COMMITTEE OF ASTIN

Hans BÜHLMANN	Honorary Chairman
Edward J. LEVAY	Chairman elsystems@premium.org
Jean H. LEMAIRE	Vice Chairman IAA-Delegate lemaire@rider.wharton.upenn.edu
Charles LEVI	Secretary clevi@guycarp.com
Freddy CORLIER	Treasurer
Andrew J.G. CAIRNS	Editor
Paul EMBRECHTS	Editor
Chris D. DAYKIN	Member
Yasuhide FUJI	Member
Henk J. KRIEK	Member
W. James MACGINNITIE	Member/IAA-Delegate
Thomas MACK	Member
Ermanno PITACCO	Member
Jukka RANTALA	Member
James N. STANARD	Member
Greg TAYLOR	Member
Stephen PHILBRICK	Co-Editor
René SCHNIEPER	Co-Editor

Neither the COMMITTEE OF ASTIN nor PEETERS s.a. are responsible for statements made or opinions expressed in the articles, criticisms and discussions published in *ASTIN BULLETIN*.

Guidelines to Authors *continued from inside back cover*.

9. Contributions must be typewritten on one side of good quality paper, with double spacings and ample margins all round.
Illustrations should be submitted as clear black and white prints. Photocopies are not acceptable. Line thickness and lettering size should be adopted to suit any likely degree of reduction.
Each contribution should obtain the author(s) full address(es), including e-mail and fax numbers when available. Authors should notify the publisher whenever their contact details change.
A copy of text should also be submitted on disk, when available. Please state clearly the type of software used, and note the filename. Contributors should also retain a copy of their article.
Two sets of proofs will be shipped to authors who should ensure that one set plus the manuscript is returned to PEETERS within one week of receipt. Authors may be charged for alterations to the original manuscript. If authors proofs are not returned by the required date, the publisher's own corrected set will be forwarded to the printer.

10. Authors will receive 50 offprints free of charge. Additional offprints may be ordered when returning corrected proofs. A scale of charges will be enclosed when the proofs are sent out.

ASTIN BULLETIN

Volume 31, No. 1

May 2001

PEETERS

ISSN 0515-0361

Volume 31, No 1 May 2001

ASTIN BULLETIN

The Journal of the ASTIN and AFIR Sections of the International Actuarial Association

EDITORS:
Andrew Cairns
Paul Embrechts

CO-EDITORS:
Stephen Philbrick
René Schnieper

EDITORIAL BOARD:
Alois Gisler
Marc Goovaerts
Mary Hardy
Ole Hesselager
Christian Hipp
Jacques Janssen
William S. Jewell
Jean Lemaire
Jukka Rantala
Axel Reich
Robert Reitano

CONTENTS

PEETERS

EDITORIAL POLICY

ASTIN BULLETIN started in 1958 as a journal providing an outlet for actuarial studies in non-life insurance. Since then a well-established non-life methodology has resulted, which is also applicable to other fields of insurance. For that reason *ASTIN BULLETIN* has always published papers written from any quantitative point of view – whether actuarial, econometric, engineering, mathematical, statistical, etc. – attacking theoretical and applied problems in any field faced with elements of insurance and risk. Since the foundation of the AFIR section of IAA, i.e. since 1988, *ASTIN BULLETIN* has opened its editorial policy to include any papers dealing with financial risk.

We especially welcome papers opening up new areas of interest to the international actuarial profession.

ASTIN BULLETIN appears twice a year (May and November), each issue consisting of at least 80 pages.

Details concerning submission of manuscripts are given on the inside back cover.

MEMBERSHIP

ASTIN and AFIR are sections of the International Actuarial Association (IAA). Membership is open automatically to all IAA members and under certain conditions to non-members also. Applications for membership can be made through the National Correspondent or, in the case of countries not represented by a national correspondent, through a member of the Committee of ASTIN.

Members of ASTIN receive *ASTIN BULLETIN* free of charge. As a service of ASTIN to the newly founded section AFIR of IAA, members of AFIR also receive *ASTIN BULLETIN* free of charge.

SUBSCRIPTION AND BACK ISSUES

Subscription price: 80 Euro.

Payments should be sent to Peeters, Bondgenotenlaan 153, B-3000 Leuven, Belgium.

To place your order or make an inquiry please contact: Peeters, Journals Department, Bondgenotenlaan 153, B-3000 Leuven, Belgium or e-mail: peeters@peeters-leuven.be

Orders are regarded as binding and payments are not refundable.

Subscriptions are accepted and entered by the volume. Claims cannot be made 4 months after publication or date of order, whichever is later.

INDEX TO VOLUMES 1-27

The Cumulative Index to Volumes 1-27 is also published for ASTIN by Peeters at the above address and is available for the price of 80 Euro.

Copyright © 2000 PEETERS

GUIDELINES TO AUTHORS

1. Papers for publication should be sent in quadruplicate to one of the Editors:

Andrew Cairns
Department of Actuarial Mathematics and Statistics
Heriot-Watt University
Edinburgh EH 14 4AS, United Kingdom
A.Cairns@ma.hw.ac.uk

Paul Embrechts
Department of Mathematics
ETHZ
CH-8092 Zurich, Switzerland.
embrechts@math.ethz.ch

or to one of the Co-Editors:

Stephen Philbrick
Swiss Re Investors
111 S. Calvert Street, Suite 1800
Baltimore, MD 21202, USA
stephen_philbrick@swissre.com

René Schnieper,
Zurich Insurance Company,
P.O. Box
CH-8022 Zurich, Switzerland
rene.schnieper@zurichre.com

Submission of a paper is held to imply that it contains original unpublished work and is not being submitted for publication elsewhere.
Receipt of the paper will be confirmed and followed by a refereeing process, which will take about three months.

2. The basic elements of the journal's style have been agreed by the Editors and Publishers and should be clear from checking a recent issue of *ASTIN BULLETIN*. If variations are felt necessary they should be clearly indicated on the manuscript.

3. Papers should be written in English or in French. Authors intending to submit longer papers (e.g. exceeding 30 pages) are advised to consider splitting their contribution into two or more shorter contributions.

4. The first page of each paper should start with the title, the name(s) of the author(s), and an abstract of the paper as well as some major keywords. An institutional affiliation can be placed between the name(s) of the author(s) and the abstract.

5. Footnotes should be avoided as far as possible.

6. References should be arranged alphabetically, and for the same author chronologically. Use a, b, c, etc. to separate publications of the same author in the same year. For journal references give author(s), year, title, journal (in italics, cf. point 9), volume (in boldface, cf. point 9), and pages. For book references give author(s), year, title (in italics), publisher, and city.
Examples

BARLOW, R.E. and PROSCHAN, F. (1975) *Mathematical Theory of Reliability and Life Testing*. Holt, Rinehart, and Winston, New York.
JEWELL, W.S. (1975a) Model variations in credibility theory. In *Credibility: Theory and Applications* (ed. P.M. KAHN), pp. 193-244, Academic Press, New York.
JEWELL, W.S. (1975b) Regularity conditions for exact credibility. *ASTIN Bulletin* **8**, 336-341.

References in the text are given by the author's name followed by the year of publication (and possibly a letter) in parentheses.

7. The address of at least one of the authors should be typed following the references.

8. Italics (boldface) should be indicated by single (wavy) underlining. Mathematical symbols will automatically be set in italics, and need not be underlined unless there is a possibility of misinterpretation. Information helping to avoid misinterpretation may be listed on a separate sheet entitled 'special instructions to the printer'. (Example of such an instruction: Greek letters are indicated with green and script letters with brown underlining, using double underlining for capitals and single underlining for lower case).

Continued overleaf.

DESIGN OF OPTIMAL BONUS-MALUS SYSTEMS WITH A FREQUENCY AND A SEVERITY COMPONENT ON AN INDIVIDUAL BASIS IN AUTOMOBILE INSURANCE

BY

NICHOLAS E. FRANGOS* AND SPYRIDON D. VRONTOS*

ABSTRACT

The majority of optimal Bonus-Malus Systems (BMS) presented up to now in the actuarial literature assign to each policyholder a premium based on the number of his accidents. In this way a policyholder who had an accident with a small size of loss is penalized unfairly in the same way with a policyholder who had an accident with a big size of loss. Motivated by this, we develop in this paper, the design of optimal BMS with both a frequency and a severity component. The optimal BMS designed are based both on the number of accidents of each policyholder and on the size of loss (severity) for each accident incurred. Optimality is obtained by minimizing the insurer's risk. Furthermore we incorporate in the above design of optimal BMS the important a priori information we have for each policyholder. Thus we propose a generalised BMS that takes into consideration simultaneously the individual's characteristics, the number of his accidents and the exact level of severity for each accident.

KEYWORDS

Optimal BMS, claim frequency, claim severity, quadratic loss function, a priori classification criteria, a posteriori classification criteria.

1. INTRODUCTION

BMS penalize the policyholders responsible for one or more claims by a premium surcharge (malus) and reward the policyholders who had a claim free year by awarding discount of the premium (bonus). In this way BMS

* Department of Statistics, Athens University of Economics and Business, Patission 76, 10434, Athens, Greece. E-mail for correspondence nef@aueb.gr and svrontos@aueb.gr
This work has been partially supported by 96SYN 3-19 on "Design of Optimal Bonus-Malus Systems in Automobile Insurance" and the General Secreteriat of Research and Technology of Greece. The authors would like to thank the referees for their valuable comments.

encourage policyholders to drive carefully and estimate the unknown risk of each policyholder to have an accident.

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. Optimal BMS can be divided in two categories: those based only on the a posteriori classification criteria and those based both on the a priori and the a posteriori classification criteria. As a posteriori classification criteria are considered the number of accidents of the policyholder and the severity of each accident. As a priori classification criteria are considered the variables whose their values are known before the policyholder starts to drive, such as characteristics of the driver and the automobile. The majority of BMS designed is based on the number of accidents disregarding their severity. Thus first let us consider the design of optimal BMS based only on the a posteriori claim frequency component.

