k_j}\left(\frac{1}{1+\alpha_l\lambda_l}\right)^{\frac{1}{\alpha_l}} \left(\frac{\alpha_l\lambda_l}{1+\alpha_l\lambda_l}\right)^{k_j} \pi_l}{\sum_{z=1}^{n} \prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_z}-1}{k_j}\left(\frac{1}{1+\alpha_z\lambda_z}\right)^{\frac{1}{\alpha_z}} \left(\frac{\alpha_z\lambda_z}{1+\alpha_z\lambda_z}\right)^{k_j} \pi_z}$$

$$= \frac{\prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_l}-1}{k_j}\left(\frac{1}{1+\alpha_l\lambda_l}\right)^{\frac{1}{\alpha_l}} \left(\frac{\alpha_l\lambda_l}{1+\alpha_l\lambda_l}\right)^{k_j} \pi_l}{\sum_{z=1}^{n} \prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_z}-1}{k_j}\left(\frac{1}{1+\alpha_z\lambda_z}\right)^{\frac{1}{\alpha_z}} \left(\frac{\alpha_z\lambda_z}{1+\alpha_z\lambda_z}\right)^{k_j} \pi_z}$$

$$= \frac{\prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_l}-1}{k_j}\left(\frac{1}{1+\alpha_l\lambda_l}\right)^{\frac{t}{\alpha_l}} \left(\frac{\alpha_l\lambda_l}{1+\alpha_l\lambda_l}\right)^{K} \pi_l}{\sum_{z=1}^{n} \prod_{j=1}^{t} \binom{k_j+\frac{1}{\alpha_z}-1}{k_j}\left(\frac{1}{1+\alpha_z\lambda_z}\right)^{\frac{t}{\alpha_z}} \left(\frac{\alpha_z\lambda_z}{1+\alpha_z\lambda_z}\right)^{K} \pi_z}.$$

Optimal Choice of $\hat{\lambda}_i^{t+1}$ As previously, under a quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}(k_1,...,k_t)$ for a policyholder with claim history $k_1,\cdots,k_t$ is the mean of the posterior structure function and is given by substituting (5.8) into (5.4).

Finite Delaporte Mixture

The Delaporte distribution can be alternatively employed for modeling the number of claims when we deal with overdispersed count data. Our portfolio consists of n categories of policyholders with probability of belonging in each category π_z and parameters of the Delaporte distribution λ_z, σ_z and ν_z, for $z = 1, \cdots, n$. Thus, the structure function is an n-point discrete distribution and the number of claims k is distributed as follows

$$\begin{aligned} P(k) &= \sum_{z=1}^{n} \pi_z \frac{e^{-\lambda_z \nu_z}}{\Gamma\left(\frac{1}{\sigma_z}\right)} \left(1 + \lambda_z \sigma_z (1 - \nu_z)\right)^{-\frac{1}{\sigma_z}} S, \\ S &= \sum_{m=0}^{k} \binom{k}{m} \frac{(\lambda_z)^k (\nu_z)^{k-m}}{k!} \left(\lambda_z + \frac{1}{\sigma_z (1 - \nu_z)}\right)^{-m} \Gamma\left(\frac{1}{\sigma_z} + m\right), \end{aligned} \tag{5.9}$$

for $k = 0, 1, ...$ and $\pi_z, \lambda_z, \sigma_z > 0$, $0 < \nu_z < 1$ and $z = 1, \cdots, n$, $\sum\limits_{z=1}^{n} \pi_z = 1$.

Posterior Probability The posterior probability of the policyholder belonging to the lth category is given by[3]

$$\pi_l(k_1, ..., k_t) = \frac{\left[\frac{e^{-t\lambda_l \nu_l}}{\left[\Gamma\left(\frac{1}{\sigma_l}\right)\right]^t} \left(1 + \lambda_l \sigma_l (1 - \nu_l)\right)^{-\frac{t}{\sigma_l}} \prod\limits_{j=1}^{t} S_{j,l}\right] \pi_l}{\sum\limits_{z=1}^{n} \left[\frac{e^{-t\lambda_z \nu_z}}{\left[\Gamma\left(\frac{1}{\sigma_z}\right)\right]^t} \left(1 + \lambda_z \sigma_z (1 - \nu_z)\right)^{-\frac{t}{\sigma_z}} \prod\limits_{j=1}^{t} S_{j,z}\right] \pi_z} \tag{5.10}$$

where

$$S_{j,l} = \sum_{m=0}^{k_j} \binom{k_j}{m} \frac{(\lambda_l)^{k_j} (\nu_l)^{k_j - m}}{k_j!} \left(\lambda_l + \frac{1}{\sigma_l (1 - \nu_l)}\right)^{-m} \Gamma\left(\frac{1}{\sigma_l} + m\right) \tag{5.11}$$

and $S_{j,z}$ is given by (5.11) for $l = z$.

Proof. Applying the Bayes theorem, we have

$$\pi_l(k_1, ..., k_t) = P(R_l | k_1, ..., k_t) = \frac{P(k_1, ..., k_t | R_l) P(R_l)}{\sum\limits_{z=1}^{n} P(k_1, ..., k_t | R_z) P(R_z)}$$

$$= \frac{\left[\prod\limits_{j=1}^{t} \frac{e^{-\lambda_l \nu_l}}{\Gamma\left(\frac{1}{\sigma_l}\right)} (1 + \lambda_l \sigma_l (1 - \nu_l))^{-\frac{1}{\sigma_l}} \sum\limits_{m=0}^{k_j} \binom{k_j}{m} \frac{(\lambda_l)^{k_j} (\nu_l)^{k_j - m}}{k_j!} \left(\lambda_l + \frac{1}{\sigma_l (1 - \nu_l)}\right)^{-m} \Gamma\left(\frac{1}{\sigma_l} + m\right)\right] \pi_l}{\sum\limits_{z=1}^{n} \left[\prod\limits_{j=1}^{t} \frac{e^{-\lambda_z \nu_z}}{\Gamma\left(\frac{1}{\sigma_z}\right)} (1 + \lambda_z \sigma_z (1 - \nu_z))^{-\frac{1}{\sigma_l}} \sum\limits_{m=0}^{k_j} \binom{k_j}{m} \frac{(\lambda_z)^{k_j} (\nu_z)^{k_j - m}}{k_j!} \left(\lambda_z + \frac{1}{\sigma_z (1 - \nu_z)}\right)^{-m} \Gamma\left(\frac{1}{\sigma_z} + m\right)\right] \pi_z}$$

[3]It should be noted that due to the existence of k_j's in Eq. (5.8) and Eq. (5.10), i.e. for the case of the finite mixture of Negative Binomial and Delaporte distributions respectively, the explicit claim frequency history determines the calculation of the posterior probabilities and thus of premium rates and not just the total number of claims as in the case of the two component Poisson mixture.

$$= \frac{\left[\frac{e^{-t\lambda_l\nu_l}}{\left[\Gamma\left(\frac{1}{\sigma_l}\right)\right]^t}(1+\lambda_l\sigma_l(1-\nu_l))^{-\frac{t}{\sigma_l}}\prod_{j=1}^{t}\sum_{m=0}^{k_j}\binom{k_j}{m}\frac{(\lambda_l)^{k_j}(\nu_l)^{k_j-m}}{k_j!}\left(\lambda_l+\frac{1}{\sigma_l(1-\nu_l)}\right)^{-m}\Gamma\left(\frac{1}{\sigma_l}+m\right)\right]\pi_l}{\sum_{z=1}^{n}\left[\frac{e^{-t\lambda_z\nu_z}}{\left[\Gamma\left(\frac{1}{\sigma_z}\right)\right]^t}(1+\lambda_z\sigma_z(1-\nu_z))^{-\frac{t}{\sigma_z}}\prod_{j=1}^{t}\sum_{m=0}^{k_j}\binom{k_j}{m}\frac{(\lambda_z)^{k_j}(\nu_z)^{k_j-m}}{k_j!}\left(\lambda_z+\frac{1}{\sigma_z(1-\nu_z)}\right)^{-m}\Gamma\left(\frac{1}{\sigma_z}+m\right)\right]\pi_z}$$

$$= \frac{\left[\frac{e^{-t\lambda_l\nu_l}}{\left[\Gamma\left(\frac{1}{\sigma_l}\right)\right]^t}(1+\lambda_l\sigma_l(1-\nu_l))^{-\frac{t}{\sigma_l}}\prod_{j=1}^{t}S_{j,l}\right]\pi_l}{\sum_{z=1}^{n}\left[\frac{e^{-t\lambda_z\nu_z}}{\left[\Gamma\left(\frac{1}{\sigma_z}\right)\right]^t}(1+\lambda_z\sigma_z(1-\nu_z))^{-\frac{t}{\sigma_z}}\prod_{j=1}^{t}S_{j,z}\right]\pi_z}.$$

Optimal Choice of $\hat{\boldsymbol{\lambda}}_{\mathbf{i}}^{\mathbf{t+1}}$ As previously, under a quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}(k_1, ..., k_t)$ for a policyholder with claim history $k_1, \cdots, k_t$ is the mean of the posterior structure function and is given by substituting (5.10) into (5.4).

5.3.2 Frequency Component Updating the Posterior Mean

Finite Negative Binomial Mixture

Generalizing the setup used by Lemaire (1995) and Frangos and Vrontos (2001), we consider a structure function given by a mixture of Gamma distributions. As previously, the portfolio is considered to be heterogeneous and all policyholders have constant but unequal underlying risks of having an accident. We assume that the number of claims $k|\lambda$ is distributed as a Poisson(λ) and that the structure function follows an n-component mixture of Gamma distributions, which has a probability density function of the form

$$u(\lambda) = \sum_{z=1}^{n}\pi_z\frac{\lambda^{\alpha_z-1}\tau_z^{\alpha_z}\exp(-\tau_z\lambda)}{\Gamma(\alpha_z)},$$

$\lambda, \alpha_z, \tau_z > 0$ for $z = 1, ..., n, \sum_{z=1}^{n}\pi_z = 1$, with mean $E(\lambda) = \sum_{z=1}^{n}\pi_z\frac{\alpha_z}{\tau_z}$. Then the unconditional distribution of the number of claims k is an n-component mixture of Negative Binomial distributions with probability density function

$$P(k) = \sum_{z=1}^{n}\pi_z\binom{k+\alpha_z-1}{k}p_z^{\alpha_z}q_z^k,\ p_z = \left(\frac{\tau_z}{1+\tau_z}\right), q_z = \left(\frac{1}{1+\tau_z}\right) \tag{5.12}$$

Proof. For $p_z = \left(\frac{\tau_z}{1+\tau_z}\right)$, $q_z = \left(\frac{1}{1+\tau_z}\right)$ and defining $\binom{k+\alpha_z-1}{k}$ as a generalized combinatorial coefficient we have that:

$$P(k) = \int_0^\infty P(k|\lambda)u(\lambda)d\lambda = \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!}\sum_{z=1}^{n}\pi_z\frac{\lambda^{\alpha_z-1}\tau_z^{\alpha_z}\exp(-\tau_z\lambda)}{\Gamma(\alpha_z)}d\lambda$$

$$= \sum_{z=1}^{n} \pi_z \frac{\tau_z^{\alpha_z}}{k!\Gamma(\alpha_z)(1+\tau_z)^{k+\alpha_z}} \int_0^{\infty} \exp\left(-\lambda\left(\tau_z+1\right)\right)\left(\lambda\left(\tau_z+1\right)\right)^{k+\alpha_z-1} d\left(\left(\lambda\left(\tau_z+1\right)\right)\right)$$

$$= \sum_{z=1}^{n} \pi_z \frac{\tau_z^{\alpha_z}}{\Gamma(k+1)\Gamma(\alpha_z)(1+\tau_z)^{k+\alpha_z}} \Gamma\left(k+\alpha_z\right) = \sum_{z=1}^{n} \pi_z \frac{\Gamma(k+\alpha_z)}{\Gamma(k+1)\Gamma(\alpha_z)} \frac{\tau_z^{\alpha_z}}{(1+\tau_z)^{k+\alpha_z}}$$

$$= \sum_{z=1}^{n} \pi_z \binom{k+\alpha_z-1}{k} \left(\frac{\tau_z}{1+\tau_z}\right)^{\alpha_z} \left(\frac{1}{1+\tau_z}\right)^{k} = \sum_{z=1}^{n} \pi_z \binom{k+\alpha_z-1}{k} p_z^{\alpha_z} q_z^{k}.$$

Note that the Negative Binomial distribution with pdf given by Eq. (3.3) in Chapter 3 is a special case of the finite Negative Binomial mixture for $n = 1$. Thus, if we let $n = 1$ in Eq. (5.12) then the proof of Eq. (3.3) follows from the proof presented above.

Posterior Structure Function Consider a policyholder with claim history $k_1, ..., k_t$ and let us denote as $K = \sum_{j=1}^{t} k_j$ the total number of claims that the policyholder had in t years, where k_j is the number of claims that the policyholder had in year $j, j = 1, ..., t$. Applying the Bayes theorem, one can find that the posterior structure function, $u\left(\lambda|k_1, ..., k_t\right)$, for a policyholder or a group of policyholders with claim history $k_1, ..., k_t$ is given by

$$u\left(\lambda|k_1, ..., k_t\right) = \sum_{z=1}^{n} \pi_z \frac{\left(\tau_z+t\right)^{K+\alpha_z} \lambda^{K+\alpha_z-1} e^{-(\tau_z+t)\lambda}}{\Gamma\left(\alpha_z+K\right)} \tag{5.13}$$

which is the probability density function of a mixture of Gamma with n components.

Proof. Considering the previous assumptions, the claim frequencies $k_1, ..., k_t$ are independent, thus we have

$$P\left(k_1, ..., k_t|\lambda\right) = P\left(k_1|\lambda\right) \cdot ... \cdot P\left(k_t|\lambda\right) = \frac{e^{-\lambda}\lambda^{k_1}}{k_1!} \cdot ... \cdot \frac{e^{-\lambda}\lambda^{k_t}}{k_t!} = \frac{e^{-t\lambda}\lambda^{K}}{\prod_{j=1}^{t} k_j!}$$

By Bayes theorem,

$$u\left(\lambda|k_1, ..., k_t\right) = \frac{P(k_1,...,k_t|\lambda)\cdot u(\lambda)}{P(k_1,...,k_t)} = \frac{P(k_1,...,k_t|\lambda)\cdot u(\lambda)}{\int_0^{\infty} P(k_1,...,k_t|\lambda)u(\lambda)d\lambda}$$

$$= \frac{\frac{e^{-t\lambda}\lambda^K}{\prod_{j=1}^{t} k_j!} \sum_{z=1}^{n} \pi_z \frac{\lambda^{\alpha_z-1}\tau_z^{\alpha_z}\exp(-\tau_z\lambda)}{\Gamma(\alpha_z)}}{\int_0^\infty \frac{e^{-t\lambda}\lambda^K}{\prod_{j=1}^{t} k_j!} \sum_{z=1}^{n} \pi_z \frac{\lambda^{\alpha_z-1}\tau_z^{\alpha_z}\exp(-\tau_z\lambda)}{\Gamma(\alpha_z)} d\lambda} = \frac{\sum_{z=1}^{n} \pi_z \lambda^{K+\alpha_z-1} e^{-(\tau_z+t)\lambda}}{\sum_{z=1}^{n} \pi_z \int_0^\infty \lambda^{K+\alpha_z-1} e^{-(\tau_z+t)\lambda} d\lambda}$$

$$= \sum_{z=1}^{n} \pi_z \frac{(\tau_z+t)^{K+\alpha_z} \lambda^{K+\alpha_z-1} e^{(\tau_z+t)\lambda}}{\int_0^\infty (\lambda(\tau_z+t))^{K+\alpha_z-1} e^{-(\tau_z+t)\lambda} d((\tau_z+t)\lambda)} = \sum_{z=1}^{n} \pi_z \frac{(\tau_z+t)^{K+\alpha_z} \lambda^{K+\alpha_z-1} e^{-(\tau_z+t)\lambda}}{\Gamma(\alpha_z+K)}.$$

Note that if we let $n = 1$ in Eq. (5.13) then the proof of Eq. (3.4) in Chapter 3 follows from the proof presented above.

Optimal Choice of $\hat{\boldsymbol{\lambda}}_{\mathbf{i}}^{\mathbf{t+1}}$ Using the quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}(k_1, ..., k_t)$ for a policyholder with claim history $k_1, ..., k_t$ is the mean of the posterior structure function, that is

$$\hat{\lambda}_{t+1}(k_1, ..., k_t) = \sum_{z=1}^{n} \pi_z \frac{K + \alpha_z}{\tau_z + t}. \tag{5.14}$$

5.3.3 Severity Component Updating the Posterior Probability

Let us consider now the severity component. Obviously, the setup for the severity component will also be a generalization of the good risk/bad risk model, as was the case for the frequency component. The portfolio is considered to be heterogeneous and we have fractions of drivers ρ_z where the risk (with respect to the mean claim size) that each policyholder of category z is imposing on the pool, $z = 1, \cdots, n$, is denoted by y_z. The probability density function (pdf) of the claim size x in each category is denoted by $f_z(x)$. Thus, the structure function is an n-point discrete distribution and the distribution of the unconditional claim size, denoted by $f(x)$, is given by

$$f(x) = \sum_{z=1}^{n} \rho_z f_z(x),$$

$k = 0, 1, ..., \rho_z > 0$, for $z = 1, \cdots, n$ and $\sum_{z=1}^{n} \rho_z = 1$. The expected value of the claim size is equal to $E(x) = \sum_{z=1}^{n} \rho_z y_z$.

We assume that a policyholder stays in the portfolio for t years and that the number of claims in year j is denoted by k_j, the total number of claims in t years is denoted by $K = \sum_{j=1}^{t} k_j$ and the claim amount for the k claim is denoted by x_k. In such a case, the information we have for their claim size history will be in the form of a vector $x_1, ..., x_K$ and the total claim amount over the t years that they stay in the portfolio is equal to $\sum_{k=1}^{K} x_k$. The risk that is imposed on the pool by the policyholder who belongs to the lth category of policyholders based on the severity of their claims is denoted by Q_l. Then the posterior probability of the policyholder belonging to the lth category is given by $\rho_l(x_1, ..., x_K)$. In order to design an optimal BMS that accounts for each claim amount, we have to find the posterior probability of belonging in each risk class, given the information we have about the claim size history for each policyholder for the time period they stay in the portfolio. Applying the Bayes theorem, the posterior probability of the policyholder belonging to the lth category is given by the following equation

$$\rho_l(x_1, ..., x_K) = \frac{f(x_1, ..., x_K | Q_l)\, \rho_l}{\sum_{z=1}^{n} f(x_1, ..., x_K | Q_z)\, \rho_z}.$$

In this way we update the posterior probability of belonging in category l, given the information we have for the claim size history of the policyholder. Using the quadratic error loss function, the optimal choice of y_{t+1} for a policyholder with claim size history $x_1, ..., x_K$,in t years is the mean of the posterior structure function, that is

$$\hat{y}_{t+1}(x_1, ..., x_K) = \sum_{z=1}^{n} y_z \rho_z(x_1, ..., x_K). \tag{5.15}$$

The mean claim size of each policyholder that belongs in the class l is considered constant. The setup described is applied to finite mixtures of Exponential, Weibull, Gamma and GB2 distributions. An extension to other finite mixture distributions can be obtained in a similar way and is straightforward.

Finite Mixture of Exponential

Let x be the size of the claim of each insured. We assume that the portfolio is heterogeneous and consists of n categories of policyholders, i.e. we have fractions of policyholders ρ_z whose claim sizes are distributed according to the Exponential distribution with parameter y_z. The parameter y_z denotes the risk that each policyholder of category z imposes on the pool, based on its mean claim size, for $z = 1, \cdots, n$. Thus, the structure function is an n-point discrete distribution and the size of loss x is distributed according to

$$f(x) = \sum_{z=1}^{n} \rho_z \frac{e^{-\frac{x}{y_z}}}{y_z}, x, \rho_z, y_z > 0, z = 1, \cdots, n \text{ and } \sum_{z=1}^{n} \rho_z = 1. \tag{5.16}$$

Posterior Probability The lth category posterior probability is equal to

$$\rho_l\left(x_1,...,x_K\right)=\rho_l\frac{e^{-\frac{\sum\limits_{k=1}^{K}x_k}{y_l}}}{y_l^K}\left[\sum_{z=1}^{n}\rho_z\frac{e^{-\frac{\sum\limits_{k=1}^{K}x_k}{y_z}}}{y_z^K}\right]^{-1} \tag{5.17}$$

Proof. Applying Bayes theorem, we have

$$\rho_l\left(x_1,...,x_k\right)=P\left(Q_l|x_1,...,x_k\right)=\frac{P(x_1,...,x_k|Q_l)P(Q_l)}{\sum\limits_{z=1}^{n}P(x_1,...,x_k|Q_z)P(Q_z)}$$

$$=\frac{\frac{\exp\left(\frac{-\sum\limits_{k=1}^{K}x_k}{y_l}\right)}{y_l^K}\rho_l}{\frac{\exp\left(\frac{-\sum\limits_{k=1}^{K}x_k}{y_1}\right)}{y_1^K}\rho_1+...+\frac{\exp\left(\frac{-\sum\limits_{k=1}^{K}x_k}{y_l}\right)}{y_l^K}\rho_l+...+\frac{\exp\left(\frac{-\sum\limits_{k=1}^{K}x_k}{y_n}\right)}{y_n^K}\rho_n}$$

$$=\frac{\frac{\exp\left(\frac{-\sum\limits_{k=1}^{K}x_k}{y_l}\right)}{y_l^K}\rho_l}{\sum\limits_{z=1}^{n}\frac{\exp\left(\frac{-\sum\limits_{k=1}^{K}x_k}{y_z}\right)}{y_z^K}\rho_z}$$

$$=\rho_l\frac{e^{-\frac{\sum\limits_{k=1}^{K}x_k}{y_l}}}{y_l^K}\left[\sum_{z=1}^{n}\rho_z\frac{e^{-\frac{\sum\limits_{k=1}^{K}x_k}{y_z}}}{y_z^K}\right]^{-1}.$$

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ Using the quadratic error loss function, the optimal choice of $\hat{y}_{t+1}(x_1, ..., x_K)$ for a policyholder with claim size history $x_1, ..., x_K$ is the mean of the posterior structure function and is given by substituting (5.17) into (5.15). We note that a mixture of exponential distributions results in a more heavy tailed distribution than the relatively tame exponential distribution.

Finite Mixture of Gamma

Similarly to the previous setup, we assume that the portfolio consists of n categories of policyholders, i.e. we have fractions of policyholders ρ_z which their claims sizes are distributed according to the Gamma distribution[4] with parameters y_z, θ_z, for $z = 1, ..., n$. Thus, the structure function is an n-point discrete distribution and the size of loss x is distributed according to

$$f(x) = \sum_{z=1}^{n} \rho_z \frac{1}{\left[\theta_z^2 y_z\right]^{\frac{1}{\theta_z^2}}} \frac{x^{\frac{1}{\theta_z^2}-1} e^{-\frac{x}{\theta_z^2 y_z}}}{\Gamma\left(\frac{1}{\theta_z^2}\right)} \tag{5.18}$$

for $x > 0$ and $\rho_z, \theta_z, y_z > 0$, for $z = 1, \cdots, n$ and $\sum_{z=1}^{n} \rho_z = 1$.

Posterior Probability The posterior probability of the policyholder belonging to the lth category is given by

$$\rho_l(x_1, ..., x_K) = \frac{\rho_l \left[\left(\theta_l^2 y_l\right)^{\frac{1}{\theta_l^2}} \Gamma\left(\frac{1}{\theta_l^2}\right)\right]^{-K} \left(\prod_{j=1}^{K} x_j\right)^{\frac{1}{\theta_l^2}-1} e^{-\frac{\sum_{j=1}^{K} x_j}{\theta_l^2 y_l}}}{\sum_{z=1}^{n} \rho_z \left[\left(\theta_z^2 y_z\right)^{\frac{1}{\theta_z^2}} \Gamma\left(\frac{1}{\theta_z^2}\right)\right]^{-K} \left(\prod_{j=1}^{K} x_j\right)^{\frac{1}{\theta_z^2}-1} e^{-\frac{\sum_{j=1}^{K} x_j}{\theta_z^2 y_z}}} \tag{5.19}$$

Proof. By means of the Bayes theorem, we have

$$\rho_l(x_1, ..., x_k) = P(Q_l | x_1, ..., x_k) = \frac{P(x_1, ..., x_k | Q_l) P(Q_l)}{\sum_{z=1}^{n} P(x_1, ..., x_k | Q_z) P(Q_z)}$$

$$= \frac{\left(\prod_{j=1}^{K} \frac{1}{\left((\theta_l)^2 y_l\right)^{\frac{1}{(\theta_l)^2}}} \frac{x_j^{\frac{1}{(\theta_l)^2}-1} e^{-\frac{x_j}{(\theta_l)^2 y_l}}}{\Gamma\left(\frac{1}{(\theta_l)^2}\right)}\right) \rho_l}{\sum_{z=1}^{n} \left(\prod_{j=1}^{K} \frac{1}{\left((\theta_z)^2 y_z\right)^{\frac{1}{(\theta_z)^2}}} \frac{x_j^{\frac{1}{(\theta_z)^2}-1} e^{-\frac{x_j}{(\theta_z)^2 y_z}}}{\Gamma\left(\frac{1}{(\theta_z)^2}\right)}\right) \rho_z}$$

[4] Using the reparameterization of Johnson et al (1994), obtained by setting $\theta^2 = \frac{1}{\alpha}$ and $y = \alpha\beta$. One can see for more Rigby and Stasinopoulos (2009).

$$= \frac{\rho_l \left[\left(\theta_l^2 y_l\right)^{\frac{1}{\theta_l^2}} \Gamma\left(\frac{1}{\theta_l^2}\right) \right]^{-K} \left(\prod_{j=1}^{K} x_j \right)^{\frac{1}{\theta_l^2}-1} e^{-\frac{\sum_{j=1}^{K} x_j}{\theta_l^2 y_l}}}{\sum_{z=1}^{n} \rho_z \left[\left(\theta_z^2 y_z\right)^{\frac{1}{\theta_z^2}} \Gamma\left(\frac{1}{\theta_z^2}\right) \right]^{-K} \left(\prod_{j=1}^{K} x_j \right)^{\frac{1}{\theta_z^2}-1} e^{-\frac{\sum_{j=1}^{K} x_j}{\theta_z^2 y_z}}}.$$

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ Using the quadratic error loss function, one can find that the optimal choice of $\hat{y}_{t+1}(x_1, ..., x_K)$ for a policyholder with claim size history $x_1, ..., x_K$ is the mean of the posterior structure function, given by substituting (5.19) into (5.15).

Finite Mixture of Weibull Type III

We assume that the portfolio consists of n categories of policyholders, with probability belonging in each category ρ_z, and their claims sizes are distributed according to the Weibull distribution with parameters y_z, θ_z, for $z = 1, ..., n$. Thus, the structure function is a n-point discrete distribution and the size of loss x is distributed according to

$$f(x) = \sum_{z=1}^{n} \rho_z \frac{\theta_z}{y_z} \Gamma\left(\frac{1}{\theta_z} + 1\right) \left[\frac{x}{y_z} \Gamma\left(\frac{1}{\theta_z} + 1\right) \right]^{\theta_z - 1} e^{-\left[\frac{x}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right]^{\theta_z}} \tag{5.20}$$

for $x > 0$ and $\rho_z, \theta_z, y_z > 0$, for $z = 1, \cdots, n$, $\sum_{z=1}^{n} \rho_z = 1$.

Posterior Probability The posterior probability of the policyholder belonging to the lth category is equal to

$$\rho_l(x_1, ..., x_K) = \frac{\rho_l \theta_l^K \left[\frac{\Gamma\left(\frac{1}{\theta_l}+1\right)}{y_l} \right]^{K\theta_l} \prod_{j=1}^{K} x_j^{\theta_l - 1} e^{-\sum_{j=1}^{K} \left[\frac{x_j}{y_l}\Gamma\left(\frac{1}{\theta_l}+1\right)\right]^{\theta_l}}}{\sum_{z=1}^{n} \rho_z \theta_z^K \left[\frac{\Gamma\left(\frac{1}{\theta_z}+1\right)}{y_z} \right]^{K\theta_z} \prod_{j=1}^{K} x_j^{\theta_z - 1} e^{-\sum_{j=1}^{K} \left[\frac{x_j}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right]^{\theta_z}}} \tag{5.21}$$

Proof. Applying the Bayes theorem, we have

$$\rho_l(x_1, ..., x_k) = P(Q_l | x_1, ..., x_k) = \frac{P(x_1,...,x_k|Q_l)P(Q_l)}{\sum_{z=1}^{n} P(x_1,...,x_k|Q_z)P(Q_z)}$$

$$= \frac{\left(\prod_{j=1}^{K} \frac{\theta_l}{y_l}\Gamma\left(\frac{1}{\theta_l}+1\right)\left(\frac{x_j}{y_l}\Gamma\left(\frac{1}{\theta_l}+1\right)\right)^{\theta_l-1} \exp\left(-\left(\frac{x_j}{y_l}\Gamma\left(\frac{1}{\theta_l}+1\right)\right)^{\theta_l}\right)\right)\rho_l}{\sum_{z=1}^{n}\left(\prod_{j=1}^{K} \frac{\theta_z}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\left(\frac{x_j}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right)^{\theta_z-1} \exp\left(-\left(\frac{x_j}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right)^{\theta_z}\right)\right)\rho_z}$$

$$= \frac{\rho_l\theta_l^K\left[\frac{\Gamma\left(\frac{1}{\theta_l}+1\right)}{y_l}\right]^{K\theta_l}\prod_{j=1}^{K} x_j^{\theta_l-1} e^{-\sum_{j=1}^{K}\left[\frac{x_j}{y_l}\Gamma\left(\frac{1}{\theta_l}+1\right)\right]^{\theta_l}}}{\sum_{z=1}^{n}\rho_z\theta_z^K\left[\frac{\Gamma\left(\frac{1}{\theta_z}+1\right)}{y_z}\right]^{K\theta_z}\prod_{j=1}^{K} x_j^{\theta_z-1} e^{-\sum_{j=1}^{K}\left[\frac{x_j}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right]^{\theta_z}}}.$$

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ As previously, using the quadratic error loss function, one can find that the optimal choice of $\hat{y}_{t+1}(x_1, ..., x_K)$ for a policyholder with claim size history $x_1, ..., x_K$ is the mean of the posterior structure function, given by substituting (5.21) into (5.15).

Finite Mixture of Generalized Beta Type II (GB2)

Finally, we consider the case of the finite GB2 mixture distribution. We assume that the portfolio consists of n categories of policyholders with probability of belonging in each category ρ_z and parameters of GB2[5] distribution $y_z, \sigma_z, \nu_z, s_z$, for $z = 1, ..., n$. Thus, the structure function is an n-point discrete distribution and the unconditional distribution of the size of claim x has a probability density function given by

$$f(x) = \sum_{z=1}^{n} \rho_z|\sigma_z|x^{\sigma_z\nu_z-1}\left\{y_z^{\sigma_z\nu_z}B(\nu_z, s_z)\left[1+\left(\frac{x}{y_z}\right)^{\sigma_z}\right]^{\nu_z+s_z}\right\}^{-1}, \tag{5.22}$$

for $x, \rho_z, y_z, \nu_z, s_z > 0$ and $-\infty < \sigma_z < \infty$,for $z = 1, ..., n$ and $\sum_{z=1}^{n} \rho_z = 1$. In this case the mean is given by

$$E(x) = \sum_{z=1}^{n} \rho_z y_z \frac{B(\nu_z + \frac{1}{\sigma_z}, s_z - \frac{1}{\sigma_z})}{B(\nu_z, s_z)}.$$

Posterior Probability The lth category posterior probability is equal to

[5] We use the parameterization of GB2 given by McDonald and Xu (1995), McDonald (1996), Rigby and Stasinopoulos (2009).

$$\rho_l(x_1,...,x_K)=\frac{\left(\frac{|\sigma_l|}{y_l^{\sigma_l\nu_l}B(\nu_l,s_l)}\right)^K\left(\prod_{j=1}^K x_j\right)^{\sigma_l\nu_l-1}\left\{\prod_{j=1}^K\left[1+\left(\frac{x_j}{y_l}\right)^{\sigma_l}\right]^{\nu_l+s_l}\right\}^{-1}\rho_l}{\sum_{z=1}^n\left(\frac{|\sigma_z|}{y_z^{\sigma_z\nu_z}B(\nu_z,s_z)}\right)^K\left(\prod_{j=1}^K x_j\right)^{\sigma_z\nu_z-1}\left\{\prod_{j=1}^K\left[1+\left(\frac{x_j}{y_z}\right)^{\sigma_z}\right]^{\nu_z+s_z}\right\}^{-1}\rho_z} \tag{5.23}$$

Proof.

Applying the Bayes theorem, we have

$$\rho_l(x_1,...,x_k)=P(Q_l|x_1,...,x_k)=\frac{P(x_1,...,x_k|Q_l)P(Q_l)}{\sum_{z=1}^n P(x_1,...,x_k|Q_z)P(Q_z)}$$

$$=\frac{\prod_{j=1}^K|\sigma_l|x_j^{\sigma_l\nu_l-1}\left\{y_l^{\sigma_l\nu_l}B(\nu_l,s_l)\left[1+\left(\frac{x_j}{y_l}\right)^{\sigma_l}\right]^{\nu_l+s_l}\right\}^{-1}\rho_l}{\sum_{z=1}^n\prod_{j=1}^K|\sigma_z|x_j^{\sigma_z\nu_z-1}\left\{y_z^{\sigma_z\nu_z}B(\nu_z,s_z)\left[1+\left(\frac{x_j}{y_z}\right)^{\sigma_z}\right]^{\nu_z+s_z}\right\}^{-1}\rho_z}$$

$$=\frac{\left(\frac{|\sigma_l|}{y_l^{\sigma_l\nu_l}B(\nu_l,s_l)}\right)^K\left(\prod_{j=1}^K x_j\right)^{\sigma_l\nu_l-1}\left\{\prod_{j=1}^K\left[1+\left(\frac{x_j}{y_l}\right)^{\sigma_l}\right]^{\nu_l+s_l}\right\}^{-1}\rho_l}{\sum_{z=1}^n\left(\frac{|\sigma_z|}{y_z^{\sigma_z\nu_z}B(\nu_z,s_z)}\right)^K\left(\prod_{j=1}^K x_j\right)^{\sigma_z\nu_z-1}\left\{\prod_{j=1}^K\left[1+\left(\frac{x_j}{y_z}\right)^{\sigma_z}\right]^{\nu_z+s_z}\right\}^{-1}\rho_z}.$$

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ Under a quadratic error loss function, the optimal choice of $\hat{y}_{t+1}(x_1,...,x_K)$ for a policyholder with claim size history $x_1,...,x_K$ is the mean of the posterior structure function, given by

$$\hat{y}_{t+1}(x_1,...,x_K)=\sum_{z=1}^n\rho_z(x_1,...,x_K)\,y_z\frac{B(\nu_z+\frac{1}{\sigma_z},s_z-\frac{1}{\sigma_z})}{B(\nu_z,s_z)},\text{for } z=1,\cdots,n, \tag{5.24}$$

where $\rho_z(x_1,...,x_K)$ is given by (5.23).