1.1. BMS based on the a posteriori claim frequency component

Lemaire (1995) developed the design of an optimal BMS based on the number of claims of each policyholder, following a game-theoretic framework introduced by Bichsel (1964) and Bühlmann (1964). Each policyholder has to pay a premium proportional to his own unknown claim frequency. The use of the estimate of the claim frequency instead of the true unknown claim frequency will incur a loss to the insurer. The optimal estimate of the policyholder's claim frequency is the one that minimizes the loss incurred. Lemaire (1995) considered, among other BMS, the optimal BMS obtained using the quadratic error loss function, the expected value premium calculation principle and the Negative Binomial as the claim frequency distribution. Tremblay (1992) considered the design of an optimal BMS using the quadratic error loss function, the zero-utility premium calculation principle and the Poisson-Inverse Gaussian as the claim frequency distribution. 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. As we see, all the BMS mentioned above take under consideration only the number of claims of each policyholder disregarding their severity.

1.2. BMS based on the a priori and the a posteriori claim frequency component

The models mentioned above are function of time and of past number of accidents and do not take into consideration the characteristics of each individual.

In this way as mentioned in Dionne and Vanasse (1989), the premiums do not vary simultaneously with other variables that affect the claim frequency distribution. The most interesting example is the age variable. Suppose that age has a negative effect on the expected number of claims, it would imply that insurance premiums should decrease with age. Premium tables derived from BMS based only on the a posteriori criteria, even though are a function of time, do not allow for a variation of age, even though age is a statistically significant variable.

Dionne and Vanasse (1989, 1992) presented a BMS that integrates a priori and a posteriori information on an individual basis. This BMS is derived as a function of the years that the policyholder is in the portfolio, of the number of accidents and of the individual characteristics which are significant for the number of accidents. Picech (1994) and Sigalotti (1994) derived a BMS that incorporates the a posteriori and the a priori classification criteria, 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) developed the setting of a Bonus-Malus scale where some rating factors are used to recognize the differentiation of underlying claim frequency by experience, but only to the extent that this differentiation is not recognized within base premiums. Pinquet (1998) developed the design of optimal BMS from different types of claims, such as claims at fault and claims not at fault.

1.3. Allowance for the severity in BMS

In the models briefly described above the size of loss that each accident incurred is not considered in the design of the BMS. Policyholders with the same number of accidents pay the same malus, irrespectively of the size of loss of their accidents. In this sense the BMS designed in the above way are unfair for the policyholders who had an accident with a small size of loss. Actually as Lemaire (1995) is pointing out all BMS in force throughout the world, with the exception of Korea, are penalizing the number of accidents without taking the severity of such claims into account. In the BMS enforced in Korea the policyholders who had a bodily injury claim pay higher maluses, depending on how severe the accident was, than the policyholders who had a property damage claim. The BMS designed to take severity into consideration include those from Picard (1976) and Pinquet (1997). Picard generalized the Negative Binomial model in order to take into account the subdivision of claims into two categories, small and large losses. In order to separate large from small losses, two options could be used: 1. The losses under a limiting amount are regarded as small and the remainder as large. 2. Subdivision of accidents in those that caused property damage and those that cause bodily injury, penalizing more severely the policyholders who had a bodily injury accident. Pinquet (1997) designed an optimal BMS which makes allowance for the severity of the claims in the following way: starting from a rating model based on the analysis of number of claims and of costs of claims, two heterogeneity components are added. They represent unobserved factors that are relevant for the explanation of the severity variables. The costs of claims are supposed to follow gamma or lognormal distribution. The rating factors, as well as the heterogeneity components are included in the scale parameter of the distribution. Considering that the heterogeneity also follows a gamma or lognnormal distribution, a credibility expression is obtained which provides a predictor for the average cost of claim for the following period.

Our first contribution in this paper 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. We assumed that the number of claims is distributed according the Negative Binomial distribution and the losses of the claims are distributed according the Pareto distribution, and we have expanded the frame that Lemaire (1995) used to design an optimal BMS based on the number of claims. Applying Bayes' theorem we find the posterior distribution of the mean claim frequency and the posterior distribution of 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. For more on this subject we refer to Vrontos (1998).

Our second contribution is the development of a generalized BMS that integrates the a priori and the a posteriori information on a individual basis. In this generalized BMS the premium will be a function of the years that the policyholder is in the portfolio, of his number of accidents, of the size of loss that each of these accidents incurred, and of the significant a priori rating variables for the number of accidents and for the size of loss that each of these claims incurred. We will do this by expanding the frame developed by Dionne and Vanasse (1989, 1992).

Pinquet (1997) is starting from a rating model and then he is adding the heterogeneity components. We design first an optimal BMS based only on the a posteriori classification criteria and then we generalize it in order to take under consideration both the a priori and the a posteriori classification criteria.

2. Design of Optimal BMS with a Frequency and a Severity Component Based on the A Posteriori Criteria

It is assumed that the number of claims of each policyholder is independent from the severity of each claim in order to deal with the frequency and the severity component separately.

2.1. Frequency component

For the frequency component we will use the same structure used by Lemaire (1995). The portfolio is considered to be heterogeneous and all policyholders have constant but unequal underlying risks to have an accident. Consider that the number of claims k, given the parameter λ, is distributed according to Poisson(λ),

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

$k = 0, 1, 2, 3, \ldots$ and $\lambda > 0$ and λ is denoting the different underlying risk of each policyholder to have an accident. Let us assume for the structure function that $\lambda \sim gamma(a, \tau)$ and λ has a probability density function of the form:

$$u(\lambda) = \frac{\lambda^{a-1}\tau^a \exp(-\tau\lambda)}{\Gamma(a)}, \quad \lambda > 0, \ a > 0, \ \tau > 0$$

with mean $E(\lambda) = a/\tau$ and variance $Var(\lambda) = a/\tau^2$. Then it can be proved that the unconditional distribution of the number of claims k will be Negative Binomial (a,τ), with probability density function

$$P(k) = \binom{k+a-1}{k}\left(\frac{\tau}{1+\tau}\right)^a \left(\frac{1}{1+\tau}\right)^k,$$

mean equal to $E(k) = a/\tau$ and variance equal to $Var(k) = (a/\tau)\,(1 + 1/\tau)$. The variance of the Negative Binomial exceeds its mean, a desirable property which is common for all mixtures of Poisson distribution and allows us to deal with data that present overdispersion.

Let us denote as $K = \sum_{i=1}^{t} k_i$ the total number of claims that a policyholder had in t years, where k_i is the number of claims that the policyholder had in the year i, $i = 1,\ldots,t$. We apply the Bayes' theorem and we obtain the posterior structure function of λ for a policyholder or a group of policyholders with claim history $k_1,\ldots.k_t$, denoted as $u(\lambda|k_1, \ldots k_t)$. It is that

$$u(\lambda|k_1,\ldots k_t) = \frac{(\tau+t)^{K+a}\lambda^{K+a-1}e^{-(t+\tau)\lambda}}{\Gamma(a+K)},$$

which is the probability density function of a $gamma(a + K, t + \tau)$. Using the quadratic error loss function the optimal choice of λ_{t+1} for a policyholder with claim history $k_1, \ldots.k_t$ will be the mean of the posterior structure function, that is

$$\lambda_{t+1}(k_1,\ldots,k_t) = \frac{a+K}{t+\tau} = \bar{\lambda}\left(\frac{a+K}{a+t\bar{\lambda}}\right), \text{ where } \bar{\lambda} = \frac{a}{\tau}. \tag{1}$$

From the above it is clear that the occurrence of K accidents in t years just necessitates an update of the parameters of gamma, from a and τ to $a + K$ and $t + \tau$ respectively and the gamma is said to have the important property of the stability of the structure function as the gamma is a conjugate family for the Poisson likelihood.

2.2. Severity component

Let us consider now the severity component. Let x be the size of the claim of each insured. We consider as y the mean claim size for each insured and we assume that the conditional distribution of the size of each claim given the mean claim size, $x|y$, for each policyholder is the one parameter exponential distribution with parameter y, and has a probability density function given by

$$f(x\,|\,y) = \frac{1}{y} \cdot e^{-\frac{x}{y}}$$

for $x > 0$ and $y > 0$. The mean of the exponential is $E(x\,|\,y) = y$ and the variance is $var(x\,|\,y) = y^2$. The mean claim size y is not the same for all the policyholders but it takes different values so it is natural our prior belief for y to be expressed in the form of a distribution. Consider that the prior distribution of the mean claim size y is Inverse Gamma with parameters s and m and probability density function, see for example Hogg and Klugman (1984) given by

$$g(y) = \frac{\frac{1}{m} \cdot e^{-\frac{m}{y}}}{\left(\frac{y}{m}\right)^{s+1} \cdot \Gamma(s)}.$$

The expected value of the mean claim size y will be:

$$E(y) = \frac{m}{s-1}.$$

The unconditional distribution of the claim size x will be equal to:

$$P(X = x) = \int_0^{\infty} f(x\,|\,y) \cdot g(y)\,dy =$$
$$= s \cdot m^s \cdot (x+m)^{-s-1}$$

which is the probability density of the Pareto distribution with parameters s and m. Thus, one way to generate the Pareto distribution is the following: if it is for the size of each claim given the mean claim size $x\,|\,y$ that $x\,|\,y \sim$ *Exponential*(y) and for the mean claim size y of each policyholder that $y \sim$ Inverse *Gamma*(s,m) then it is for the unconditional distribution of the claim size x in the portfolio that $x \sim$ *Pareto*(s,m). In this way, the relatively tame exponential distribution gets transformed in the heavy-tailed Pareto distribution and instead of using the exponential distribution which is often inappropriate for the modelling of claim severity we are using the Pareto distribution which is often a good candidate for modelling the claim severity. Taking the mean claim size y distributed according the Inverse Gamma, we incorporate in the model the heterogeneity that characterizes the severity of the claims of different policyholders. We should note here that such a generation of the Pareto distribution does not appear for the first time in the actuarial literature. Such a use can be found for example in Herzog (1996). To the best of our knowledge it is the first time it is used in the design of an optimal BMS.

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 distribution of the mean claim size y given the information we have about the claim size history for each policyholder for the time period he is in the portfolio. Consider that the policyholder is in the portfolio for t years and that the number of claims he had in the year i is denoted with k_i, by $K = \sum_{i=1}^{t} k_i$ is denoted the total number of claims he has, and by x_k is denoted the claim amount for the k claim. Then the information we have for his claim size history will be in the form of a vector $x_1, x_2, ..., x_k$ and the total claim amount for the specific policyholder over the t years that he is in the portfolio will be equal to $\sum_{k=1}^{K} x_k$. Applying Bayes' theorem we find the posterior distribution of the mean claim size y given the claims size history of the policyholder $x_1, ..., x_k$ and it is that:

$$g(y \mid x_1, ..., x_k) =$$

$$= \frac{\dfrac{1}{(m + \sum_{k=1}^{K} x_k)} \cdot e^{-\frac{m + \sum_{k=1}^{K} x_k}{y}}}{\left(\dfrac{y}{(m + \sum_{k=1}^{K} x_k)}\right)^{K+s+1} \cdot \Gamma(K+s)}$$

which is the probability density function of the Inverse Gamma($s + K$, $m + \sum_{k=1}^{K} x_k$). This means that the occurrence of K claims in t years with aggregate claim amount equal to $\sum_{k=1}^{K} x_k$ just necessitates for the distribution of the mean claim size an update of the parameters of the Inverse Gamma from s and m to $s + K$ and $m + \sum_{k=1}^{K} x_k$ respectively and the Inverse Gamma distribution is said to have the important property of being conjugate with the exponential likelihood. The mean of the posterior distribution of the mean claim size will be:

$$E(x \mid y) = \frac{m + \sum_{k=1}^{K} x_k}{s + K - 1} \tag{2}$$

and the predictive distribution of the size of the claim of each insured x will be also a member of the Pareto family.

2.3. Calculation of the Premium according the Net Premium Principle

As shown, the expected number of claims $\lambda_{t+1}(k_1, ..., k_t)$ for a policyholder or a group of policyholders who in t years of observation have produced K claims with total claim amount equal to $\sum_{k=1}^{K} x_k$ is given by (1) and the expected claim severity $y_{t+1}(x_1, ..., x_K)$ is given by (2).

Thus, the net premium that must be paid from that specific group of policyholders will be equal to the product of $\lambda_{t+1}(k_1, ..., k_t)$ and $y_{t+1}(x_1, ..., x_K)$, i.e. it will be equal to

$$Premium = \frac{a+K}{t+\tau} \cdot \frac{m+\sum_{k=1}^{K} x_k}{s+K-1} \tag{3}$$

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

1. the parameters of the Negative Binomial distribution a and τ, (see Lemaire (1995) for the estimation of the parameters of the Negative Binomial)
2. the parameters of the Pareto distribution s and m (see Hogg and Klugman (1984) for the estimation of the parameters of the Pareto distribution)
3. the number of years t that the policyholder is under observation,
4. his number of claims K and
5. his total claim amount $\sum_{k=1}^{K} x_k$.

All of these can be obtained easily and taking under consideration that the negative binomial is often used as a claim frequency distribution and the Pareto as a claim severity distribution this enlarges the applicability of the model.

2.4. 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 his claim frequency and his claim severity, taking into account, through the Bayes' theorem, all the information available for the time that he is in our portfolio both for the number of his claims and the loss that these claims incur. 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. Every single year the average of all premiums collected from all policyholders remains constant and equal to

$$P = \frac{a}{\tau} \frac{m}{s-1} \tag{4}$$

In order to prove this it is enough to show, considering that the claim frequency and the claim severity are independent components, that:

$$E_\Lambda[\Lambda] = E[E[\lambda | k_1, ..., k_t]] = \frac{a}{\tau}$$

and that

$$E_Y[Y] = E[E[y | x_1, ..., x_k]] = \frac{m}{s-1}.$$

A proof of the first can be found in Lemaire (1995), and of the second in Vrontos (1998).

3. In the beginning all the policyholders are paying the same premium which is equal to (4).
4. The more accidents caused and the more the size of loss that each claim incurred the higher the premium.
5. The premium always decreases when no accidents are caused.
6. The drivers who had a claim with 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 with somebody which had an accident with a big loss. In this way the phenomenon of bonus hunger will have a decrease and the estimate of the actual claim frequency will be more accurate.
7. The severity component is introduced in the design of a BMS which from a practical point of view is more crucial than the number of claims for the insurer since it is the component that determines the expenses of the insurer from the accidents and thus the premium that must be paid.
8. The estimator of the mean of severity may not be robust and therefore it is prone to be affected by variation. For practical use a more robust estimator could be used. (i.e. cutting of the data, M-estimator).

3. Design of Optimal BMS with a Frequency and a Severity Component Based Both on the A Priori and the A Posteriori Criteria

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, of the number of accidents and of the significant – for the number of accidents – individual characteristics. We extend this model by introducing the severity component. We propose a generalized BMS that integrates a priori and a posteriori information on an individual basis based both on the frequency and the severity component. This generalized BMS will be derived as a function of the years that the policyholder is in the portfolio, of the number of accidents, of the exact size of loss that each one of these accidents incurred, and of the significant individual characteristics for the number of accidents and for the severity of the accidents. Some of the a priori rating variables that could be used are the age, the sex and the place of residence of the policyholder, the age, the type and the cubic capacity of the car, etc. As already said one of the reasons for the development of a generalized model which integrates a priori and a posteriori information is that premiums should vary simultaneously with the variables that affect the distribution of the number of claims and the size of loss distribution.

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

$$Premium = GBM_F * GBM_S \tag{5}$$

where GBM_F denotes the generalized BMS obtained when only the frequency component is used and GBM_S denotes the generalized BMS obtained when only the severity component is used.

3.1. Frequency Component

The generalized BMS obtained with the frequency component GBM_F will be developed according to Dionne and Vanasse (1989, 1992). Consider an individual i with an experience of t periods. Assume that the number of claims of the individual i for period j, denoted as K_i^j, follows the Poisson distribution with parameter λ_i^j, and K_i^j are independent. The expected number of claims of the individual i for period j is then denoted by λ_i^j and consider that it is a function of the vector of h individual's characteristics, denoted as $c_i^j = (c_{i,1}^j, ..., c_{i,h}^j)$, which represent different a priori rating variables. Specifically assume that $\lambda_i^j = \exp(c_i^j \beta^j)$, where β^j is the vector of the coefficients. The non-negativity of λ_i^j is implied from the exponential function. The probability specification becomes

$$P\left(K_i^j = k\right) = \frac{e^{-\exp(c_i^j \beta^j)}(\exp(c_i^j \beta^j))^k}{k!}.$$

In this model we assume that the h individual characteristics provide enough information for determining the expected number of claims. The vector of the parameters β^j can be obtained by maximum likelihood methods, see Hausmann, Hall and Griliches (1984) for an application. 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. As Gourieroux, Montfort and Trognon (1984a), (1984b) suggested, we can write

$$\begin{aligned} \lambda_i^j &= \exp(c_i^j \beta^j + \varepsilon_i) = \\ &= \exp(c_i^j \beta^j) u_i, \end{aligned}$$

where $u_i = \exp(\varepsilon_i)$, yielding a random λ_i^j. If we assume that u_i follows a gamma distribution with $E(u_1) = 1$ and $Var(u_i) = 1/a$, the probability specification becomes

$$P(K_i^j = k) = \frac{\Gamma(k+a)}{k!\Gamma(a)} \left[\frac{\exp(c_i^j \beta^j)}{a}\right]^k \left[1 + \frac{\exp(c_i^j \beta^j)}{a}\right]^{-(k+a)},$$

which is a negative binomial distribution with parameters a and $\exp(c_i^j \beta^j)$. 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 $E(\varepsilon_i) = 0$. Then

$$E(K_i^j) = \exp(c_i^j \beta^j) \text{ and } Var(K_i^j) = \exp(c_i^j \beta^j) \left[1 + \frac{\exp(c_i^j \beta^j)}{a}\right].$$

The interesting reader can see for more on the Negative Binomial regression Lawless (1987). The insurer needs to calculate the best estimator of the expected number of accidents at period $t+1$ using the information from past experience for the claim frequency over t periods and of known individual characteristics over the t+1 periods. Let us denote this estimator as $\hat{\lambda}_i^{t+1}(K_i^1,..., K_i^t; c_i^1,..., c_i^{t+1})$. Using the Bayes theorem one finds that the posterior structure function for a policyholder with $K_i^1,..., K_i^t$ claim history and $c_i^1,..., c_i^{t+1}$ characteristics is gamma with updated parameters $(\alpha + \sum_{j=1}^{t} K_i^j, \frac{\alpha}{\exp(c_i^j \beta^j)} + t)$. Using the classical quadratic loss function one can find that the optimal estimator given the observation of $K_i^1,..., K_i^t$ and $C_i^1,..., C_i^{t+1}$, is equal to:

$$\hat{\lambda}_i^{t+1}(K_i^1,..., K_i^t; c_i^1,..., c_i^{t+1})$$

$$= \int_0^{\infty} \lambda_i^{t+1}(K_i^{t+1}, u_i) f\left(\lambda_i^{t+1} \middle| K_i^i,..., K_i^t;\, c_i^1,..., c_i^t\right) d\lambda_i^{t+1} =$$

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

where $\sum_{j=1}^{t} K_i^j$ denotes the total number of claims of policyholder i in t periods. When $t = 0$, $\hat{\lambda}_i^1 \equiv \exp(C_i^1 \beta^j)$ 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 the well-known univariate without regression component model, see Lemaire (1995), Ferreira (1974).