5.3.4 Severity Component Updating the Posterior Mean

Finite Mixture of Pareto

Generalizing the structure proposed by Frangos and Vrontos (2001), we consider a heterogeneous portfolio with respect to the mean claim size of each policyholder. Assume that claim severity given the mean claim severity, $x|y$, is distributed according to an exponential distribution and that the structure function follows an n-component mixture of Inverse Gamma distributions, with a pdf given by

$$g(y)=\sum_{z=1}^{n}\rho_z\frac{\frac{1}{m_z}\exp\left(-\frac{m_z}{y}\right)}{\left(\frac{y}{m_z}\right)^{s_z+1}\Gamma(s_z)},$$

$y>0, s_z>0, m_z>0$,for $z=1,...,n, \sum_{z=1}^{n}\rho_z=1$, with mean $E(y)=\sum_{z=1}^{n}\rho_z\frac{m_z}{s_z-1}$. Then the unconditional distribution of the claim severity x will be an n-component mixture of Pareto distributions with pdf

$$f(x)=\sum_{z=1}^{n}\rho_z s_z m_z^{s_z}(x+m_z)^{-s_z-1} \tag{5.25}$$

Proof. Considering the assumptions of the model we have

$$f(x)=\int_0^\infty f(x|y)\,g(y)\,dy=\int_0^\infty\frac{e^{-\frac{x}{y}}}{y}\sum_{z=1}^{n}\rho_z\frac{\frac{1}{m_z}\exp\left(-\frac{m_z}{y}\right)}{\left(\frac{y}{m_z}\right)^{s_z+1}\Gamma(s_z)}dy$$

$$=\sum_{z=1}^{n}\rho_z(m_z)^{s_z}(x+m_z)^{-s_z-1}s_z\int_0^\infty\frac{\left(\exp\left(-\frac{(x+m_z)}{y}\right)\right)\left(\frac{(x+m_z)}{y}\right)^{s_z+1}}{y\Gamma(s_z+1)}dy.$$

The integrand of the above expression is of the same form as an Inverse Gamma with parameters s_z+1 and m_z+x, therefore

$$\int_0^\infty\frac{\left(\exp\left(-\frac{(x+m_z)}{y}\right)\right)\left(\frac{(x+m_z)}{y}\right)^{s_z+1}}{y\Gamma(s_z+1)}dy=1.$$

Thus we have

$$f(x)=\sum_{z=1}^{n}\rho_z s_z m_z^{s_z}(x+m_z)^{-s_z-1}.$$

Note that the Pareto distribution with pdf given by Eq. (4.3) in Chapter 4, is a special case of the finite Pareto mixture for $n=1$. Thus, if we let $n=1$ in Eq. (5.25) then the proof of Eq. (4.3) in Chapter 4 follows from the proof presented above.

Posterior Structure Function Consider that a policyholder stays in the portfolio for t years, the number of claims they had in the year j is denoted by k_j , the total number of claims that they had in t years is denoted by $K=\sum_{j=1}^{t}k_j$ and by x_k is denoted the claim amount for the k claim. Then the information we have for their claim size history will be in the form of a vector $x_1, ..., x_K$ and the total claim amount for that specific policyholder over the t years that they are in the portfolio will be equal to $\sum_{k=1}^{K} x_k$. Applying the Bayes theorem, we find that the posterior structure function of the mean claim size y, given the policyholder's claim size history $x_1, ..., x_K$ denoted as $g(y|x_1, ..., x_K)$ is given by

$$g(y|x_1, ..., x_K) = \sum_{z=1}^{n} \rho_z \frac{\left(m_z + \sum_{k=1}^{K} x_k\right)^{K+s_z} e^{-\frac{\left(m_z + \sum_{k=1}^{K} x_k\right)}{y}}}{y^{K+s_z+1} \Gamma(K + s_z)}, \tag{5.26}$$

which is the pdf of a mixture of Inverse Gamma with n components.

Proof. The claim sizes $x_1, ..., x_k$ are independent and hence

$$f(x_1, ..., x_k|y) = f(x_1|y) \cdot ... \cdot f(x_k|y) = \frac{\exp\left(-\frac{\sum_{k=1}^{K} x_k}{y}\right)}{y^K}$$

By Bayes theorem,

$$g(y|x_1, ..., x_k) = \frac{f(x_1,...,x_k|y)g(y)}{f(x_1,...,x_k)} = \frac{f(x_1,...,x_k|y)g(y)}{\int_0^\infty f(x_1,...,x_k|y)g(y)dy}$$

$$= \frac{\frac{\exp\left(-\frac{\sum_{k=1}^{K} x_k}{y}\right)}{y^K} \sum_{z=1}^{n} \rho_z \frac{\frac{1}{m_z}\exp\left(-\frac{m_z}{y}\right)}{\left(\frac{y}{m_z}\right)^{s_z+1}\Gamma(s_z)}}{\int_0^\infty \frac{\exp\left(-\frac{\sum_{k=1}^{K} x_k}{y}\right)}{y^K} \sum_{z=1}^{n} \rho_z \frac{\frac{1}{m_z}\exp\left(-\frac{m_z}{y}\right)}{\left(\frac{y}{m_z}\right)^{s_z+1}\Gamma(s_z)} dy}$$

$$= \frac{\sum_{z=1}^{n} \rho_z \frac{\frac{1}{m_z} \exp\left(-\frac{\left(m_z + \sum_{k=1}^{K} x_k \right)}{y} \right)}{y^K \left(\frac{y}{m_z} \right)^{s_z+1} \Gamma(s_z)}}{\sum_{z=1}^{n} \rho_z \int_0^\infty \frac{\frac{1}{m_z} \exp\left(-\frac{\left(m_z + \sum_{k=1}^{K} x_k \right)}{y} \right)}{y^K \left(\frac{y}{m_z} \right)^{s_z+1} \Gamma(s_z)} dy}$$

$$= \frac{\sum_{z=1}^{n} \rho_z \frac{\left(\frac{1}{m_z}\right)^{s_z} \exp\left(-\frac{\left(m_z + \sum_{k=1}^{K} x_k \right)}{y} \right)}{(y)^{K+s_z+1} \Gamma(s_z)}}{\sum_{z=1}^{n} \rho_z \left(\frac{1}{m_z}\right)^{s_z} \left(m_z + \sum_{k=1}^{K} x_k \right)^{-K-s_z} \frac{\Gamma(K+s_z)}{\Gamma(s_z)} \int_0^\infty \frac{\exp\left(-\frac{\left(m_z + \sum_{k=1}^{K} x_k \right)}{y} \right) \left(\frac{\left(m_z + \sum_{k=1}^{K} x_k \right)}{y} \right)^{K+s_z}}{y \Gamma(K+s_z)} dy}.$$

The integrand of the above expression is of the same form as an Inverse Gamma with parameters $K + s_z$ and $m_z + \sum_{k=1}^{K} x_k$, therefore

$$\int_0^\infty \frac{\exp\left(-\frac{\left(m_z + \sum_{k=1}^{K} x_k \right)}{y} \right) \left(\frac{\left(m_z + \sum_{k=1}^{K} x_k \right)}{y} \right)^{K+s_z}}{y \Gamma(K+s_z)} = 1.$$

Thus we have

$$g\left(y|x_1,...,x_k\right)=\sum_{z=1}^{n}\rho_z\frac{\frac{1}{m_z+\sum_{k=1}^{K}x_k}\exp\left(-\frac{\left(m_z+\sum_{k=1}^{K}x_k\right)}{y}\right)}{\left(\frac{y}{m_z+\sum_{k=1}^{K}x_k}\right)^{K+s_z+1}\Gamma(K+s_z)}$$

$$=\sum_{z=1}^{n}\rho_z\frac{\left(m_z+\sum_{k=1}^{K}x_k\right)^{K+s_z}e^{-\frac{\left(m_z+\sum_{k=1}^{K}x_k\right)}{y}}}{y^{K+s_z+1}\Gamma(K+s_z)}.$$

Note that if we let $n=1$ in Eq. (5.26) then the proof of Eq. (4.4) in Chapter 4 follows from the proof presented above.

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ Consequently, by using the quadratic error loss function, the optimal choice of $\hat{y}_{t+1}\left(x_1,...,x_K\right)$ for a policyholder with claim size history $x_1,...,x_K$ is the mean of the posterior structure function, that is

$$\hat{y}_{t+1}\left(x_1,...,x_K\right)=\sum_{z=1}^{n}\rho_z\frac{m_z+\sum_{k=1}^{K}x_k}{K+s_z-1}. \tag{5.27}$$

5.3.5 Calculation of the Premiums According to the Net Premium Principle

We calculate the premiums based on the net premium principle for the set of the distributions that were presented in the previous sections. Consider a policyholder or a group of policyholders who in t years have produced K claims with total claim amount equal to $\sum_{k=1}^{K}x_k$. The net premium that should be paid by that specific group of policyholders is equal to the product of their annual expected number of claims at fault for period $t+1$, $\hat{\lambda}_{t+1}(k_1,...,k_t)$ and their expected claim severity, $\hat{y}_{t+1}(x_1,...,x_K)$.

- In the case where we update the posterior probability, the premium is given by

$$Premium=e\hat{\lambda}_{t+1}(k_1,...,k_t)\hat{y}_{t+1}(x_1,...,x_K)=e\sum_{z=1}^{n}\pi_z\left(k_1,...,k_t\right)\lambda_z\sum_{z=1}^{n}\rho_z\left(x_1,...,x_K\right)y_z, \tag{5.28}$$

where $e = \frac{1}{3.5}$ denotes the exposure to risk, since as mentioned in the previous chapters all policyholders were observed for 3.5 years and where $\pi_z(k_1, \cdots, k_t)$ in (5.28) is given by the Eqs (5.6, 5.8, 5.10) for the case of Poisson, Negative Binomial and Delaporte respectively and $\rho_z(x_1, ..., x_K)$ is given by the Eqs (5.17, 5.21 and 5.19) for the case of Exponential, Weibull and Gamma respectively. However, $\hat{y}_{t+1}(x_1, ..., x_K) = \sum_{z=1}^{n} \rho_z(x_1, ..., x_K)\, y_z \frac{B(\nu_z + \frac{1}{\sigma_z}, s_z - \frac{1}{\sigma_z})}{B(\nu_z, s_z)}$ where $\rho_z(x_1, ..., x_K)$ is given by Eq. (5.23), for the case of the GB2.

- In the case where we update the posterior mean, using Eqs (5.14 and 5.27), the premium is equal to

$$Premium = e\hat{\lambda}_{t+1}(k_1, ..., k_t)\hat{y}_{t+1}(x_1, ..., x_K) = e\sum_{z=1}^{n} \pi_z \frac{K + \alpha_z}{\tau_z + t} \sum_{z=1}^{n} \rho_z \frac{m_z + \sum_{k=1}^{K} x_k}{K + s_z - 1}. \tag{5.29}$$

5.3.6 Properties of the Optimal BMS with a Frequency and a Severity Component

1. The system is fair as each insured pays a premium proportional to their claim frequency and their claim severity, taking into account, through the Bayes theorem, all the information available for the time that they are in our portfolio both for the number of their claims and the loss that these claims incurred. We use the exact loss x_k that is incurred by each claim in order to have a differentiation in the premium for policyholders with the same number of claims, not just a scaling with the average claim severity of the portfolio.

2. The system is financially balanced. Every single year,the average premium per policyholder remains constant and is equal to

$$P = e\sum_{z=1}^{n} \pi_z \lambda_z \sum_{z=1}^{n} \rho_z y_z, \tag{5.30}$$

where e is the corresponding exposure to risk. When we update the posterior mean, the average of all premiums is again constant and is equal to

$$P = e\sum_{z=1}^{n} \pi_z \frac{\alpha_z}{\tau_z} \sum_{z=1}^{n} \rho_z \frac{m_z}{s_z - 1} \tag{5.31}$$

The above is proved considering that the claim frequency and the claim severity are independent components and that

$$E_\Lambda[\Lambda] = E[E[\lambda|k_1, ...k_t]],$$

$$E_Y[Y] = E[E[y|x_1, ..., x_K]].$$

3. In the beginning, all the policyholders are paying the same premium which is equal to (5.30), when we update the posterior probability, and equal to (5.31), when we update the posterior mean.

4. Using finite mixture distributions we are able to deal with heterogeneous portfolios more efficiently in comparison with simpler models.

5.4 The Design of an Optimal BMS Based Both on the a Priori and the a Posteriori Criteria

As we mentioned in Chapter 3, Dionne and Vanasse (1989, 1992) presented a BMS that integrates risk classification and experience rating based on the number of claims of each policyholder. This BMS is derived as a function of the years that the policyholder is in the portfolio, the number of accidents and the statistically significant individual characteristics for the number of accidents. Furthermore, as we mentioned in Chapter 4, Frangos and Vrontos (2001) extended this model by introducing a generalized BMS that integrates a priori and a posteriori information on an individual basis based on both the frequency and the severity component. This generalized BMS was derived as a function of the years that the policyholder is in the portfolio, the number of accidents, the exact size of loss that each one of these accidents incurred, and the statistically significant individual characteristics for the number of accidents and for the severity of the accidents.

We extend these models by considering three finite mixture regression models for the frequency component; Poisson, Negative Binomial and Delaporte and four finite mixture regression models for the severity component; Exponential, Weibull, Gamma and GB2. These models are derived by updating the posterior probability of the policyholder belonging to a specific risk category. Furthermore, we consider updating the posterior mean claim frequency by employing a finite Negative Binomial regression model and the posterior mean claim severity by employing a finite Pareto regression model.

5.4.1 Frequency Component Updating the Posterior Probability

Consider a policyholder i with an experience of t periods whose number of claims for period j, denoted as K_i^j are independent. We assume that the portfolio consists of n categories of policyholders classified with respect to the risk they impose on the pool and that the expected number of claims of the individual i who belongs to the zth category, for period j is denoted by $\lambda_{z,i}^j$. Employing an n-point discrete finite mixture to model the number of claims K_i^j, produces fractions of policyholders π_z, $z = 1, ..., n$, with mean claim frequency $\lambda_{z,i}^j$. The expected number of claims, $\lambda_{z,i}^j$, is allowed to be a function of the vector $c_{z,i}^j \left(c_{z,i,1}^j, ..., c_{z,i,h}^j\right)$ of h individual's characteristics, which correspond to different a priori rating variables. To ensure non-negativity of $\lambda_{z,i}^j$ we assume that $\lambda_{z,i}^j = \exp\left(c_{z,i}^j \beta_z^j\right)$, for $z = 1, ..., n$, with β_z^j denoting the vector of coefficients.

The distribution of the number of claims k in each category is denoted by $P_z\left(k;\lambda_{z,i}^{j}\right)$. As previously, the structure function is an n-point discrete distribution and the distribution of the unconditional number of claims is denoted by $P\left(k\right)$ given by

$$P\left(K_i^j=k\right)=\sum_{z=1}^{n}\pi_z P_z\left(k;\lambda_{z,i}^{j}\right), k=0,1,...,\pi_z>0, \text{ for } z=1,\cdots,n \text{ and } \sum_{z=1}^{n}\pi_z=1.$$

Note also that the mean of the number of claims is given by

$$E\left(k\right)=\sum_{z=1}^{n}\pi_z\lambda_{z,i}^{j}=\sum_{z=1}^{n}\pi_z\exp\left(c_{z,i}^{j}\beta_z^{j}\right).$$

Let us denote with $K=\sum_{j=1}^{t}K_i^j$ the total number of claims of policyholder i in t years and R_l the risk the policyholder imposes on the pool if we assume that they belong to the lth category of policyholders. The insurer has to calculate the best estimator of the expected number of accidents at period $t+1$, $\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)$, for $z=1,...,n$ categories using the past information over t periods of claim frequency and known individual characteristics over $t+1$ periods. We apply the Bayes theorem in order to obtain the posterior probability $\pi_l\left(K_i^1,...,K_i^t;c_{l,i}^1,...,c_{l,i}^{t+1}\right)$ that the policyholder belongs to the lth category

$$\pi_l\left(K_i^1,...,K_i^t;c_{l,i}^1,...,c_{l,i}^{t+1}\right)=\frac{P(K_i^1,...,K_i^t;c_{l,i}^1,...,c_{l,i}^{t+1}|R_l)\pi_l}{\sum_{z=1}^{n}P(K_i^1,...,K_i^t;c_{z,1}^1,...,c_{z,n}^{t+1}|R_z)\pi_z}.$$

Similarly to the case where only the a posteriori criteria are taken into account, using the quadratic error loss function the optimal choice of $\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)$ is the mean of the posterior structure function

$$\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)=\sum_{z=1}^{n}\pi_z\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)\exp\left(c_{z,i}^{t+1}\beta_z^{t+1}\right). \tag{5.32}$$

When $t=0$, $\hat{\lambda}_i^1(c_{z,i}^1)=\sum_{z=1}^{n}\pi_z\exp\left(c_{z,i}^1\beta_z^1\right)$, which implies that only a priori rating is used in the first period.

Finite Poisson Mixture Regression Model

We fit an n-point discrete finite Poisson mixture to model the number of claims K_i^j and we have fractions of drivers π_z with Poisson parameter $\lambda_{z,i}^j$, $z=1,...,n$. Consider that $\lambda_{z,i}^j$ is a function of the vector $c_{z,i}^j\left(c_{z,i,1}^j,...,c_{z,i,h}^j\right)$ of h individual's characteristics, which represent different a priori rating variables. Specifically, assuming that $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$, for $z=1,...,n$, non-negativity of $\lambda_{z,i}^j$ is ensured from the exponential function and β_z^j is the vector of the coefficients. Thus, the probability specification will be a $n-$component Poisson regression mixture model of the following form

$$P\left(K_i^j = k\right) = \sum_{z=1}^{n} \pi_z \frac{\exp\left(-\exp\left(c_{z,i}^j \beta_z^j\right)\right)\left[\exp\left(c_{z,i}^j \beta_z^j\right)\right]^k}{k!} \tag{5.33}$$

with $k = 0, 1, ...$ and $\pi_z, \lambda_{z,i}^j > 0$,for $z = 1, ..., n$ and $\sum_{z=1}^{n} \pi_z = 1$.

Posterior Probability The posterior probability of the policyholder belonging to the lth category of drivers is given by[6]

$$\pi_l\left(K_i^1, ..., K_i^t; c_{l,i}^1, ..., c_{l,i}^{t+1}\right) = \frac{\left[\exp\left(c_{l,i}^j \beta_l^j\right)\right]^K e^{-t\exp\left(c_{l,i}^j \beta_l^j\right)} \pi_l}{\sum_{z=1}^{n} \left[\exp\left(c_{z,i}^j \beta_z^j\right)\right]^K e^{-t\exp\left(c_{z,i}^j \beta_z^j\right)} \pi_z}. \tag{5.34}$$

Proof. Eq. (5.34) can be obtained by letting $\lambda_{z,i}^j = \exp\left(c_{z,i}^j \beta_z^j\right)$ into Eq. (5.6).

Optimal Choice of $\hat{\boldsymbol{\lambda}}_{\mathbf{i}}^{\mathbf{t+1}}$ Employing the quadratic error loss function, the optimal choice of $\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_{z,i}^1, ..., c_{z,i}^{t+1}\right)$ for $z = 1, ..., n$ categories, is the mean of the posterior structure function given substituting (5.34) into (5.32).

Finite Negative Binomial Type I Mixture Regression Model

We fit an n-point discrete finite Negative Binomial mixture to model the number of claims K_i^j and we have fractions of policyholders π_z with Negative Binomial parameters $\lambda_{z,i}^j$, $\alpha_{z,i}^j$, for $z = 1, ..., n$. Consider that $\lambda_{z,i}^j$ is a function of the vector $c_{z,i}^j\left(c_{z,i,1}^j, ..., c_{z,i,h}^j\right)$ of h individual's characteristics, which represent different a priori rating variables. Specifically, assuming that $\lambda_{z,i}^j = \exp\left(c_{z,i}^j \beta_z^j\right)$, for $z = 1, ..., n$, non-negativity of $\lambda_{z,i}^j$ is ensured from the exponential function and β_z^j is the vector of the coefficients. Thus, the probability specification is an $n-$component Negative Binomial regression mixture model of the following form

$$P\left(K_i^j = k\right) = \sum_{z=1}^{n} \pi_z \binom{k + \frac{1}{\alpha_{z,i}^j} - 1}{k} \left(\frac{1}{1 + \alpha_{z,i}^j \exp\left(c_{z,i}^j \beta_z^j\right)}\right)^{\frac{1}{\alpha_{z,i}^j}} \left(\frac{\alpha_{z,i}^j \exp\left(c_{z,i}^j \beta_z^j\right)}{1 + \alpha_{z,i}^j \exp\left(c_{z,i}^j \beta_z^j\right)}\right)^k, \tag{5.35}$$

for $k = 0, 1, 2, 3, ...$ where $\alpha_{z,i}^j > 0, \lambda_{z,i}^j > 0$ for $z = 1, ..., n$ and $\sum_{z=1}^{n} \pi_z = 1$.

[6] Note that Eq. (5.34) cannot be obtained directly from Eq. (5.6), i.e. by letting $\lambda_{z,i}^j = \exp\left(c_{z,i}^j \beta_z^j\right)$ into (5.6).

Posterior Probability The lth category posterior probability is equal to

$$\pi_l\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)$$
$$= \frac{\prod_{j=1}^{t}\binom{K_i^j+\frac{1}{\alpha_{l,i}^j}-1}{K_i^j}\left(\frac{1}{1+\alpha_{l,i}^j\exp\left(c_{l,i}^j\beta_l^j\right)}\right)^{\frac{t}{\alpha_{l,i}^j}}\left(\frac{\alpha_{l,i}^j\exp\left(c_{l,i}^j\beta_l^j\right)}{1+\alpha_{l,i}^j\exp\left(c_{l,i}^j\beta_l^j\right)}\right)^K\pi_l}{\sum_{z=1}^{n}\prod_{j=1}^{t}\binom{K_i^j+\frac{1}{\alpha_{z,i}^j}-1}{K_i^j}\left(\frac{1}{1+\alpha_{z,i}^j\exp\left(c_{z,i}^j\beta_z^j\right)}\right)^{\frac{t}{\alpha_{z,i}^j}}\left(\frac{\alpha_{z,i}^j\exp\left(c_{z,i}^j\beta_z^j\right)}{1+\alpha_{z,i}^j\exp\left(c_{z,i}^j\beta_z^j\right)}\right)^K\pi_z} \quad (5.36)$$

Proof. Eq. (5.36) can be obtained by letting $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$ into Eq. (5.8).

Optimal Choice of $\hat{\boldsymbol{\lambda}}_{\mathbf{i}}^{\mathbf{t+1}}$ Employing the quadratic error loss function, the optimal choice of $\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)$ for $z=1,...,n$ categories, is the mean of the posterior structure function given by substituting (5.36) into (5.32).

Finite Delaporte Mixture Regression Model

We fit an n-point discrete finite Delaporte mixture to model the number of claims $k=K_i^j$ and we have fractions of drivers π_z with Delaporte parameters $\lambda_{z,i}^j,\sigma_{z,i}^j,\nu_{z,i}^j$, for $z=1,...,n$. Consider again that $\lambda_{z,i}^j$ is a function of the vector $c_{z,i}^j\left(c_{z,i,1}^j,...,c_{z,i,h}^j\right)$ of h individual's characteristics, which represent different a priori rating variables. Specifically, assuming that $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$, for $z=1,...,n$, non-negativity of $\lambda_{z,i}^j$ is ensured from the exponential function and β_z^j is the vector of the coefficients. Thus, the probability specification given the parameters $\lambda_{1,i}^j,...,\lambda_{n,i}^j$ will be an $n-$component Delaporte regression mixture model of the following form

$$P\left(K_i^j=k\right)=\sum_{z=1}^{n}\pi_z\frac{e^{-\left(\exp\left(c_{z,i}^j\beta_z^j\right)\right)\nu_{z,i}^j}}{\Gamma\left(\frac{1}{\sigma_{z,i}^j}\right)}\left[1+\left(\exp\left(c_{z,i}^j\beta_z^j\right)\right)\sigma_{z,i}^j\left(1-\nu_{z,i}^j\right)\right]^{-\frac{1}{\sigma_{z,i}^j}}S, \quad (5.37)$$

where

$$S=\sum_{m=0}^{k}\binom{k}{m}\frac{\left(\exp\left(c_{z,i}^j\beta_z^j\right)\right)^k\left(\nu_{z,i}^j\right)^{k-m}}{k!}\left(\exp\left(c_{z,i}^j\beta_z^j\right)+\frac{1}{\sigma_{z,i}^j\left(1-\nu_{z,i}^j\right)}\right)^{-m}\Gamma\left(\frac{1}{\sigma_{z,i}^j}+m\right) \quad (5.38)$$

with $k=0,1,...$ and $\pi_z,\lambda_{z,i}^j>0, 0<\nu_{z,i}^j<1$, $z=1,...,n$ and $\sum_{z=1}^{n}\pi_z=1$.

Posterior Probability The posterior probability of the policyholder belonging to the lth category of drivers is equal to

$$\pi_l\left(K_i^1,...,K_i^t;c_{l,i}^1,...,c_{l,i}^{t+1}\right)$$

$$= \frac{\left[\frac{e^{-t\exp\left(c_{l,i}^j\beta_l^j\right)\nu_{l,i}^j}}{\left[\Gamma\left(\frac{1}{\sigma_{l,i}^j}\right)\right]^t}\left[1+\exp\left(c_{l,i}^j\beta_l^j\right)\sigma_{l,i}^j\left(1-\nu_{l,i}^j\right)\right]^{-\frac{t}{\sigma_{l,i}^j}}\prod_{j=1}^t S_{j,l}\right]\pi_l}{\sum_{z=1}^n\left[\frac{e^{-t\exp\left(c_{z,i}^j\beta_z^j\right)\nu_{z,i}^j}}{\left[\Gamma\left(\frac{1}{\sigma_{z,i}^j}\right)\right]^t}\left[1+\exp\left(c_{z,i}^j\beta_z^j\right)\sigma_{z,i}^j\left(1-\nu_{z,i}^j\right)\right]^{-\frac{t}{\sigma_{z,i}^j}}\prod_{j=1}^t S_{j,z}\right]\pi_z}, \tag{5.39}$$

where

$$S_{j,l}=\sum_{m=0}^{K_i^j}\binom{K_i^j}{m}\frac{\left(\exp\left(c_{l,i}^j\beta_l^j\right)\right)^{K_i^j}\left(\nu_{l,i}^j\right)^{K_i^j-m}}{K_i^j!}\left(\exp\left(c_{l,i}^j\beta_l^j\right)+\frac{1}{\sigma_{l,i}^j\left(1-\nu_{l,i}^j\right)}\right)^{-m}\Gamma\left(\frac{1}{\sigma_{l,i}^j}+m\right)$$

and $S_{j,z}$ is given for $z=l$.

Proof. Eq. (5.39) can be obtained by letting $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$ into Eq. (5.10).

Optimal Choice of $\hat{\lambda}_i^{t+1}$ Employing again the quadratic error loss function, the optimal choice of $\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)$ for $z=1,...,n$ categories, is the mean of the posterior structure function given by substituting (5.39) into (5.32).

5.4.2 Frequency Component Updating the Posterior Mean

Finite Negative Binomial Type I Mixture Regression Model

In this case, the generalized BMS obtained for the frequency component will be derived as a generalization of the structure used by Dionne and Vanasse (1989, 1992). We consider a policyholder i with an experience of t periods whose number of claims for period j, denoted as K_i^j are independent. We assume that K_i^j follows Poisson distribution with parameter λ^j. We consider a heterogeneous portfolio of n categories of policyholders, with expected number of claims of the individual i who belongs to the zth category denoted as $\lambda_{z,i}^j$, $z=1,...,n$. We allow λ_z^j vary from one individual to another. Let $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$, where $c_{z,i}^j\left(c_{z,i,1}^j,...,c_{z,i,h}^j\right)$ is the vector of h individual's characteristics and β_z^j is the vector of the coefficients. The conditional to $c_{z,i}^j$ probability that policyholder i will be involved in k accidents during the period j will become

$$P\left(K_i^j=k|c_{z,i}^j\right)=\frac{e^{-\exp\left(c_{z,i}^j\beta_z^j\right)}\left(\exp\left(c_{z,i}^j\beta_z^j\right)\right)^k}{k!},$$

for $k=0,1,2,3,...$ and $\lambda_i^j>0$ where $E(K_i^j|c_{z,i}^j)=Var(K_i^j|c_{z,i}^j)=\lambda_i^j=\exp\left(c_{z,i}^j\beta_z^j\right)$. For the determination of the expected number of claims in this model we assume that the h individual characteristics provide enough information. However, if one assumes that the a priori rating variables do not contain all the significant information for the expected number of claims then a random variable ε_i has to be introduced into the regression component. According to Gourieroux, Montfort and Trognon (1984 a), (1984 b) we can write

$$\lambda_i^j = \exp\left(c_{z,i}^j \beta_z^j + \varepsilon_i\right) = \exp\left(c_{z,i}^j \beta_z^j\right) u_i,$$

where $u_i = \exp(\varepsilon_i)$, yielding a random λ_i^j. Assume that u_i follows an $n-$component Gamma mixture distribution with probability density function

$$\upsilon(u_i) = \sum_{z=1}^{n} \pi_z \frac{u_i^{\frac{1}{\alpha_z}-1} \frac{1}{\alpha_z}^{\frac{1}{\alpha_z}} \exp\left(-\frac{1}{\alpha_z} u_i\right)}{\Gamma\left(\frac{1}{\alpha_z}\right)}, \tag{5.40}$$

$u_i > 0, \alpha_z > 0$ for $z = 1, ..., n, \sum_{z=1}^{n} \pi_z = 1$ with mean $E(u_i) = 1$. Under this assumption the conditional distribution of $K_i^j | c_{z,i}^j$ becomes

$$P\left(K_i^j = k | c_{z,i}^j\right) = \sum_{z=1}^{n} \pi_z \binom{k + \frac{1}{\alpha_z} - 1}{k} \frac{\left[\alpha_z \exp\left(c_{z,i}^j \beta_z^j\right)\right]^k}{\left[1 + \alpha_z \exp\left(c_{z,i}^j \beta_z^j\right)\right]^{k+\frac{1}{\alpha_z}}}, \tag{5.41}$$

which is an n-component Negative Binomial mixture distribution with parameters α_z and $\exp\left(c_{z,i}^j \beta_z^j\right)$, with $E(K_i^j | c_{z,i}^j) = \sum_{z=1}^{n} \pi_z \exp\left(c_{z,i}^j \beta_z^j\right)$.

Proof. Considering the assumptions of the model we have

$$P\left(K_i^j = k | c_{z,i}^j\right) = \int_0^\infty \frac{e^{-\exp\left(c_{z,i}^j \beta_z^j\right) u_i} \left(\exp\left(c_{z,i}^j \beta_z^j\right) u_i\right)^k}{k!} \upsilon(u_i)\, du_i$$

$$= \int_0^\infty \frac{e^{-\exp\left(c_{z,i}^j \beta_z^j\right) u_i} \left[\exp\left(c_{z,i}^j \beta_z^j\right) u_i\right]^k}{k!} \sum_{z=1}^{n} \pi_z \frac{u_i^{\frac{1}{\alpha_z}-1} \frac{1}{\alpha_z}^{\frac{1}{\alpha_z}} \exp\left(-\frac{1}{\alpha_z} u_i\right)}{\Gamma\left(\frac{1}{\alpha_z}\right)} du_i =$$

$$= \sum_{z=1}^{n} \pi_z \frac{\frac{1}{\alpha_z}^{\frac{1}{\alpha_z}} \left[\exp\left(c_{z,i}^j \beta_z^j\right)\right]^k}{k! \Gamma\left(\frac{1}{\alpha_z}\right)} \int_0^\infty \exp\left[-\left(\frac{1}{\alpha_z} + \exp\left(c_{z,i}^j \beta_z^j\right)\right) u_i\right] u_i^{k+\frac{1}{\alpha_z}-1} du_i =$$

$$= \sum_{z=1}^{n} \pi_z \frac{\frac{1}{\alpha_z}^{\frac{1}{\alpha_z}} \left[\exp\left(c_{z,i}^j \beta_z^j\right)\right]^k}{k! \Gamma\left(\frac{1}{\alpha_z}\right) \left[\frac{1}{\alpha_z} + \exp\left(c_{z,i}^j \beta_z^j\right)\right]^{k+\frac{1}{\alpha_z}}} \int_0^\infty \exp\left[-\left(\frac{1}{\alpha_z} + \exp\left(c_{z,i}^j \beta_z^j\right)\right) u_i\right] \cdot$$

$$\cdot \left[\left(\frac{1}{\alpha_z} + \exp\left(c_{z,i}^j \beta_z^j\right)\right) u_i\right]^{k+\frac{1}{\alpha_z}-1} d\left(\left(\frac{1}{\alpha_z} + \exp\left(c_{z,i}^j \beta_z^j\right)\right) u_i\right) =$$

$$= \sum_{z=1}^{n} \pi_z \frac{\frac{1}{\alpha_z}^{\frac{1}{\alpha_z}} \left[\exp\left(c_{z,i}^{j}\beta_z^j\right)\right]^k}{k!\Gamma\left(\frac{1}{\alpha_z}\right)\left[\frac{1}{\alpha_z}+\exp\left(c_{z,i}^{j}\beta_z^j\right)\right]^{k+\frac{1}{\alpha_z}}} \Gamma\left(k+\frac{1}{\alpha_z}\right) =$$

$$= \sum_{z=1}^{n} \pi_z \frac{\Gamma\left(k+\frac{1}{\alpha_z}\right)}{\Gamma(k+1)\Gamma\left(\frac{1}{\alpha_z}\right)} \frac{\left[\alpha_z \exp\left(c_{z,i}^{j}\beta_z^j\right)\right]^k}{\left[1+\alpha_z \exp\left(c_{z,i}^{j}\beta_z^j\right)\right]^{k+\frac{1}{\alpha_z}}} =$$

$$= \sum_{z=1}^{n} \pi_z \binom{k+\frac{1}{\alpha_z}-1}{k} \frac{\left[\alpha_z \exp\left(c_{z,i}^{j}\beta_z^j\right)\right]^k}{\left[1+\alpha_z \exp\left(c_{z,i}^{j}\beta_z^j\right)\right]^{k+\frac{1}{\alpha_z}}}.$$

Note that the NBI regression model given by Eq. (3.19) in Chapter 3, is a special case of the finite Negative Binomial mixture regression model for $n = 1$. Thus, if we let $n = 1$ in Eq. (5.41) then the proof of Eq. (3.19) in Chapter 3 follows from the proof presented above. Note also that if α of the NBI model is reparameterized to $\frac{\sigma}{\mu}$ then the NBII model given by Eq. (2.15) in Chapter 2 can be obtained.

Posterior Structure Function We are going to build a generalized optimal BMS based on the number of past claims and on an individual's characteristics in order to adjust that individual's premiums over time. The problem is to determine, at the renewal of the policy, the expected claim frequency of the policyholder i for the period $t+1$ given the observation of the reported accidents in the preceding t periods and observable characteristics in the preceding $t+1$ periods and the current period.

Consider a policyholder i with $K_i^1, ..., K_i^t$ claim history and $c_i^1, ..., c_i^{t+1}$ characteristics and denote as $K = \sum_{j=1}^{t} K_i^j$ the total number of claims that they had. The mean claim frequency of the individual i for period $t+1$ is $\lambda_i^{t+1}\left(c_i^{t+1}, u_i\right)$ a function of both the vector of the individual's characteristics and a random factor u_i with pdf $u(u_i)$. The posterior distribution of the mean claim frequency λ_i^{t+1} for an individual i observed over $t+1$ periods with $K_i^1, ..., K_i^t$ claim history and $c_{i,z}^1, ..., c_{i,z}^{t+1}$ characteristics is obtained using Bayes theorem and is given by an n-component Gamma mixture with updated parameters $\frac{1}{\alpha_z} + K$ and $S_{i,z}^j$, with pdf

$$f\left(\lambda_i^{t+1} | K_i^1, ..., K_i^t; c_{i,z}^1, ..., c_{i,z}^t\right) = \sum_{z=1}^{n} \pi_z \frac{\left(S_{i,z}^j\right)^{K+\frac{1}{\alpha_z}} \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1} \exp\left[-S_{i,z}^j \lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha_z}+K\right)}, \quad (5.42)$$

where $S_{i,z}^j = \frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t}\exp\left(c_{z,i}^j\beta_z^j\right)}{\exp\left(c_{z,i}^{t+1}\beta_z^{t+1}\right)}$ with $\lambda_i^{t+1} > 0, \alpha_z > 0$ and $z = 1, ..., n$ and $\sum_{z=1}^{n} \pi_z = 1$. Let us consider, as a special case, the situation in which the vector of the individual characteristics remains the same from one year to the next, i.e. $c_{i,z}^1 = c_{i,z}^2 = ... = c_{i,z}^{t+1} = c_{z,i}$ and $\beta_z^1 = \beta_z^2 = ... = \beta_z^t = \beta_z$. Then the posterior distribution of the mean claim frequency λ_i^{t+1} is simplified to

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)=\sum_{z=1}^{n}\pi_z\frac{\left(Z_{i,z}^j\right)^{K+\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1}\exp\left[-Z_{i,z}^j\lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha_z}+K\right)},$$

where $Z_{i,z}=\left[\frac{1}{\exp(c_{z,i}\beta_z)\alpha_z}+t\right]^{K+\frac{1}{\alpha_z}}$ with $\lambda_i^{t+1}>0,\alpha_z>0$ and $z=1,...,n,\sum_{z=1}^{n}\pi_z=1$.

Proof. In the following we are going to provide the proof of Eq. (5.42).

First, using the Theorem 9 in Chapter 3 we are going to calculate $f(\lambda_i^{t+1})$ which represents the pdf of $\lambda_i^{t+1}\left(c_i^{t+1},u_i\right)$, i.e. the mean claim frequency of the individual i for period $t+1$, called the structure function. Based on this theorem, one can find that $f\left(\lambda_i^{t+1}\right)$ is given by

$$f\left(\lambda_i^{t+1}\right)=\sum_{z=1}^{n}\pi_z\frac{\left(\frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}e^{-\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)\alpha_z}}}{\Gamma\left(\frac{1}{\alpha_z}\right)},\tag{5.43}$$

$\lambda_i^{t+1}>0,\alpha_z>0$ and $z=1,...,n,\sum_{z=1}^{n}\pi_z=1$, which is a n-component Gamma mixture distribution with parameters α_z and $\exp\left(c_i^{t+1}\beta^{t+1}\right)$. If we Let $g\left(u_i\right)=\exp\left(c_i^{t+1}\beta^{t+1}\right)u_i$, then g is a strictly increasing function. Also, as we have already mentioned, u_i follows a n-point continuous distribution the mixture of n Gamma, denoted as $\upsilon\left(u_i\right)$ whose pdf is given by (5.40).

Note that here the support sets:

$$X=\{u_i:\upsilon\left(u_i\right)>0\}\text{ and }Y=\left\{\lambda_i^{t+1}:\lambda_i^{t+1}=g\left(u_i\right)\text{ for some }x\in X\right\}$$

are both the interval $(0,\infty)$. From Eq.(5.40) we can easily see that the pdf of u_i, $\upsilon\left(u_i\right)$ is continuous on X. If we let $\lambda_i^{t+1}=g\left(u_i\right)$, then:

$$g^{-1}\left(\lambda_i^{t+1}\right)=\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\text{ and }\frac{d}{d\lambda_i^{t+1}}g^{-1}\left(\lambda_i^{t+1}\right)=\frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}$$

and g^{-1} is continuous on Y. Applying Theorem 9 in Chapter 3, for $\lambda_i^{t+1}\in(0,\infty)$, we get:

$$f\left(\lambda_i^{t+1}\right)=\upsilon\left(g^{-1}\left(\lambda_i^{t+1}\right)\right)\left|\frac{d}{d\lambda_i^{t+1}}g^{-1}\left(\lambda_i^{t+1}\right)\right|$$

$$=\sum_{z=1}^{n}\pi_z\frac{\left(\frac{1}{\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)\alpha_z}\right)}{\Gamma\left(\frac{1}{\alpha_z}\right)}\frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}=$$

$$=\sum_{z=1}^{n}\pi_z\frac{\left(\frac{1}{\exp\left(c_i^{t+1}\beta^{t+1}\right)\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\exp\left(c_i^{t+1}\beta^{t+1}\right)\alpha_z}\right)}{\Gamma\left(\frac{1}{\alpha_z}\right)}.$$

Next, using Bayes theorem, we will prove that $f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)$ is given by (5.42). By Bayes rule

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)=\frac{P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)}{\bar{P}\left(K_i^1,...,K_i^t|c_i^1,...,c_i^t\right)} \tag{5.44}$$

and where by definition

$$\bar{P}\left(\left(K_i^1,...,K_i^t\right)|c_i^1,...,c_i^t\right)=\int_0^\infty P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)d\lambda_i^{t+1}. \tag{5.45}$$

From (5.44) and (5.45) we have that

$$\begin{aligned}&f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)\\ =&\frac{P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)}{\int_0^\infty P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)f\left(\lambda_i^{t+1}\right)d\lambda_i^{t+1}}.\end{aligned} \tag{5.46}$$

The probability of the sequence $K_i^1,...,K_i^t$ given the frequency of accidents at $t+1$ and the individual's characteristics over the t periods $c_i^1,...,c_i^t$, will be a t-dimension Poisson distribution:

$$P\left(K_i^1,...,K_i^t|\lambda_i^{t+1};c_i^1,...,c_i^t\right)=\frac{e^{-\sum_{j=1}^t\lambda_i^j}\prod_{j=1}^t\left(\lambda_i^j\right)^{K_i^j}}{\prod_{j=1}^t K_i^j!}. \tag{5.47}$$

If we let $\lambda_i^j=\exp\left(c_{z,i}^j\beta_z^j\right)u_i\equiv\dot{\lambda}_i^j u_i$ then from (5.43) we get:

$$f\left(\lambda_i^{t+1}\right)=\sum_{z=1}^n\pi_z\frac{\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}e^{-\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}\alpha_z}}}{\Gamma\left(\frac{1}{\alpha_z}\right)}, \tag{5.48}$$

with $u_i>0,\alpha_z>0$ and $z=1,...,n,\sum_{z=1}^n\pi_z=1.$
By substituting (5.47) and (5.48) into (5.46), we get:

$$f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right)=$$

$$=\frac{\dfrac{e^{-\sum_{j=1}^{t}\lambda_i^j}\prod_{j=1}^{t}\left(\lambda_i^j\right)^{K_i^j}}{\prod_{j=1}^{t}K_i^j!}\sum_{z=1}^{n}\pi_z\dfrac{\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}\alpha_z}\right)}{\Gamma\left(\frac{1}{\alpha_z}\right)}}{\displaystyle\int_0^{\infty}\dfrac{e^{-\sum_{j=1}^{t}\lambda_i^j}\prod_{j=1}^{t}\left(\lambda_i^j\right)^{K_i^j}}{\prod_{j=1}^{t}K_i^j!}\sum_{z=1}^{n}\pi_z\dfrac{\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}\alpha_z}\right)}{\Gamma\left(\frac{1}{\alpha_z}\right)}d\lambda_i^{t+1}}$$

$$^{7}=\frac{e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j u_i}\prod_{j=1}^{t}\dot{\lambda}_i^j u_i^K\sum_{z=1}^{n}\pi_z\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}\alpha_z}\right)}{\displaystyle\int_0^{\infty}e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j u_i}\prod_{j=1}^{t}\dot{\lambda}_i^j u_i^K\sum_{z=1}^{n}\pi_z\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}\alpha_z}\right)d\lambda_i^{t+1}}$$

$$^{8}=\frac{e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}}\left(\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}\right)^K\sum_{z=1}^{n}\pi_z\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}\alpha_z}\right)}{\displaystyle\int_0^{\infty}e^{-\sum_{j=1}^{t}\dot{\lambda}_i^j\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}}\left(\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}}\right)^K\sum_{z=1}^{n}\pi_z\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\left(\lambda_i^{t+1}\right)^{\frac{1}{\alpha_z}-1}\exp\left(-\frac{\lambda_i^{t+1}}{\dot{\lambda}_i^{t+1}\alpha_z}\right)d\lambda_i^{t+1}}$$

$$=\frac{\left(\frac{1}{\dot{\lambda}_i^{t+1}}\right)^K\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\sum_{z=1}^{n}\pi_z\left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1}\exp\left[-\left(\dfrac{\frac{1}{\alpha_z}+\sum_{j=1}^{t}\dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)\lambda_i^{t+1}\right]}{\left(\frac{1}{\dot{\lambda}_i^{t+1}}\right)^K\left(\frac{1}{\dot{\lambda}_i^{t+1}\alpha_z}\right)^{\frac{1}{\alpha_z}}\sum_{z=1}^{n}\pi_z\displaystyle\int_0^{\infty}\left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1}\exp\left[-\left(\dfrac{\frac{1}{\alpha_z}+\sum_{j=1}^{t}\dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)\lambda_i^{t+1}\right]d\lambda_i^{t+1}}$$

[7] If we let $\lambda_i^j \equiv \dot{\lambda}_i^j \cdot u_i$ and $K=\sum_{j=1}^{t}K_i^j$

[8] If we let $\lambda_i^{t+1} \equiv \dot{\lambda}_i^{t+1} \cdot u_i$

$$= \frac{\sum_{z=1}^{n} \pi_z \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1} \left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)^{K+\frac{1}{\alpha_z}} \exp\left[-\left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)\lambda_i^{t+1}\right]}{\int_0^\infty \sum_{z=1}^{n} \pi_z \left[\left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)\lambda_i^{t+1}\right]^{K+\frac{1}{\alpha_z}-1} \exp\left[-\left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)\lambda_i^{t+1}\right] d\left(\left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)\lambda_i^{t+1}\right)}$$

$$= \sum_{z=1}^{n} \pi_z \frac{\left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1} \left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)^{K+\frac{1}{\alpha_z}} \exp\left[-\left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \dot{\lambda}_i^j}{\dot{\lambda}_i^{t+1}}\right)\lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha_z}+K\right)}$$

$$= \sum_{z=1}^{n} \pi_z \frac{\left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \exp\left(c_{z,i}^j \beta_z^j\right)}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)^{K+\frac{1}{\alpha_z}} \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1} \exp\left[-\left(\frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \exp\left(c_{z,i}^j \beta_z^j\right)}{\exp\left(c_i^{t+1}\beta^{t+1}\right)}\right)\lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha_z}+K\right)}$$

$$= \sum_{z=1}^{n} \pi_z \frac{\left(S_{i,z}^j\right)^{K+\frac{1}{\alpha_z}} \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1} \exp\left[-S_{i,z}^j \lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha_z}+K\right)},$$

where $S_{i,z}^j = \frac{\frac{1}{\alpha_z}+\sum_{j=1}^{t} \exp\left(c_{z,i}^j \beta_z^j\right)}{\exp\left(c_{z,i}^{t+1}\beta_z^{t+1}\right)}$ with $u_i > 0, \alpha_z > 0$ and $z = 1, ..., n$ and $\sum_{z=1}^{n} \pi_z = 1$.

When the vector of the individual characteristics remains the same from one year to the next we have that $\exp\left(c_{z,i}^j \beta_z^j\right) \equiv \exp\left(c_{z,i}\beta_z\right)$ and it can be easily verified that $S_{i,z}^j$ is simplified to $Z_{i,z} = \left[\frac{1}{\exp(c_{z,i}\beta_z)\alpha_z} + t\right]^{K+\frac{1}{\alpha_z}}$ thus

$$f\left(\lambda_i^{t+1} | K_i^1, ..., K_i^t; c_i^1, ..., c_i^t\right) = \sum_{z=1}^{n} \pi_z \frac{\left(Z_{i,z}^j\right)^{K+\frac{1}{\alpha_z}} \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1} \exp\left[-Z_{i,z}^j \lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha_z}+K\right)}.$$

Note that if we let $n = 1$ in Eq. (5.42) then the proof of Eq. (3.21) in Chapter 3 follows from the proof presented above.

Optimal Choice of $\hat{\lambda}_i^{t+1}$ Using the quadratic loss function, in the general case, one can find that the optimal estimator of $\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)$ is the mean of the posterior structure function given by

$$\begin{aligned}\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{i,z}^1,...,c_{i,z}^{t+1}\right) &= \int_0^\infty \lambda_i^{t+1}(c_i^{t+1},u_i) f\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_i^1,...,c_i^t\right) d\lambda_i^{t+1} \\ &= \sum_{z=1}^n \pi_z \exp\left(c_{z,i}^{t+1}\beta_z^{t+1}\right)\left[\frac{\frac{1}{\alpha_z}+\sum_{j=1}^t K_i^j}{\frac{1}{\alpha_z}+\sum_{j=1}^t \exp\left(c_{z,i}^j\beta_z^j\right)}\right] \qquad (5.49)\end{aligned}$$

This estimator defines the premium and corresponds to the multiplicative tariff formula where the base premium is the a priori frequency $\exp\left(c_{z,i}^{t+1}\beta_z^{t+1}\right)$ and where the Bonus-Malus factor is represented by the expression in brackets. When the vector of the individual characteristics remains the same from one year to the next $\hat{\lambda}_i^{t+1}$ is simplified to

$$\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{i,z}^1,...,c_{i,z}^{t+1}\right) = \sum_{z=1}^n \pi_z \exp\left(c_{z,i}\beta_z\right)\left[\frac{\frac{1}{\alpha_z}+\sum_{j=1}^t K_i^j}{\frac{1}{\alpha_z}+t\exp\left(c_{z,i}\beta_z\right)}\right]$$

When $t=0$, $\hat{\lambda}_i^1\left(c_i^1\right) = \sum_{z=1}^n \pi_z \exp\left(c_{z,i}^1\beta_z\right)$ which implies that only a priori rating is used in the first period. Moreover, when the regression component is limited to a constant $\beta_{z,0}$ one obtains

$$\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t\right) = \sum_{z=1}^n \pi_z \exp\left(\beta_{z,0}\right)\left[\frac{\frac{1}{\alpha_z}+\sum_{j=1}^t K_i^j}{\frac{1}{\alpha_z}+t\exp\left(\beta_{z,0}\right)}\right],$$

which corresponds to the 'univariate' ,without regression component, model.

5.4.3 Severity Component Updating the Posterior Probability

Let us consider now the severity component. We will model the severity component using a finite mixture of Exponential, Gamma, Weibull and GB2 regression models. Consider a policyholder i with an experience of t periods where their number of claims for period j are independent and are denoted as $K_i^j = k$, their total number of claims over t periods is denoted as $K = \sum_{j=1}^{t} K_i^j$ and by $X_{i,k}^j$ is denoted the loss incurred from their claim k for the period j. Then, the information we have for their claim size history will be in the form of a vector $X_{i,1}, X_{i,2}, ..., X_{i,K}$ and the total claim amount for the specific policyholder over the t periods that they are in the portfolio will be equal to $\sum_{k=1}^{K} X_{i,k}$. We will assume that the portfolio consists of n categories of drivers based on their claims severity and that the expected claim severity of the individual i who belongs to the zth category $z = 1, ..., n$, for period j is denoted by $y_{z,i}^j$. Furthermore, consider that the expected claim severity $y_{z,i}^j$ is a function of the vector $d_{z,i}^j \left(d_{z,i,1}^j, ..., d_{z,i,h}^j\right)$ of h individual's characteristics, which represent different a priori rating variables. Specifically, assume that $y_{z,i}^j = \exp\left(d_{z,i}^j \gamma_z^j\right)$, for $z = 1, ..., n$, non-negativity of $y_{z,i}^j$ is implied from the exponential function and γ_z^j is the vector of the coefficients. Let us denote as Q_l the risk that it is imposed on the insurance company if we assume that the policyholder belongs to the lth category of drivers sorted by the amount of loss that their accidents produce. Then the posterior probability of the policyholder belonging to the lth category is given by $\rho_l \left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_{l,i}^1, ..., d_{l,i}^{t+1}\right)$. The portfolio is considered to be heterogeneous and we have fractions of drivers ρ_z where the risk that each policyholder of category z is imposing on the pool, $z = 1, \cdots, n$ is denoted by y_z. The pdf of the claim size $X_{i,k}^j = x$ in each category is denoted by $f_z(x)$. Thus, the structure function is a n-point discrete distribution and the pdf of the unconditional claim size of the claim k of the policyholder i in period j is denoted by $f(x)$ and has a pdf of the following form

$$f(x) = \sum_{z=1}^{n} \rho_z f_z \left(x; y_{z,i}^j\right), k = 0, 1, ..., \rho_z > 0, \text{ for } z = 1, \cdots, n \text{ and } \sum_{z=1}^{n} \rho_z = 1.$$

Note also that the expected value of the claim size is equal to

$$E(x) = \sum_{z=1}^{n} \rho_z y_{z,i}^j = \sum_{z=1}^{n} \rho_z \exp\left(d_{z,i}^j \gamma_z^j\right).$$

In order to design an optimal BMS that will take into account the size of loss of each claim, we have to find the posterior probability of belonging to each risk class given the information we have about the claim size history for each policyholder for the time period they are in the portfolio. Applying the Bayes theorem, the posterior probability of the policyholder belonging to the lth category is given by the following equation

$$\rho_l \left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_{l,i}^1, ..., d_{l,i}^{t+1}\right) = \frac{f\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_{l,i}^1, ..., d_{l,i}^{t+1} | Q_l\right) \rho_l}{\sum_{z=1}^{n} f\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1} | Q_z\right) \rho_z}$$

In this way, we update the posterior probability of belonging to category l given the information we have for the claim size history of the policyholder. Using the quadratic error loss function the optimal choice of $\hat{y}_i^{t+1}\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right)$ for $z = 1, ..., n$ categories, given the observation of $X_{i,1,}X_{i,2},...,X_{i,K}$ and $d_{z,i}^{1},...,d_{z,i}^{t+1}$ for $z=1,...,n$,will be the mean of the posterior structure function

$$\hat{y}_i^{t+1}\left(X_{i,1,}X_{i,2},...,X_{i,K};d_{z,i}^{1},...,d_{z,i}^{t+1}\right) = \sum_{z=1}^{n}\rho_z\left(X_{i,1,}X_{i,2},...,X_{i,K};d_{z,i}^{1},...,d_{z,i}^{t+1}\right)\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right). \tag{5.50}$$

When $t=0$, $\hat{y}_i^1\left(d_{z,i}^1\right)=\sum_{z=1}^{n}\rho_z\exp\left(d_{z,i}^1\gamma_z^1\right)$, which implies that only a priori rating is used in the first period.

Finite Exponential Mixture Regression Model

The finite exponential mixture regression model will be the first model we present in order to deal with the generalized Bonus-Malus factor obtained with the use of the severity component. We fit an n-point continuous finite Exponential mixture to model claim size $X_{i,k}^{j}$ and we have fractions of drivers ρ_z with Exponential parameters $y_{z,i}^{j}$ for $z=1,...,n$. Furthermore, as we have already mentioned, if we assume that $y_{z,i}^{j}=\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)$, for $z=1,...,n$ then the probability specification of $X_{i,k}^{j}$ will be an $n-$component Exponential mixture regression model of the following form

$$f(x)=\sum_{z=1}^{n}\rho_z\frac{e^{-\frac{x}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}}}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)} \tag{5.51}$$

for $X_{i,k}^{j}>0,\rho_z$, $y_{z,i}^{j}>0$, $z=1,...,n$ and $\sum_{z=1}^{n}\rho_z=1$.

Posterior Probability The lth category posterior probability is equal to

$$\rho_l\left(X_{i,1,}X_{i,2},...,X_{i,K};d_{l,i}^{1},...,d_{l,i}^{t+1}\right)= \frac{\dfrac{\exp\left(\dfrac{-\sum_{k=1}^{K}x_k}{\exp\left(d_{l,i}^{j}\gamma_l^{j}\right)}\right)}{\left(\exp\left(d_{l,i}^{j}\gamma_l^{j}\right)\right)^K}\rho_l}{\sum_{z=1}^{n}\dfrac{\exp\left(\dfrac{-\sum_{k=1}^{K}x_k}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}\right)}{\left(\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)\right)^K}\rho_z}. \tag{5.52}$$

Proof. Eq. (5.52) can be obtained by letting $y_{z,i}^j = \exp\left(d_{z,i}^j \gamma_z^j\right)$ into Eq. (5.17).

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ Using the quadratic error loss function the optimal choice of $\hat{y}_i^{t+1}\left(X_{i,1,}X_{i,2},...,X_{i,K};d_{z,i}^1,...,d_{z,i}^{t+1}\right)$, given the observation of $X_{i,1,}X_{i,2},...,X_{i,K}$ and $d_{z,i}^1,...,d_{z,i}^{t+1}$ for $z=1,...,n$ categories, will be the mean of the posterior structure function and it is given by substituting (5.52) into (5.50).

Finite Gamma Mixture Regression Model

In this case, we obtain the generalized Bonus-Malus factor by fitting an n-point continuous finite Gamma mixture regression to model for assessing claim size $X_{i,k}^j$. We have fractions of drivers ρ_z with Gamma parameters ($y_{z,i}^j,\theta_{z,i}^j$) for $z=1,...,n$. Suppose that $y_{z,i}^j=\exp\left(d_{z,i}^j\gamma_z^j\right)$, for $z=1,...,n$ then the probability specification of $X_{i,k}^j$ will be an $n-$component Gamma regression mixture model of the following form

$$f(x)=\sum_{z=1}^{n}\rho_z\frac{1}{\left[\left(\theta_{z,i}^j\right)^2\exp\left(d_{z,i}^j\gamma_z^j\right)\right]^{\frac{1}{\left(\theta_{z,i}^j\right)^2}}}\frac{x^{\frac{1}{\left(\theta_{z,i}^j\right)^2}-1}e^{-\frac{x}{\left(\theta_{z,i}^j\right)^2\exp\left(d_{z,i}^j\gamma_z^j\right)}}}{\Gamma\left(\frac{1}{\left(\theta_{z,i}^j\right)^2}\right)} \tag{5.53}$$

for $X_{i,k}^j>0,\rho_z$, $y_{z,i}^j,\theta_{z,i}^j>0, z=1,...,n$ and $\sum_{z=1}^{n}\rho_z=1$.

Posterior Probability The posterior probability of the policyholder belonging to the lth category is given by

$$\rho_l\left(X_{i,1,}X_{i,2},...,X_{i,K};d_{z,i}^1,...,d_{z,i}^{t+1}\right)=$$

$$\frac{\left[\left[\left(\theta_{l,i}^j\right)^2\exp\left(d_{l,i}^j\gamma_l^j\right)\right]^{\frac{1}{\left(\theta_{l,i}^j\right)^2}}\Gamma\left(\frac{1}{\left(\theta_{l,i}^j\right)^2}\right)\right]^{-K}\left(\prod_{k=1}^{K}x_k\right)^{\frac{1}{\left(\theta_{l,i}^j\right)^2}-1}\exp\left(\frac{-\sum_{k=1}^{K}x_k}{\left(\theta_{l,i}^j\right)^2\exp\left(d_{l,i}^j\gamma_l^j\right)}\right)\rho_l}{\sum_{z=1}^{n}\left[\left[\left(\theta_{z,i}^j\right)^2\exp\left(d_{z,i}^j\gamma_z^j\right)\right]^{\frac{1}{\left(\theta_{z,i}^j\right)^2}}\Gamma\left(\frac{1}{\left(\theta_{z,i}^j\right)^2}\right)\right]^{-K}\left(\prod_{k=1}^{K}x_k\right)^{\frac{1}{\left(\theta_{z,i}^j\right)^2}-1}\exp\left(\frac{-\sum_{k=1}^{K}x_k}{\left(\theta_{z,i}^j\right)^2\exp\left(d_{z,i}^j\gamma_z^j\right)}\right)\rho_z} \tag{5.54}$$

Proof. Eq. (5.54) can be obtained by letting $y_{z,i}^j=\exp\left(d_{z,i}^j\gamma_z^j\right)$ into Eq. (5.19).

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ Using the quadratic error loss function the optimal choice of $\hat{y}_i^{t+1}\left(X_{i,1,}X_{i,2},...,X_{i,K};d_{z,i}^1,...,d_{z,i}^{t+1}\right)$, given the observation of $X_{i,1,}X_{i,2},...,X_{i,K}$ and $d_{z,i}^1,...,d_{z,i}^{t+1}$ for $z=1,...,n$ categories, will be the mean of the posterior structure function and it is given by substituting (5.54) into (5.50).

Finite Weibull Type III Mixture Regression Model

Next we fit an n-point continuous finite Weibull mixture to model $X_{i,k}^{j}$ and we have fractions of drivers ρ_z with Weibull parameters $y_{z,i}^{j},\theta_{z,i}^{j}$ for $z=1,...,n$. If we assume that $y_{z,i}^{j}=\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)$, for $z=1,...,n$ then the probability specification of $X_{i,k}^{j}$ will be a $n-$component Weibull mixture regression model of the following form

$$f(x) = \sum_{z=1}^{n}\rho_z\frac{\theta_{z,i}^{j}}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}\Gamma\left(\frac{1}{\theta_{z,i}^{j}}+1\right) \left[\frac{x}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}\Gamma\left(\frac{1}{\theta_{z,i}^{j}}+1\right)\right]^{\theta_{z,i}^{j}-1} e^{-\left[\frac{x}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}\Gamma\left(\frac{1}{\theta_{z,i}^{j}}+1\right)\right]^{\theta_{z,i}^{j}}} \quad (5.55)$$

for $X_{i,k}^{j}>0,\rho_z$, $y_{z,i}^{j},\theta_{z,i}^{j}>0, z=1,...,n$ and $\sum_{z=1}^{n}\rho_z=1$.

Posterior Probability The lth category posterior probability is given by

$$\rho_l\left(X_{i,1},X_{i,2},...,X_{i,K};d_{l,i}^{1},...,d_{l,i}^{t+1}\right)$$

$$= \frac{\left(\theta_{l,i}^{j}\right)^{K}\left[\frac{\Gamma\left(\frac{1}{\theta_{l,i}^{j}}+1\right)}{\exp\left(d_{l,i}^{j}\gamma_l^{j}\right)}\right]^{K\theta_{l,i}^{j}}\prod_{k=1}^{K}x_k^{\theta_{l,i}^{j}-1}\left[\exp\left(-\sum_{k=1}^{K}\left(\frac{x_k}{\exp\left(d_{l,i}^{j}\gamma_l^{j}\right)}\Gamma\left(\frac{1}{\theta_{l,i}^{j}}+1\right)\right)^{\theta_{l,i}^{j}}\right)\right]\rho_l}{\left(\theta_{z,i}^{j}\right)^{K}\left[\frac{\Gamma\left(\frac{1}{\theta_{z,i}^{j}}+1\right)}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}\right]^{K\theta_{z,i}^{j}}\prod_{k=1}^{K}x_k^{\theta_{z,i}^{j}-1}\left[\exp\left(-\sum_{k=1}^{K}\left(\frac{x_k}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}\Gamma\left(\frac{1}{\theta_{z,i}^{j}}+1\right)\right)^{\theta_{z,i}^{j}}\right)\right]\rho_z} \quad (5.56)$$

Proof. Eq. (5.56) can be obtained by letting $y_{z,i}^{j}=\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)$ into Eq. (5.21).

Optimal Choice of $\mathbf{\hat{y}_i^{t+1}}$ Using the quadratic error loss function the optimal choice of $\hat{y}_i^{t+1}\left(X_{i,1},X_{i,2},...,X_{i,K};d_{z,i}^{1},...,d_{z,i}^{t+1}\right)$, given the observation of $X_{i,1},X_{i,2},...,X_{i,K}$ and $d_{z,i}^{1},...,d_{z,i}^{t+1}$ for $z=1,...,n$ categories, will be the mean of the posterior structure function and it is given by substituting (5.56) into (5.50).

Finite Generalized Beta Type II Mixture Regression Model

Finally, we consider the case of the n-point continuous finite Generalized Beta type 2 mixture regression model for assessing claim severity $X_{i,k}^{j}$. We fit an n-point continuous finite GB2 mixture to model $X_{i,k}^{j}$ and have fractions of drivers ρ_z with GB2 parameters $y_{z,i}^{j},\sigma_{z,i}^{j},\nu_{z,i}^{j},s_{z,i}^{j}$, for $z=1,...,n$. Suppose that $y_{z,i}^{j}=\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)$, for $z=1,...,n$ then the probability specification of $X_{i,k}^{j}$ will be a $n-$component GB2 mixture regression model of the following form

$$f(x) = \sum_{z=1}^{n} \rho_z |\sigma^j_{z,i}| x^{\sigma^j_{z,i}\nu^j_{z,i}-1}$$
$$\left\{ \left(\exp\left(d^j_{z,i}\gamma^j_z\right)\right)^{\sigma^j_{z,i}\nu^j_{z,i}} B(\nu^j_{z,i}, s^j_{z,i}) \left[1 + \left(\frac{x}{\exp\left(d^j_{z,i}\gamma^j_z\right)}\right)^{\sigma^j_{z,i}}\right]^{\nu^j_{z,i}+s^j_{z,i}} \right\}^{-1} \quad (5.57)$$

for $X^j_{i,k} > 0, \rho_z > 0, y^j_{z,i} > 0, -\infty < \sigma^j_{z,i} < \infty$ and $\nu^j_{z,i}, s^j_{z,i} > 0$,for $z = 1, ..., n$ and $\sum_{z=1}^{n} \rho_z = 1$. In this case the mean is

$$E(x) = \sum_{z=1}^{n} \rho_z \exp\left(d^j_{z,i}\gamma^j_z\right) \frac{B(\nu^j_{z,i} + \frac{1}{\sigma^j_{z,i}}, s^j_{z,i} - \frac{1}{\sigma^j_{z,i}})}{B(\nu^j_{z,i}, s^j_{z,i})}$$

Posterior Probability The lth category posterior probability is equal to

$$\rho_l\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d^1_{z,i}, ..., d^{t+1}_{z,i}\right) =$$

$$\frac{\left[\frac{|\sigma^j_{l,i}|}{\left[\exp\left(d^j_{l,i}\gamma^j_l\right)\right]^{\sigma^j_{l,i}\nu^j_{l,i}} B(\nu^j_{l,i}, s^j_{l,i})}\right]^K \left(\prod_{j=1}^{K} x_j\right)^{\sigma^j_{l,i}\nu^j_{l,i}-1} \left\{\prod_{j=1}^{K}\left[1 + \left[\frac{x_j}{\exp\left(d^j_{l,i}\gamma^j_l\right)}\right]^{\sigma^j_{l,i}}\right]^{\nu^j_{l,i}+s^j_{l,i}}\right\}^{-1} \rho_l}{\sum_{z=1}^{n}\left[\frac{|\sigma^j_{z,i}|}{\left[\exp\left(d^j_{z,i}\gamma^j_z\right)\right]^{\sigma^j_{z,i}\nu^j_{z,i}} B(\nu^j_{z,i}, s^j_{z,i})}\right]^K \left(\prod_{j=1}^{K} x_j\right)^{\sigma^j_{z,i}\nu^j_{z,i}-1} \left\{\prod_{j=1}^{K}\left[1 + \left[\frac{x_j}{\exp\left(d^j_{z,i}\gamma^j_z\right)}\right]^{\sigma^j_{z,i}}\right]^{\nu^j_{z,i}+s^j_{z,i}}\right\}^{-1} \rho_z} \quad (5.58)$$

Proof. Eq. (5.58) can be obtained by letting $y^j_{z,i} = \exp\left(d^j_{z,i}\gamma^j_z\right)$ into Eq. (5.23).

Optimal Choice of $\hat{\mathbf{y}}^{\mathbf{t+1}}_{\mathbf{i}}$ Using the quadratic error loss function the optimal choice of $\hat{y}^{t+1}_i\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d^1_{z,i}, ..., d^{t+1}_{z,i}\right)$, given the observation of $X_{i,1}, X_{i,2}, ..., X_{i,K}$ and $d^1_{z,i}, ..., d^{t+1}_{z,i}$ for $z = 1, ..., n$ categories, will be the mean of the posterior structure function, given by

$$\hat{y}^{t+1}_i\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d^1_{z,i}, ..., d^{t+1}_{z,i}\right)$$

$$= \sum_{z=1}^{n} \rho_z\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d^1_{z,i}, ..., d^{t+1}_{z,i}\right) \exp\left(d^j_{z,i}\gamma^j_z\right) \frac{B(\nu^j_{z,i} + \frac{1}{\sigma^j_{z,i}}, s^j_{z,i} - \frac{1}{\sigma^j_{z,i}})}{B(\nu^j_{z,i}, s^j_{z,i})} \quad (5.59)$$

where $\rho_z\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d^1_{z,i}, ..., d^{t+1}_{z,i}\right)$ is given by (5.58). When $t = 0$,

$$\hat{y}_i^1\left(d_{z,i}^1\right)=\sum_{z=1}^{n}\rho_z\exp\left(d_{z,i}^1\gamma_z^1\right)\frac{B(\nu_{z,i}^1+\frac{1}{\sigma_{z,i}^1},s_{z,i}^1-\frac{1}{\sigma_{z,i}^1})}{B(\nu_{z,i}^1,s_{z,i}^1)},$$

which implies that only a priori rating is used in the first period.