Now we will deal with the generalized bonus-malus factor obtained when the severity component is used. It will be developed in the following way.

3.2. Severity Component

Consider an individual i with an experience of t periods. Assume that the number of claims of the individual i for period j is denoted as K_i^j, the total number of claims of the individual i is denoted as K and by $X_{i,k}^j$ is denoted the loss incurred from his claim k for the period j. Then, the information we have for his 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 he is in the portfolio will be equal to $\sum_{k=1}^{K} X_{i,k}$. We assume that $X_{i,k}^j$ follows an exponential distribution with parameter y_i^j. The parameter y_i^j denotes the mean or the expected claim severity of a policyholder i in period j. As we have already said, all policyholders do not have the same expected claim severity, their cost for the insurer is different and thus it is fair each policyholder to pay a premium proportional to his mean claim severity. Consider that the expected claim severity is a function of the vector of the h individual's characteristics, denoted as $d_i^j = (d_{i,1}^j,..., d_{i,h}^j)$, which represent different a priori rating variables. Specifically assume that $y_i^j = \exp(d_i^j \gamma^j)$, where γ is the vector of the coefficients. The non-negativity of y_i^j is implied from the exponential function. The probability specification becomes

$$P\left(X_{i,k}^j = x\right) = \frac{1}{\exp(d_i^j \gamma^j)} \cdot e^{-\frac{x}{\exp(d_i^j \gamma^j)}}.$$

In this model we assume that the h individual characteristics provide enough information for determining the expected claim severity. However if one assumes that the a priori rating variables do not contain all the significant information for the expected claim severity then a random variable z_i has to be introduced into the regression component. Thus we can write

$$\begin{aligned} y_i^j &= \exp(d_i^j \gamma^j + z_i) = \\ &= \exp(d_i^j \gamma^j) w_i, \end{aligned}$$

where $w_i = \exp(z_i)$, yielding a random y_i^j. If we assume that w_i follows an inverse gamma$(s, s-1)$ distribution with

$$E(w_i) = 1 \text{ and } Var(w_i) = \frac{1}{s-2}, \quad s > 2,$$

then y_i^j follows inverse gamma$(s,(s-1)\exp(c_i^j \gamma^j))$ and the probability specification for $X_{i,k}^j$ becomes

$$P(X_{i,k}^j = x) = s \cdot \left[(s-1)\exp(d_i^j \gamma^j)\right]^s \cdot (x + (s-1)\exp(d_i^j \gamma^j))^{-s-1}$$

which is a Pareto distribution with parameters s and $(s-1)\exp(c_i^j \gamma^j)$. It can be shown that the above parameterization does not affect the results if there is a constant term in the regression. We choose $E(w_i) = 1$ in order to have $E(z_i) = 0$. We also have

$$E(X_{i,k}^j) = \exp(d_i^j \gamma^j) \text{ and } Var(X_{i,k}^j) = \frac{\left[(s-1)\exp(d_i^j \gamma^j)\right]^2}{s-1} \left(\frac{2}{s-2} - \frac{1}{s-1}\right).$$

The insurer needs to calculate the best estimator of the expected claim severity at period $t + 1$ using the information from past experience for the claim severity over t periods and of known individual characteristics over the $t + 1$ periods. Let us denote this estimator as $\hat{y}_i^{t+1}(X_{i,1},..., X_{i,K}; d_i^1,..., d_i^{t+1})$. Using the Bayes theorem the posterior distribution of the mean claim severity for a policyholder with claim sizes $X_{i,1},..., X_{i,K}$ in t periods and characteristics $d_i^1,..., d_i^{t+1}$ is inverse gamma with the following updated parameters:

$$IG(s + K, (s-1)\exp(d_i^j \gamma^j) + \sum_{k=1}^{K} X_{i,k}).$$

Using the classical quadratic loss function one can find that the optimal estimator of the mean claim severity for the period $t + 1$ given the observation of

$X_{i,1},\dots,X_{i,K}$ and $d_i^1,\dots,d_i^{t+1}$, is the mean of the posterior inverse gamma and thus it is equal to

$$\hat{y}_i^{t+1}(X_{i,1},\dots,X_{i,K};d_i^1,\dots,d_i^{t+1}) =$$

$$= \int_0^{\infty} y_i^{t+1}\left(X_i^{t+1},w_i\right) f\left(y_i^{t+1}\middle| X_{i,1},\dots,X_{i,K};d_i^1,\dots,d_i^{t+1}\right) dy_i^{t+1} =$$

$$= \frac{(s-1)\frac{1}{t}\sum_{j=1}^{t}\exp(d_i^j\gamma^j) + \sum_{k=1}^{K} X_{i,k}}{s+K-1}$$

When $t = 0$, which implies that only a priori rating is used in the first period it is $\hat{y}_i^1 = \exp(d_i^1\gamma^j)$.

3.3. 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. Thus it will be

$$Premium = GBM_F * GBM_S =$$

$$= \frac{1}{t}\sum_{j=1}^{t}\exp(c_i^j\beta^j)\left[\frac{a+\sum_{j=1}^{t}K_i^j}{a+t\exp(c_i^j\beta)}\right]\frac{(s-1)\frac{1}{t}\sum_{j=1}^{t}\exp(d_i^j\gamma^j)+\sum_{k=1}^{K}X_{i,k}}{s+K-1}. \tag{6}$$

3.4. Properties of the Generalized BMS

1. It is fair since it takes into account the number of claims, the significant a priori rating variables for the number of claims, the claim severity and the significant a priori rating variables for the claim severity for each policyholder.
2. It is financially balanced for the insurer. Each year the average premium will be equal to

$$P = \exp(c_i^{t+1}\beta^{t+1})\exp(d_i^{t+1}\gamma^{t+1}) \tag{7}$$

In order to prove the above equation and assuming that claim frequency and the claim severity component are independent it is sufficient to show that

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

and that

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

3. All the properties we mentioned for generalized 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 (7).
4. The more accidents are caused and the more the size of loss that each claim incurred the higher is the premium.
5. The premium always decreases when no accidents are caused.
6. This generalized BMS could lead to a decrease of the phenomenon of bonus hunger.
7. The severity component, which is more crucial than the number of claims for the insurer, is introduced in the design of the generalized BMS.
8. The premiums vary simultaneously with the variables that affect the distribution of the number of claims and the size of loss distribution.

3.5. Estimation

The premiums will be calculated according (6). We have to know the number of the years t that the policyholder is in the portfolio, his total number of accidents in t years and his aggregate claim amount in t years.

For the frequency component of the generalized BMS we have to estimate the parameters of the negative binomial regression model, that is the dispersion parameter a and the vector β. This can be done using the maximum likelihood method. For more on the negative binomial regression the interested reader can see Lawless (1987), Gourieroux, Montfort and Trognon (1984a) and Gourieroux, Montfort and Trognon (1984b).

For the severity component of the generalized BMS we have to estimate s and γ^j. We will achieve this using the quasi-likelihood and according to Renshaw (1994). Renshaw is using the generalized linear models as a modelling tool for the study of the claim process in the presence of covariates. He is giving special attention to the variety of probability distributions that are available and to the parameter estimation and model fitting techniques that can be used for the claim frequency and the claim severity process based on the concepts of quasi-likelihood and extended quasi-likelihood.

Following Renshaw (1994) consider the following scheme. The mean claim severity is denoted by y_i, categorized over a set of units u. The data take the form (u, k_u, x_u) where x_u denotes the claim average in cell u based on n_u claims. Independence of n_u and x_u is assumed. The units $u \equiv (i_1, i_2,...)$ are a cross-classified grid of cells defined for preselected levels of appropriate covariates, often rating factors. Denoting the underlying expected claim severity in cell u by μ_u and assuming the independence of individual claim amounts, the cell means are modelled as the responses of a GLM with $E(x_u) = \mu_u$ and $Var(X_u) = \varphi V(\mu_u)/n_u$. Covariates defined on {u} enter through a linear predictor, linked to the mean μ_u. For those unfamiliar with the generalized linear models we refer to the classical text of McCullagh and Nelder (1989). In McCullagh and Nelder (1983, 1989) a re-analysis of the celebrated car insurance data of Baxter, Coutts and Ross (1979), based on independent gamma distributed claim amounts can be found.

Let us focus now on the Pareto distribution with parameters s and $(s-1)\exp(c_i^j \gamma^j)$ and density

$$P(X_{i,k}^j = x) = s \cdot \left[(s-1)\exp\left(d_i^j \gamma^j\right)\right]^s \cdot \left(x + (s-1)\exp(d_i^j \gamma^j)\right)^{-s-1}$$

We have that

$$E(X_{i,k}^j) = \exp(d_i^j \gamma^j) \text{ and } Var(X_{i,k}^j) = \frac{\left[(s-1)\exp(d_i^j \gamma^j)\right]^2}{s-1} \left(\frac{2}{s-2} - \frac{1}{s-1}\right), \quad s > 2.$$

Introducing the reparameterisation:

$$\mu = \exp(d_i^j \gamma^j) \text{ and } \varphi = \frac{s}{s-2},$$

a 1:1 mapping $(s,(s-1)\exp(c_i^j \gamma^j)) \to (\mu, \varphi)$ with domain $R_{>2} \times R_{>0}$ and image set $R_{>2} \times R_{>1}$ implies that we can construct a GLM based on independent Pareto distributed claim amounts for which the mean responses, X_u, satisfy mean $\mu_u = E(X_u)$, variance function $V(\mu_u) = \mu_u^2$, scale parameter $\varphi > 1$ and weights n_u so that $Var(X_u) = \varphi V(\mu_u)/n_u$. Apart from the mild extra constraint on the scale parameter, these details are identical to those of the GLM based on independent gamma responses and the two different modelling assumptions lead to essentially identical GLMs. They differ only in the parameter estimation method. In the case of gamma response we use maximum likelihood method and in the case of Pareto response we use maximum quasi-likelihood.