5.4.4 Severity Component Updating the Posterior Mean

Finite Pareto Mixture Regression Model

In this case, the generalized Bonus-Malus factor for the severity component is derived as a generalization of the structure used by Frangos and Vrontos (2001). Consider a policyholder i with an experience of t periods. Assume that number of claims of the individual i for period j are independent and is denoted as K_i^j and by $X_{i,k}^j$ is denoted the loss incurred from their claim k for the period j. We consider that $X_{i,k}^j$ follows the Exponential distribution with mean claim severity for period j, y^j. We allow the y^j parameter to vary from one individual to another assuming that the expected claim severity $y_{z,i}^j$ is a function of the vector $d_{z,i}^j\left(d_{z,i,1}^j,...,d_{z,i,h}^j\right)$ of h individual's characteristics. Since policyholders have different mean claim severity, it is fair for each policyholder to pay a premium proportional to the risk that they impose on the pool. Specifically, we assume that $y_{z,i}^j=\exp\left(d_{z,i}^j\gamma_z^j\right)$, where $d_{z,i}^j\left(d_{z,i,1}^j,...,d_{z,i,h}^j\right)$ is the $1\times h$ vector of h individual's characteristics, which represent different a priori rating variables and γ_z^j is the vector of the coefficients. Then, the conditional to $d_{z,i}^j$ pdf of the claim size $X_{i,k}^j$, for a claim k of a policyholder i in period j will become

$$f\left(X_{i,k}^j|d_{z,i}^j\right)=\frac{e^{-\frac{x}{\exp\left(d_{z,i}^j\gamma_z^j\right)}}}{\exp\left(d_{z,i}^j\gamma_z^j\right)},$$

for $X_{i,k}^j>0$ and $y_i^j>0$ where $E(X_{i,k}^j|d_{z,i}^j)=y_i^j=\exp\left(d_{z,i}^j\gamma_z^j\right)$ and $Var(X_{i,k}^j|d_{z,i}^j)=\left(y_i^j\right)^2=\left(\exp\left(d_{z,i}^j\gamma_z^j\right)\right)^2$. For the determination of the expected claim severity in this model we assume that the h individual characteristics provide enough information. Nevertheless, if one assumes that the a priori rating variables do not contain all the significant information for the mean claim severity, then a random variable ξ_i has to be introduced into the regression component. Thus we can write

$$y_i^j=\exp\left(d_{z,i}^j\gamma_z^j+\xi_i\right)=\exp\left(d_{z,i}^j\gamma_z^j\right)w_i,$$

where $w_i=\exp\left(\xi_i\right)$, yielding a random y_i^j. We will assume that w_i follows an n-component Inverse Gamma mixture distribution with probability density function

$$w\left(w_i\right)=\sum_{z=1}^{n}\rho_z\frac{\frac{1}{(s_z-1)}\exp\left(-\frac{(s_z-1)}{w_i}\right)}{\left(\frac{w_i}{s_z-1}\right)^{s_z+1}\Gamma\left(s_z\right)}, \qquad (5.60)$$

$w_i > 0, s_z > 0$ for $z = 1, ..., n, \sum_{z=1}^{n} \rho_z = 1$ with mean $E(w_i) = 1$. It can be shown that the above parameterization does not affect the results if there is a constant term in the regression. We chose $E(w_i) = 1$ in order to have $E(\xi_i) = 0$. Under this assumption the conditional distribution of $X_{i,k}^j | d_{z,i}^j$ becomes

$$f\left(X_{i,k}^j | d_{z,i}^j\right) = \sum_{z=1}^{n} \rho_z s_z \frac{\left((s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)^{s_z}}{\left(x + (s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)^{s_z+1}}, \tag{5.61}$$

which is an n-component Pareto$\left(s_z, (s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)$ mixture distribution and has a mean equal to $E(X_{i,k}^j | d_{z,i}^j) = \sum_{z=1}^{n} \rho_z \exp\left(d_{z,i}^j \gamma_z^j\right)$.

Proof. Considering the assumptions of the model we have

$$f\left(X_{i,k}^j | d_{z,i}^j\right) = \int_0^\infty \frac{e^{-\frac{x}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}}}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i} w\left(w_i\right) dw_i$$

$$= \int_0^\infty \frac{e^{-\frac{x}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}}}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i} \sum_{z=1}^{n} \rho_z \frac{\frac{1}{(s_z-1)} \exp\left(-\frac{(s_z-1)}{w_i}\right)}{\left(\frac{w_i}{s_z-1}\right)^{s_z+1} \Gamma(s_z)} dw_i =$$

$$= \int_0^\infty \frac{e^{-\frac{x}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}}}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i} \sum_{z=1}^{n} \rho_z \frac{\left(\frac{s_z-1}{w_i}\right)^{s_z} \exp\left(-\frac{(s_z-1)}{w_i}\right)}{w_i \Gamma(s_z)} dw_i$$

$$= \sum_{z=1}^{n} \rho_z \left(s_z - 1\right)^{s_z} \int_0^\infty \frac{\exp\left(-\frac{\left(x + (s_z-1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}\right) \frac{1}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i^{s_z+1}}}{w_i \Gamma(s_z+1) \frac{1}{s_z}} dw_i$$

$$= \sum_{z=1}^{n} \rho_z s_z \left((s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)^{s_z} \left(x + (s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)^{-s_z-1}$$
$$\int_0^\infty \frac{\exp\left(-\frac{\left(x + (s_z-1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}\right) \left(\frac{\left(x + (s_z-1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}\right)^{s_z+1}}{w_i \Gamma(s_z+1)} dw_i.$$

The integrand of the above expression is of the same form as an Inverse Gamma with parameters $s_z + 1$ and $(s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right) + x$ thus we have

$$\int_0^\infty \frac{\exp\left(-\frac{\left(x + (s_z-1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}\right) \left(\frac{\left(x + (s_z-1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)}{\exp\left(d_{z,i}^j \gamma_z^j\right) w_i}\right)^{s_z+1}}{w_i \Gamma(s_z+1)} dw_i = 1,$$

and the probability specification becomes

$$f\left(X_{i,k}^{j}|d_{z,i}^{j}\right)=\sum_{z=1}^{n}\rho_z\frac{\left((s_z-1)\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)\right)^{s_z}}{\left(x+(s_z-1)\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)\right)^{s_z+1}}.$$

Note that the Pareto regression model given by Eq. (4.15) in Chapter 4 is a special case of the finite Pareto mixture regression model for $n=1$. Thus if we let $n=1$ in Eq. (5.61) then the proof of Eq. (4.15) in Chapter 4 follows from the proof presented above.

Posterior Structure Function Our goal is to construct a generalized optimal BMS based on the past claim size history and on an individual's characteristics in order to adjust that individual's premiums over time. Thus the problem is to determine, at the renewal of the policy, the expected claim severity of the policyholder i for the period $t+1$ given the observation of the reported claim sizes in the preceding t periods and observable characteristics in the preceding $t+1$ periods and the current period. Consider a policyholder i with $X_{i,1}^{1},X_{i,2}^{2},...,X_{i,K_i^t}^{t}$ claim size history in t periods and $d_i^1,...,d_i^{t+1}$ characteristics. The total number of claims for this specific policyholder in the preceding t periods will be denoted as $K=\sum_{j=1}^{t}K_i^j$ and the total claim amount that the accidents they were at fault produced will be equal to $\sum_{j=1}^{t}\sum_{k=1}^{K_i^j}X_{i,k}^{j}=\sum_{k=1}^{K}X_{i,k}$. The mean claim severity of the policyholder i for period $t+1$ is $y_i^{t+1}\left(d_{z,i}^{t+1},w_i\right)$ a function of both the vector of individual's characteristics and a random factor w_i with pdf $w(w_i)$. The posterior pdf of the mean claim severity y_i^{t+1} for an individual i observed over $t+1$ periods, with $X_{i,1}^{1},X_{i,2}^{2},...,X_{i,K_i^t}^{t}$ claim size history and $d_i^1,...,d_i^{t+1}$ characteristics, is obtained by applying Bayes theorem and is an n-component Inverse Gamma mixture with updated parameters, given by

$$g\left(y_i^{t+1}|X_{i,1}^{1},X_{i,2}^{2},...,X_{i,K_i^t}^{t};d_{i,z}^{1},...,d_{i,z}^{t+1}\right)=\sum_{z=1}^{n}\rho_z\frac{\frac{1}{C_{i,z}^{j}}\exp\left(-\frac{C_{i,z}^{j}}{y_i^{j}}\right)}{\left(\frac{y_i^{j}}{C_{i,z}^{j}}\right)^{K+s_z+1}\Gamma\left(s_z+K\right)},\tag{5.62}$$

for $y_i^{t+1}>0, s_z>0$ and $z=1,...,n,\sum_{z=1}^{n}\rho_z=1$, where

$$C_{i,z}^{j}=\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j}X_{i,k}^{j}}{\exp\left(d_{z,i}^{j}\gamma_z^{j}\right)}\right]\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right).$$

As previously, in the case that the individual characteristics remains constant, i.e. $d_{i,z}^1 = d_{i,z}^2 = ... = d_{i,z}^{t+1} = d_{z,i}$ and $\gamma_z^1 = \gamma_z^2 = ... = \gamma_z^t = \gamma_z$ the posterior pdf of the mean claim severity is simplified to

$$g\left(y_i^{t+1}|X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_{i,z}^1, ..., d_{i,z}^{t+1}\right) = \sum_{z=1}^{n} \rho_z \frac{\frac{1}{d_{z,i}} \exp\left(-\frac{D_{z,i}}{y_i^j}\right)}{\left(\frac{y_i^j}{D_{z,i}}\right)^{K+s_z+1} \Gamma\left(s_z + K\right)},$$

where $D_{z,i} = (s_z - 1)\exp\left(d_{z,i}\gamma_z\right) + \sum_{j=1}^{t} \sum_{k=1}^{K_i^j} X_{i,k}^j$.

Proof. In the following we are going to provide the proof of (5.62).

First, using the Theorem 9 in Chapter 3 we are going to calculate $g(y_i^{t+1})$ which represents the distribution of the mean claim severity for the individual i for period $t+1$, called the structure function. Based on this theorem one can find that y_i^{t+1} has a pdf of the following form:

$$g\left(y_i^{t+1}\right) = \sum_{z=1}^{n} \rho_z \frac{\left(\frac{(s_z-1)\exp\left(d_{z,i}^{t+1}\gamma^{t+1}\right)}{y_i^{t+1}}\right)^{s_z} \exp\left(-\frac{(s_z-1)\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)}{y_i^{t+1}}\right)}{y_i^{t+1}\Gamma\left(s_z\right)}, \tag{5.63}$$

$y_i^j > 0, s_z > 0$ and $z = 1, ..., n, \sum_{z=1}^{n} \rho_z = 1$, which is an n-component Inverse Gamma mixture distribution with parameters s_z and $\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)$. If we let $\omega(w_i) = \exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right) w_i$, then ω is a strictly increasing function. Also, as we have already mentioned, w_i follows a n-component Gamma mixture distribution, denoted as $w\left(w_i\right)$ with the pdf given by (5.60).

The support sets:

$$X = X = \{w_i : w\left(w_i\right) > 0\} \text{ and } Y = \left\{y_i^{t+1} : y_i^{t+1} = \omega\left(w_i\right) \text{ for some } x \in X\right\}$$

are both the interval $(0, \infty)$. From (4) we can easily see that the pdf of w_i, $w\left(w_i\right)$ is continuous on X. If we let $y_i^{t+1} = \omega\left(w_i\right)$, then:

$$\omega^{-1}\left(y_i^{t+1}\right) = \frac{y_i^{t+1}}{\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)} \text{ and } \frac{d}{dy_i^{t+1}}\omega^{-1}\left(y_i^{t+1}\right) = \frac{1}{\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)}$$

and g^{-1} is continuous on Y. Applying Theorem 9, for $y_i^{t+1} \in (0, \infty)$, we get:

$$g\left(y_i^{t+1}\right) = w\left(\omega^{-1}\left(y_i^{t+1}\right)\right)\left|\frac{d}{dy_i^{t+1}}\omega^{-1}\left(y_i^{t+1}\right)\right|$$

$$= \sum_{z=1}^{n} \rho_z \frac{\frac{1}{(s_z-1)} \exp\left(-\frac{(s_z-1)}{\exp\left(-d_{z,i}^{t+1}\gamma_z^{t+1}\right)y_i^{t+1}}\right)}{\left(\frac{\exp\left(-d_{z,i}^{t+1}\gamma_z^{t+1}\right)y_i^{t+1}}{s_z-1}\right)^{s_z+1} \Gamma(s_z)} \left(\exp\left(-d_{z,i}^{t+1}\gamma_z^{t+1}\right)\right)$$

$$= \sum_{z=1}^{n} \rho_z \frac{\left(\frac{(s_z-1)\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)}{y_i^{t+1}}\right)^{s_z} \exp\left(-\frac{(s_z-1)\exp\left(\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)\right)}{y_i^{t+1}}\right)}{y_i^{t+1}\Gamma(s_z)}.$$

Then the conditional distribution $g\left(y_i^{t+1}|X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right)$ represents the posterior distribution of the mean claim severity y_i^{t+1} for an individual i observed over $t+1$ periods with periods with $X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t$ claim size history and $d_i^1, ..., d_i^{t+1}$ characteristics. Applying Bayes theorem, one can find that the pdf of the posterior distribution of the mean claim severity is given by (5.62).

By Bayes rule

$$g\left(y_i^{t+1}|X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right) = \frac{f\left(X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t|y_i^{t+1}; d_i^1, ..., d_i^{t+1}\right) g\left(y_i^{t+1}\right)}{\bar{f}\left(X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t|d_i^1, ..., d_i^{t+1}\right)} \tag{5.64}$$

and where by definition

$$\bar{f}\left(X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t|d_i^1, ..., d_i^{t+1}\right) = \int_0^\infty f\left(X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t|y_i^{t+1}; d_i^1, ..., d_i^{t+1}\right) g\left(y_i^{t+1}\right) dy_i^{t+1}. \tag{5.65}$$

From (5.64), (5.65) we have that

$$g\left(y_i^{t+1}|X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t; d_i^1, ..., d_i^{t+1}\right) = \frac{f\left(X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t|y_i^{t+1}; d_i^1, ..., d_i^{t+1}\right) g\left(y_i^{t+1}\right)}{\int_0^\infty f\left(X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t|y_i^{t+1}; d_i^1, ..., d_i^{t+1}\right) g\left(y_i^{t+1}\right) dy_i^{t+1}}. \tag{5.66}$$

The probability of the sequence $X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t$ given the severity of accidents at $t+1$ and the individual's characteristics over the t periods $d_i^1, ..., d_i^{t+1}$, will be a t-dimension Exponential distribution:

$$f\left(X_{i,1,}^{1}X_{i,2}^{2},...,X_{i,K_i^t}^{t}|y_i^{t+1};d_i^1,...,d_i^{t+1}\right)=\frac{\exp\left(-\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j}X_{i,k}^j}{y_i^j}\right)}{\prod_{j=1}^{t}\prod_{k=1}^{K_i^j}y_i^j}. \tag{5.67}$$

If we let $y_i^j=\exp\left(d_{z,i}^j\gamma_z^j\right)w_i\equiv\dot{y}_i^jw_i,$ then $g\left(y_i^{t+1}\right)$ becomes

$$g\left(y_i^{t+1}\right)=\sum_{z=1}^{n}\rho_z\frac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}\Gamma\left(s_z\right)}, \tag{5.68}$$

with $y_i^j>0, s_z>0$ and $z=1,...,n,\sum_{z=1}^{n}\rho_z=1.$

By substituting (5.67) and (5.68) into (5.66), we get:

$$g\left(y_i^{t+1}|X_{i,1,}^{1}X_{i,2}^{2},...,X_{i,K_i^t}^{t};d_i^1,...,d_i^{t+1}\right)$$

$$=\frac{\frac{\exp\left(-\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j}X_{i,k}^j}{y_i^j}\right)}{\prod_{j=1}^{t}\prod_{k=1}^{K_i^j}y_i^j}\sum_{z=1}^{n}\rho_z\frac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}\Gamma(s_z)}}{\int_0^\infty\frac{\exp\left(-\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j}X_{i,k}^j}{y_i^j}\right)}{\prod_{j=1}^{t}\prod_{k=1}^{K_i^j}y_i^j}\sum_{z=1}^{n}\rho_z\frac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}\Gamma(s_z)}dy_i^{t+1}}$$

$$
{}^{9}\!= \frac{\dfrac{\exp\left(-\sum_{j=1}^{t}\dfrac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j w_i}\right)}{\prod_{j=1}^{t}\prod_{k=1}^{K_i^j}\dot{y}_i^j (w_i)^K}\sum_{z=1}^{n}\rho_z \dfrac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}}}{\int\limits_0^\infty \dfrac{\exp\left(-\sum_{j=1}^{t}\dfrac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j w_i}\right)}{\prod_{j=1}^{t}\prod_{k=1}^{K_i^j}\dot{y}_i^j (w_i)^K}\sum_{z=1}^{n}\rho_z \dfrac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}} dy_i^{t+1}}
$$

$$
{}^{10}\!= \frac{\dfrac{\exp\left(-\sum_{j=1}^{t}\dfrac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j \frac{y_i^{t+1}}{\dot{y}_i^{t+1}}}\right)}{\left(\frac{y_i^{t+1}}{\dot{y}_i^{t+1}}\right)^K}\sum_{z=1}^{n}\rho_z \dfrac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}}}{\int\limits_0^\infty \dfrac{\exp\left(-\sum_{j=1}^{t}\dfrac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j \frac{y_i^{t+1}}{\dot{y}_i^{t+1}}}\right)}{\left(\frac{y_i^{t+1}}{\dot{y}_i^{t+1}}\right)^K}\sum_{z=1}^{n}\rho_z \dfrac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}} dy_i^{t+1}}
$$

[9] If we let $y_i^j \equiv \dot{y}_i^j \cdot w_i$ and $K = \sum_{j=1}^{t} K_i^j$

[10] If we let $y_i^{t+1} \equiv \dot{y}_i^{t+1} \cdot w_i$

$$
= \frac{\frac{\exp\left(-\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j \frac{y_i^{t+1}}{\dot{y}_i^{t+1}}}\right)}{\left(y_i^{t+1}\right)^K}\sum_{z=1}^{n}\rho_z \frac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}}}{\int\limits_0^\infty \frac{\exp\left(-\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j \frac{y_i^{t+1}}{\dot{y}_i^{t+1}}}\right)}{\left(y_i^{t+1}\right)^K}\sum_{z=1}^{n}\rho_z \frac{\left(\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)^{s_z}\exp\left(-\frac{(s_z-1)\dot{y}_i^{t+1}}{y_i^{t+1}}\right)}{y_i^{t+1}}\, dy_i^{t+1}}
$$

$$
= \frac{\sum_{z=1}^{n}\rho_z \frac{\frac{1}{(s_z-1)\dot{y}_i^{t+1}}\exp\left[-\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\frac{1}{y_i^{t+1}}\right]}{\left(y_i^{t+1}\right)^K\left[\frac{y_i^{t+1}}{(s_z-1)\dot{y}_i^{t+1}}\right]^{s_z+1}}}{\sum_{z=1}^{n}\rho_z \int\limits_0^\infty \frac{\frac{1}{(s_z-1)\dot{y}_i^{t+1}}\exp\left[-\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\frac{1}{y_i^{t+1}}\right]}{\left(y_i^{t+1}\right)^K\left[\frac{y_i^{t+1}}{(s_z-1)\dot{y}_i^{t+1}}\right]^{s_z+1}}\, dy_i^{t+1}}
$$

$$
= \frac{\sum_{z=1}^{n}\rho_z \left[\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\right]^{K+s_z} \frac{\exp\left[-\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\frac{1}{y_i^{t+1}}\right]}{\Gamma(K+s_z)\left(y_i^{t+1}\right)^{K+s_z+1}}}{\sum_{z=1}^{n}\rho_z \int\limits_0^\infty \frac{\left[\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\right]^{K+s_z}\exp\left[-\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\frac{1}{y_i^{t+1}}\right]}{y_i^{t+1}\Gamma(K+s_z)}\, dy_i^{t+1}}.
$$

The integrand of the above expression is of the same form as an Inverse Gamma with parameters $s_z + K$ and $\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}$, thus we have

$$\int_0^\infty \frac{\left[\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\right]^{K+s_z}\exp\left[-\left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\dot{y}_i^j}\right]\dot{y}_i^{t+1}\frac{1}{y_i^{t+1}}\right]}{y_i^{t+1}\Gamma(K+s_z)}dy_i^{t+1}=1.$$

For $\dot{y}_i^j \equiv \exp\left(d_{z,i}^j\gamma_z^j\right)$ we get

$$g\left(y_i^{t+1}|X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_{i,z}^1, ..., d_{i,z}^{t+1}\right)=\sum_{z=1}^{n}\rho_z\frac{\frac{1}{C_{i,z}^j}\exp\left(-\frac{C_{i,z}^j}{y_i^j}\right)}{\left(\frac{y_i^j}{C_{i,z}^j}\right)^{K+s_z+1}\Gamma(s_z+K)},$$

for $y_i^j > 0, s_z > 0$, where $z = 1, ..., n$ and $\sum_{z=1}^{n}\rho_z = 1$, and where

$$C_{i,z}^j = \left[(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_{z,i}^j\gamma_z^j\right)}\right]\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right).$$

When the vector of the individual characteristics remains the same from one year to the next, we have that $\exp\left(d_{z,i}^j\beta_z^j\right)\equiv\exp\left(d_{z,i}\beta_z\right)$ and it can be easily verified that $C_{i,z}^j$ is simplified to $D_{z,i}=(s_z-1)\exp\left(d_{z,i}\gamma_z\right)+\sum_{j=1}^{t}\sum_{k=1}^{K_i^j}X_{i,k}^j$ thus,

$$g\left(y_i^{t+1}|X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_{i,z}^1, ..., d_{i,z}^{t+1}\right)=\sum_{z=1}^{n}\rho_z\frac{\frac{1}{d_{z,i}}\exp\left(-\frac{D_{z,i}}{y_i^j}\right)}{\left(\frac{y_i^j}{D_{z,i}}\right)^{K+s_z+1}\Gamma(s_z+K)}.$$

Note that if we let $n = 1$ in Eq. (5.62) then the proof of Eqs (4.17 and 4.18) in Chapter 4 follows from the proof presented above.

Optimal Choice of $\hat{\mathbf{y}}_{\mathbf{i}}^{\mathbf{t+1}}$ In the general case, using the quadratic loss function the optimal estimator of $\hat{y}_i^{t+1}\left(X_{i,1},X_{i,2},...,X_{i,K};d_{z,i}^1,...,d_{z,i}^{t+1}\right)$, given the observation of $X_{i,1},X_{i,2},...,X_{i,K}$ and $d_{z,i}^1,...,d_{z,i}^{t+1}$ for $z=1,...,n$ categories, will be the mean of the posterior structure function and is given by

$$\begin{aligned}
&\hat{y}_i^{t+1}\left(X_{i,1}^1,X_{i,2}^2,...,X_{i,K_i^t}^t;d_{i,z}^1,...,d_{i,z}^{t+1}\right) \\
=& \int_0^\infty y_i^{t+1}(d_{z,i}^{t+1},w_i)g\left(y_i^{t+1}|X_{i,1}^1,X_{i,2}^2,...,X_{i,K_i^t}^t;d_i^1,...,d_i^{t+1}\right)dy_i^{t+1} \\
=& \sum_{z=1}^n \rho_z \exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)\left[\frac{(s_z-1)+\sum_{j=1}^t \frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_{z,i}^j\gamma_z^j\right)}}{s_z+K-1}\right]
\end{aligned} \tag{5.69}$$

This estimator defines the premium and corresponds to the multiplicative tariff formula where the base premium is the a priori severity $\exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)$ and where the Bonus-Malus factor is represented by the expression in brackets. When the vector of the individual characteristics remains the same for all years the optimal estimator is simplified to

$$\hat{y}_i^{t+1}\left(X_{i,1}^1,X_{i,2}^2,...,X_{i,K_i^t}^t;d_{i,z}^1,...,d_{i,z}^{t+1}\right)=\sum_{z=1}^n \rho_z \frac{(s_z-1)\exp\left(d_{z,i}\gamma_z\right)+\sum_{j=1}^t\sum_{k=1}^{K_i^j} X_{i,k}^j}{s_z+K-1}$$

When $t=0$, $\hat{y}_i^1\left(d_i^1\right)=\sum_{z=1}^n \rho_z \exp\left(d_{z,i}^1\gamma_z\right)$ which implies that only a priori rating is used in the first period. Moreover, when the regression component is limited to a constant $\gamma_{z,0}$ one obtains

$$\hat{y}_i^{t+1}\left(X_{i,1}^1,X_{i,2}^2,...,X_{i,K_i^t}^t\right)=\sum_{z=1}^n \rho_z \frac{(s_z-1)\exp\left(\gamma_{z,0}\right)+\sum_{j=1}^t\sum_{k=1}^{K_i^j} X_{i,k}^j}{K+s_z-1},$$

which corresponds to the 'univariate' ,without regression component, model.

5.4.5 Calculation of the Premiums of the Generalized BMS

Now we are able to compute the premiums of the generalized optimal BMS based both on the frequency and the severity component. As we said, the premiums of the generalized optimal BMS will be given from the product of the generalized BMS based on the frequency component and of the generalized BMS based on the severity component. Consider a policyholder i that belongs to a group of policyholders who in t years have produced K claims with total claim amount equal to $\sum_{k=1}^{K} X_{i,k}$. The net premium that should be paid from that specific group of policyholders is calculated via the product of their annual expected claim frequency for period $t+1$, $\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_{z,i}^1, ..., c_{z,i}^{t+1}\right)$ and their expected claim severity, $\hat{y}_i^{t+1}\left(X_{i,1,}X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right)$.

- In the case where we update the posterior probability, the premium is given by

$$Premium = e\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_{z,i}^1, ..., c_{z,i}^{t+1}\right)\ \hat{y}_i^{t+1}\left(X_{i,1,}X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right)$$

$$= e\sum_{z=1}^{n} \pi_l\left(K_i^1, ..., K_i^t; c_{z,i}^1, ..., c_{z,i}^{t+1}\right)\lambda_z \sum_{z=1}^{n} \rho_z\left(X_{i,1,}X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right) y_z, \tag{5.70}$$

where $e = \frac{1}{3.5}$ is the corresponding risk exposure and where $\pi_l\left(K_i^1, ..., K_i^t; c_{z,i}^1, ..., c_{z,i}^{t+1}\right)$ in (5.70) is given by the Eqs (5.34,5.36,5.39) for the case of Poisson, Negative Binomial and Delaporte respectively and where $\rho_z\left(X_{i,1,}X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right)$ is given by the Eqs (5.52, 5.56 and 5.54) for the case of Exponential, Weibull and Gamma respectively, when we update the posterior probability. However,

$$\begin{aligned} &\hat{y}_i^{t+1}\left(X_{i,1,}X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right) \\ = \ &\rho_z\left(X_{i,1,}X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right) \cdot \exp\left(d_{z,i}^j \gamma_z^j\right) \frac{B(\nu_{z,i}^j + \frac{1}{\sigma_{z,i}^j}, s_{z,i}^j - \frac{1}{\sigma_{z,i}^j})}{B(\nu_{z,i}^j, s_{z,i}^j)}, \end{aligned}$$

where $\rho_z\left(X_{i,1,}X_{i,2}, ..., X_{i,K}; d_{z,i}^1, ..., d_{z,i}^{t+1}\right)$ is given by Eq. (5.58), for the case of the GB2.

- In the case where we update the posterior mean, using Eqs (5.49 and 5.69) the premium is equal to

$$Premium = e\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_{z,i}^1, ..., c_{z,i}^{t+1}\right)\hat{y}_i^{t+1}\left(X_{i,1,}^1 X_{i,2}^2, ..., X_{i,K_i^t}^t; d_{z,i}^1, ..., d_{z,i}^{t+1}\right)$$

$$= e\sum_{z=1}^{n} \pi_z \exp\left(c_{z,i}^{t+1}\beta_z^{t+1}\right)\left[\frac{\frac{1}{\alpha_z} + \sum_{j=1}^{t} K_i^j}{\frac{1}{\alpha_z} + \sum_{j=1}^{t} \exp\left(c_{z,i}^j \beta_z^j\right)}\right] \sum_{z=1}^{n} \rho_z \exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right)\left[\frac{(s_z - 1) + \sum_{j=1}^{t} \frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_{z,i}^j \gamma_z^j\right)}}{s_z + K - 1}\right]. \tag{5.71}$$

5.5 Application

We will use the same data set we used in Chapter 2 the descriptive statistics of which can be found in Tables 2.1 and 2.2 of Chapter 2. The available a priori rating variables we employ are the Bonus Malus (BM) category, the horsepower (HP) of the car and gender of the driver. However, as mentioned in the previous chapters, it is important to note that gender has recently been ruled out by the European Court as a rating factor and that records for fleet data were not available for the case of the claim counts. Recall also that this Bonus-Malus System has 20 classes and the transition rules are described as follows: Each claim free year is rewarded by one class discount and each claim in given year is penalized by one class. Claim counts are modelled for all 15641 policies. The Bonus-Malus category consists of five categories of neighboring BM classes : C1 = "drivers who belong to BM classes 1 and 2", C2 = "drivers who belong to BM classes 3-5", C3 = "drivers who belong to BM classes 6-9", C4 = "drivers who belong to BM class 10" and C5 = "drivers who belong to BM classes 11-20". The horsepower of the car consists of four categories: C1 = "drivers who had a car with a hp between 0-33", C2 = "drivers who had a car with a hp between 34-66", C3 = "drivers who had a car with a hp between 67-99", C4 ="drivers who had a car with a hp between 100-132". Regarding the amount paid for each claim, there were 5590 observations that met our criteria. Also, both private cars and fleet vehicles have been considered in this sample and the available a priori rating variables we employ are the Bonus Malus class, the horsepower of the car and gender of the driver. The Bonus-Malus category consists of five categories of neighboring BM classes: C1 = "drivers who belong to BM classes 1 and 2", C2 = "drivers who belong to BM classes 3-5", C3 = "drivers who belong to BM classes 6-9", C4 = "drivers who belong to BM class 10" and C5 = "drivers who belong to BM classes 11-20". The horsepower of the car consists of eleven categories: C1 = "drivers who had a car with a hp between 0-33", C2 = "drivers who had a car with a hp between 34-44", C3 = "drivers who had a car with a hp between 45-55", C4 = "drivers who had a car with a hp between 56-66", C5 = "drivers who had a car with a hp between 67-74", C6 = "drivers who had a car with a hp between 75-82", C7 = "drivers who had a car with a hp between 83-90", C8 = "drivers who had a car with a hp between 91-99", C9 ="drivers who had a car with a hp between 100-110", C10 = "drivers who had a car with a hp between 111-121" and C11 = "drivers who had a car with a hp between 122-132". Finally, the gender consists of three categories: M = "male", F = "female" and B = "both", since in this case, data for fleet vehicles used by either male or female drivers were also available, i.e. shared use.

In our application we fit the Poisson and Negative Binomial and their two component mixtures on the number of claims and the Exponential, Gamma, Weibull, GB2 and Pareto and their two and three component mixtures on the claim sizes. Furthermore, we introduce a regression component in the above models and we include risk classifying characteristics in order to use all the available information in the estimation of the claim frequency and severity distributions. We give emphasis on both the analysis of the claim frequency and severity using two and/or three component mixtures of distributions and generalized linear models (glm) as these methods have not been extensively studied in the BMS literature. The location and weight of these components are estimated from the data employing the EM algorithm. The number of components was chosen based on the information we had from the frequency and severity data respectively[11]. Based on the current methodology as presented in Sections 5.3 and 5.4 we will find the premiums determined from the optimal BMS based on the a posteriori frequency and severity component and the premiums determined from the optimal BMS with a frequency and a severity component based both on the a priori and the a posteriori criteria.

5.5.1 Modelling Results

This subsection is divided into two parts. The first part describes the modelling results of the models that have been applied to model claim frequency and the second part provides the results of the claim severity analysis. Employing the methods described in Rigby and Stasinopoulos (2001, 2005, 2009), we estimated the parameters of the above models[12]. Variable selection techniques were applied in order to find the variables that are considered as better predictors. For this purpose we used the function step.GAIC within the GAMLSS package in software R, which performs the stepwise model selection using a Generalized Akaike information criterion (GAIC). The final claim frequency and severity models we selected were the best fitted models.

Claim Frequency Models

Firstly, the Poisson, the Negative Binomial Type I (NBI) and their two component mixture distributions, given by Eqs (5.5 and 5.7), for $n = 1$ and $n = 2$ respectively, were fitted on the number of claims. The maximum likelihood estimators of the parameters for these models are presented in Table 5.1.

Table 5.1: Results of the Fitted Claim Frequency Distributions

Poisson	Two Component Poisson C1	C 2	NBI	Two Component NBI C 1	C 2
λ	π_1	π_2	λ	π_1	π_2
0.4847	0.8666	0.1334	0.4847	0.5655	0.4345
-	λ_1	λ_2	α	λ_1	λ_2
-	0.3033	1. 6693	0.9178	0.3891	0.5583
-	-	-	-	α_1	α_2
-	-	-	-	0.3195	0.1519

Note that the two component Negative Binomial(τ_z, α_z) mixture derived by updating the posterior mean, with pdf given by Eq. (5.12), for $n = 2$, is given from a reparameterization of the pdf of the NBI(λ_z, σ_z) distribution if we let $\lambda_z = \frac{\alpha_z}{\tau_z}$ and $\sigma_z = \frac{1}{\alpha_z}$, for $z = 1, 2$. Thus, the maximum likelihood estimators of the parameters of this model are $\hat{\pi}_1 = 0.5655, \hat{\pi}_2 =$

[11] In principle one could use more components regarding the data set examined and then select the best models.

[12] One can see for more on generalized additive models de Jong and Heller (2008) and Heller et al (2007).

$0.4345, \hat{\alpha}_1 = \frac{1}{0.3195} = 3.1299, \hat{\alpha}_2 = \frac{1}{0.1519} = 6.579, \hat{\tau}_1 = \frac{3.1299}{0.3891} = 8.0439$ and $\hat{\tau}_2 = \frac{6.5790}{0.5583} = 11.7840$.

Secondly, the Poisson, Negative Binomial Type I (NBI) and their two component mixture regression models, given by Eqs(5.33, 5.35 and 5.41), for $n = 1$ and $n = 2$ respectively, have been applied to our insurance data. The results are displayed in Table 5.2. For brevity, a $*$ in Table 5.2 indicates the estimated values which are statistically significant at a 5% threshold.

Table 5.2: Results of the Fitted Claim Frequency Regression Models

	Poisson	Two Component Poisson		NBI	Two Component NBI	
		C1	C2		C1	C2
		π_1	π_2		π_1	π_2
		0.9007	0.0993		0.8907	0.1093
Variable						
Intercept	−0.8514*	−1.3593*	0.7364*	−0.8366*	−1.2913*	0.5794*
Bonus-Malus						
Category 1	0	0	0	0	0	0
Category 2	0.6081*	0.8116*	0.1242	0.6084*	0.8067*	−0.0026
Category 3	0.8488*	0.8979*	0.7251*	0.8467*	0.8936*	0.7089*
Category 4	−0.9420*	−1.3351	−0.5028*	−0.9402*	−1.3588	−0.4427*
Category 5	1.9628*	2.1803*	1.1903*	1.9670*	2.1865*	1.1194*
Horsepower						
Category 1	0	0	0	0	0	0
Category 2	−0.2033	−0.0634*	−0.4403*	−0.2235	−0.0665*	−0.4794*
Category 3	−0.0367	0.1036*	−0.2859*	−0.0537	0.1106*	−0.3419*
Category 5	0.0380	0.0455*	0.0201*	0.0151	0.0357*	0.0056*
Gender						
Male	0	0	0	0	0	0
Female	0.0699*	0.1443*	−0.1076*	0.0794*	0.1561*	−0.1429*
Parameter	-	-	-	α	α_1	α_2
-	-	-	-	0.6556*	0.2068*	0.2496*

Claim Severity Models

Firstly, the Exponential, Gamma, Weibull, GB2 and Pareto[13] and their two and three component mixture distributions, given by Eqs (5.16, 5.18, 5.20, 5.22 and 5.25), for $n = 1$, $n = 2$ and $n = 3$ respectively, were fitted on the costs of claims. The maximum likelihood estimators of the parameters for these models are presented in Tables 5.3 and 5.4 respectively.

[13]Recall that $E(y) = \sum_{z=1}^{n} \rho_z \frac{m_z}{s_z - 1}$ for the the case of the finite mixture of Pareto distributions.

Table 5.3: Results of the Fitted Finite Mixture of Severity Distributions With One, Two and Three Components, Update of the Posterior Probability

EXP	Two Component EXP		Three Component EXP		
	C1	C2	C1	C 2	C 3
y	ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
328. 00	0.4720	0.5280	0.3292	0.3317	0.3391
-	y_1	y_2	y_1	y_2	y_3
-	328. 00	328. 00	328. 00	328. 00	328. 00
Gamma	Two Component Gamma		Three Component Gamma		
	C1	C2	C1	C 2	C 3
y	ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
328. 00	0.4560	0.5440	0.5268	0.1496	0.3236
θ	y_1	y_2	y_1	y_2	y_3
1.9883	247. 89	395.04	243.47	418.22	423.69
-	θ_1	θ_2	θ_1	θ_2	θ_3
-	0.1497	0.5906	0.1615	0.9260	0.2967
Weibull	Two Component Weibull		Three Component Weibull		
	C1	C2	C1	C 2	C 3
y	ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
328. 65	0.4721	0.5279	0.4753	0.3774	0.1473
θ	y_1	y_2	y_1	y_2	y_3
1. 7686	243.71	403.03	239.61	389.16	455.32
-	θ_1	θ_2	θ_1	θ_2	θ_3
-	2.0180	0.5294	2.0320	1.2030	0.1140
GB2	Two Component GB2		Three Component GB2		
	C1	C2	C1	C 2	C 3
y	ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
271. 97	0.4926	0.5074	0.03663	0.63287	0.3305
σ	y_1	y_2	y_1	y_2	y_3
5.2460	237.46	499.70	36.05	247.65	377.29
ν	σ_1	σ_2	σ_1	σ_2	σ_3
0.7029	10.1600	4.752	3.862	9.475	5.623
s	ν_1	ν_2	ν_1	ν_2	ν_3
0.6061	1.2058	0.3940	1.6093	0.9103	1.5923
-	s_1	s_2	s_1	s_2	s_3
-	0.9461	0.9305	0.5728	1.0131	0.7191

Table 5.4: Results of the Fitted Finite Pareto Mixture Distributions With One, Two and Three Components, Update of the Posterior Mean

Pareto	Two Component Pareto		Three Component Pareto		
	C1	C2	C1	C 2	C 3
m	ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
28001.13	0.4948	0.5052	0.3502	0.3066	0.3432
s	m_1	m_2	m_1	m_2	m_3
85. 798	24834.77	25084.36	25084.36	25084.36	25084.36
-	s_1	s_2	s_1	s_2	s_3
-	76.325	77.169	76.784	76.784	76.784

Secondly, the Exponential, Gamma, Weibull, GB2, Pareto and the two and three component Exponential, Gamma, Weibull, GB2 and Pareto[14] mixture regression models have been applied to claim severity analysis. The results are summarized in the Tables 5.5, 5.6, 5.7, 5.8 and 5.9 respectively. In these Tables, for brevity, a * indicates the estimated values which are statistically significant at a 5% threshold.

Table 5.5: Results of the Fitted Finite Exponential Mixture Regression Models With One, Two and Three Components, Update of the Posterior Probability

	EXP	Two Component EXP		Three Component EXP		
		C1	C2	C1	C2	C3
		ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
		0.5093	0.4907	0.3093	0.3553	0.3354
Variable						
Intercept	5.7458*	5.7460*	5.7456*	5.7458*	5.7457*	5.7460*
Bonus-Malus						
Category 1	0	0	0	0	0	0
Category 2	-0.0233	-0.0233	-0.0233	-0.0233	-0.0233	-0.0232
Category 3	0.1128*	0.1128	0.1128	0.1128	0.1128	0.1128
Category 4	-0.7019*	-0.7028*	-0.7010*	-0.7019*	-0.7017*	-0.7021*
Category 5	0.4105	0.4105	0.4106	0.4105	0.4105	0.4105
Horsepower						
Category 1	0	0	0	0	0	0
Category 2	-0.2101	-0.2100	-0.2102	-0.2101	-0.2101	-0.2101
Category 3	-0.2012	-0.2012	-0.2013	-0.2012	-0.2012	-0.2012
Category 4	-0.0164	-0.0162	-0.0165	-0.0163	-0.0164	-0.0164
Category 5	0.0028	0.0029	0.0028	0.0029	0.0028	0.0028
Category 6	0.1381	0.1382	0.1380	0.1381	0.1381	0.1381
Category 7	0.1574	0.1575	0.1574	0.1574	0.1574	0.1575
Category 8	0.3388*	0.3389	0.3387*	0.3388	0.3387	0.3388
Category 9	0.4445*	0.4445*	0.4444*	0.4444	0.4444	0.4445
Category 10	0.6563*	0.6563*	0.6563*	0.6563*	0.6563*	0.6563*
Category 11	1.0899*	1.0899*	1.0909*	1.0880*	1.0883*	1.0935*
Gender						
Both	0	0	0	0	0	0
Male	-0.0783*	-0.0786	-0.0779	-0.0782	-0.0781	-0.0785
Female	-0.0233	-0.0236	-0.0229	-0.0232	-0.0231	-0.0234

[14]Note that the GAMLSS package allows us to find the maximum likelihood estimators of the parameters of the regression model where the distribution of the response variable is the Pareto2 (m', s') distribution, with pdf given by $f(x) = s'm'^{s'}(x+m')^{-s'-1}$. The Pareto$(m, s)$ response distribution can be derived from a reparameterization of the pdf of the Pareto2 (m', s') distribution with $s' = s$ and $m' = (s'-1)m$. Thus $\hat{s} = \hat{s}'$ and $\hat{m} = \frac{\hat{m}'}{\hat{s}'-1}$.

Table 5.6: Results of the Fitted Finite Gamma Mixture Regression Models With One, Two and Three Components, Update of the Posterior Probability

	Gamma	Two Component Gamma		Three Component Gamma		
		C1	C2	C1	C2	C3
		ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
		0.4861	0.5139	0.0123	0.4666	0.5211
Variable						
Intercept	5.7458*	5.8652*	5.2486*	5.0943*	5.2467*	5.8900*
Bonus-Malus						
Category 1	0	0	0	0	0	0
Category 2	-0.0233	-0.0533*	0.0725*	0.0946*	0.0694*	-0.0416*
Category 3	0.1128*	0.0705*	0.1852*	0.2090	0.1817*	0.0821
Category 4	-0.7019*	-0.5515*	-1.6775*	-1.6603*	-1.6629*	-0.5460*
Category 5	0.4105*	0.3834*	0.3176*	0.2476	0.3341*	0.4137
Horsepower						
Category 1	0	0	0	0	0	0
Category 2	-0.2101*	-0.1612	-0.0094	0.4797	-0.0197	-0.1486*
Category 3	-0.2012*	-0.1854*	0.0061	0.3914	-0.0050*	-0.1740*
Category 4	-0.0164	-0.0318	0.1713*	1.6349	0.1597*	-0.0200
Category 5	0.0028	0.0018	0.2045*	1.7018	0.1920*	0.0119
Category 6	0.1381*	0.1637*	0.2780*	1.7060	0.2699*	0.1690
Category 7	0.1574*	0.1791*	0.3194*	0.4072	0.3069*	0.1830*
Category 8	0.3388*	0.3250*	0.3641*	3.0272	0.3519*	0.3277*
Category 9	0.4445*	0.4143*	0.4155*	0.3602*	0.4030*	0.4240*
Category 10	0.6563*	0.6261*	0.6094*	1.0707*	2.2276*	0.3562*
Category 11	1.0899*	1.0567*	0.6894*	0.7790*	0.6805*	1.0533*
Gender						
Both	0	0	0	0	0	0
Male	-0.0783*	0.0073	0.0032	-1.3113	0.0170*	-0.0300*
Female	-0.0233	0.0645	0.0168	-1.3072	0.0310*	0.0193
Parameter	θ	θ_1	θ_2	θ_1	θ_2	θ_3
	0.4268*	0.0958*	0.4568*	9.3062e-08*	0.0961*	0.4531*

Table 5.7: Results of the Fitted Finite Weibull Mixture Regression Models With One, Two and Three Components, Update of the Posterior Probability

	Weibull	Two Component Weibull		Three Component Weibull		
		C1	C2	C1	C2	C3
		ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
		0.4561	0.5439	0.3727	0.4957	0.1316
Variable						
Intercept	5.7429*	6.5734*	5.6778*	5.3201*	5.7228*	6.3215*
Bonus-Malus						
Category 1	0	0	0	0	0	0
Category 2	-0.0411*	0.0731*	-0.0528*	0.0710*	-0.0491*	0.0703*
Category 3	0.0930*	0.1834*	0.0587*	0.1805*	0.0131	0.7435*
Category 4	-0.5021*	-1.6594*	-0.4330*	0.6471	-0.7017*	-0.0328
Category 5	0.4053*	0.3033	0.4070*	0.2557	-0.9300	0.3192
Horsepower						
Category 1	0	0	0	0	0	0
Category 2	-0.1990*	0.6084	-0.4225*	0.5130	-0.2831	-0.5771
Category 3	-0.2344*	0.0010	-0.2069*	0.5227	-0.3358*	-0.5610
Category 4	-0.0425	0.1663	-0.0048*	0.2072	0.3042	-0.3585
Category 5	-0.0179	0.1964	0.0108*	0.1035	0.0791	0.2537
Category 6	0.1367*	0.2756*	0.1909*	0.2072*	0.3042	-0.3585
Category 7	0.1905*	0.3020*	0.2470*	0.2094*	0.1574	-0.0103
Category 8	0.3833*	0.3714*	0.3895*	0.3053*	0.3269*	0.0370
Category 9	0.4904*	0.4341*	0.4436*	0.3052	0.5155*	0.0370
Category 10	0.7596*	0.5963*	0.7526*	0.5412*	0.8680*	0.4232*
Category 11	1.3134*	1.5216*	1.0665*	0.5680*	1.4816*	0.3616*
Gender						
Both	0	0	0	0	0	0
Male	-0.0782*	-1.3277	0.1527*	0.0183*	-0.0116	-0.4919
Female	-0.0264	-1.3126	0.2009*	0.0357*	0.0558	-0.5092
Parameter	θ	θ_1	θ_2	θ_1	θ_2	θ_3
	2.2344*	12.4410*	2.2338*	12.3172*	2.1557*	14.3106*

Table 5.8: Results of the Fitted Finite GB2 Mixture Regression Models With One, Two and Three Components, Update of the Posterior Probability

	GB2	Two Component GB2		Three Component GB2		
		C1	C2	C1	C2	C3
		ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
		0.6187	0.3813	0.6037	0.2016	0.1947
Variable						
Intercept	5.4618*	5.2816*	5.9881*	5.2832*	5.7465*	6.4478*
Bonus-Malus						
Category 1	0	0	0	0	0	0
Category 2	0.0204	0.0714*	-0.0499*	0.0676*	0.0594*	-0.1503
Category 3	0.1486*	0.1834*	0.0886*	0.1815*	0.1978*	-0.0742
Category 4	-1.2790*	-1.6719*	-0.6281*	-0.3594*	-2.2511*	-0.3700*
Category 5	0.4202*	0.3100*	0.3184*	0.3180*	0.2087*	0.6076*
Horsepower						
Category 1	0	0	0	0	0	0
Category 2	-0.1850*	-0.0240	-0.1346	-0.0242	0.0898	-0.3642*
Category 3	-0.1394*	-0.0083	-0.1657*	-0.0029	0.0865	-0.4538*
Category 4	0.0435	0.1586*	-0.0001	0.1686*	0.2525*	-0.2700*
Category 5	0.0576	0.1906*	0.0228	0.1977*	0.2625*	-0.2204
Category 6	0.1604*	0.2666*	0.1782*	0.2765*	0.3100*	0.0841
Category 7	0.1973*	0.3039*	0.1661*	0.3137*	0.2982*	0.1044
Category 8	0.2951*	0.3533*	0.3141*	0.3592*	0.3500*	0.3176*
Category 9	0.3843*	0.4018*	0.3996	0.4160*	0.4077*	0.4160*
Category 10	0.5821*	0.6024*	0.5642	0.6310*	0.5246*	0.6743*
Category 11	0.8585*	0.6711*	0.8588	0.8002*	0.7676*	1.1760*
Gender						
Both	0	0	0	0	0	0
Male	-0.1087*	0.0012	0.0208	-0.0174	0.0403	-0.0639
Female	-0.0668*	0.0151	0.0663*	-0.0014	0.0230	0.0538
Parameter	σ	σ_1	σ_2	σ_1	σ_2	σ_3
	6.155*	17.3500*	5.4240*	15.1000*	11.7900*	4.4820*
Parameter	ν	ν_1	ν_2	ν_1	ν_2	ν_3
	1.0063*	0.8781*	0.5224*	0.9317*	1.0896*	0.3979*
Parameter	s	s_1	s_2	s_1	s_2	s_3
	0.5112*	1.1364*	0.9354*	0.9936*	0.9150*	1.1457*

Table 5.9: Results of the Fitted Finite Pareto Mixture Regression Models With One, Two and Three Components, Update of the Posterior Mean

	Pareto	Two Component Pareto		Three Component Pareto		
		C1	C2	C1	C2	C3
		ρ_1	ρ_2	ρ_1	ρ_2	ρ_3
		0.5261	0.4739	0.3625	0.3098	0.3277
Variable						
Intercept	8.6786*	7.6775*	7.6773*	7.6770*	7.6776*	7.6772*
Bonus-Malus						
Category 1	0	0	0	0	0	0
Category 2	-0.0216	-0.0195	-0.0195	-0.0195	-0.0194	-0.0195
Category 3	0.1148*	0.1172	0.1172	0.1173	0.1172	0.1172
Category 4	-0.7272*	-0.7653*	-0.7649*	-0.7642*	-0.7665*	-0.7649*
Category 5	0.4113	0.4124	0.4124	0.4124	0.4124	0.4124
Horsepower						
Category 1	0	0	0	0	0	0
Category 2	-0.2105	-0.2105	-0.2105	-0.2105	-0.2105	-0.2105
Category 3	-0.1975	-0.1919	-0.1919	-0.1920	-0.1918	-0.1919
Category 4	-0.0127	-0.0072	-0.0072	-0.0073	-0.0070	-0.0072
Category 5	0.0057	0.0101	0.0101	0.0100	0.0102	0.0101
Category 6	0.1392	0.1414	0.1413	0.1413	0.1415	0.1413
Category 7	0.1564	0.1564	0.1564	0.1564	0.1565	0.1564
Category 8	0.3354*	0.3314	0.3314	0.3313	0.3315	0.3314
Category 9	0.4405*	0.4354*	0.4354*	0.4353	0.4354	0.4353
Category 10	0.6463*	0.6334*	0.6336*	0.6335*	0.6334*	0.6335*
Category 11	1.0669*	1.0402*	1.0387*	1.0391*	1.0397*	1.0392*
Gender						
Both	0	0	0	0	0	0
Male	-0.0807	-0.0848	-0.0846	-0.0843	-0.0852	-0.0846
Female	-0.0259	-0.0301	-0.0299	-0.0297	-0.0305	-0.0299
Parameter	s'	s'_1	s'_2	s'_1	s'_2	s'_3
	2.9620*	2.0070*	2.0070*	2.0070*	2.0070*	2.0070*

5.5.2 Models Comparison

So far, we have several competing models for the claim frequency and claim severity component. As seen in the previous chapters, the differences between models produce different premiums. Consequently, to distinguish between these models, this section will purpose to compare them in order to select the optimal one for each case.

Claim Frequency Models

The models we have for the claim frequency component are all nested, i.e., Poisson is nested to Negative Binomial model and to the mixture of Poisson with two components, Negative Binomial is nested to the mixture of Negative Binomial model with two components, and the mixture of Poisson with two components is nested to the mixture of Negative Binomial model with two components. In order to accept or reject some models, classical hypothesis/specification tests for nested models can be used (see, Boucher et al., 2007, 2008). The three standard tests are the log-likelihood ratio (LR), Wald, and Score (or Lagrange Multiplier, LM) tests, which are all asymptotically equivalent. One problem with standard specification tests (Wald or LR tests) arises when the null hypothesis is on the boundary of the parameter space. When a parameter is bounded by the H_0 hypothesis, the estimate is also bounded, and the asymptotic normality of the MLE no longer holds under H_0. Consequently a correction must be done, and a mixture of a probability mass of one half on the boundary and $\frac{1}{2}X^2_{1-2\theta}$ (rather than $X^2_{1-\theta}$) should be used for the distribution of the LR statistic. Another standard method of comparing nested models (and also non-nested models) is to use the information criteria, such as the AIC or the SBC.

Firstly, we compare the non-nested claim frequency distributions presented in Section 5.3. The results are depicted in Table 5.10 (Panels A and B). Our findings suggest that the best fit is given by the two component Negative Binomial mixture distribution.

Table 5.10: Claim Frequency Distributions Comparison

Panel A: Based on Likelihood Ratio Test				
Null Hypothesis	Alternative Hypothesis	Value	Sig. Level	Decision
Poisson	NBI	1032.2	0.00	Reject
Poisson	Poisson ($C = 2$)	986.5	0.00	Reject
NBI	NBI ($C = 2$)	1043	0.00	Reject
Poisson ($C = 2$)	NBI ($C = 2$)	1088.7	0.00	Reject
Panel B: Based on AIC, SBC				
Model	df		AIC	SBC
Poisson	1		30368.8	30376.4
NBI	2		29338.6	29353.9
Poisson (C=2)	3		29386.3	29409.3
NBI (C=2)	5		29307.3	29345.6

Secondly, we compare the non-nested claim frequency regression models presented in Section 5.4 employing the LR test and the Global Deviance, AIC and SBC criteria (as suggested by Rigby and Stasinopoulos, 2009). Table 5.11 reports our results with respect to the aforementioned nested comparisons. Specifically, from Panel A and Panel B we observe the superiority of mixture models with two components vs the models with one component and the superiority of the Negative Binomial distribution vs the Poisson distribution. Overall, the best fit is given by the Negative Binomial mixture regression model with two components as suggested by the

LR test and the Global Deviance, AIC and SBC criteria.

Table 5.11: Claim Frequency Regression Models Comparison

Panel A: Based on Likelihood Ratio Test				
Null Hypothesis	Alternative Hypothesis	Value	Sig. Level	Decision
Poisson	NBI	649.2	0.00	Reject
Poisson	Poisson ($C = 2$)	784.2	0.00	Reject
NBI	NBI ($C = 2$)	158.3	0.00	Reject
Poisson ($C = 2$)	NBI ($C = 2$)	23.3	0.00	Reject
Panel B: Based on Global Deviance, AIC, SBC				
Model	df	Global Deviance	AIC	SBC
Poisson	9	29067.1	29085.1	29154.0
NBI	10	28417.9	28437.9	28514.5
Poisson (C=2)	19	28282.9	28320.9	28466.4
NBI (C=2)	21	28259.6	28301.6	28462.4

Claim Severity Models

The Exponential model is nested to Gamma, Weibull, and also to the finite mixtures of two or three Exponential, Gamma and Weibull models. The Pareto model is nested to GB2 and also to the finite mixtures of Pareto, GB2 models with two or three components. Also there are non-nested model comparisons, such as Exp - GB2, Exp - Pareto, Gamma - Pareto, Gamma - GB2, Weibull - Gamma , Weibull - GB2 and Weibull - Pareto for one, two and three components.

Firstly, we compare the claim severity distributions presented in Section 5.3. Our severity distributions comparison is based on the likelihood ratio test (LR) for the nested distributions and AIC, SBC for non-nested distributions. Moreover, for non-nested comparisons we employ the Vuong test (Vuong, 1989). Table 5.12 reports our results with respect to the aforementioned nested comparisons. We observe that there is a superiority of mixture distributions of Gamma, Weibull and GB2 with two or three components vs Gamma, Weibull, and GB2 distributions respectively, while for Exponential and Pareto distributions we do not reject the null hypothesis. However, when we compare the Exponential distribution with one component vs the Gamma distribution with two or three components, Weibull distribution with two or three components, and GB2 distribution with two or three components, we can conclude that mixture distributions are superior to the simple one. Also, the finite mixtures of GB2 distributions employing two and three components provided better fitting performances compared to the Pareto distribution with one component.

Table 5.12: Nested Severity Distributions Comparison Based on Likelihood Ratio Test

Null Hypothesis	Alternative Hypothesis	Value	p-value	Decision
Exponential	Gamma ($C = 1$)	4004.4	0.00	Reject
Exponential	Gamma ($C = 2$)	5844.3	0.00	Reject
Exponential	Gamma ($C = 3$)	6340.8	0.00	Reject
Exponential	Weibull ($C = 1$)	2893.4	0.00	Reject
Exponential	Weibull ($C = 2$)	5480.9	0.00	Reject
Exponential	Weibull ($C = 3$)	69784.8	0.00	Reject
Exponential	Exponential ($C = 2$)	0.00	1.00	No Reject
Exponential	Exponential ($C = 3$)	0.00	1.00	No Reject
Exponential ($C = 2$)	Exponential ($C = 3$)	0.00	1.00	No Reject
Gamma	Gamma ($C = 2$)	1839.9	0.00	Reject
Gamma	Gamma ($C = 3$)	2336.4	0.00	Reject
Gamma ($C = 2$)	Gamma ($C = 3$)	496.5	0.00	Reject
Weibull	Weibull ($C = 2$)	357.4	0.00	Reject
Weibull	Weibull ($C = 3$)	3282.3	0.00	Reject
Weibull ($C = 2$)	Weibull ($C = 3$)	694.8	0.00	Reject
GB2	GB2 ($C = 2$)	1095.5	0.00	Reject
GB2	GB2 ($C = 3$)	1350.5	0.00	Reject
GB2 ($C = 2$)	GB2 ($C = 3$)	255	0.00	Reject
Pareto	GB2 ($C = 1$)	5158.1	0.00	Reject
Pareto	GB2 ($C = 2$)	6253.6	0.00	Reject
Pareto	GB2 ($C = 3$)	6508.6	0.00	Reject
Pareto	Pareto ($C = 2$)	4.9	0.17	No Reject
Pareto	Pareto ($C = 3$)	4.8	0.56	No Reject
Pareto ($C = 2$)	Pareto ($C = 3$)	0.1	0.99	No Reject

In Table 5.13 (Panels A and B) we compare the non-nested severity distributions. Overall, the best distribution according to AIC, SBC and the Vuong test is the GB2 distribution when one, two or three components are used.

Table 5.13: Non - Nested Severity Distributions Comparison

Panel A: Based on AIC, BIC

Model	df	AIC	SBC
Exponential	1	75946.5	75953.1
Gamma	2	71944.1	71957.3
Weibull	2	73055.1	73068.3
GB2	4	70835.2	70861.7
Pareto	2	75989.3	76002.5
Exp (C=2)	3	75950.5	75970.4
Gamma (C=2)	5	70110.2	70143.4
Weibull (C=2)	5	70473.6	70506.8
GB2 (C=2)	9	69749.7	69809.4
Pareto (C=2)	5	76000.2	76033.3
Exp (C=3)	5	75954.5	75987.6
Gamma (C=3)	8	69619.7	69672.7
Weibull (C=3)	8	69784.8	69837.8
GB2 (C=3)	14	69504.7	69597.5
Pareto (C=3)	8	76006.1	76059.2

Panel B: Based on Vuong test

Model 1	Model 2	Vuong Test Statistic	p-value	Decision
Exponential	GB2	-33.81	0.00	GB2
Exponential	Pareto	17.38	0.00	Exp
Gamma	GB2	-9.05	0.00	GB2
Gamma	Pareto	23.64	0.00	Gamma
Weibull	Gamma	-14.38	0.00	Gamma
Weibull	GB2	-12.43	0.00	GB2
Weibull	Pareto	14.73	0.00	Weibull
Exp (C=2)	GB2 (C=2)	-43.96	0.00	GB2
Exp (C=2)	Pareto (C=2)	16.56	0.00	Exp
Gamma (C=2)	GB2 (C=2)	-8.29	0.00	GB2
Gamma (C=2)	Pareto (C=2)	38.48	0.00	Gamma
Weibull (C=2)	Gamma (C=2)	-4.67	0.00	Gamma
Weibull (C=2)	GB2 (C=2)	-6.45	0.00	GB2
Weibull (C=2)	Pareto (C=2)	32.39	0.00	Weibull
Exp (C=3)	GB2 (C=3)	-47.99	0.00	GB2
Exp (C=3)	Pareto (C=3)	16.65	0.00	Exp
Gamma (C=3)	GB2 (C=3)	-33.84	0.00	GB2
Gamma (C=3)	Pareto (C=3)	47.59	0.00	Gamma
Weibull (C=3)	Gamma (C=3)	-27.98	0.00	Gamma
Weibull (C=3)	GB2 (C=3)	13.79	0.00	Weibull
Weibull (C=3)	Pareto (C=3)	45.65	0.00	Weibull

Secondly, we compare the claim severity regression models presented in Section 5.4, employing the LR test for nested models and Global Deviance, AIC, SBC and the Vuong test for non-nested models. As we observe from Table 5.14, there is a superiority of Gamma mixture models with two or three components vs the simple Gamma model, and the same holds for Weibull and GB2 mixture models. When comparing the Exponential model with one component vs the one with two or three components we cannot conclude that there is superiority of mixture models versus the simple model. However, the Gamma, Weibull and GB2 models and their finite mixtures employing two or three components provided better fitting performances than the Exponential model with one component. The two and three component Pareto mixture models are superior to the Pareto model with one component. Nevertheless, when we compare the two component Pareto mixture model vs the one with three components we do not reject the null hypothesis. Also, the two and three component GB2 mixture models are superior to the Pareto model with one component.

Table 5.14: Nested Severity Regression Models Comparison Based on Likelihood Ratio Test

Null Hypothesis	Alternative Hypothesis	Value	p-value	Decision
Exponential	Gamma ($C = 1$)	5544.9	0.00	Reject
Exponential	Gamma ($C = 2$)	9011.0	0.00	Reject
Exponential	Gamma ($C = 3$)	10283.5	0.00	Reject
Exponential	Weibull ($C = 1$)	4639.8	0.00	Reject
Exponential	Weibull ($C = 2$)	8215.6	0.00	Reject
Exponential	Weibull ($C = 3$)	7785.0	0.00	Reject
Exponential	Exponential ($C = 2$)	0.00	1.00	No Reject
Exponential	Exponential ($C = 3$)	0.00	1.00	No Reject
Exponential ($C = 2$)	Exponential ($C = 3$)	0.00	1.00	No Reject
Gamma	Gamma ($C = 2$)	3466.1	0.00	Reject
Gamma	Gamma ($C = 3$)	4738.6	0.00	Reject
Gamma ($C = 2$)	Gamma ($C = 3$)	1272.5	0.00	Reject
Weibull	Weibull ($C = 2$)	3575.8	0.00	Reject
Weibull	Weibull ($C = 3$)	3145.2	0.00	Reject
Weibull ($C = 2$)	Weibull ($C = 3$)	430.6	0.00	Reject
GB2	GB2 ($C = 2$)	2088	0.00	Reject
GB2	GB2 ($C = 3$)	2807	0.00	Reject
GB2 ($C = 2$)	GB2 ($C = 3$)	719	0.00	Reject
Pareto	GB2 ($C = 1$)	6925	0.00	Reject
Pareto	GB2 ($C = 2$)	9013	0.00	Reject
Pareto	GB2 ($C = 3$)	9732	0.00	Reject
Pareto	Pareto ($C = 2$)	367.2	0.00	Reject
Pareto	Pareto ($C = 3$)	232.3	0.00	Reject
Pareto ($C = 2$)	Pareto ($C = 3$)	0.1	1.00	No Reject

Next, we compare the non-nested severity regression models employing Global Deviance, AIC, SBC (Table 5.15, Panel A) and the Vuong test (Table 5.15, Panel B). As mentioned before, these are non-nested model comparisons, such as Exp - GB2, Exp - Pareto, Gamma - Pareto, Gamma - GB2, Weibull - Gamma , Weibull - GB2 and Weibull - Pareto for one, two and three components. Our findings suggest that when one component is used, GB2 is superior to the Exponential, Weibull, Gamma and Pareto distributions. However, when two or three components are used, the Gamma mixture model is superior to the Exponential, Weibull, GB2 and Pareto mixture models. With respect to the Vuong test results, when one component is employed the best model is GB2, while for mixtures with two components our findings suggest that both Gamma and GB2 are superior to the other models. When comparing three component mixtures, the Gamma mixture turns out to be superior to the remaining specifications.

Table 5.15: Non - Nested Severity Regression Models Comparison

Panel A: Based on Global Deviance, AIC, BIC				
Model	df	Global Deviance	AIC	SBC
Exponential	17	75520.6	75554.6	75667.3
Gamma	18	69975.7	70011.7	70131.0
Weibull	18	70880.8	70916.8	71036.1
GB2	20	68821.9	68861.9	68994.5
Pareto	18	75746.9	75782.9	75902.2
Exp (C=2)	35	75520.6	75590.6	75822.6
Gamma (C=2)	37	66509.6	66583.6	66828.8
Weibull (C=2)	37	67305.0	67379.0	67624.3
GB2 (C=2)	41	66733.9	66815.9	67087.7
Pareto (C=2)	37	76114.1	76188.1	76433.4
Exp (C=3)	53	75520.6	75626.6	75978.0
Gamma (C=3)	56	65237.1	65349.1	65720.3
Weibull (C=3)	56	67735.6	67847.6	68218.8
GB2 (C=3)	62	66014.9	66138.9	66549.9
Pareto (C=3)	56	76114.2	76226.2	76597.4

Panel B: Based on Vuong test				
Model 1	Model 2	Vuong Test Statistic	p-value	Decision
Exponential	GB2	-35.49	0.00	GB2
Exponential	Pareto	99.87	0.00	Exp
Gamma	GB2	-10.76	0.00	GB2
Gamma	Pareto	36.79	0.00	Gamma
Weibull	Gamma	-10.71	0.00	Gamma
Weibull	GB2	-11.80	0.00	GB2
Weibull	Pareto	34.39	0.00	Weibull
Exp (C=2)	GB2 (C=2)	-47.35	0.00	GB2
Exp (C=2)	Pareto (C=2)	110.19	0.00	Exp
Gamma (C=2)	GB2 (C=2)	-0.45	0.32	None
Gamma (C=2)	Pareto (C=2)	51.19	0.00	Gamma
Weibull (C=2)	Gamma (C=2)	-6.00	0.00	Gamma
Weibull (C=2)	GB2 (C=2)	-6.42	0.00	GB2
Weibull (C=2)	Pareto (C=2)	49.73	0.00	Weibull
Exp (C=3)	GB2 (C=3)	-37.31	0.00	GB2
Exp (C=3)	Pareto (C=3)	110.12	0.00	Exp
Gamma (C=3)	GB2 (C=3)	8.57	0.00	Gamma
Gamma (C=3)	Pareto (C=3)	49.83	0.00	Gamma
Weibull (C=3)	Gamma (C=3)	-5.84	0.00	Gamma
Weibull (C=3)	GB2 (C=3)	-3.18	0.00	Weibull
Weibull (C=3)	Pareto (C=3)	51.04	0.00	Weibull

Note that the same conclusions for the claim frequency and severity models may not necessarily apply to another observed portfolio.

5.5.3 Optimal BMS Based on the a Posteriori Criteria

In this subsection we consider the premiums determined by the optimal BMS with a frequency and a severity component based on the a posteriori criteria. As we have already mentioned in the previous chapters, all the policies were in force for 3.5 years thus the expected claim frequencies must be multiplied by the exposure to risk $e = \frac{1}{3.5}$ in order to calculate the premiums provided by the claim frequency models. Also, the premiums resulting from both the claim frequency and severity distributions will be divided by the premium when $t = 0$, since we are not so much interested in the absolute premium values as in the differences between various classes. We will present the results so that the premium for a new policyholder is 100.

We consider first the optimal BMS based on the a posteriori frequency component. For the two component Poisson mixture we assume that a policyholder who belongs to the first category is a good risk while a policyholder who belongs to the second category is a bad risk. The maximum likelihood estimation for this model led to a portfolio consisting of $\hat{\pi}_1 = 86.66\%$ of good drivers with claim frequency $\hat{\lambda}_1 = \frac{1}{3.5} \cdot 0.303 = 0.086$ and $\hat{\pi}_2 = 13.34\%$ of bad risks with claim frequency $\hat{\lambda}_2 = \frac{1}{3.5} \cdot 1.669 = 0.476$. In the following example we consider that the specific policyholder is a bad risk[15]. We can now estimate the posterior probability of the second component, $\hat{\pi}_2 (k_1, ..., k_t)$, i.e. that the policyholder is a bad risk, by substituting the estimated parameters of the two component Poisson mixture distribution into (5.6), for $l = n = 2$. Note also that $\hat{\pi}_1 (k_1, ..., k_t) = 1 - \hat{\pi}_2 (k_1, ..., k_t)$. The optimal BMS resulting from this model is obtained by substituting $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\pi}_1 (k_1, ..., k_t)$ and $\hat{\pi}_2 (k_1, ..., k_t)$ into equation Eq. (5.4), for $n = 2$. In Table 5.16 we present the posterior probability that the policyholder is a bad risk and the scaled premiums that must be paid for various number of claims when the age of the policy is up to $t = 7$ years. From Table 5.16, (Panel A) we observe that if the policyholder has a claim free year the probability of being a bad risk is reduced whereas if the policyholder has one or more claims the probability of being a bad risk is increased. From Table 5.16, (Panel B) we see that if the policyholder has a claim free year the premium is reduced while if the policyholder has one or more claims their premium is increased. For example, if the policyholder has one claim in the first year, the posterior probability of being a bad risk increases to 36.44% from 13.33% and they face a malus of 65% in their premium.

[15] The analogous procedure can be applied for a policyholder who belongs in the first category.