4. Application

4.1. Description of the Data

The models discussed are applied in a data set that one Greek insurance company provided us. The data set consists of 46420 policyholders. The mean of the claim frequency is 0.0808 and the variance is 0.10767. The a priori rating variables were age and sex of the driver, BM class and the horsepower of the car. The drivers were divided in three categories according their age. Those aged between 28-45, those between 46-55 and those aged between 18-27 or higher than 55. The drivers were also divided in three categories according the horsepower of their car. Those who had a car with a horsepower between 0-33, between 34-66 and between 67-99. The drivers were also divided in three categories according their BM class. The current Greek BMS has 16 classes, from 5 to 20. The malus zone includes classes from 12 to 20, the bonus zone includes classes 5 to 8 and the neutral zone includes classes from 9 to 11.

We fitted the Negative Binomial distribution on the number of claims and the Pareto distribution on the claim sizes. We will find the premiums determined from the optimal BMS based on the a posteriori frequency component, 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.

4.2. Optimal BMS based on the a posteriori frequency component

We apply the Negative Binomial distribution. The maximum likelihood estimators of the parameters are $\hat{a} = 0.228$ and $\hat{\tau} = 2.825$. We will find first the optimal BMS based only on the frequency component following Lemaire (1995). The BMS will be defined from (1) and is presented in Table 1. This optimal BMS can be considered generous with good drivers and strict with bad drivers. For example, the bonuses given for the first claim free year are 26% of the basic premium. Drivers who have one accident over the first year will have to pay a malus of 298% of the basic premium.

TABLE 1.

OPTIMAL BMS BASED ON THE A POSTERIORI FREQUENCY COMPONENT

Year	*Number of claims*					
t	0	1	2	3	4	5
0	100					
1	74	398	722	1046	1370	1693
2	59	315	572	829	1086	1342
3	48	261	474	687	899	1112
4	41	223	404	586	768	949
5	36	194	353	511	669	828
6	32	172	313	453	594	734
7	29	155	281	407	533	659

4.3. Optimal BMS based on the a posteriori frequency and severity component

Let us see the implementation of an optimal BMS based both on the frequency and the severity component. We fit the Pareto distribution to the claim sizes and we find the maximum likelihood estimates of s and m. It is $\hat{s} = 2.382$ and $\hat{m} = 493927.087$. In order to find the premium that must be paid we have to know the age of the policy, the number of claims he has done in these years and the aggregate claim amount. The steps that must be followed in order to find this optimal BMS are:

1. We find the age of the policy t.
2. We find the total number of claims k that the policyholder has done in t years.

3. We find the aggregate claim amount for the policyholder, $\sum_{k=1}^{K} x_k$
4. We compute the premiums using (3).
5. We go to the table with the specific total claim amount and we find the premium that corresponds to k claims in t years of observation.

The Bonus-Malus System determined in the above way is presented in the following tables. Here we will illustrate only the cases that the aggregate claim amount of a policyholder is equal to 250000 drs, and 1000000 drs. It is obvious that we can use the above formula with any value that the aggregate claim amount can have. We use these values of the aggregate claim amount for brevity. In the following tables we will use the actual values, the premiums are not divided with the premium when t = 0, as it will be interesting to see the variation of the premiums paid for various number of claims and claim sizes in comparison not with the premium paid when t = 0 but with the specific claim sizes. This is the basic advantage of this BMS in comparison with the one that takes under consideration only the frequency component, the differentiation according the severity of the claim. Of course the percentage change in the premium after on or more claims could be also interesting.

Let's see an example in order to understand better how such BMS work. In Table 3 we can see the premiums that must be paid for various number of claims when the age of the policy is up to 7 years. For example a policyholder with one accident of claim size 250000 drs in the first year of observation will pay 100259 drs (see Table 2). If the second year of observation he has an accident with claim size 750000 drs, then, a surcharge will be enforced and he will have to pay 203964 drs, which is the premium for two accidents of aggregate claim amount 1000000 drs in two years of observation (see Table 3). If in the third year he does not have an accident, he will have a reduction in the premium because he had a claim free year and he will pay 168947 drs, which is the premium for two accidents of aggregate claim amount 1000000 drs in three years of observation (see Table 3).

TABLE 2.

OPTIMAL BMS BASED ON THE A POSTERIORI FREQUENCY AND SEVERITY COMPONENT – TOTAL CLAIM SIZE OF 250000.

Year	***Number of claims***					
t	0	1	2	3	4	5
0	28841					
1	21300	100259	128122	143269	152788	159323
2	16886	79479	101567	113575	121121	126302
3	13987	65834	84130	94076	100327	104618
4	11937	56188	71803	80292	85626	89289
5	10412	49007	62627	70031	74683	77878
6	9232	43454	55530	62095	66220	69053
7	8292	39031	49878	55775	59480	62025

TABLE 3.

OPTIMAL BMS BASED ON THE A POSTERIORI FREQUENCY AND SEVERITY COMPONENT – TOTAL CLAIM SIZE OF 1000000.

Year	*Number of claims*					
t	0	1	2	3	4	5
0	28841					
1	21300	201336	257290	287708	306823	319947
2	16886	159607	203964	228077	243230	253634
3	13987	132206	168947	188921	201472	210091
4	11937	112834	144192	161239	171952	179307
5	10412	98414	125765	140633	149976	156392
6	9232	87262	111513	124697	132982	138670
7	8292	78380	100163	112005	119446	124556

It is obvious that this optimal BMS allows the discrimination of the premium with respect to the severity of the claims. Table 4 shows the premiums that must be paid when the policyholder is observed for the first year of his presence in the portfolio, his number of accidents range from 1 to 5 and the aggregate claim amount of his accidents ranges from 250000 to 4000000 dr. A policyholder who had one claim with claim size 250000 will have to pay a premium of 100259 drs, a policyholder who had one claim with claim size 500000 will have to pay a premium of 133951 drs and a policyholder who had one claim with claim size 1000000 will have to pay a premium of 201336 drs.

TABLE 4.

COMPARISON OF PREMIUMS FOR VARIOUS NUMBER OF CLAIMS AND CLAIM SIZES IN THE FIRST YEAR OF OBSERVATION.

	Number of claims				
Claim Size	1	2	3	4	5
250000	100259	181903	263547	345191	426835
500000	133951	243032	352113	461194	570275
1000000	201336	365291	529246	693201	857155
2000000	336106	609808	883511	1157213	1430915
3000000	470876	854326	1237775	1621225	2004675
4000000	605646	1098843	1592040	2085237	2578434

For more on such a system the interesting reader can see Vrontos (1998).

4.4. Generalized optimal BMS with a frequency and a severity component based both on the a priori and the a posteriori classification criteria

Let us calculate now the premiums of the generalized optimal BMS based both on the frequency and the severity component when both the a priori and the a posteriori rating variables are used. 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, GBM_F, and of the generalized BMS based on the severity component, GBM_S.

Implementing the negative binomial regression model we estimate the dispersion parameter a and the vector β of the significant a priori rating variables for the number of claims. We found that many a priori rating variables are significant for the number of claims. These are the BM class, the age and the sex of the driver and the interaction between age and sex. In the multivariate model $\hat{a} = 47.96$ is larger, than in the univariate negative binomial model where we had $\hat{a} = 0.228$. This result indicates that part of the variance is explained by the a priori rating variables in the multivariate model. The estimates of the vector β can be found in the appendix. The parameters of GBM_S, that is the parameter of the Pareto s, and the vector parameter γ of the significant for the claim severity a priori rating variables d_i^j, are found using the quasi-likelihood method. The significant a priori characteristics for the claim severity are the age and the sex of the driver, the BM class, the horsepower of the car, the interaction between age and sex and the interaction between age and class. The premiums are calculated using (6). Below we can see the premiums for different categories of policyholders.

Let us examine two groups of policyholders which have the following common characteristics. They belong in the malus zone, their car's horsepower is between 67 and 99, and their age is between 28 and 45. If the policyholder is a man he will have to pay the following premiums after one or more accidents of total claim amount 500000 in the first year.

TABLE 5.

MEN, AGE 28-45, MALUS-ZONE, HORSEPOWER 67-99.

Year	***Number of claims***					
t	0	1	2	3	4	5
0	100413					
1	39609	279852	357626	399906	426474	444718
2	24670	174304	222745	249079	265627	276990
3	17914	126568	161743	180865	192881	201132
4	14063	99358	126970	141981	151414	157891
5	11574	81777	104503	116858	124622	129953
6	9834	69482	88792	99289	105886	110415
7	8549	60401	77187	86313	92047	95985

If the policyholder is a woman with the above characteristics she will have to pay the following premiums after one or more accidents of total claim amount 500000 in the first year.

TABLE 6.

WOMEN, AGE 28-45, MALUS-ZONE, HORSEPOWER 67-99.

Year	*Number of claims*					
t	0	1	2	3	4	5
0	75096					
1	28632	248946	318131	355742	379376	395605
2	17688	153791	196531	219766	234367	244392
3	12797	111263	142184	158994	169557	176810
4	10024	87160	111383	124551	132826	138508
5	8240	71641	91551	102374	109176	113846
6	6994	60813	77713	86901	92674	96639
7	6076	52828	67510	75491	80506	83950

We notice that men have to pay higher premiums than women. We saw an example of premiums obtained with generalized optimal BMS with a frequency and a severity component based both on the a priori and the a posteriori classification criteria. Other combinations of a priori characteristics could be used and also higher total claim amounts.

It is interesting to compare this BMS with the one obtained when the only the a posteriori frequency and severity component are used. Using this BMS we saw from Table 4 that a policyholder with one accident with claim size of 500000 drs in one year has to pay 133951 drs. Using the generalized optimal BMS with a frequency and a severity component based both on the a priori and the a posteriori classification criteria, a man, age 28-45, who belongs to the malus zone, with a car with horsepower between 67-99 for one accident of claim size 500000 drs in one year will has to pay 279852 drs, while a woman, age 28-45, who belongs to the malus zone, with a car with horsepower between 67-99 for one accident of claim size 500000 drs in one year will has to pay 248946 drs. This system is more fair since it considers all the important a priori and a posteriori information for each policyholder both for the frequency and the severity component in order to estimate his risk to have an accident and thus it permits the differentiation of the premiums for various number of claims and for various claim amounts 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.

5. Conclusions

We developed in this paper the design of an optimal BMS based both on the a posteriori frequency and the a posteriori severity component. We did this by fitting the Negative Binomial distribution in the claim frequency and the Pareto distribution on the claim severity, extending the – classical in the BMS literature – model of Lemaire (1995) which used the Negative Binomial distribution. The optimal BMS obtained has all the attractive properties of the optimal BMS designed by Lemaire, furthermore it allows the differentiation of the premiums according to the claim severity and in this way it is more fair for the policyholders and it is obtained in a very natural context according to our opinion.

Moreover, we developed the design of a generalized optimal BMS with a frequency and a severity component based both on the a priori and the a posteriori classification criteria extending the model developed by Dionne and Vanasse (1989, 1992) which was based only on the frequency component. The BMS obtained has all the attractive properties of the one obtained by Dionne and Vanasse (1989, 1992) and furthermore it allows the differentiation of the premiums utilizing the severity component in a very natural context. This generalized BMS takes into consideration simultaneously the important individual's characteristics for the claim frequency, the important individual's characteristics for the claim severity, the claim frequency and the claim severity of each accident for each policyholder.

An interesting topic for further research could be the extension of the two above BMS for different claim frequency and claim severity distributions.

References

Baxter, L.A., Coutts S.M. and Ross G.A.F. (1979) *Applications of Linear Models in Motor Insurance*. 21st International Congress of Actuaries.

Bichsel, F. (1964) Erfahrung-Tarifieung in der Motorfahrzeug-haftplichtversiherung, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 119-129.

Buhlmann, H. (1964) Optimale Pramienstufensysteme, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 193-213.

Coene, G. and Doray, L.G. (1996) A Financially Balanced Bonus-Malus System. *Astin Bulletin* **26**, 107-115.

Dionne, G. and Vanasse, C. (1989) A generalization of actuarial automobile insurance rating models: the negative binomial distribution with a regression component, *Astin Bulletin* **19**, 199-212.

Dionne, G. and Vanasse, C. (1992) Automobile insurance ratemaking in the presence of asymmetrical information, *Journal of Applied Econometrics* **7**, 149-165.

Ferreira, J. (1974) The long term effect of merit-rating plans on individual motorists, *Operations Research*, **22**, 954-978.

Gourieroux C., Montfort A. and Trognon A. (1984a) Pseudo maximum likelihood methods: theory, *Econometrica*, **52**, 681-700.

Gourieroux C., Montfort A. and Trognon A. (1984b) Pseudo maximum likelihood methods: application to Poisson models, *Econometrica*, **52**, 701-720.

Hausmann J.A., Hall B.H. and Griliches Z. (1984) Econometric models for count data with an application to the patents-R&D relationship, *Econometrica*, **46**, 1251-1271.

HERZOG, T. (1996) *Introduction to Credibility Theory*. Actex Publications, Winstead.
HOGG, R.V. and KLUGMAN, S.A. (1984) *Loss Distributions*, John Wiley & Sons, New York.
MCCULLAGH, P. and NELDER, J.A. (1983, 1989) Generalized Linear Models. Chapman & Hall.
LAWLESS, J.F. (1987) Negative Binomial Distribution and Mixed Poisson Regression, *Canadian Journal of Statistics*, **15**, **3**, 209-225.
LEMAIRE, J. (1995) *Bonus-Malus Systems in Automobile Insurance*, Kluwer Academic Publishers, Massachusetts.
PICARD, P. (1976) Generalisation de l'etude sur la survenance des sinistres en assurance automobile, *Bulletin Trimestriel de l'Institute des Actuaires Francais*, 204-267.
PICECH, L. (1994) The Merit-Rating Factor in a Multiplicating Rate-Making model. *Astin Colloquium*, Cannes.
PINQUET (1997) Allowance for Costs of Claims in Bonus-Malus Systems, *Astin Bulletin*, **27**, 33-57.
PINQUET J. (1998) Designing Optimal Bonus-Malus Systems From Different Types of Claims. *Astin Bulletin*, **28**, 205-220.
RENSHAW A.E. (1994) Modelling The Claims Process in the Presence of Covariates. *Astin Bulletin*, **24**, 265-285.
SIGALOTTI (1994) Equilibrium Premiums in a Bonus-Malus System, *Astin Colloquium*, Cannes.
TAYLOR G. (1997) Setting A Bonus-Malus Scale in the Presence of Other Rating Factors. *Astin Bulletin*, **27**, 319-327.
TREMBLAY, L. (1992) Using the Poisson Inverse Gaussian in Bonus-Malus Systems, *Astin Bulletin* **22**, 97-106.
VRONTOS S. (1998) *Design of an Optimal Bonus-Malus System in Automobile Insurance*. Msc. Thesis, Department of Statistics, Athens University of Economics and Business, ISBN: 960-7929-21-7.
WALHIN J.F. and PARIS J. (1999) Using Mixed Poisson distributions in connection with Bonus-Malus Systems. *Astin Bulletin* **29**, 81-99.

NICHOLAS E. FRANGOS AND SPYRIDON D. VRONTOS
Department of Statistics
Athens University of Economics and Business
Patission 76, 10434, Athens, Greece
E-mail: nef@aueb.gr and svrontos@aueb.gr

INSURANCE PREMIUM CALCULATIONS WITH ANTICIPATED UTILITY THEORY

By

Cuncun Luan

Department of Mathematics
Nanjing University, Nanjing, China

Abstract

This paper examines an insurance or risk premium calculation method called the mean-value-distortion pricing principle in the general framework of anticipated utility theory. Then the relationship between comonotonicity and independence is explored. Two types of risk aversion and optimal reinsurance contracts are also discussed in the context of the pricing principle.

Keywords

Mean-value-distortion pricing principle, anticipated utility theory, comonotonicity, risk aversion, and optimal reinsurance.

1. Introduction

The calculation of insurance or risk premiums has been an essential and active topic in actuarial literature, which has attracted the attention of actuaries such as Bühlmann (1970), Goovaerts et al. (1984) and Hürlimann (1997, 1998). Recently, modern theory of risk and economic choice under uncertainty has played an important role in studying insurance premium calculations (Wang et al., 1997, Wang and Young, 1998, Young, 1998). Hürlimann (1998) makes a brief, yet comprehensive summary about the development of insurance premium calculations. He emphasizes desirable and reasonable properties that insurance premiums should satisfy. In fact, most modern pricing principles, other than the distortion pricing principle, are presented in an expected utility framework, while Wang et al. (1997) applies Yaari' dual theory. However, both expected utility theory and Yaari' dual theory are special cases of anticipated utility theory (Puppe, 1991).

In this paper, the mean-value-distortion pricing principle is presented under anticipated utility theory as an approach to insurance premium calculations. This kind of premium calculation can be found in Denuit et al. (1999), which refers to Chateauneuf et al. (1997). An outline of the paper is as follows. In section 2, main properties of the mean-value-distortion pricing principle are investigated. It is shown that these properties are consistent with those of the mean value principle. Section 3 shows the relationship between independence and comonotonicity. Here risk aversion and optimal reinsurance are also discussed.

2. Properties of the mean-value-distortion pricing principle

2.1. The Mean-Value-Distortion Pricing Principle

Quiggin (1982) first discussed anticipated utility theory. Subsequently, Segal (1989) proposed an axiomatization of this theory, where the ordinal independence axiom substituted the independence axiom of expected utility theory. Analogous to Segal (1989), define "risk" as a non-negative random variable $X \in \Omega$ with distribution function $F_X(x)$ and survival function $S_X(x)$, where $x \geq 0$ and $\Omega = \{X: X \geq 0, 0 \leq EX \leq \infty\}$. The insurance premium calculation is a non-negative real function $\pi: \Omega \to R$. The premium of risk X is denoted by $\pi(X)$.

Risks are restricted to bounded random variables and $\Pi = [0, M]$ is the domain of risks. Further, let $\preccurlyeq$ be a binary preference relation.

Axiom 1 (Weak Order): The relation $\preccurlyeq$ is weak order.

Axiom 2 (Continuity): For every risk X, the sets $\{F_Y(x) : X \preccurlyeq Y\}$ and $\{F_Y(x) : Y \preccurlyeq X\}$ are closed in the topology of weak convergence.

Axiom 3 (Monotonicity): For all risks X and Y, if $F_X(x) \geq F_Y(x)$, $x \geq 0$, then $X \prec Y$.

Axiom 4 (Ordinal Independence): For all risks X, X', Y and Y', if $F_X(x) = F_{X'}(x)$, $F_Y(x) = F_{Y'}(x)$ on $[0, c)$ (respectively on $[c, M]$) and $F_X(x) = F_Y(x)$, $F_{X'}(x) = F_{Y'}(x)$ on $[c, M]$ (respectively $[0, c)$), then $X \preccurlyeq Y \Leftrightarrow X' \preccurlyeq Y'$.