Table 5.16: Optimal BMS, Two Component Poisson Mixture Model

Panel A: Posterior Probability of the Second Component

Year	Number of Claims						
	k						
t	0	1	2	3	4	5	6
0	0.1333	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0943	0.3644	0.7593	0.9455	0.9896	0.9981	0.9997
2	0.0658	0.2796	0.6811	0.9216	0.9850	0.9972	0.9995
3	0.0455	0.2080	0.5911	0.8884	0.9777	0.9959	0.9993
4	0.0312	0.1509	0.4946	0.8434	0.9673	0.9939	0.9988
5	0.0214	0.1074	0.3984	0.7847	0.9525	0.9910	0.9983
6	0.0145	0.0753	0.3096	0.7116	0.9314	0.9870	0.9976
7	0.0099	0.0523	0.2328	0.6255	0.9018	0.9806	0.9964

Panel B: Optimal BMS

Year	Number of Claims						
	k						
t	0	1	2	3	4	5	6
0	100.00	0.00	0.00	0.00	0.00	0.00	0.00
1	89.02	165.00	276.11	328.50	340.90	343.28	343.72
2	81.00	141.13	254.10	321.76	339.53	343.03	343.67
3	75.28	121.00	228.78	312.41	337.53	342.65	343.61
4	71.27	104.94	201.62	299.76	334.63	342.10	343.50
5	68.49	92.69	174.57	283.25	330.46	341.29	343.35
6	66.57	83.66	149.56	262.68	324.52	340.10	343.13
7	65.26	77.17	127.97	238.45	316.21	338.36	342.81

As we have already mentioned, the two component Negative Binomial mixture can be derived in two alternative (not equivalent) ways, either by updating the posterior probability (given by Eq. (5.7), for $n = 2$) or by updating the posterior mean (given by Eq. (5.12), for $n = 2$). Firstly, we consider the two component Negative Binomial mixture model derived by updating the posterior probability. The maximum likelihood estimation for this model led to a portfolio consisting of $\hat{\pi}_1 = 56.55\%$ of good drivers with $\hat{\lambda}_1 = \frac{1}{3.5} \cdot 0.389 = 0.111\,, \hat{\alpha}_1 = 0.319$ and $\hat{\pi}_2 = 43.45\%$ of bad risks with $\hat{\lambda}_2 = \frac{1}{3.5} \cdot 0.558 = 0.159, \hat{\alpha}_2 = 0.152$. The posterior probability $\hat{\pi}_2\left(k_1, ..., k_t\right)$ that the policyholder is a bad risk is calculated according to Eq. (5.8), for $l = n = 2$. The premiums resulting from this model are calculated by substituting $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\pi}_1\left(k_1, ..., k_t\right)$ and $\hat{\pi}_2\left(k_1, ..., k_t\right)$ into equation Eq. (5.4), for $n = 2$. Note that due to the existence of the $k_j^{'}$s in Eq. (5.8), the explicit claim frequency history determines the calculation of the posterior probabilities and thus of premium rates and not just the total number of claims as in the case of the two component Poisson mixture. For this reason, in Table 5.17 we specify the exact order of the claims history in order to derive the posterior probability of the second component and the premiums rates using different scenarios for the claim frequency history when the age of the policy is up to $t = 2$ years. For example, considering a bad risk policyholder we observe that they have at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ (i.e. total number of claims $K = 2$ at $t = 2$) then the posterior probability of being a bad risk is increased from 43.45% to 56.16% and their premium increases from 100 to 104.64 while if they have $k_1 = 1, k_2 = 1$ claim frequency history (i.e. total number of claims $K = 2$ at $t = 2$) then the posterior probability of being a bad risk is increased from 43.45% to 59.47% and their premium is increased from 100 to 105.85.

Table 5.17: Optimal BMS, Two Component Negative Binomial Mixture Model, Update of the Posterior Probability

Year	Number of Claims k_t	Posterior Probability of the 2nd component	Optimal BMS
$t = 0$	$k_0 = 0$	0.4345	100
	$k_1 = 0$	0.4227	99.56
$t = 1$	$k_1 = 1$	0.5149	102.94
	$k_1 = 2$	0.5734	105.08
	$k_1 = 0, k_2 = 0$	0.4109	99.13
$t = 2$	$k_1 = 0, k_2 = 1$	0.5029	102.50
	$k_1 = 0, k_2 = 2$	0.5616	104.64
	$k_1 = 1, k_2 = 0$	0.5029	102.50
$t = 2$	$k_1 = 1, k_2 = 1$	0.5947	105.85
	$k_1 = 1, k_2 = 2$	0.6501	107.88
	$k_1 = 2, k_2 = 0$	0.5616	104.64
$t = 2$	$k_1 = 2, k_2 = 1$	0.6501	107.88
	$k_1 = 2, k_2 = 2$	0.7017	109.77

Secondly, we consider the two component Negative Binomial mixture derived by updating the posterior mean. In this case, the maximum likelihood estimators of the parameters are $\hat{\pi}_1 = 0.5655, \hat{\pi}_2 = 0.4345, \hat{\alpha}_1 = 3.1299, \hat{\alpha}_2 = 6.579, \hat{\tau}_1 = \frac{3.1299}{\frac{1}{3.5}\cdot 0.3891} = 28.154$ and $\hat{\tau}_2 = \frac{6.5790}{\frac{1}{3.5}\cdot 0.5583} = 41.244$. In the following example we assume that a policyholder belongs to a group of policyholders observed for 7 years whose number of claims range from 1 to 6. The optimal BMS resulting from this model will be defined by substituting the above values into Eq. (5.14), for $n = 2$ and is presented in Table 5.18. From Table 5.18 we observe that this system can be considered generous with good drivers and strict with bad drivers. For instance, the bonuses given for the first claim free year are 2.88% of the basic premium and drivers who have one accident over the first year will have to pay a malus of 19.58% of the basic premium.

Table 5.18: Optimal BMS, Two Component Negative Binomial Mixture Model, Update of the Posterior Mean

	Number of Claims						
Year				k			
t	0	1	2	3	4	5	6
0	100.00	0.00	0.00	0.00	0.00	0.00	0.00
1	97.12	119.58	142.04	164.49	186.95	209.41	231.86
2	94.42	116.21	138.00	159.79	181.58	203.37	225.16
3	91.86	113.03	134.19	155.35	176.51	197.67	218.84
4	89.45	110.01	130.59	151.16	171.73	192.30	212.87
5	87.15	107.17	127.18	147.20	167.21	187.22	207.24
6	84.99	104.47	123.95	143.44	162.92	182.41	201.89
7	82.92	101.90	120.88	139.87	158.86	177.84	196.83

The BMSs presented in Tables 5.16, 5.17 and 5.18 are financially balanced and do not differ much. Also, the system presented in Table 5.18 shows much less extreme premiums in comparison with the one obtained from the traditional Negative Binomial distribution, which was presented in Table 3.6 of Chapter 3. For example, the bonuses given for the first claim free year are 11.28% and 2.88% of the basic premium in the case of the Negative Binomial (Table 3.6, Chapter3) and the two component Negative Binomial mixture (Table 5.18) models respectively. On the contrary, policyholders who had one claim over the first year of observation will have to pay a malus of 70.14% and 19.58% of the basic premium in the case of the Negative Binomial and the two component Negative Binomial models respectively.

In terms of the a posteriori claim severity component, we consider first the case of the two and three component Exponential, Gamma, Weibull and GB2 mixture models derived by updating the posterior probability. The estimation of the parameters of these models led to the following results:

- In the case of the two component Exponential mixture, the resulting portfolio consists of $\hat{\rho}_1 = 47.20\%$ of good drivers with claim severity $\hat{y}_1 = 327.99$ and $\hat{\rho}_2 = 52.80\%$ of bad risks with claim severity $\hat{y}_2 = 327.99$.

- In the case of the two component Gamma mixture, the resulting portfolio consists of $\hat{\rho}_1 = 45.60\%$ of good drivers with claim severity $\hat{y}_1 = 247.89$, $\hat{\theta}_1 = 0.1497$ and $\hat{\rho}_2 = 54.40\%$ of bad risks with claim severity $\hat{y}_2 = 395.04$, $\hat{\theta}_2 = 0.5906$.

- In the case of the two component Weibull mixture, the resulting portfolio consists of $\hat{\rho}_1 = 47.21\%$ of good drivers with claim severity $\hat{y}_1 = 243.71$, $\hat{\theta}_1 = 2.018$and $\hat{\rho}_2 = 52.79\%$ of bad risks with claim severity $\hat{y}_2 = 403.03$, $\hat{\theta}_2 = 0.5294$.

- In the case of the two component GB2 mixture, the resulting portfolio consists of $\hat{\rho}_1 = 49.26\%$ of good drivers with claim severity $\hat{y}_1 = 237.46$, $\hat{\sigma}_1 = 10.1600$, $\hat{\nu}_1 = 1.2058$, $\hat{s}_1 = 0.9461$ and $\hat{\rho}_2 = 50.74\%$ of bad risks with claim severity $\hat{y}_2 = 499.70$, $\hat{\sigma}_2 = 4.752$, $\hat{\nu}_2 = 0.3940$, $\hat{s}_2 = 0.9305$.

- In the case of the three component Exponential mixture, the resulting portfolio consists of $\hat{\rho}_1 = 32.92\%$ of good risks with claim severity $\hat{y}_1 = 328.00$, $\hat{\rho}_2 = 33.17\%$ of average risks with claim severity $\hat{y}_2 = 328.00$ and $\hat{\rho}_3 = 33.91\%$ of bad risks with claim severity $\hat{y}_3 = 328.00$.

- In the case of the three component Gamma mixture, the resulting portfolio consists of $\hat{\rho}_1 = 52.68\%$ of good risks with claim severity $\hat{y}_1 = 243.47$, $\hat{\theta}_1 = 0.1615$, $\hat{\rho}_2 = 14.96\%$ of average risks with claim severity $\hat{y}_2 = 418.22$, $\hat{\theta}_2 = 0.9260$ and $\hat{\rho}_3 = 32.36\%$ of bad risks with claim severity $\hat{y}_3 = 423.69$, $\hat{\theta}_3 = 0.2967$.

- In the case of the three component Weibull mixture, the resulting portfolio consists of $\hat{\rho}_1 = 47.53\%$ of good risks with claim severity $\hat{y}_1 = 239.61$, $\hat{\theta}_1 = 2.032$, $\hat{\rho}_2 = 37.74\%$ of average risks with claim severity $\hat{y}_2 = 389.16$, $\hat{\theta}_2 = 1.203$and $\hat{\rho}_3 = 14.73\%$ of bad risks with claim severity $\hat{y}_3 = 455.32$, $\hat{\theta}_3 = 0.114$.

- In the case of the three component GB2 mixture, the resulting portfolio consists of $\hat{\rho}_1 = 3.66\%$ of good risks with claim severity $\hat{y}_1 = 36.05$, $\hat{\sigma}_1 = 3.862$, $\hat{\nu}_1 = 1.6093$, $\hat{s}_1 = 0.5728$, $\hat{\rho}_2 = 63.28\%$ of average risks with claim severity $\hat{y}_2 = 247.65$, $\hat{\sigma}_2 = 9.475$, $\hat{\nu}_2 = 0.9103$, $\hat{s}_2 = 1.0131$ and $\hat{\rho}_3 = 33.06\%$ of bad risks with claim severity $\hat{y}_3 = 377.29$, $\hat{\sigma}_3 = 5.623$, $\hat{\nu}_3 = 1.5923$, $\hat{s}_3 = 0.7191$.

Based on the above estimates we are now able to derive the optimal BMSs resulting from these models. Let us see an example in order to understand better how such systems work. Consider that a policyholder is observed over the first year their presence in the portfolio, has one claim, i.e. $K = 1$, and the claim amount, $x1$, of their accident ranges from 150 to 500 euros. Firstly, we consider the case of the two component mixture models. The posterior probability $\rho_2(x_1)$ that the policyholder is a bad risk is given by the Eqs (5.17, 5.21, 5.19 and 5.23), for $l = n = 2$, for the case of the two component Exponential, Weibull, Gamma and GB2 mixtures respectively[16]. The results are displayed in Table 5.19.

[16]Note that $\rho_1(x_1) = 1 - \rho_2(x_1)$.

Table 5.19: Posterior Probability of the Second Component, Two Component Mixture Models for Assessing Claim Severity, One Claim in the First Year of Observation

Claim Size	Exponential	Gamma	Weibull	GB2
150	0.5280776	0.9291123	0.6072640	0.8272659
175	0.5280776	0.6120109	0.3866989	0.5091249
200	0.5280776	0.3098473	0.2391092	0.2487356
225	0.5280776	0.1978124	0.1695440	0.1538288
250	0.5280776	0.1853316	0.1662397	0.1581051
275	0.5280776	0.2420095	0.2736654	0.2369350
300	0.5280776	0.3878595	0.6741644	0.3887037
325	0.5280776	0.6271392	0.9839598	0.5776653
350	0.5280776	0.8511934	0.9999589	0.7419573
375	0.5280776	0.9601340	1.0000000	0.8519773
400	0.5280776	0.9918783	1.0000000	0.9159169
425	0.5280776	0.9986305	1.0000000	0.9512052
450	0.5280776	0.9998013	1.0000000	0.9706653
475	0.5280776	0.9999747	1.0000000	0.9816513
500	0.5280776	0.9999971	1.0000000	0.9880639

Secondly, we consider the case of the three component mixture models. The posterior probability $\rho_2(x_1)$ that the policyholder is an average risk, i.e. they belong to the second category of risks, is given by the Eqs (5.17, 5.21, 5.19 and 5.23), for $l = 2$ and $n = 3$ and the posterior probability $\rho_3(x_1)$ that the policyholder is a bad risk, i.e. they belong to the third category of risks, is given by the Eqs (5.17, 5.21, 5.19 and 5.23), for $l = 3$ and $n = 3$, for the case of the three component Exponential, Weibull, Gamma and GB2 mixtures respectively, using the same categories as before[17]. The results are depicted in Table 5.20 (Panels A and B).

[17] Note also that $\rho_1(x_1) = 1 - \rho_2(x_1) - \rho_3(x_1)$.

Table 5.20: Posterior Probability of the Second and the Third Component, Three Component Mixture Models for Assessing Claim Severity, One Claim in the First Year of Observation

Claim Size	Exponential	Gamma	Weibull	GB2
Panel A: Second component				
150	0.3316923	0.5404953	0.2723886	0.9327530
175	0.3316923	0.1672704	0.1934542	0.9799350
200	0.3316923	0.0640814	0.1400271	0.9875993
225	0.3316923	0.0392743	0.1207626	0.9854583
250	0.3316923	0.0364328	0.1489229	0.9728115
275	0.3316923	0.0460964	0.3086360	0.9319897
300	0.3316923	0.0689915	0.7227783	0.8220258
325	0.3316923	0.1001418	0.8722634	0.6141273
350	0.3316923	0.1207824	0.8822294	0.3766626
375	0.3316923	0.1247956	0.8854009	0.2056107
400	0.3316923	0.1228099	0.8844872	0.1109158
425	0.3316923	0.1219009	0.8792389	0.0625614
450	0.3316923	0.1240281	0.8689408	0.0374914
475	0.3316923	0.1295363	0.8522981	0.0238178
500	0.3316923	0.1385600	0.8272656	0.0159125
Panel B: Third component				
150	0.3391230	0.0704857	0.1927821	0.0063898
175	0.3391230	0.0575825	0.0948670	0.0073214
200	0.3391230	0.0471946	0.0503026	0.0088190
225	0.3391230	0.0527043	0.0333150	0.0130332
250	0.3391230	0.0784905	0.0329080	0.0261612
275	0.3391230	0.1439094	0.0568181	0.0669633
300	0.3391230	0.2868004	0.1151328	0.1766946
325	0.3391230	0.5163146	0.1248151	0.3843838
350	0.3391230	0.7270201	0.1177686	0.6218652
375	0.3391230	0.8324215	0.1145991	0.7930958
400	0.3391230	0.8674829	0.1155128	0.8879790
425	0.3391230	0.8761783	0.1207611	0.9364717
450	0.3391230	0.8756294	0.1310592	0.9616290
475	0.3391230	0.8704078	0.1477020	0.9753513
500	0.3391230	0.8614317	0.1727344	0.9832778

The results presented in Tables 5.19 and 5.20 indicate that the behavior of the posterior probabilities is irregular for small claim sizes. Thus the behavior of the premium formulas is also expected to be irregular. For this purpose we consider that for claim sizes smaller than 235 euros (for all models) the policyholder always pays the same premium calculated for an accident of claim amount 235 euros. This value was chosen since for claim sizes up to 235 euros both the posterior probabilities in the case of the two and three component mixture models indicate an irregular behavior. Furthermore, due to the irregular behavior of the posterior probabilities in the case of the three component Weibull mixture, it was found that there was a very small drop in the premium rates which range from 350 to 400 compared to the premium rate of 325. For this reason, we assumed that these values are equal to that of 325 euros. The premium rates resulting from the two component mixture models are derived by Eq. (5.15), for $n = 2$, and those resulting from the three component mixtures are obtained by Eq. (5.15), for $n = 3$. The results are displayed in Table 5.21 (Panels A and B). More specifically, in Panel A we report the premiums for the two component mixture models for a bad risk, i.e. belonging to the second category of risks and in Panel B we depict the premiums for the three component mixture models for a bad risk, i.e. belonging to the third category of risks using the same categories as before. From Table 5.21 we observe that the premium is equal to 100, the basic premium, in the case of the two and three component Exponential mixture, revealing the unnecessity of the two and three components. As expected, in the case of the two and three component Gamma, Weibull and GB2 mixtures the premium values increase proportionally to the claim severity. For instance, for one claim size of 325 in the first year the premium increases from 100 to 103.73, 122.16, 103.19, in the two component mixture Gamma, Weibull and GB2 models, respectively. For the three component mixtures, we have to note that for one claim size of 325 in the first year the premium increases from 100 to 107.96, 121.09 and 106.16, in the case of the Gamma, Weibull and GB2 models, respectively. For claim sizes up to 275 euros all models reward the policyholder with a bonus[18].

[18]Note that the mean claim size is 327.95 euros.

Table 5.21: Optimal BMS, Two and Three Component Mixture Models for Assessing Claim Severity, Update of the Posterior Probability, One Claim in the First Period of Observation

Panel A: Two component				
Claim Size	Exponential	Gamma	Weibull	GB2
150	100	83.85582	82.11809	83.63431
175	100	83.85582	82.11809	83.6343
200	100	83.85582	82.11809	83.6343
225	100	83.85582	82.11809	83.6343
250	100	83.90725	82.42309	84.1708
275	100	86.45047	87.64366	87.74360
300	100	92.99496	107.10672	94.62224
325	100	103.73179	122.16185	103.18658
350	100	113.78542	122.93936	110.63282
375	100	118.67374	122.94136	115.61927
400	100	120.09815	122.94136	118.51722
425	100	120.40114	122.94136	120.11659
450	100	120.45367	122.94136	120.99859
475	100	120.46145	122.94136	121.49651
500	100	120.46246	122.94136	121.78715
Panel B: Three component				
150	100	79.48091	80.81860	77.17606
175	100	79.48091	80.81860	77.17606
200	100	79.48091	80.81860	77.17606
225	100	79.48091	80.81860	77.17606
250	100	80.49958	82.04732	77.94160
275	100	84.60968	90.90667	81.15704
300	100	93.68239	113.63694	89.79472
325	100	107.95547	121.09342	106.15590
350	100	120.63485	121.09342	124.87792
375	100	126.64114	121.09342	138.38609
400	100	128.46215	121.09342	145.87617
425	100	128.89157	121.14488	149.70645
450	100	128.97476	121.35269	151.69440
475	100	128.98133	121.68854	152.77880
500	100	128.96888	122.19369	153.40482

Let us now consider the two and the three component Pareto mixture distributions for assessing claim severity. These models were derived by updating the posterior mean claim severity and the maximum likelihood estimates of their parameters can be found in the preceding. In the following example, we calculate the premiums that must be paid by a policyholder who is observed for the first year of her presence in the portfolio, has one accident and the claim amount of her accident ranges from 150 to 1000 euros. The optimal BMSs resulting from the two and three component Pareto mixture distributions will be defined by Eq. (5.27), for $n = 2$ and $n = 3$ respectively, and are presented in Table 5.22. It is obvious that these optimal BMSs as well allow the discrimination of the premium with respect to the severity of the claims. These systems are financially balanced and do not differ much with the one obtained in Chapter 4 when the Pareto distribution was used (see Table 4.2, Chapter 4). For instance, for one claim size of 500 in the first year the premium increases from 100 to 100.59, 100.67 and 100.66 in the case of the Pareto, two component Pareto mixture and three component Pareto mixture models respectively. Compared to the BMSs provided by the two and three component mixture models derived by updating the posterior probability, (see Table 5.21), the two and three component Pareto mixture models show much less extreme premiums. The systems presented in Tables 5.21 and 5.22 are fair because each insured pays a premium proportional to her claim severity taking into account through the Bayes theorem all the information available for her claim size history.

Table 5.22: Optimal BMS, Two and Three Component Pareto Mixture Models for Assessing Claim Severity, One Claim in the First Year of Observation

Claim Size	Optimal BMS Two Component Pareto	Optimal BMS Three Component Pareto
150	99.29018	99.28784
175	99.38903	99.38621
200	99.48789	99.48458
225	99.58674	99.58294
250	99.68559	99.68131
275	99.78445	99.77967
300	99.88330	99.87804
325	99.98216	99.97640
350	100.08101	100.07477
375	100.17987	100.17314
400	100.27872	100.27150
425	100.37758	100.36987
450	100.47643	100.46823
475	100.57529	100.56660
500	100.67414	100.66496
525	100.77299	100.76333
550	100.87185	100.86170
575	100.97070	100.96006
600	101.06956	101.05843
625	101.16841	101.15679
650	101.26727	101.25516
675	101.36612	101.35352
700	101.46498	101.45189
725	101.56383	101.55026
750	101.66269	101.64862
775	101.76154	101.74699
800	101.86040	101.84535
825	101.95925	101.94372
850	102.05810	102.04208
875	102.15696	102.14045
900	102.25581	102.23882
925	102.35467	102.33718
950	102.45352	102.43555
975	102.55238	102.53391
1000	102.65123	102.63228

Finally, in Tables 5.23 and 5.24 we present the optimal BMS based on the a posteriori frequency and severity component. As we have mentioned before, the resulting premium rates are calculated via the product of the expected claim frequency and the expected claim severity with independence between the two components assumed. We observe that for one claim size of 350 in the first year the premium increases from 100 to 165.00, 187.75, 202.85, 182.54, 165.13 in the two component Poisson mixture model and the corresponding two component severity models (Table 5.23), to 102.94, 117.13, 126.55, 113.88, 103.02 in the two component Negative Binomial mixture model (updating the posterior probability) and the corresponding two component severity models (Table 5.24, Panel A), to 119.58, 136.07, 147.01, 132.30, 119.68 in the two component Negative Binomial mixture model (updating the posterior mean) and the corresponding two component severity models (Table 5.24, Panel B).

Table 5.23: Optimal BMS Based on the Two Component Poisson Mixture Model for Assessing Claim Frequency and the Various Two Component Mixture Models for Assessing Claim Severity, One Claim in the First Year of Observation

Two Component Poisson Mixture Model					
Claim Size	POIS-EXP	POIS-GA	POIS-WEI	POIS-GB2	POIS-PA
150	165.00	138.3621	135.4948	137.9966	163.8288
175	165.00	138.3621	135.4948	137.9966	163.9919
200	165.00	138.3621	135.4948	137.9966	164.1550
225	165.00	138.3621	135.4948	137.9966	164.3181
250	165.00	138.4469	135.9981	138.8818	164.4812
275	165.00	142.6433	144.6120	144.7769	164.6443
300	165.00	153.4418	176.7261	156.1266	164.8075
325	165.00	171.1578	201.5671	170.2579	164.9706
350	165.00	187.7459	202.8499	182.5441	165.1337
375	165.00	195.8117	202.8532	190.7718	165.2968
400	165.00	198.1620	202.8532	195.5534	165.4599
425	165.00	198.6619	202.8532	198.1924	165.6230
450	165.00	198.7486	202.8532	199.6477	165.7861
475	165.00	198.7614	202.8532	200.4692	165.9492
500	165.00	198.7631	202.8532	200.9488	166.1123

Table 5.24: Optimal BMS Based on the Alternative Two Component Negative Binomial Mixture Models for Assessing Claim Frequency and the Various Two Component Mixture Models for Assessing Claim Severity, One Claim in the First Year of Observation

Panel A: Two Component Negative Binomial Mixture Model (Update of the Posterior Probability)

Claim Size	NB-EXP	NB-GA	NB-WEI	NB-GB2	NB-PA
150	102.94	86.32118	84.53236	86.09316	102.2093
175	102.94	86.32118	84.53236	86.09316	102.3111
200	102.94	86.32118	84.53236	86.09316	102.4128
225	102.94	86.32118	84.53236	86.09316	102.5146
250	102.94	86.37412	84.84633	86.64540	102.6164
275	102.94	88.99211	90.22038	90.32327	102.7181
300	102.94	95.72901	110.25566	97.40414	102.8199
325	102.94	106.78150	125.75341	106.22027	102.9216
350	102.94	117.13071	126.55377	113.88542	103.0234
375	102.94	122.16275	126.55583	119.01848	103.1252
400	102.94	123.62904	126.55583	122.00162	103.2269
425	102.94	123.94093	126.55583	123.64802	103.3287
450	102.94	123.99501	126.55583	124.55595	103.4304
475	102.94	124.00302	126.55583	125.06851	103.5322
500	102.94	124.00405	126.55583	125.36769	103.6340

Panel B: Two Component Negative Binomial Mixture Model (Update of the Posterior Mean)

Claim Size	NB-EXP	NB-GA	NB-WEI	NB-GB2	NB-PA
150	119.58	100.2748	98.19681	100.0099	118.7312
175	119.58	100.2748	98.19681	100.0099	118.8494
200	119.58	100.2748	98.19681	100.0099	118.9676
225	119.58	100.2748	98.19681	100.0099	119.0858
250	119.58	100.3363	98.56153	100.6514	119.2040
275	119.58	103.3775	104.80429	104.9238	119.3222
300	119.58	111.2034	128.07821	113.1493	119.4405
325	119.58	124.0425	146.08114	123.3905	119.5587
350	119.58	136.0646	147.01088	132.2947	119.6769
375	119.58	141.9101	147.01327	138.2575	119.7951
400	119.58	143.6134	147.01327	141.7229	119.9133
425	119.58	143.9757	147.01327	143.6354	120.0315
450	119.58	144.0385	147.01327	144.6901	120.1497
475	119.58	144.0478	147.01327	145.2855	120.2679
500	119.58	144.0490	147.01327	145.6331	120.3861

5.5.4 Optimal BMS Based Both on the a Priori and the a Posteriori Criteria

In this subsection we consider the premiums determined by the generalized optimal BMS with a frequency and a severity component when both the a priori and the a posteriori rating variables are used. The expected claim frequencies are multiplied again by the exposure to risk $e = \frac{1}{3.5}$ in order to derive the generalized premiums. The premiums determined by the claim frequency and severity models are divided again by the premium when $t = 0$, in order to observe the percentage change in the premiums after one or more claims.

In terms of the claim frequency component we consider first the two component Poisson mixture regression model for assessing claim frequency. In the following example Table 5.25 (Panels A and B), we examine a group of policyholders who share the following common characteristics: The policyholder i is a woman, who belongs to the first Bonus-Malus category and has a car with horsepower between 0-33. The estimation for this group led to a portfolio consisting of $\hat{\pi}_1 = 90.07\%$ of good drivers with annual expected claim frequency $\hat{\lambda}_{1,i}^{j} = e \exp\left(c_{1,i}^{j} \beta_1^{j}\right) = \frac{1}{3.5} \cdot \exp\left(-1.359 + 0.144\right) = 0.084$ and $\hat{\pi}_2 = 9.93\%$ of bad risks with claim frequency $\hat{\lambda}_{2,i}^{j} = e \exp\left(c_{2,i}^{j} \beta_2^{j}\right) = \frac{1}{3.5} \cdot \exp\left(0.736 + (-0.108)\right) = 0.535$, where j represents the age of the policy, We assume that the specific policyholder belongs to the second category of risks, her number of claims range from 1 to 6 and that the age of the policy j is up to 7 years. The posterior probability $\hat{\pi}_2\left(K_i^1, ..., K_i^t; c_{2,i}^1, ..., c_{2,i}^{t+1}\right)$, i.e. that she is a bad risk, will be estimated according to Eq. (5.34), for $n = 2$. Note also that $\hat{\pi}_1\left(K_i^1, ..., K_i^t; c_{1,i}^1, ..., c_{1,i}^{t+1}\right) = 1 - \hat{\pi}_2\left(K_i^1, ..., K_i^t; c_{2,i}^1, ..., c_{2,i}^{t+1}\right)$. The optimal BMS resulting from this model is obtained by substituting $\hat{\lambda}_{1,i}^{j}, \hat{\lambda}_{2,i}^{j}, \hat{\pi}_1\left(K_i^1, ..., K_i^t; c_{1,i}^1, ..., c_{1,i}^{t+1}\right)$ and $\hat{\pi}_2\left(K_i^1, ..., K_i^t; c_{2,i}^1, ..., c_{2,i}^{t+1}\right)$ into Eq. (5.32), for $n = 2$. From Table 5.25, (Panel A) we observe that if the policyholder has a claim free year the probability of being a bad risk is reduced whereas if the policyholder has one or more claims the probability of being a bad risk is increased. From Table 5.25, (Panel B) we see that if the policyholder has a claim free year her premium is reduced while if the policyholder has one or more claims the premium is increased. For example, if the policyholder i has one claim in the first year, the posterior probability of being a bad risk increases to 30.73% from 9.92% and she faces a malus of 72.44% in her premium. Furthermore, it is interesting to compare this BMS with the one obtained when only the a posteriori frequency component is used. Using the two component Poisson mixture we saw in Table 5.16 that a policyholder with one claim in the first year faces a malus of 65% of the basic premium. Using the two component Poisson mixture regression model, a woman who belongs to the first Bonus-Malus category, with a car with horsepower 0-33 and one claim in the first year faces a malus of 72.44% in her premium. This system is more fair since it considers all the important a priori and a posteriori information for each policyholder in order to estimate her risk of having an accident and thus it permits the differentiation of the premiums for various number of claims based on the expected claim frequency of each policyholder as these are estimated both from the a priori and the a posteriori classification criteria.

Table 5.25: Optimal BMS, Two Component Poisson Mixture Regression Model

Panel A: Posterior Probability of the Second Component

Year	Number of Claims						
	k						
t	0	1	2	3	4	5	6
0	0.0992	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0655	0.3073	0.7371	0.9466	0.9915	0.9985	0.9997
2	0.0427	0.2203	0.6410	0.9186	0.9861	0.9977	0.9996
3	0.0276	0.1525	0.5322	0.8779	0.9784	0.9965	0.9994
4	0.0178	0.1028	0.4201	0.8207	0.9661	0.9945	0.9991
5	0.0114	0.0680	0.3157	0.7447	0.9485	0.9914	0.9986
6	0.0073	0.0444	0.2271	0.6501	0.9215	0.9867	0.9978
7	0.0046	0.0287	0.1577	0.5420	0.8820	0.9792	0.9966

Panel B: Optimal BMS

Year	Number of Claims						
	k						
t	0	1	2	3	4	5	6
0	100.00	0.00	0.00	0.00	0.00	0.00	0.00
1	88.27	172.44	322.10	395.03	410.54	413.13	413.55
2	80.33	142.15	288.65	385.29	408.81	412.85	413.50
3	75.07	118.54	250.74	371.12	406.13	412.42	413.43
4	71.64	101.24	211.73	351.23	402.00	411.73	413.32
5	69.41	89.13	175.39	324.74	395.71	410.66	413.15
6	67.98	80.91	144.54	291.80	386.30	409.00	412.88
7	67.06	75.45	120.34	254.16	372.57	406.41	412.46

The two component Negative Binomial mixture regression model can be derived in two alternative (not equivalent) ways, either by updating the posterior probability (given by Eq. (5.35), for $n = 2$) or by updating the posterior mean (given by Eq. (5.41), for $n = 2$). In what follows we present the optimal BMSs obtained by the two component Negative Binomial mixture regression model for both cases, i.e. for the case of updating the posterior probability and for the case of updating the posterior mean. Let us see an example in order to understand better how these systems work. Firstly, we consider the case of the two component Negative Binomial mixture regression model derived by updating the posterior probability. The estimation for the same group of policyholders we described before led to a portfolio consisting of $\hat{\pi}_1 = 89.07\%$ of good drivers with $\hat{\lambda}^j_{1,i} = e \exp\left(c^j_{1,i}\beta^j_1\right) = \exp(-1.291 + 0.156) = 0.091$, $\hat{\alpha}^j_{1,i} = 0.207$ and $\hat{\rho}_2 = 10.93\%$ of bad risks with $\hat{\lambda}^j_{2,i} = e \exp\left(c^j_{2,i}\beta^j_2\right) = \frac{1}{3.5} \cdot \exp\left(0.579 + (-0.143)\right) = 0.442$, $\hat{\alpha}^j_{2,i} = 0.250$. The posterior probability $\hat{\pi}_2\left(K^1_i, ..., K^t_i; c^1_{2,i}, ..., c^{t+1}_{2,i}\right)$ will be estimated for a bad risk according to Eq. (5.36), for $n = 2$. Secondly, we consider the case of the two component Negative Binomial mixture regression model derived by updating the posterior mean. Consider now a new group of policyholders who share the following common characteristics: The policyholder

i is a woman, who has a car with horsepower 0-33 and her Bonus-Malus class varies over time, starting from BM class 1[19]. The mean (or location) parameter of this model is given by $E(K_i^j|c_{z,i}^j) = e\sum_{z=1}^{2}\pi_z \exp\left(c_{z,i}^j\beta_z^j\right)$, where $e = \frac{1}{3.5}$. Due to the fact that the drivers were divided into five categories according to their BM class, the estimation for this group led to the following results presented in Table 5.26.

Table 5.26: Women, Horse Power 0-33

Bonus-Malus Category	$e\exp\left(c_{1,i}^j\beta_1^j\right)$	$e\exp\left(c_{2,i}^j\beta_2^j\right)$
1	0.091	0.442
2	0.206	0.440
3	0.224	0.898
4	0.023	0.284
5	0.818	1.353

Based on the above estimates for these two different groups of individuals we are now able to calculate the premium rates according to Eqs (5.32 and 5.49), for $n = 2$, for the case of updating the posterior probability and the case of updating the posterior mean respectively. The results are displayed in Table 5.27. Note that in both cases the claim frequency history plays an important role. For the case of updating the posterior probability, this follows from the existence of the $K_i^{j\prime}$s in Eq. (5.36) and for the case of updating the posterior mean this follows from Eq. (5.49) due to the fact that Bonus-Malus class varies substantially from one period to another depending on the number of claims K_i^j of each policyholder i for period j. For this reason in Table 5.27 we specify the exact order of the claims history in order to calculate the premiums that must be paid by these two different groups of policyholders. For instance, for the case of updating the posterior probability and considering a bad risk policyholder, we observe that if she has at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ (i.e. total number of claims $K = 2$ at $t = 2$) then the posterior probability of being a bad risk is increased from 10.93% to 56.26% and her premium increases from 100 to 222.03 while if she has $k_1 = 1, k_2 = 1$ claim frequency history (i.e. total number of claims $K = 2$ at $t = 2$) then the posterior probability of being a bad risk is increased from 10.93% to 55.40% and her premium is increased from 100 to 219.72. For the case of updating the posterior mean we see that if she has at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ then her premium increases from 100 to 237.37, while if she has $k_1 = 1, k_2 = 1$ claim frequency history her premium reaches 233.12.