Preference relation $\preccurlyeq$ satisfies axioms 1, 2, 3 and 4 if and only if there exists a continuous measure ω on $\Pi \times [0,1]$ with $\omega(A) > 0$ for every non-empty open set $A \in \Pi \times [0,1]$ such that

$$X \preccurlyeq Y \Leftrightarrow \omega(e_X) \leq \omega(e_Y)$$

where epigraph e_X is the closure of set $\{(x, p) \in \Pi \times [0,1]: p \geq F_X(x)\}$. The generalized utility function is defined by

$$\bar{v}(x, p) = \omega([0, x] \times [1\text{-}p, 1]) \text{ for all } (x, p) \in \Pi \times [0,1].$$

If the corresponding relative utility index $\frac{\bar{v}(x, \delta p)}{\bar{v}(x, p)}$ is independent of x for all $\delta \in [0,1]$, then the preference relation $\preccurlyeq$ can be expressed by a real-valued functional:

$$V(X) = -\int_{\Pi} v(t) dg(S_X(t))$$

where $v(x)$ and $g(x)$ are non-decreasing functions. Accordingly, the mean-value-distortion pricing principle π satisfies the following equation:

$$v(\pi(X)) = -\int_{\Pi} v(t)dg(S_X(t)). \tag{1}$$

If $\lim_{d\to\infty} \pi(\min(X,d)) = \pi(X)$, then the extension of equation (1) to Ω is given by

$$v(\pi(X)) = -\int_0^\infty v(t)dg(S_X(t)). \tag{2}$$

Integrating equation (2) by parts and assuming $\lim_{t\to\infty} v(t)g(S_X(t)) = 0$ gives

$$v(\pi(X)) = \int_0^\infty g(S_X(t))dv(t). \tag{2'}$$

Obviously, if $g(x) = x$, equation (2) results in the mean value principle and if $v(x) = x$, equation (2') results in the distortion pricing principle. Since

$$v(\pi(X)) = -\int_0^\infty v(t)dg(S_{\pi(X)}(t)) = v(\pi(\pi(X))).$$

Hence equation (2) displays the certainty equivalent principle in which $\pi(X)$ is the sure payment leading to indifference. In the next part, the properties of mean-value-distortion pricing principle are developed.

2.2. Properties

Suppose that $v(x)$ is an increasing convex function, i.e., $v'(x) > 0$, $v''(x) \geq 0$, and $g(x)$ is an increasing concave and distortion function on [0,1] such that $g(0) = 0$, $g(1) = 1$ and $g(x) \geq x$.

Theorem 2.1 (*Non-Negative Loading*): $\pi(X) \geq EX$ for all $X \in \Omega$.
Proof: Since $g(x) \geq x$ and $v(x)$ is convex,

$$v(\pi(X)) = \int_0^\infty g(S_X(t))dv(t) \geq \int_0^\infty S_X(t)dv(t) = E(v(X)) \geq v(EX). \tag{3}$$

Thus, $\pi(X) \geq EX$. □

Theorem 2.2 (*Non-Excessive Loading*): $\pi(X) \leq sup(X)$ for all $X \in \Omega$.
This result is obvious.

Theorem 2.3 (*Scale Invariant*): $\pi(kX) = k\pi(X)$ for all $k > 0$ if and only if $v(x) = a + bx^\theta$, where $a \in R$, $b > 0$ *and* $\theta > 0$.

Theorem 2.4 (*Translation Invariant*): $\pi(X + c) = \pi(X) + c$ for all $c \in R$ if and only if $v(x) = x$ or $v(x) = e^{rx}$ where $r > 0$.

Before proving the prior two theorems, several lemmas from Goovaerts et al. (1984) are generalized.

Lemma 2.1: Suppose $v(x)$ and $\tilde{v}(x)$ are continuous and increasing functions. For bounded risks, the sufficient and necessary condition such that $v(x)$ and $\tilde{v}(x)$ have the same solutions with respect to equation (1) is $\tilde{v}(x) = \alpha + \beta v(x)$, for all $x \in \Pi$ where $\alpha, \beta \in R$.

Proof: Assume $\tilde{\pi}(X)$ is a solution of equation (1) corresponding to $\tilde{v}(x)$. If for all $x \in \Pi$, $\tilde{v}(x) = \alpha + \beta v(x)$, and $\alpha, \beta \in R,$ then

$$\tilde{v}(\tilde{\pi}(X)) = \alpha + \beta v(\pi(X)) \text{ and } \tilde{v}(\tilde{\pi}(X)) = \alpha + \beta v(\tilde{\pi}(X)).$$

So $v(x)$ and $\tilde{v}(x)$ have the same solutions. Conversely, let X be a two-point random variable, i.e., $X = M$ with probability q, $X = 0$ with probability 1-q, where $0 < q < 1$. According to equation (2),

$$\tilde{v}(\pi(X)) = g(q)\tilde{v}(M) + (1-g(q))\tilde{v}(0) \text{ and } v(\pi(X)) = g(q)v(M) + (1-g(q))v(0)$$

Since $1-g(q) \neq 0$, comparing the two equations above gives

$$\frac{\tilde{v}(\pi(X)) - \tilde{v}(M)}{v(\pi(X)) - v(M)} = \frac{\tilde{v}(M) - \tilde{v}(0)}{v(M) - v(0)}.$$

This implies $\tilde{v}(x) = \alpha + \beta v(x)$ for $x \in [0, M]$, where $\alpha = \tilde{v}(M) - \frac{\tilde{v}(M) - \tilde{v}(0)}{v(M) - v(0)} v(M)$ and $\beta = \frac{\tilde{v}(M) - \tilde{v}(0)}{v(M) - v(0)}$. □

Lemma 2.2: If $\lim_{d \to \infty} \pi(\min(X, d)) = \pi(X)$ for all $d \geq 0$, lemma 2.1 also holds for risk $X \in \Omega$.

Proof of theorem 2.3: The "if" part is easy to verify. It suffices to prove the "only if" part. Since $\pi(kX) = k\pi(X)$ for all $k > 0$,

$$v(k\pi(X)) = v(\pi(kX)) = -\int_0^\infty v(kt)dg(S_X(t)).$$

Let $\tilde{v}(x) = v(kx)$, it follows that

$$\tilde{v}(\pi(X)) = v(k\pi(X)) = -\int_0^\infty v(kt)dg(S_X(t)) = -\int_0^\infty \tilde{v}(t)dg(S_X(t)).$$

According to lemma 2.2, $\tilde{v}(x) = \alpha(k) + \beta(k)v(x)$ where $\alpha(k), \beta(k) \in R$ are dependent on k. Let $x = 0$, $\tilde{v}(0) = \alpha(k) + \beta(k)v(0)$ and

$$v(kx) - v(0) = \beta(k)[v(x) - v(0)].$$

Differentiating the above equation with respect to variable k,

$$xv'(kx) = \beta'(k)v(x) \text{ and } x^2v''(kx) = \beta''(k)v(x).$$

Finally, if $k = 1$, then $\frac{v''(x)}{v'(x)} = \frac{\beta''(1)}{\beta'(1)}x$. This implies that $v(x)$ can be represented as $v(x) = a + bx^\theta$, where $a \in R$, $b > 0$ and $\theta > 0$. □

Proof of theorem 2.4: The "if" part is obvious by calculation. Conversely, assume $\pi(X+c) = \pi(X)+c$ for all $c \in R$. If $\tilde{v}(x) = v(x+c)$, then

$$\tilde{v}(\pi(X)) = v(\pi(X)+c) = v(\pi(X+c)) = -\int_0^\infty v(t+c)dg(S_X(t)) = -\int_0^\infty \tilde{v}(t)dg(S_X(t)).$$

According to lemma 2.2, $\tilde{v}(x) = \alpha(c)+\beta(c)v(x)$ where $\alpha(c), \beta(c) \in R$ are dependent on c. Let $x = 0$, $\tilde{v}(0) = \alpha(c)+\beta(c)v(0)$ and

$$v(x+c) - v(0) = \beta(c)[v(x) - v(0)].$$

Differentiating the above equation with respect to variable x,

$$v'(x+c) = \beta(c)v'(x) \text{ and } v''(x+c) = \beta(c)v''(x).$$

Finally if $x = 0$, then $\frac{v''(c)}{v'(c)} = \frac{v''(0)}{v'(0)}$ and $v''(0) = 0$. It implies $v(x) = x$ otherwise, $v(x) = e^{rx}$ where $r > 0$. □

Theorem 2.5 (*Independent Additive*): If risks X and Y are independent, $\pi(X+Y) = \pi(X)+\pi(Y)$ if and only if $v(x) = x$ or $v(x) = e^{rx}$ where $r > 0$ and $g(x) = x$.

Proof: The "if" part has been proved by Goovaerts et al. (1984). To prove the "only if" part, first note that the independent additive property implies that $\pi(X)$ satisfies translation invariance. Hence $v(x) = x$ or $v(x) = e^{rx}$ where $r > 0$. If $v(x) = x$, then $\pi(X) = \int_0^\infty g(S_X(t))dt$. Let risk $X \sim B(1, q)$, risk $Y \sim B(1, p)$ where $0 \le p, q \le 1$ and let risks X and Y be independent. Thus, $\pi(X) = g(q)$, $\pi(Y) = g(p)$, $\pi(X+Y) = g(p+q-pq)+g(pq)$ and

$$g(p+q-pq)+g(pq) = g(q)+g(p). \tag{4}$$

Differentiating equation (4) by argument p and then q,

$$g''(p+q-pq)(1-p)(1-q) - g'(p+q-pq) + g''(pq)pq + g'(pq) = 0.$$

If $q = 0$, then $g''(p)(1-p) - g'(p) + g'(0) = 0$ and $g'(0) = g'(1)$. (5)

If $q = 1$, then $g''(p)p - g'(1) + g'(p) = 0$. (6)

Comparing (5) and (6), $g''(p) = 0$ for all $p \in [0,1]$ which implies $g(x) = x$. Similarly, $g(x) = x$ if $v(x) = e^{rx}$ where $r > 0$. □

Theorem 2.6 (*Comonotonic Additive*): If risks X and Y are comonotonic, $\pi(X+Y) = \pi(X)+\pi(Y)$ if and only if $v(x) = x$.