[19] Recall that BM class 1 corresponds to BM category 1 according to the grouping of the levels of the BM class explanatory variable.

Table 5.27: Optimal BMS, Two Component Negative Binomial Mixture Regression Model

Year	Number of Claims k_t	Posterior Probability of the 2nd component	Optimal BMS Posterior Prob.	Optimal BMS Posterior Mean
$t = 0$	$k_0 = 0$	0.1093	100	100
	$k_1 = 0$	0.0811	92.43	95.14
$t = 1$	$k_1 = 1$	0.2807	146.16	208.54
	$k_1 = 2$	0.6410	243.16	245.46
	$k_1 = 0, k_2 = 0$	0.0598	86.67	90.99
$t = 2$	$k_1 = 0, k_2 = 1$	0.2194	129.65	201.72
	$k_1 = 0, k_2 = 2$	0.5626	222.03	237.37
	$k_1 = 1, k_2 = 0$	0.2194	129.65	109.49
$t = 2$	$k_1 = 1, k_2 = 1$	0.5540	219.72	233.12
	$k_1 = 1, k_2 = 2$	0.8503	299.50	342.75
	$k_1 = 2, k_2 = 0$	0.5626	222.03	233.12
$t = 2$	$k_1 = 2, k_2 = 1$	0.8503	299.50	342.75
	$k_1 = 2, k_2 = 2$	0.9629	329.81	388.13

The BMSs obtained by the two component Negative Binomial mixture regression model (Table 5.27) for both cases do not differ much from the system provided by the two component Poisson mixture regression model (Table 5.25). Nevertheless, for the case of updating the posterior mean, the BMS determined by the two component Negative Binomial mixture regression model differs from the one obtained by the Negative Binomial regression model, which was presented in Table 3.10 of Chapter 3. For example, a woman who at $t = 2$ has claim frequency history $k_1 = 0, k_2 = 2$ faces a malus of 261.15% and 137.37% of the basic premium in the case of the Negative Binomial (see Table 3.10, Chapter 3) and the two component Negative Binomial mixture models respectively, while a woman who at $t = 2$ has claim frequency history $k_1 = 1, k_2 = 1$ faces a malus of 239.90%, 133.12% of the basic premium in the case of the Negative Binomial and two component Negative Binomial mixture regression models respectively. Let us now compare the BMSs presented in Table 5.27 with the systems presented in Tables 5.17 and 5.18, when only the a posteriori classification criteria are used. Firstly, we consider the case of updating the posterior probability. Using the two component Negative Binomial mixture derived by updating the posterior probability, we saw from Table 5.17 that a bad risk policyholder who at $t = 2$ has claim frequency history $k_1 = 0, k_2 = 2$ faces a malus of 4.64% of the basic premium while a policyholder who has $k_1 = 1, k_2 = 1$ claim frequency history faces a malus of 5.85% of the basic premium. Using the two component Negative Binomial mixture regression model derived by updating the posterior probability and considering a bad risk policyholder, we observe that if she has at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ she faces a malus of 122.03% of the basic premium, while if she has $k_1 = 1, k_2 = 1$ claim frequency history then she has to pay a malus of 119.72% of the basic premium. Secondly, we consider the case of updating the posterior mean. Using the two component Negative Binomial mixture derived by updating the posterior mean, we saw from Table 5.18 that a policyholder who at $t = 2$ has two claims faces a malus 34.36% of the basic premium. Using the two component

Negative Binomial mixture regression model derived by updating the posterior mean, consider a woman who has a car with horsepower between 0-33 and her Bonus-Malus class varies over time. As mentioned before, if at $t = 2$ she has claim frequency history $k_1 = 0, k_2 = 2$ she faces a malus of 137.37% of the basic premium, while if she has $k_1 = 1, k_2 = 1$ claim frequency history then she faces a malus of 133.12% of the basic premium.

Regarding the claim severity component, we consider first the two and three component Exponential, Gamma, Weibull and GB2 mixture regression models derived by updating the posterior probability. Let us see an example in order to understand better how the BMSs resulting from these models work. Consider a group of policyholders who share the following common characteristics: The policyholder i is a woman, she belongs to the first Bonus-Malus category and her car has horsepower 0-33. The estimation for this group led to the following results:

- In the case of the two component Exponential mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 50.93\%$ of good drivers with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right)$ $= \exp(5.746034 + (-0.023606)) = 305.64$ and $\hat{\rho}_2 = 49.07\%$ of bad risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(\ 5.745604 + (-0.022953)) = 305.71$, where j represents the age of the policy.

- In the case of the two component Gamma mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 48.61\%$ of good drivers with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right) =$ $\exp(5.248615 + 0.01684) = 193.53$, $\hat{\theta}^j_{1,i} = 0.0958$ and $\hat{\rho}_2 = 51.39\%$ of bad risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(5.865226 + 0.064557) = 376.07$, $\hat{\theta}^j_{2,i} = 0.4568$, where j represents the age of the policy.

- In the case of the two component Weibull mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 45.61\%$ of good drivers with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right)$ $= \exp(6.573425 + (-1.312589)) = 192.64$, $\hat{\theta}^j_{1,i} = 12.4410$ and $\hat{\rho}_2 = 54.39\%$ of bad risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(5.677825 + 0.200917) = 357.36$, $\hat{\theta}^j_{2,i} = 2.2338$, where j represents the age of the policy.

- In the case of the two component GB2 mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 61.87\%$ of good drivers with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right) =$ $\exp(5.13552 + 0.0522) = 179.06$, $\hat{\sigma}^j_{1,i} = 17.3500$, $\hat{\nu}^j_{1,i} = 0.8781$, $\hat{s}^j_{1,i} = 1.1364$ and $\hat{\rho}_2 =$ 38.13% of bad risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(6.25452 + (-0.01857)) =$ 510.79 , $\hat{\sigma}^j_{2,i} = 5.424$, $\hat{\nu}^j_{2,i} = 0.5224$, $\hat{s}^j_{2,i} = 0.9354$, where j represents the age of the policy.

- In the case of the three component Exponential mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 30.93\%$ of good risks with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right)$ $= \exp(5.745773 + (-0.023272)) = 305.6684$, $\hat{\rho}_2 = 35.53\%$ of average risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(\ 5.745709 + (-0.023174)) = 305.68$ and $\hat{\rho}_3 = 33.54\%$ of bad risks with claim severity $\hat{y}^j_{3,i} = \exp\left(d^j_{3,i}\gamma^j_3\right) = \exp(5.745988 + (-0.023415)) = 305.69$, where j represents the age of the policy.

- In the case of the three component Gamma mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 1.23\%$ of good risks with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right) = \exp(5.09428+(-1.30716)) = 44.13$, $\hat{\theta}^j_{1,i} = 9.3062e-08$, $\hat{\rho}_2 = 46.67\%$ of average risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(\ 5.246655+0.030974) = 195.91$, $\hat{\theta}^j_{2,i} = 0.0961$ and $\hat{\rho}_3 = 52.10\%$ of bad risks with claim severity $\hat{y}^j_{3,i} = \exp\left(d^j_{3,i}\gamma^j_3\right) = \exp(5.88956+0.01927) = 368.27$, $\hat{\theta}^j_{3,i} = 0.4531$, where j represents the age of the policy.

- In the case of the three component Weibull mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 37.27\%$ of good risks with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right) = \exp(5.3201+0.03568) = 211.83$, $\hat{\theta}^j_{1,i} = 12.3172$, $\hat{\rho}_2 = 49.58\%$ of average risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(\ 5.72276+0.05575) = 323.27$, $\hat{\theta}^j_{2,i} = 2.1557$ and $\hat{\rho}_3 = 13.15\%$ of bad risks with claim severity $\hat{y}^j_{3,i} = \exp\left(d^j_{3,i}\gamma^j_3\right) = \exp(6.32151+(-0,5092)) = 334.39$, $\hat{\theta}^j_{3,i} = 14.3106$, where j represents the age of the policy.

- In the case of the three component GB2 mixture regression model, the resulting portfolio consists of $\hat{\rho}_1 = 60.37\%$ of good risks with claim severity $\hat{y}^j_{1,i} = \exp\left(d^j_{1,i}\gamma^j_1\right) = \exp(\ 5.28321+(-0.00138)) = 196.73$, $\hat{\sigma}^j_{1,i} = 15.1000$, $\hat{\nu}^j_{1,i} = 0.9317$, $\hat{s}^j_{1,i} = 0.9936$, $\hat{\rho}_2 = 20.16\%$ of average risks with claim severity $\hat{y}^j_{2,i} = \exp\left(d^j_{2,i}\gamma^j_2\right) = \exp(5.74645+0.02297) = 320.35$, $\hat{\sigma}^j_{2,i} = 11.7900$, $\hat{\nu}^j_{2,i} = 1.0896$, $\hat{s}^j_{2,i} = 0.9150$ and $\hat{\rho}_3 = 19.47\%$ of bad risks with claim severity $\hat{y}^j_{3,i} = \exp\left(d^j_{3,i}\gamma^j_3\right) = \exp(6.44784+0.05378) = 666.22$, $\hat{\sigma}^j_{3,i} = 4.4820$, $\hat{\nu}^j_{3,i} = 0.3979$, $\hat{s}^j_{3,i} = 1.1457$, where j represents the age of the policy.

Based on the above results for this group of individuals, we are now able to derive the premiums determined by the two and three component mixture regression models respectively. In what follows we assume that the policyholder i is observed for the first year of her presence in the portfolio, has one accident, i.e. $K = 1$,and the claim amount of her accident $X_{i,1}$ ranges from 150 to 500 euros. Firstly, we consider the case of the two component mixture regression models. The posterior probability $\rho_2\left(X_{i,1}; d^1_{2,i}, ..., d^{t+1}_{2,i}\right)$ that the policyholder is a bad risk is given by the Eqs (5.52, 5.56, 5.54 and 5.58), for $l = n = 2$, for the case of the two component Exponential, Weibull, Gamma and GB2 mixture regression models respectively[20]. The results are summarized in Table 5.28.

[20] Note that $\rho_1\left(X_{i,1}; d^1_{1,i}, ..., d^{t+1}_{1,i}\right) = 1 - \rho_2\left(X_{i,1}; d^1_{2,i}, ..., d^{t+1}_{2,i}\right)$.

Table 5.28: Posterior Probability of the Second Component, Two Component Mixture Regression Models for Assessing Claim Severity, One Claim in the First Year of Observation

Claim Size	Exponential	Gamma	Weibull	GB2
150	0.4906756	0.5511852	0.3853065	0.0469568
175	0.4906802	0.1097841	0.1368834	0.0183314
200	0.4906847	0.0957214	0.1209398	0.0734445
225	0.4906893	0.3187215	0.8847980	0.4688375
250	0.4906938	0.8827866	1.0000000	0.9008195
275	0.4906984	0.9971102	1.0000000	0.9870134
300	0.4907030	0.9999733	1.0000000	0.9980976
325	0.4907075	0.9999999	1.0000000	0.9996733
350	0.4907121	1.0000000	1.0000000	0.9999350
375	0.4907166	1.0000000	1.0000000	0.9999852
400	0.4907212	1.0000000	1.0000000	0.9999962
425	0.4907258	1.0000000	1.0000000	0.9999989
450	0.4907303	1.0000000	1.0000000	0.9999996
475	0.4907349	1.0000000	1.0000000	0.9999998
500	0.4907394	1.0000000	1.0000000	0.9999999

Secondly, we consider the case of the three component mixture regression models. The posterior probability $\rho_2\left(X_{i,1}; d_{2,i}^1, ..., d_{2,i}^{t+1}\right)$ that the policyholder is an average risk is given by the Eqs (5.52, 5.56, 5.54 and 5.58), for $l = 2$ and $n = 3$ and the posterior probability $\rho_3\left(X_{i,1}; d_{3,i}^1, ..., d_{3,i}^{t+1}\right)$ that the policyholder is a bad risk is given by the Eqs (5.52, 5.56, 5.54 and 5.58), for $l = 3$ and $n = 3$, for the case of the three component Exponential, Weibull, Gamma and GB2 mixtures respectively, using the same categories as before[21]. The results are depicted in Table 5.29 (Panels A and B).

[21] Note that $\rho_1\left(X_{i,1}; d_{1,i}^1, ..., d_{1,i}^{t+1}\right) = 1 - \rho_2\left(X_{i,1}; d_{2,i}^1, ..., d_{2,i}^{t+1}\right) - \rho_3\left(X_{i,1}; d_{3,i}^1, ..., d_{3,i}^{t+1}\right)$.

Table 5.29: Posterior Probability of the Second and the Third Component, Three Component Mixture Regression Models For Assessing Claim Severity, One Claim in the First Year of Observation

Panel A: Second component				
Claim Size	Exponential	Gamma	Weibull	GB2
150	0.3552789	0.3565165	0.7233854	0.0006681
175	0.3552788	0.8653749	0.3544217	0.0008067
200	0.3552788	0.8985533	0.1613521	0.0024792
225	0.3552787	0.7056917	0.2070387	0.0253259
250	0.3552786	0.1524158	0.8850081	0.2363544
275	0.3552786	0.0047511	0.7753841	0.6537234
300	0.3552785	5.375790e-05	0.5464771	0.8246564
325	0.3552785	2.946738e-07	0.3791533	0.8532723
350	0.3552784	8.747647e-10	0.4223180	0.8183945
375	0.3552783	1.530722e-12	0.8691040	0.7292184
400	0.3552783	1.690473e-15	0.9998387	0.5909865
425	0.3552782	1.245893e-18	1.0000000	0.4308295
450	0.3552781	6.418062e-22	1.0000000	0.2869196
475	0.3552781	2.402189e-25	1.0000000	0.1806378
500	0.3552780	6.750378e-29	1.0000000	0.1112596
Panel B: Third component				
150	0.3353838	0.6434836	8.405750e-05	0.1260292
175	0.3353848	0.1346251	0.0002891	0.0276356
200	0.3353858	0.1014467	0.0007288	0.0194449
225	0.3353867	0.2943083	0.0043252	0.0549678
250	0.3353877	0.8475842	0.0738791	0.1709929
275	0.3353887	0.9952489	0.2246159	0.1996051
300	0.3353897	0.9999462	0.4535229	0.1493001
325	0.3353907	0.9999997	0.6208467	0.1402153
350	0.3353916	1.0000000	0.5776820	0.1791491
375	0.3353926	1.0000000	0.1308960	0.2695859
400	0.3353936	1.0000000	0.0001613	0.4083779
425	0.3353946	1.0000000	6.839931e-12	0.5688361
450	0.3353956	1.0000000	1.741381e-28	0.7129113
475	0.3353965	1.0000000	1.265132e-63	0.8192788
500	0.3353975	1.0000000	1.053979e-134	0.8886994

Based on the estimates of Tables 5.28 and 5.29 we can now derive the premiums according to the Eq. (5.50) for $n = 2$ and $n = 3$ for the case two and three component mixture regression models respectively. The results are displayed in Table 5.30 (Panels A and B). More specifically, in Panel A we report the premiums for the two component mixture models for a bad risk, i.e. belonging to the second category of risks and in Panel B we depict the premiums for the three component mixture models for a bad risk, i.e. belonging to the third category of risks using the same categories as before. We consider that for claim sizes smaller than 200 euros (for all models) the policyholder always pays the same premium calculated for claim sizes equal to 200 euros, and receives a bonus due to the fact that the cost of the claims that the insurance company has to pay is not significant, and also due to the irregular behavior of the posterior probabilities and thus premium formulas for extremely small and large claim sizes. This value was chosen since for claim sizes up to 200 euros both the posterior probabilities in the case of the two and three component mixture models indicate an irregular behavior. Furthermore, because of the irregular behavior of the posterior probabilities, in the case of the three component Weibull mixture regression model, it was found that there was a very small drop in the premium rates ranging from 350 to 400 compared to the premium rate of 325. For this reason, we assumed that these values are equal to that of 325 euros. Form Table 5.30 we observe that the premiums are equal in the case of Exponential, revealing the lack of necessity of the two and three components. As expected, for the other models the higher the claim size the higher the premium, revealing the appropriateness of the modelling technique. For example, for one claim size of 325 in the first year, the premium increases from 100 to 130.88, 126.62, 161.20, in the two component mixture Gamma, Weibull and GB2 models, respectively. For the three component mixtures, we have to note that for one claim size of 325 in the first year, the premium increases from 100 to 129.74, 116.59 and 125.90, in the case of the Gamma, Weibull and GB2 models, respectively. For small claim sizes (i.e. 150, 175, 200) all models reward the policyholder with a bonus. Furthermore, for small differences in the claim sizes the premiums are equal or almost equal.

Table 5.30: Optimal BMS, Two and Three Component Mixture Regression Models for Assessing Claim Severity, Update of the Posterior Probability, One Claim in the First Year of Observation

Panel A: Two component				
Claim Size	Exponential	Gamma	Weibull	GB2
150	100	73.43644	75.31598	69.52824
175	100	73.43644	75.31598	69.52824
200	100	73.43644	75.31598	69.52824
225	100	87.60332	119.89696	108.66376
250	100	123.43757	126.62049	151.42081
275	100	130.70039	126.62049	159.95219
300	100	130.88228	126.62049	161.04929
325	100	130.88397	126.62049	161.20526
350	100	130.88397	126.62049	161.23116
375	100	130.88397	126.62049	161.23613
400	100	130.88397	126.62049	161.23721
425	100	130.88397	126.62049	161.23748
450	100	130.88397	126.62049	161.23755
475	100	130.88397	126.62049	161.23758
500	100	130.88397	126.62049	161.23758
Panel B: Three component				
150	100	75.18024	81.17901	72.36503
175	100	75.18024	81.17901	72.36503
200	100	75.18024	81.17901	72.36503
225	100	86.89235	83.13255	77.12034
250	100	120.49171	112.82262	99.24548
275	100	129.45910	115.03205	122.36008
300	100	129.74436	115.93033	125.45200
325	100	129.74761	116.58695	125.90145
350	100	129.74763	116.58695	128.21717
375	100	129.74763	116.58695	133.20173
400	100	129.74763	116.58695	140.78501
425	100	129.74763	116.58695	149.53536
450	100	129.74763	116.58695	157.38720
475	100	129.74763	116.58695	163.18229
500	100	129.74763	116.58695	166.96378

Let us now consider the two and three component Pareto mixture regression models derived by updating the posterior mean claim severity. In the following example we consider a group of policyholders who share the following common characteristics: The policyholder i is a woman who has a car with horsepower between 0-33 and her Bonus-Malus class varies over time, starting from BM class 1. The mean (or location) parameter of the two and three component Pareto mixture models are given by $E(X_{i,k}^{j}|d_{z,i}^{j}) = \sum_{z=1}^{n} \rho_z \exp\left(d_{z,i}^{j}\gamma_z^{j}\right)$, for $n = 2$ and $n = 3$ respectively, where $d_{z,i}^{j}\left(d_{z,i,1}^{j}, ..., d_{z,i,h}^{j}\right)$ is the $1 \times h$ vector of h individual's characteristics, which represent different a priori rating variables and γ_z^{j} is the vector of the coefficients. The estimation of the vector γ_z^{j} led to the following results presented in Table 5.31.

Table 5.31: Women, Horse Power 0-33

Bonus-Malus	Two Component Pareto		Three Component Pareto		
Category	$\exp\left(d_{1,i}^{j}\gamma_1^{j}\right)$	$\exp\left(d_{2,i}^{j}\gamma_2^{j}\right)$	$\exp\left(d_{1,i}^{j}\gamma_1^{j}\right)$	$\exp\left(d_{2,i}^{j}\gamma_2^{j}\right)$	$\exp\left(d_{3,i}^{j}\gamma_3^{j}\right)$
1	325.4869	325.1616	325.4869	325.1616	325.4869
2	319.0418	318.7230	319.0418	319.0418	319.3610
3	365.8862	365.5205	365.8862	365.5205	365.8862
4	151.4601	151.3087	151.4601	151.1575	151.4601
5	491.4314	490.9402	491.4314	490.9402	491.4314

The generalized optimal BMSs obtained by the two and three component Pareto mixture models will be defined by Eq. (5.69), for $n = 2$ and $n = 3$ respectively. Assume that the policyholder i is observed for the first year of her presence in the portfolio, has one accident and the claim amount of her accident ranges from 150 to 1000 euros. Table 5.32 demonstrates the resulting premium rates for these models. These systems do not differ much from the one obtained in Chapter 4 when the Pareto regression model was used (see Table 4.5, Chapter 4). For instance, consider the same group of policyholders we described before. For one claim size of 500 in the first year the premium increases from 100 to 100.96, 105.09 and 105.16 in the case of the Pareto, two component Pareto mixture and three component Pareto mixture regression models respectively.

Table 5.32: Optimal BMS, Two and Three Component Pareto Mixture Regression Models for Assessing Claim Severity, One Claim in the First Year of Observation

Claim Size	Optimal BMS Two Component Pareto	Optimal BMS Three Component Pareto
150	90.92049	90.97747
175	91.93276	91.99022
200	92.94503	93.00297
225	93.95731	94.01572
250	94.96958	95.02848
275	95.98185	96.04123
300	96.99412	97.05398
325	98.00639	98.06673
350	99.01867	99.07948
375	100.03094	100.09223
400	101.04321	101.10499
425	102.05548	102.11774
450	103.06776	103.13049
475	104.08003	104.14324
500	105.09230	105.15599
525	106.10457	106.16874
550	107.11684	107.18150
575	108.12912	108.19425
600	109.14139	109.20700
625	110.15366	110.21975
650	111.16593	111.23250
675	112.17821	112.24525
700	113.19048	113.25801
725	114.20275	114.27076
750	115.21502	115.28351
775	116.22729	116.29626
800	117.23957	117.30901
825	118.25184	118.32176
850	119.26411	119.33452
875	120.27638	120.34727
900	121.28866	121.36002
925	122.30093	122.37277
950	123.31320	123.38552
975	124.32547	124.39827
1000	125.33774	125.41103

Compared to the BMSs provided by the two and three component Weibull, Gamma and GB2 mixture regression models derived by updating the posterior probability (see Table 5.30) the systems resulting from the two and three component Pareto mixture models are always much cheaper. It interesting to compare the BMSs based both on the a priori and the a posteriori severity component (see Tables 5.30 and 5.32) with those based only on the a posteriori severity component (see Tables 5.21 and 5.22). Firstly, we consider the case of updating the posterior probability. Using the BMSs presented in Table 5.21 we observed that a bad risk policyholder with one accident size of 500 euros in one year has to pay a malus of 20.46%, 22.94%, 21.78% and 28.97%, 22.19%, 53.40% of the basic premium in the case of the two and three component Gamma, Weibull and GB2 mixture distributions respectively. Using the systems depicted in Table 5.30 we saw that a woman who belongs to the first Bonus-Malus category, her car has horsepower 0-33 and is a bad risk, for one accident of 500 euros in one year, faces a malus of 30.88%, 26.62%, 61.23% and 29.74%, 14.15%, 66.96% of the basic premium in the case of the two and three component Gamma, Weibull and GB2 mixture regression models respectively. Secondly, we consider the case of updating the posterior mean. In this case we observed from Table 5.22 that a policyholder with one accident size of 500 euros in one faces a malus 0.67% and 0.66% of the basic premium in the case of the two and three component Pareto mixtures. Using the generalized optimal BMSs with a severity component based both on the a priori and the a posteriori classification criteria, we observed from Table 5.32 that a woman who has a car with horsepower between 0-33 and her Bonus-Malus class varies over time, for one accident of 500 euros in one year, will have to pay a malus of 05.09% and 05.16% of the basic premium in the case of the two and three component Pareto mixture regression models.

Finally, we consider the generalized premiums based both on the a priori and the a posteriori frequency and severity component. In Tables 5.33 and 5.34 we report these generalized premiums. We observe that for one claim size of 250 in the first year the premium increases from 100 to 212.85, 218.34, 261.11, in the two component Poisson mixture regression model and the corresponding two component severity models (Table 5.33), to 180.41, 185.06, 221.31 in the two component Negative Binomial mixture regression model (updating the posterior probability) and the corresponding two component severity models (Table 5.34, Panel A), to 208.54, 257.42, 264.06, 315.78 in the two component Negative Binomial mixture regression model (updating the posterior mean) and the corresponding two component severity models (Table 5.34, Panel B). The generalized BMSs are more fair than the systems based only on the a posteriori classification criteria (see Tables 5.23 and 5.24) since they consider all the important a priori and a posteriori information for each policyholder for the frequency and the severity component in order to estimate their risk of having an accident.

Table 5.33: Optimal BMS Based on the Two Component Poisson Mixture Regression Model for Assessing Claim Frequency and the Various Two Component Mixture Regression Models for Assessing Claim Severity, One Claim in the First Year of Observation

Two Component Poisson Mixture					
Claim Size	POIS-EXP	POIS-GA	POIS-WEI	POIS-GB2	POIS-PA
150	172.44	126.6338	129.8749	119.8945	156.7833
175	172.44	126.6338	129.8749	119.8945	158.5289
200	172.44	126.6338	129.8749	119.8945	160.2744
225	172.44	151.0632	206.7503	187.3798	162.0200
250	172.44	212.8557	218.3444	261.1100	163.7655
275	172.44	225.3798	218.3444	275.8216	165.5111
300	172.44	225.6934	218.3444	277.7134	167.2567
325	172.44	225.6963	218.3444	277.9824	169.0022
350	172.44	225.6963	218.3444	278.0270	170.7478
375	172.44	225.6963	218.3444	278.0356	172.4934
400	172.44	225.6963	218.3444	278.0375	174.2389
425	172.44	225.6963	218.3444	278.0379	175.9845
450	172.44	225.6963	218.3444	278.0380	177.7300
475	172.44	225.6963	218.3444	278.0381	179.4756
500	172.44	225.6963	218.3444	278.0381	181.2212

Table 5.34: Optimal BMS Based on the Alternative Two Component Negative Binomial Type I Mixture Regression Models for Assessing Claim Frequency and the Various Two Component Mixture Regression Models for Assessing Claim Severity, One Claim in the First Year of Observation

Panel A: Two Component Negative Binomial Mixture Regression Model (Update of the Posterior Probability)					
Claim Size	NB-EXP	NB-GA	NB-WEI	NB-GB2	NB-PA
150	146.16	107.3347	110.0818	95.9020	156.7833
175	146.16	107.3347	110.0818	95.9020	158.5289
200	146.16	107.3347	110.0818	101.6224	160.2744
225	146.16	128.0410	175.2414	158.8229	162.0200
250	146.16	180.4164	185.0685	221.3166	163.7655
275	146.16	191.0317	185.0685	233.7861	165.5111
300	146.16	191.2975	185.0685	235.3896	167.2567
325	146.16	191.3000	185.0685	235.6176	169.0022
350	146.16	191.3000	185.0685	235.6555	170.7478
375	146.16	191.3000	185.0685	235.6627	172.4934
400	146.16	191.3000	185.0685	235.6643	174.2389
425	146.16	191.3000	185.0685	235.6647	175.9845
450	146.16	191.3000	185.0685	235.6648	177.7300
475	146.16	191.3000	185.0685	235.6648	179.4756
500	146.16	191.3000	185.0685	235.6649	181.2212

Panel B: Two Component Negative Binomial Mixture Regression Model (Update of the Posterior Mean)					
Claim Size	NB-EXP	NB-GA	NB-WEI	NB-GB2	NB-PA
150	208.54	153.1444	157.0639	144.9942	189.6056
175	208.54	153.1444	157.0639	144.9942	191.7166
200	208.54	153.1444	157.0639	144.9942	193.8276
225	208.54	182.6880	250.0331	226.6074	195.9386
250	208.54	257.4167	264.0544	315.7730	198.0496
275	208.54	272.5626	264.0544	333.5643	200.1605
300	208.54	272.9419	264.0544	335.8522	202.2715
325	208.54	272.9454	264.0544	336.1774	204.3825
350	208.54	272.9454	264.0544	336.2315	206.4935
375	208.54	272.9454	264.0544	336.2418	208.6045
400	208.54	272.9454	264.0544	336.2441	210.7155
425	208.54	272.9454	264.0544	336.2446	212.8265
450	208.54	272.9454	264.0544	336.2448	214.9375
475	208.54	272.9454	264.0544	336.2448	217.0485
500	208.54	272.9454	264.0544	336.2449	219.1595

Chapter 6

Conclusion

The research projects presented in this dissertation lie on the frontiers of actuarial science and statistics. We deal with the important actuarial pillars of a priori risk classification and experience rating in motor third-party liability insurance and link our work with contributions in this area. The idea behind risk classification is to divide an insurance portfolio into different classes that consist of risks with a similar profile and to design a fair tariff for each of them. For the construction of a fair tariff structure actuaries apply risk classification based on regression techniques. When the explanatory variables used as risk factors express measurable information about the policyholder the system is an a priori classification scheme. However, many important factors cannot be taken into account a priori when pricing motor third party liability insurance products. For instance, swiftness of reflexes, aggressiveness behind the wheel or knowledge of the highway code are difficult to integrate into a priori risk classification. Consequently, tariff cells are still quite heterogeneous despite the use of many classification variables. Experience rated or Bonus-Malus Systems (BMSs) re-evaluate the premiums by taking the history of claims of the insured into account. A basic interest of actuarial literature is the construction of an optimal or 'ideal' BMS defined as a system obtained through Bayesian analysis. An optimal BMS is financially balanced for the insurer and fair for the policyholder. Optimal BMSs can be broadly derived in two ways; based only on the a posteriori classification criteria and based on both the a priori and the a posteriori classification criteria. In this thesis (Chapters 3, 4 and 5), we focus on the study of optimal BMSs based on different statistical models for assessing claim frequency and claim severity.

In Chapter 1 we discussed a literature review of the statistical techniques that can be practically implemented for pricing risks through ratemaking based on a priori risk classification and Bonus-Malus Systems.

In Chapter 2 we extended recent actuarial literature research which uses generalized linear models, GLM, for pricing risks through ratemaking based on a priori risk classification (see, for instance, Denuit et al., 2007 & Boucher et al., 2007, 2008). This was achieved by using the generalized additive models for location, scale and shape (GAMLSS). The (GAMLSS) were introduced by Rigby and Stasinopoulos (2001, 2005) and Akantziliotou, Rigby, and Stasinopoulos (2002) as a way of overcoming some of the limitations associated with the popular generalized linear models, GLM, and generalized additive models, GAM. The GAMLSS can be seen as an extension to the conventional GLM and generalized additive models, GAM. In the GAMLSS the distribution for the response variable can be selected from a very general family of distributions including highly skew or kurtotic continuous and discrete distributions. Moreover, the GAMLSS regress not only the expected mean but every distribution parameter (e.g. location, scale and shape) to a set of covariates. Therefore, both mean and variance may be assessed by choosing a marginal distribution and building a predictive model using ratemaking factors as independent variables. In the setup we considered, risk heterogeneity was modeled as the distribution of frequency and/or severity of claims changes between clusters by a function of the level of ratemaking factors underlying the analyzed clusters. GAMLSS were used to model the frequency and the severity of claims. Specifically, within the framework of the GAMLSS we assumed that the number of claims was distributed according to the Poisson, Negative Binomial Type II, the Delaporte, Sichel and Zero-Inflated Poisson GAMLSS and that the losses were distributed according to the Gamma, Weibull, Weibull Type III, Generalized Gamma and Generalized Pareto GAMLSS respectively. These classification models were calibrated employing a Generalized Akaike Information Criterion (GAIC) which is valid for both nested or non-nested model comparisons (as suggested by Rigby and Stasinopoulos, 2005 & 2009). The best fitted claim frequency model was the Negative Binomial Type II model, followed closely by the Sichel and Delaporte models while regarding the claim severity models, the best fitting performances were provided by the Generalized Gamma model followed by the Generalized Pareto and Gamma models. Furthermore, the difference between these models was analyzed through the mean and the variance of the annual number of claims and the severity of claims of the policyholders, who belong to different risk classes. The resulting a priori premiums rates were calculated via the expected value and standard deviation principles with independence between the claim frequency and severity components assumed.

In Chapter 3 we developed the design of an optimal BMS assuming that the number of claims was distributed according to a Sichel distribution. This system was proposed as an alternative to the optimal BMS resulting from the traditional Negative Binomial distribution, which cannot handle data with a long tail efficiently (see Lemaire, 1995). We also considered the optimal BMS provided by the Poisson-Inverse Gaussian distribution (PIG), which is a special case of the Sichel distribution. These systems were obtained by updating the posterior mean claim frequency. Furthermore, we presented a generalized BMS that integrates the a priori and the a posteriori information on a individual basis extending the framework developed by Dionne and Vanasse (1989, 1992). This was achieved by using the Sichel GAMLSS to approximate the number of claims as an alternative to the Negative Binomial regression model used by Dionne and Vanasse (1989, 1992). The new model offers the advantage of being able to model count data with high dispersion. We also considered the generalized system derived by PIG GAMLSS for assessing claim frequency. With the aim of constructing an optimal BMS by updating the posterior mean claim frequency, we adopted the parametric linear formulation of these models and we allowed only their mean parameter to be modelled as a function of the significant explanatory variables for the number of claims. The models were calibrated with respect to Global Deviance, AIC, SBC information criteria and the Vuong test. The modeling results showed that the Sichel distribution and the Sichel GAMLSS provided the best fitting performances for the data set examined in this thesis. The optimal BMSs obtained had all the attractive properties of the BMSs developed by Lemaire (1995) and Dionne and Vanasse (1989, 1992).

In Chapter 4 we integrated claim severity into the optimal BMSs based on the a posteriori criteria of Chapter 3. For this purpose we assumed that the losses were distributed according to a Pareto distribution, following the framework proposed by Frangos and Vrontos (2001). The BMS resulting from the Sichel and Pareto models and that derived from the PIG and Pareto models were compared to the system provided by the Negative Binomial and Pareto models (see Frangos and Vrontos, 2001). The basic advantage of the optimal BMSs based on the a posteriori frequency and severity component, in comparison with those that take into consideration only the frequency component, is the differentiation of the premiums according to the severity of the claim. We also presented the development of a generalized BMS with a frequency and a severity component based both on the a priori and the a posteriori criteria. For the frequency component we considered that the number of claims was distributed according to the Negative Binomial Type I, Poisson Inverse Gaussian and Sichel GAMLSS, following the current methodology as presented in Chapter 3. For the severity component we assumed that the costs of claims were distributed according to a Pareto GAMLSS. These systems were derived as functions of the years that the policyholder was in the portfolio, their number of accidents, the size of loss of each of these accidents and of the statistically significant a priori rating variables for the number of accidents and for the size of loss that each of these claims incurred. Furthermore, we presented a generalized form of the system obtained in Frangos and Vrontos (2001).

In Chapter 5 we presented the development of an optimal BMS using finite mixtures of distributions and regression models (see Mclachlan and Peel, 2000, and Rigby and Stasinopoulos, 2009). The finite mixture models are a popular statistical modelling technique, given that they constitute a flexible and easily extensible model class for approximating general distribution functions in a semi-parametric way and accounting for unobserved heterogeneity. Finite mixture models have been widely applied in many areas, such as biology, biometrics, genetics, medicine and marketing. However, they have not been extensively studied in BMS literature, with the exception of Lemaire(1995). The framework we considered focused on both the analysis of the claim frequency and severity components. For the frequency component we assumed that the number of claims was distributed according to a finite Poisson, Delaporte and Negative Binomial mixture, and for the severity component we considered that the losses were distributed according to a finite Exponential, Gamma, Weibull and GB2 mixture. These optimal BMS were obtained by updating the posterior probability of the policyholder's risk class. Furthermore, we extended the setup of Frangos and Vrontos (2001) for Negative Binomial and Pareto mixtures and designed an optimal BMS based on posterior distribution of both the mean claim frequency and the mean claim size, given the information we have about the claim frequency history and the claim size history for each policyholder. We have also developed a generalized BMS that integrates the a priori and the a posteriori information on an individual basis, extending the framework developed by Dionne and Vanasse (1989, 1992) and Frangos and Vrontos (2001) using finite mixtures of regression models. In our application, there were both nested and non-nested distributions/regression models comparisons. The models were calibrated with respect to Global Deviance, AIC, SBC information criteria and the LR and Vuong tests. The modelling results showed that the mixture of distributions/regression models was extremely important as it provided a superior fit. Using this formulation the heterogeneity in the data was accounted for in two ways. Firstly, the population heterogeneity was accounted for by choosing a finite number of unobserved latent components, each of which may be regarded as a sub-population. This was a discrete representation of heterogeneity in the data since the mean rate was approximated by a finite number of support points. Secondly, depending on the choice of the component distribution, heterogeneity was also accommodated within each component by including the explanatory variables in the mean rate function. The designs presented in Chapters 2, 3, 4 and 5 can be employed by insurance companies which are free to set up their own tariff structures and rating policies.

Challenges for future research are omnipresent in this thesis with several ideas for further work presenting themselves. In Chapters 2, 3 and 4 different count and loss distributions within the framework of GAMLSS models can be fitted. Moreover, these models are parametric and it would also be useful to explore the semiparametric approach and go through the ratemaking (a priori and a posteriori) exercise when functional forms other than the linear are included. In this case the problem of the choice between models becomes more acute and should be the topic of further research. In Chapter 5 another possible topic for further research is the design of optimal BMS using different claim frequency and severity models within the framework of finite mixtures of GAMLSS models and different premium calculation principles. Furthermore, the application of all the models used throughout this thesis in health insurance as well as in other fields of general insurance is the topic of ongoing research. In Chapters 2, 4 and 5, in light of the importance of large claims in actuarial science, a combination of GAMLSS models, Bayesian hierarchical regression models and finite mixtures of GAMLSS models with ideas from extreme value analysis should receive more attention in the future. In Chapters 3, 4 and 5 an important line of further research is to apply the same mixtures to all the contracts of the same insured so a dependence between the contracts can be modelled. This kind of model is called longitudinal data (see, for instance, Boucher, Denuit and Guillen, 2007). Another possible line of future research is the implementation of the Generalized Linear Mixed Models, GLMMs, to the data set used in this thesis, for a posteriori ratemaking (see Breslow and Clayton, 1993, and Antonio and Beirlant, 2007 for an insurance application). For distributions from the exponential family, GLMMs extend GLMs by including random effects in the linear predictor. The random effects not only determine the correlation structure between observations on the same subject (i.e. contracts), but also take heterogeneity among subjects, due to unobserved characteristics, into account. The use of hierarchical generalized linear models (HGLMs) in a posteriori risk classification, which are GLMMs with random effects having non–normal distributions (see Nelder, 1996, and Lee and Nelder, 2001) is another topic of ongoing research.

Bibliography

[1] **Abramowitz, M., and Stegun, I.A. (1974).** Handbook of Mathematical Functions. Dover, New York.

[2] **Akantziliotou, C., Rigby, R.A., and Stasinopoulos, D.M. (2002).** The R Implementation of Generalized Additive Models for Location, Scale and Shape. In M. Stasinopoulos and G. Touloumi (eds.), Statistical Modelling in Society: Proceedings of the 17th International Workshop on Statistical Modelling, pp. 75-83. Chania, Greece.

[3] **Albrecht, P. (1980).** On the Correct Use of the Chi-Square Goodness-of -Fit Test. Scandinavian Actuarial Journal. 149-170. Comment in Scandinavian Actuarial Journal 1982\ 168-170.

[4] **Albrecht, P. (1982a).** On Some Statistical Methods Connected With the Mixed Poisson Distribution. Scandinavian Actuarial Journal. 1-14.

[5] **Albrecht, P. (1982b).** Testing the Goodness of Fit of a Mixed Poisson Process. Insurance: Mathematics and Economics. 1, 27-33.

[6] **Antonio, K., and Beirlant J, (2007).** Actuarial statistics with generalized linear mixed models. Insurance: Mathematics and Economics. 40(1):58-76.

[7] **Baxter, L.A., Coutts, S.M., and Ross G.A.F. (1979).** Applications of Linear Models in Motor Insurance. 21st International Congress of Actuaries.

[8] **Beirlant, J., Derveaux, V., de Meyer, A.M, Goovaerts, M., Labie, E., and Maenhout, B. (1991).** Statistical Risk Evaluation Applied to (Belgian) Car Insurance. Insurance: Mathematics and Economics. 10, 289-302.

[9] **Besson, J. L., and Partrat, C. (1992).** Trend et systemes de bonus-maius. ASTIN Bulletin. 22. 11-31.

[10] **Bichsel, F. (1960).** Une methode pour calculer une ristoume adequate pour annees sans sinistres. ASTIN Bulletin. 1, 106-112.

[11] **Bichsel, F. (1964).** Erfahrung-Tarifierung in der Motorfahrzeug-haftplichtversicherung.Mitteiluneen der Vereinigung Schweizerischer Versicherungs-mathematiker. 119-129.

[12] **Bolance, C. Guillen, M., and Pinquet, J. (2003).** Time-varying credibility for frequency risk models: Estimation and tests for autoregressive specification on the random effect. Insurance: Mathematics and Economics 33, 273–282.

[13] **Bonsdorff, H. (1992).** On the Convergence Rate of Bonus-Malus Systems. ASTIN Bulletin. 22, 217-223.

[14] **Borgan, Hoem, 0. J., and Norberg, R. (1981).** A Non Asymptotic Criterion for the Evaluation of Automobile Bonus Systems. Scandinavian Actuarial Journal. 165-178.

[15] **Boucher, J. P., and Denuit, M. (2006).** Fixed Versus Random Effects in Poisson Regression Models for Claim Counts: Case Study with Motor Insurance, ASTIN Bulletin 36, pp. 285-301.

[16] **Boucher, J. P., Denuit, M., and Guillen, M. (2007).** Risk Classification for Claim Counts: A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models. North American Actuarial Journal, 11, 4, 110-131.

[17] **Boucher, J. P., Denuit, M., and Guillen M. (2008).** Models of Insurance Claim Counts with Time Dependence Based on Generalization of Poisson and Negative Binomial Distributions. Variance, 2, 1, 135-162.

[18] **Boulanger, F. (1994).** Systeme de Bonus-MaJus multi-garanties. ASTIN Colloquium, Cannes.

[19] **Boyer, M., Dionne G., and Vanasse, C. (1992).** Econometric Models of Accident Distribution. In Contributions to Insurance Economics, ed. G. Dionne, pp. xx–yy. Boston: Kluwer.

[20] **Breslow, N.E., and Clayton, D.G. (1993).** Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association 88 (421): 9–25. doi:10.2307/2290687. JSTOR 2290687.

[21] **Brockman, M.S., and Wright, TS. (1992).** Statistical Motor Rating: Making Effective Use of Your Data. Journal of the Institute of Actuaries. 119 III.

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

[23] **Buhlmann, H. (1964).** Optimale Pramienstufensysteme. Mitteiluneen der Vereiniaung Schweizerischer Versicherungsmathematiker. 193-213.

[24] **Buhlmann, H. (1970).** Mathematical Methods in Risk Theory. Berlin: Springer.

[25] **Cebrian, A., Denuit, M., and Lambert, Ph. (2003).** Generalized Pareto fit to the society of Actuaries' large claims database. North American Actuarial Journal 7, 18–36.

[26] **Centeno, L. e Andrade e Silva, J. (2001).** Bonus Systems in Open Portfolio. Insurance Mathematics e Economics, pp. 341-350.

[27] **Coene, G., and Doray, L.G. (1996).** A Financially Balanced Bonus-Malus System. ASTIN Bulletin 26, 107-115.

[28] **Cole, T.J., and Green, P.J. (1992).** Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood. Statistics in Medicine, 11, 1305-1319.

[29] **Consul, P.C. (1990).** A Model for Distributions of Injuries in Auto Accidents. Mitteilungen der Vereinieung Schweizerischer Versicherungsmathematiker. 163-168.

[30] **Corlier, F., Lemaire, J., and Muhokoio, D. (1979c).** Simulation of an Automobile Portfolio. Geneva Papers on Risk and Insurance. 40-46.

[31] **Cousy, H., and Claassens, H. (1994).** Ex Post Control of Insurance Policies in Belgium. Geneva Papers on Risk and Insurance. 70, 46-59.

[32] **Dean, C., Lawless, J.F., and Willmot, G.E. (1989).** A mixed Poisson-inverse-Gaussian regression model. Canadian Journal of Statistics 17 (2), 171-181.

[33] **de Jong, P., and Heller, G. Z. (2008).** Generalized linear models for insurance data. Cambridge University Press.

[34] **Delaporte, P. (1960).** Un probleme de tarification de I'assurance accidents d'automobiles examine par la statistique mathematique. Proceedings of the Sixteenth International Congress of Actuaries, Bruxelles, 2, 121-135.

[35] **Delaporte, P. (1965).** Tarification du risque individuel d'accidents automobiles par la prime modeiee sur le risque. ASTIN Bulletin, 3, 251-271.

[36] **Delaporte, P. (1972a).** Les mathematiques de I'assurance automobile. ASTIN Bulletin. 6, 185-190.

[37] **Delaporte, P. (1972b).** Construction d'un tarif d'assurance automobile base sur le principe de Ja prime modeiee sur le risque. Mitteitungen der Vereinigung Schweizeriseher Versicherungsmathematiker. 101-113.

[38] **De Leve, G., and Weeda, P.J. (1968).** Driving with Markov-Programming. ASTIN Bulletin, 5, 62-86.

[39] **Dellaert, N.P., Frenk, JBG, Kouwenhoven. A., and Van der Laan, B.S.. (1990).** Optimal Claim Behaviour for Third-Party Liability Insurances or to Claim or Not to Claim: That Is the Question. Insurance: Mathematics and Economics, 9, 59-76.

[40] **Dellaert, N.P., Frenk, JBG, and Voshol., E. (1991).** Optimal Claim Behaviour for Third-Party Liability Insurances with Perfect Information. Insurance: Mathematics and Economics, 10, 145-151.

[41] **Dempster, A., Laird, N., and Rubin, D. (1977).** Maximum likelihood from incomplete data via EM algorithm (with discussion). J. R. Statist. Soc., 39: 1-38.

[42] **Denuit, M., and Lang, S. (2004).** Nonlife Ratemaking with Bayesian GAM's. Insurance: Mathematics and Economics 35: 627–47.

[43] **Denuit, M., Marechal, X., Pitrebois, S., and Walhin, J. F. (2007).** Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. Wiley.

[44] **De Pril, N. (1978).** The Efficiency of a Bonus-Malus System. ASTIN Bulletin, Vol. 10 Part 1, pp. 59-72.

[45] **De Pril, N. (1979).** Optimal Claim Decisions for a Bonus-Malus System: A Continuous Approach. ASTIN Bulletin, 10, 215-222.

[46] **De Pril, N., and M. Goovaerts. (1983).** Bounds for the Optimal Critical Claim Size of a Bonus System. Insurance: Mathematics and Economics. 2, 27-32.

[47] **Dionne, G., and Vanasse, C. (1989).** A Generalization of Automobile Insurance Rating Models: The Negative Binomial Distribution with a Regression Component. ASTIN Bulletin. 19, 199-212.

[48] **Dionne, G., and Vanasse, C. (1992).** Automobile insurance ratemaking in the presence of asymmetrical information. Journal of Applied Econometrics 1, 149-165.

[49] **Dionnne, G., Artis, M., and Guillen, M. (1996).** Count Data Models for a Credit Scoring System. Journal of Empirical Finance 3: 303-25.

[50] **Dufresne, F. (1988).** Distributions stationnaires d'un systeme bonus-malus et probability de ruine. ASTIN Bulletin. 18, 31-46.

[51] **Dufresne, F. (1995).** The Efficiency of the Swiss Bonus-Malus System. Bulletin of the Swiss Actuaries, 1995(1) 29-41.

[52] **Dureuil, G., and Geoffrey, C. (1994).** Modelisation bivariee de frequences de sinistres dependantes. ASTIN Colloquium, Cannes.

[53] **Evans, D. A. (1953).** Experimental evidence concerning contagious distributions in ecology. Biometrika, 40: 186-211.

[54] **Feller, W. (1971).** An Introduction to Probability Theory and its Applications. New York: Wiley.

[55] **Ferreira, J. (1974).** The Long-term Effect of Merit-Rating Plans on Individual Motorists. Operations Research. 22, 954-978.

[56] **Ferreira, J. (1977).** Identifying Equitable Insurance Premiums for Risk Classes: An Alternative to the Classical Approach. Twenty-third International Meeting of the Institute of Management Sciences, Athens.

[57] **Frangos, N., and Vrontos, S. (2001).** Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. ASTIN Bulletin, Vol. 31, No.1, pp. 1-22.

[58] **Frangos, N., and Karlis, D. (2004).** Modelling Losses using a Exponential-Inverse Gaussian Distribution. Inusrance: Mathematics and Economics, 35, 53-67.

[59] **Geller, H. (1979).** An introduction Mathematical Risk Theory. University of Pennsylvania.

[60] **Gilde, V., and Sundt., B. (1989a).** On Bonus Systems with Credibility Scales. Scandinavian Actuarial Journal. 13-22.

[61] **Goovaerts, M., De Vijlder, F., and Haezendonck, J. (1984).** Insurance Premiums: Theory and Applications. Amsterdam: North Holland.

[62] **Gourieroux, C., Montfort, A., and Trongoton., A. (1984 a).** Pseudo maximum likelihood methods: theory. Econometrica, 52, 681-700.

[63] **Gourieroux, C., Montfort, A., and Trongoton., A. (1984 b).** Pseudo maximum likelihood methods: application to Poisson models. Econometrica, 52, 701-720.

[64] **Gourieroux, C., and Jasiak, J. (2004).** Heterogeneous INAR(1) Model with Application to Car Insurance. Insurance: Mathematics and Economics 34: 177–92.

[65] **Green, P. J., and Silverman, B.W. (1994).** Nonparametric Regression and Generalized Linear Models. Chapman and Hall, London.

[66] **Gregorius, F. (1982).** Development of the Study. In New Motor Rating Structure in the Netherlands. ASTIN Groep Nederland, 15-37.

[67] **Grenander, U. (1957a).** On the heterogeneity in non-life insurance. Scandinavian Actuarial Journal, 153–179.

[68] **Grenander, U. (1957b).** Some remarks on bonus systems in automobile insurance. Scandinavian Actuarial Journal, 180–197.

[69] **Guerreiro, G. R., and Mexia, J. T. (2004).** An alternative approach to bonus-malus. Discuss. Math. Probab. Stat.24, no.2.

[70] **Gurtler, M. (1963).** Bonus ou malus. ASTIN Bulletin. 3, 43-61.

[71] **Haberman, and Renshaw, A.E. (1996).** Generalized linear models and actuarial science. The Statistician, 45(4):407–436.

[72] **Hastie, T.J., and Tibshirani, R.J. (1990).** Generalized Additive Models. Chapman and Hall, London.

[73] **Hausmann, J.A., Hall, B.H., and Griuches, Z. (1984).** Econometric models for count data with an application to the patents-R&D relationship. Econometrica, 46, 1251-1271.

[74] **Healy, M. (1986).** Matrices for Statistics. Oxford Science Publications.

[75] **Herzog, T. (1996).** Introduction to Credibility Theory. Actex Publications, Winstead.

[76] **Heller, G. Z., Stasinopoulos, D.M., Rigby, R.A, and de Jong P. (2007).** Mean and dispersion modeling for policy claims costs. Scandinavian Actuarial Journal, 4, 281-292.

[77] **Hilbe, J. M. (2011).** Negative Binomial Regression Extensions. Cambridge University Press.

[78] **Hogg, R.V., and Klugman, S.A. (1984).** Loss Distributions. John Wiley & Sons, New York.

[79] **Holtan, J. (1994).** Bonus Made Easy. ASTIN Bulletin. 24, 61-74.

[80] **Hulin, L. (1999).** Modelisation de la survenance d'accidents.G.E.M.M.E, 9924, Universite de Liege.

[81] **Hurlimann, W. (1990).** On Maximum Likelihood Estimation for Count Data Models. Insurance Mathematics and Economics, 9 39-49.

[82] **Islam, M.N., and Consul, P.C. (1992).** A Probabilistic Model for Automobile Claims. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker. 85-93.

[83] **Johnson, N. L. (1949).** Systems of frequency curves generated by methods of translation. Biometrika, 36: 149-176.

[84] **Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994).** Continuous Univariate Distributions, Volume I, 2nd edn. Wiley, New York.

[85] **Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995).** Continuous Univariate Distributions, Volume II, 2nd edn. Wiley, New York.

[86] **Johnson, N. L., Kotz S., and Kemp A.W. (2005).** Univariate Discrete Distributions. Wiley.

[87] **Jones, M. C. (2005).** In discussion of Rigby, R. A., and Stasinopoulos, D. M. (2005) Generalized additive models for location, scale and shape,. Applied Statistics, 54: 507-554.

[88] **Jørgensen, B. (1982).** Statistical Properties of the Generalized Inverse Gaussian Distribution. In Lecture Notes in Statistics, 9. Springer-Verlag, New York.

[89] **Jørgensen, B. (1997).** The Theory of Dispersion Models. Chapman and Hall, London.

[90] **Jørgensen, B., and Paes de Souza, M.C. (1994).** Fitting Tweedie's compound Poisson model to insurance claims data. Scandinavian Actuarial Journal, 69-93.

[91] **Justens, D. (1996).** Construction interpreted de distributions de probability a valeurs dans les naturels. G.E.M.M.E, 9603, Universite de Liege.

[92] **Kalb, G.R.J., Kofman, P., and Vorst, T.C.F. (1996).** Mixtures of tails in clustered automobile collision claims. Insurance, Mathematics and Economics, 18, pp 89-107.

[93] **Karlis, D., and Xekalaki, E. (2006).** The Polygonal Distribution. Paper presented at the International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo, University of Castilla-La Mancha.

[94] **Kestmont, R. M., and Paris, J. (1985).** Sur la Ajustement du Nombre de Sinistres. Bulletin of the Swiss Actuaries, 85 157-164.

[95] **Klugman, S., Panjer, H., and Willmot, G. (2004).** Loss Models: From Data to Decisions. New York: Wiley.

[96] **Kuha, J. (2004).** AIC and BIC Comparisons of Assumptions and Performance. Sociological Methods and Research 33: 188–229.

[97] **Lambert, D. (1992).** Zero-inflated Poisson Regression with an application to defects in Manufacturing. Technometrics, 34: 1-14.

[98] **Lawless, J.F. (1987).** Negative Binomial Distribution and Mixed Poisson Regression. Canadian Journal of Statistics, 15, 3, 209-225.

[99] **Lemaire, J. (1975).** Si les assures connaissaient la programmation dynamique. Bulletin de la Association Rovale des Actuaires Beiges. 54-63.

[100] **Lemaire, J. (1976).** Driver Versus Company: Optimal Behaviour of the Policyholder. Scandinavian Actuarial Journal. 209-219.

[101] **Lemaire, J. (1977).** Critique du tariff automobile responsablite civil belge. Association Rovale des Actuaires Beiges, pp. 93-109.

[102] **Lemaire, J. (1977a).** Lasoif du bonus. ASTIN Bulletin. 9, 181-190.

[103] **Lemaire, J. (1977b).** Selection Procedures of Regression Analysis Applied to Automobile Insurance. Mitteilungen der Vereinigunq Schweizerischer Versiefaerengsroatheroatiker. Part I. 1977, 143-160.

[104] **Lemaire, J. (1977c).** Critique du tarif automobile beige. Bulletin de i'Association Rovale des Actuaires Beiges. 93-109.

[105] **Lemaire, J. (1979a).** Selection Procedures of Regression Analysis Applied to Automobile Insurance. Mitteilungen der Vereinieung Schweizerischer Versieherungsmathematiker. Part II, 1979, 65-72.

[106] **Lemaire, J. (1979b).** How to Define a Bonus-Malus System With an Exponential Utility Function. ASTIN Bulletin. 10, 274-282.

[107] **Lemaire, J. (1984).** The Effect of Expense Loadings on the Fairness of a Tariff. ASTIN Bulletin. 14, 165-171.

[108] **Lemaire, J. (1984).** Automobile insurance: Actuarial models. Kluwer-Nijhoff, Netherlands.

[109] **Lemaire, J. (1988a).** Construction of the New Belgian Motor Third Party Tariff Structure. ASTIN Bulletin. 18. 99-111.

[110] **Lemaire, J. (1988b).** A Comparative Analysis of Most European and Japanese Bonus-Malus Systems. Journal of Risk and Insurance. 55, 660-681.

[111] **Lemaire, J. (1992).** Negative Binomial or Poisson-Inverse Gaussian. Proceedings of the Twenty-fourth International Congress of Actuaries. Montreal.

[112] **Lemaire, J. (1993).** Selecting a Fitting Distribution for Taiwanese Automobile Losses. Unpublished manuscript.

[113] **Lemaire, J., and Zi., H. (1994a).** High Deductibles Instead of Bonus-Malus: Can it Work?ASTIN Bulletin. 24, 75-86.

[114] **Lemaire, J., and Zi., H. (1994b).** A Comparative Analysis of 30 Bonus-Malus Systems. ASTIN Bulletin. 24, 287-309.

[115] **Lemaire, J. (1995).** Bonus-Malus Systems in Automobile Insurance. Kluwer Academic Publishers.

[116] **Lindsay, B. (1995).** Mixture Models Theory, Geometry and Applications. Pennsylvania State University.

[117] **Loimaranta, K. (1972).** Some Asymptotic Properties of Bonus Systems. ASTIN Bulletin. 6, 233-245.

[118] **Lopatatzidis, A., and Green, P. J. (2000).** Nonparametric quantile regression using the gamma distribution. submitted for publication.

[119] **Mack, T. (1991).** A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. ASTIN Bulletin., vol. 21, no.l, pp. 93-109.

[120] **Mahmoudvand, R., and Hassani, H. (2009).** Generalized Bonus-Malus systems with a frequency and a severity component on an individual basis in automobile insurance. ASTIN Bulletin, 39, 307-315.

[121] **Martin, D.B. (1960).** Automobile Insurance: Canadian Accident-Free Classification System. ASTIN Bulletin. 1, 123-133.

[122] **Martin-Lof, A. (1973).** A Method for Finding the Optimal Deductible Rule for a Policyholder of an Insurance Company With a Bonus System. Scandinavian Actuarial Journal. 23-29.

[123] **McCullagh, P., and Nelder, J. A. (1989).** Generalized Linear Models (2nd ed.).London: Chapman and Hall.

[124] **McDonald, J. B., and Xu, Y. J. (1995).** A generalization of the beta distribution with applications. Journal of Econometrics, 66, 133-152.

[125] **McDonald, J. B. (1996).** Probability Distributions for Financial Models. In Maddala,G. S. and C. R.Rao (eds.). Handbook of Statistics, 14, 427-460. Elsevier.

[126] **McLachlan, G., and Peel, D. (2000).** Finite Mixture Models. John Wiley & Sons.

[127] **Mert, M., and Saykan, Y. (2005).** On a bonus-malus system where the claim frequency distribution is Geometric and the claim severity distribution is Pareto. Hacettepe Journal of Mathematics and Statistics Vol. 34, pp.75-81.

[128] **Nelder, J.A., and Wedderburn, R.W.M. (1972).** Generalized Linear Models. Journal of the Royal Statistical Society A, 135, 370-384.

[129] **Neuhaus, W. (1988).** A Bonus-Malus System in Automobile Insurance. Insurance: Mathematics and Economics. 7, 103- 112.

[130] **Nonnan, J.M., and Shearns, D.C.S. (1980).** Optimal Claiming on Vehicle Insurance Revisited. Journal of the Operational Research Society. 31, 181-186.

[131] **Norberg, R. (1976).** A Credibility Theory for Automobile Bonus Systems. Scandinavian Actuarial Journal. 92-107.

[132] **Ortiz, E. (1990).** A Stochastic Model of the Distribution of Wars in Time. American Statistical Association Annual Meeting. Los Angeles.

[133] **Panjer, H.H. (1987).** Models of Claim Frequency. In Advances in the Statistical Sciences: Actuarial Science. VI, 115-125. LB. MacNeill and G J. Umphrey (eds.).

[134] **Panjer, H.H., and Willmot, G.E. (1992).** Insurance Risk Models. Schaumburg, III.: Society of Actuaries.

[135] **Panjer, H.H., and Wtllmot, G.E. (1988a).** Motivating Claims Frequency Models. Proceedings of the Twenty-third International Congress of Actuaries, Helsinki, 3, 269-284.

[136] **Partrat, C. (1993).** Compound Poisson Models for Two Types of Claims. ASTIN Colloquium, Cambridge.

[137] **Picard, P. (1976).** Generalisation de P etude sur la survenance des sinistres en assurance automobile. Bulletin Trimestriel de PInstitut des Actuaires Francais. 204-267.

[138] **Picech, L. (1994).** The Merit-Rating Factor in a Multiplicative Rate-Making Model. ASTIN Colloquium, Cannes.

[139] **Pinquet, J. (1997).** Allowance for Costs of Claims in Bonus-Malus Systems. ASTIN Bulletin, 27, 33-57.

[140] **Pinquet, J. (1998).** Designing Optimal Bonus-Malus Systems From Different Types of Claims. ASTIN Bulletin, 28, 205-220.

[141] **Pinquet, J., Guillen M., and Bolance, C. (2001).** Long-range contagion in automobile insurance data: estimation and implications for experience rating. ASTIN Bulletin, 31(2), 337-348.

[142] **Pitrebois, S., Denuit, M., and Walhin, J.F. (2003a).** Fitting the Belgian Bonus-Malus system. Belgian Actuarial Bulletin 3, 58–62.

[143] **Pitrebois, S., Denuit, M., and Walhin, JF. (2003b).** Setting a bonus-malus scale in the presence of other rating factors: Taylor's work revisited. ASTIN Bulletin 33, 419–436.

[144] **Pitrebois, S., Denuit, M., and Walhin, JF. (2004).** Bonus-malus scales in segmented tariffs: Gilde & Sundt's work revisited. Australian Actuarial Journal 10, 107–125.

[145] **Pitrebois, S., Denuit, M., and Walhin, JF. (2005).** Bonus-malus systems with varying deductibles. ASTIN Bulletin 35, 261–274.

[146] **Pitrebois, S., Denuit, M., and Walhin, JF. (2006a).** Multi-event bonus-malus scales. Journal of Risk and Insurance 73, 517–528.

[147] **Pitrebois, S., Denuit, M., and Walhin, JF. (2006b).** An actuarial analysis of the French bonus-malus system. Scandinavian Actuarial Journal, 247–264.

[148] **Pitrebois, S., Walhin, J.F., and Denuit, M. (2006c).** How to transfer policyholders from one bonus-malus scale to the other? German Actuarial Bulletin 27, 607–618.

[149] **Renshaw, A.E. (1994).** Modelling The Claims Process in the Presence of Covariates. ASTIN Bulletin, 24, 265-285.

[150] **Rigby, R.A., and Stasinopoulos, D.M. (1996a).** A Semi-parametric Additive Model for Variance Heterogeneity. Statistal Computing, 6, 57-65.

[151] **Rigby, R.A., and Stasinopoulos, D.M. (1996b).** Mean and Dispersion Additive Models. In W. Hardle and M.G. Schimek (eds.), Statistical Theory and Computational Aspects of Smoothing, pp. 215-230. Physica, Heidelberg.

[152] **Rigby, R.A., and Stasinopoulos, D. M. (2001).** The GAMLSS project: a flexible approach to statistical modelling. In B. Klein and L. Korsholm (eds.), New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling, 249–256, Odense, Denmark.

[153] **Rigby, R.A., and Stasinopoulos, D.M. (2004).** Smooth Centile Curves for Skew and Kurtotic data Modelled Using the Box-Cox Power Exponential Distribution. Statistics in Medicine, 23, 3053-3076.

[154] **Rigby, R. A., and Stasinopoulos, D. M (2005).** Generalized additive models for location, scale and shape, (with discussion).Applied Statistics, 54, 507–554.

[155] **Rigby, R.A., and Stasinopoulos, D.M. (2006)**. Using the Box-Cox t Distribution in GAMLSS to Model Skewness and Kurtosis. Statistical Modelling, 6, 209-229.

[156] **Rigby, R.A., Stasinopoulos, D.M., and Akantziliotou, C. (2008).** A framework for modeling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. Computational Statistics and Data Analysis, 53, 381–393.

[157] **Rigby, R. A., and Stasinopoulos, D. M. (2009).** A flexible regression approach using GAMLSS in R.

[158] **Ruohonen, M. (1988).** A Model for the Claim Number Process. ASTIN Bulletin. 18, 57-68.

[159] **Seneta, E. (1981).** A Non-Negative Matrices and Markov Chains. Springer- Verlag.

[160] **Schiesinger, H. (1981).** The Optimal Level of Deducibility in Insurance Contracts. Journal of Risk and Insurance. 48, 465-481.

[161] **Sichel, H. (1985).** A bibliometric distribution which really works. Journal of the American society for information science, 36(5), 314-321.

[162] **Sichel, H. (1971).** On a Family of Discrete Distributions Particularly Suited to Represent Long-Tailed Frequency Data. Proceedings of the Third Symposium on Mathematical Statistics, N. Loubsher (ed.), Pretoria.

[163] **Sigalotti, L. (1994).** Equilibrium Premiums in a Bonus-Malus System. ASTIN Colloquium, Cannes.

[164] **Sharif, A.H., and Panjer, H.H. (1993).** A Probabilistic Model for Automobile Claims: A Comment on the Article by M.N. Islam and P.C. Consul. Mitteilungen der Vereinigung Sehweizeriseher Versicherungsmathematiker. 279-292.

[165] **Stasinopoulos, D. M. (2006).** Contribution to the discussion of the paper by Lee and Nelder, Double hierarchical generalized linear models. Appl. Statist., 55: 171-172.

[166] **Stasinopoulos, D. M., Rigby, R. A., and Akantziliotou, C. (2007).** Instructions on how to use the GAMLSS package in R, Second Edition. Technical Report 01/08, STORM Research Centre, London Metropolitan University, London.

[167] **Stasinopoulos, D. M., Rigby, R. A., and Fahrmeir, L. (2000).** Modelling rental guide data using mean and dispersion additive models. Statistician, 49: 479-493.

[168] **Stein, G. Z., Zucchini, W., and Juritz, J. M. (1987).**Parameter Estimation of the Sichel Distribution and its Multivariate Extension. Journal of American Statistical Association, 82: 938-944.

[169] **Stuart, C. (1983).** Pareto-Optimal Deductibles in Property-Liability Insurance: The Case of Homeowner Insurance in Sweden. Scandinavian Actuarial Journal. 227-238.

[170] **Sundt, B. (1984).** An Introduction to Non-Life Insurance Mathematics. Karlsruhe: Verlag Versicherungswirtschaft.

[171] **Sundt, B. (1989b).** Bonus Hunger and Credibility Estimators with Geometric Weights. Insurance: Mathematics and Economics. 8, 119-126.

[172] Sundt, B. and W. Jewell. (1981). Further Results on Recursive Evaluation of Compound Distributions. ASTIN Bulletin. 12, 27-39.

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

[174] **Tremblay, L. (1992)**. Using the Poisson Inverse Gaussian in Bonus-Malus Systems. ASTIN Bulletin. 22. 97-106.

[175] **Vandenbroek, M. (1993).** Bonus-malus system or partial coverage to oppose moral hazard problems., Insurance: Mathematics and Economics, 13, pp 1-5.

[176] **Venezia, I., and Levy, H. (1980).** Optimal Claims in Automobile Insurance. Review of Economic Studies, 47, 539-549.

[177] **Venezia, I., and Levy, H. (1983).** Optimal Multi-Period Insurance Contracts. Insurance: Mathematics and Economics. 2, 199-208.

[178] **Venter, G. (1990).** Credibility. In Foundations of Casualty Actuarial Science, Washington, D.C.: Casualty Actuarial Society.

[179] **Venter, G. (1991a).** Effects of Variations from Gamma-Poisson Assumptions. Proceedings of the Casualty Actuarial Society. 78, 41-55.

[180] **Venter, G. (1991b).** A Comparative Analysis of Most European and Japanese Bonus-malus Systems: Extension. Journal of Risk and Insurance. 58, 542-547.

[181] **Vepsafainen, S. (1972).** Applications to a Theory of Bonus Systems. ASTIN Bulletin. 6, 212-221.

[182] **Vuong, Q. (1989).** Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57, 307-333.

[183] **Wald, A., and Wolfowitz, J. (1951).** Bayes Solutions of Sequential Decision Problems. Annals of Mathematical Statistics. 53, 82-99.

[184] **Walhin, J. F., and Paris, J. (1997).** Using mixed Poisson process in connection with bonus-malus systems. ASTIN Bulletin, Vol 29, No I 1999, pp. 81-99.

[185] **Walhin, J. F., and Paris, J. (2000).** The true claim amount and frequency distribution within a bonus-malus system. ASTIN Bulletin, Vol 30,2000, pp. 391-403.

[186] **Wiilmot, G.E. (1990).** AsymptoticTail Behaviour of Poisson Mixtures with Applications. Advances in Applied Probability. 22, 147-159.

[187] **Wiilmot, G.E. (1993).** On Recursive Evaluation of Mixed Poisson Probabilities and Related Quantities. Scandinavian Actuarial Journal, 114-133.

[188] **Willmot, G.E. (1986).** Mixed Compound Poisson Distributions. ASTIN Bulletin. 16-S, 59-79.

[189] **Willmot, G.E. (1987).** The Poisson-Inverse Gaussian Distribution as an Alternative to the Negative Binomial. Scandinavian Actuarial Journal. 113-127.

[190] **Willmot, G.E. (1988b).** Sundt and Jewell's Family of Discrete Distributions. ASTIN Bulletin. 18, 17-29.

[191] **Yip, K., and Yau, K. (2005).** On Modeling Claim Frequency Data in General Insurance with Extra Zeros. Insurance: Mathematics and Economics 36: 153-63.