Proof: The comonotonic additive property implies $\pi(X)$ preserves the scale and translation invariant properties. Therefore, it follows that $v(x) = x$ by theorem 2.3 and theorem 2.4. □

Lemma 2.3 (Wang, 1998): For two comonotonic risks X and Y, $Cov(X, Y) \geq 0$.

Let $E_g(X) = \int_0^\infty t d[1 - g(S_X(t))]$ and $Cov_g(X, Y) = E_g(XY) - E_g(X)E_g(Y)$.

Lemma 2.4: For two comonotonic risks X and Y, $Cov_g(X, Y) \geq 0$.
The proof of this lemma is omitted since it roughly resembles that of lemma 2.3.

Theorem 2.7 (*Sub-Additive*): For all risks X and Y, $\pi(X+Y) \leq \pi(X) + \pi(Y)$ if and only if $v(x) = x$.

Proof: The "if" part has been proved by Hürlimann (1998). To prove the "only if" part, first note that the sub-additive property implies $\pi(X+c) \leq \pi(X) + \pi(c)$ for all X and c. In addition, $\pi(c) = c$ by equation (2) and $\pi(X) = \pi(X+c-c) \leq \pi(X+c) - \pi(c)$. Hence $\pi(X)$ is translation invariant. By theorem 2.4, $v(x) = x$ or $v(x) = e^{rx}$ where $r > 0$. If $v(x) = e^{rx}$, then $\pi(X) = \frac{1}{r} \log\left[E_g(e^{rX})\right]$ where $r > 0$. Assuming risks X and Y are comonotonic, so are e^{rX} and e^{rY}. According to lemma 2.4 and theorem 2.6,

$$E_g(e^{rX} e^{rY}) > E_g(e^{rX})E_g(e^{rY}).$$

That is, $\pi(X+Y) > \pi(X) + \pi(Y)$ for comonotonic risks. □

Theorem 2.8 (*Stop-Loss Order Preserving*): If $X \prec_{sl} Y$, then $\pi(X) \leq \pi(Y)$.
The proof of this theorem refers to the third part of Hürlimann (1998). However, two points should be noted. One is that if $u = S_X(t)$ and $t = F_X^{-1}(1-u)$, equation (2) can be rewritten by

$$v(\pi(X)) = -\int_0^\infty v(t) dg(S_X(t)) = -\int_0^1 v(F_X^{-1}(1-u)) dg(u) = \int_0^1 v(F_X^{-1}(u)) d\gamma(u), \qquad (7)$$

where $\gamma(u) = 1\text{-}g(1\text{-}u)$. The second point is since $v(x)$ is an increasing convex function, then

$$E(v(X)) \leq E(v(Y)). \qquad (8)$$

Equation (3.5) from Hürlimann (1998), equations (7) and (8) from above all combine to show that the mean-value-distortion pricing principle regarding $\pi(X)$ preserves stop-loss order.

It is shown that essential properties of the mean-value-distortion pricing principle are consistent with corresponding properties of the mean value principle. These properties are more closely related to $v(x)$ than $g(x)$ because under anticipated utility theory, the effect of loss severity and loss probability is multiplicatively separable and

$$v(\pi(X)) = -\int_0^\infty v(t) dg(S_X(t)) = \int_0^\infty v(t) d[1 - g(S_X(t))]. \qquad (9)$$

The right-hand side of equation (9) can be viewed as the expected value of $v(X)$ with respect to $1-g(S_X(x))$ instead of $F_X(x)$. Obviously $P_g = \{1-g(S_X(x))\}$ is a probability space denoted as the distort-probability space. Here $\pi(X)$ may be regarded as the mean value premium of risk X on P_g. In this light, the properties of the mean-value-distortion pricing principle should be different little from those of the mean value principle. It is believed that the distort-probability space P_g is non-additive for independent risks, unless $g(x) = x$, but additive for comonotonic risks. Therefore, different additive properties among risks should be defined in different probability spaces when describing practical insurance operations.

3. Some related conclusions and comments

3.1. The Relationship Between Independence and Comonotonicity

In expected utility theory, independence is an important concept. In Yaari' dual theory, comonotonicity is stressed because of theoretical work and practical meanings. According to theorems 2.5 and 2.6, the mean-value-distortion pricing principle of $\pi(X)$ satisfies the independent additive and comonotonic additive properties if and only if $v(x) = x$ and $g(x) = x$. The following theorem (theorem 3.1) presents an alternative interpretation of the aforementioned result. Here the description of comonotonicity in Denneberg (1994) is applied. For further discussion regarding comonotonicity, one should refer to Schmeidler (1986) and Yaari (1987).

Lemma 3.1 (Denneberg, 1994**):** Risks X and Y are said to be comonotonic if there exist a risk Z and increasing real-valued functions $f_1(x)$, $f_2(x)$ such that

$$X = f_1(Z) \text{ and } Y = f_2(Z).$$

Theorem 3.1: Risks X and Y are both independent and comonotonic if and only if one of them is a degenerate random variable.

Proof: Without loss of generality, let risk X be a degenerate random variable. Obviously,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) = \min\{F_X(x), F_Y(y)\}.$$

Thus risks X and Y are independent and comonotonic. Conversely, assume risks X and Y are not degenerate random variables. By lemma 3.1, there exists a non-degenerate random variable Z and increasing real-valued functions $f_1(x)$ and $f_2(x)$ such that $X = f_1(Z)$ and $Y = f_2(Z)$. Hence, $F_X(x) = F_Z(f_1^{-1}(x))$ and $F_Y(y) = F_Z(f_2^{-1}(y))$. Since risks X and Y are independent and comonotonic,

$$\begin{aligned} F_X(x)F_Y(y) &= F_Z(f_1^{-1}(x))F_Z(f_2^{-1}(y)) = \min\left\{F_Z(f_1^{-1}(x)), F_Z(f_2^{-1}(y))\right\} \\ &= F_Z\left(\min\left\{f_1^{-1}(x), f_2^{-1}(y)\right\}\right). \end{aligned} \tag{10}$$

Let $t_1 = f_1^{-1}(x)$ and $t_2 = f_2^{-1}(y)$, it follows that

$$F_Z(t_1)F_Z(t_2) = F_Z(\min\{t_1, t_2\}). \tag{11}$$

Finally assume $t_1 \geq t_2$ and $t_2 \to a$, where a is a lower bound of risk Z, $0 \leq a \leq \infty$. It follows that

$$F_Z(t_1) = 1,\ t_1 \geq a \text{ and } F_Z(t_1) = 0,\ t_1 < a.$$

This implies risk Z is a degenerate random variable, which contradicts the assumption. □

The above theorem illustrates that if non-degenerate risks are comonotonic, they must not be independent and vise versa. The next theorem provides a sufficient condition for determining whether risks are comonotonic or independent. Further, if risks are comonotonic, their sum may be easily obtained.

Theorem 3.2: If risks X and Y are comonotonic, their sum may be simplified as the addition of real-value functions, i.e.,

$$X + Y = f_1(Z) + f_2(Z) = (f_1 + f_2)(Z).$$

Proof: Let risks X and η have identical distributions, i.e. $F_X(x) = F_\eta(x)$ for all x. Thus,

$$F_{X,\eta}(x, y) = \min\{F_X(x),\ F_\eta(y)\} = F_X(\min\{x, y\}).$$

If $x \leq y$ and $F_{X,\eta}(x, y) = F_X(x)$, then η is constant and independent of risk X. It follows that

$$P(X + f(\eta) \leq z) = \int_{x+f(x)\leq z} dF_{X,Y}(x, y) = \int_{x+f(x)\leq z} dF_X(x) = F_X\left[(1+f)^{-1} z\right]. \tag{12}$$

Analogously, if $x > y$, equation (12) also exists. Thus, if $\eta = X$, then $f = f_2 \circ f_1^{-1}$. □

To examine a collective risk model, let

$$X(t) = \sum_{i=1}^{N(t)} X_i \text{ and } X(0) = 0,$$

where $\{X_i\}_1^\infty \in \Omega$ are independent claim sizes. $N(t)$ is the number of claims in the interval $[0,t]$ with $N(0) = 0$ and $t > 0$ independent of $\{X_i\}_1^\infty$. Risks X_1, X_2, ... may generate comonotonic risks $F_{X_1}^{-1}(\xi), F_{X_2}^{-1}(\xi), \ldots$, which have the same marginal distribution functions as risks $X_1, X_2, \ldots$, where $\xi \sim U(0,1)$. Risks X_i and $F_{X_i}^{-1}(\xi)$ belong to the same individual risk group and $\sum_{i=1}^{N(t)} X_i \prec_{sl} \sum_{i=1}^{N(t)} F_{X_i}^{-1}(\xi)$, then $\pi(\sum_{i=1}^{N(t)} X_i) \leq \pi(\sum_{i=1}^{N(t)} F_{X_i}^{-1}(\xi))$. Hence, the portfolio consisting of comonotonic risks determines an upper bound of insurance premiums that may be viewed as a market price. Insurance companies should not price risks above this market price.

3.2. Risk Aversion

Wang (1996), Wang and Young (1998) distinguish between two types of risk aversion. One type is based on an individual's attitude towards wealth under expected utility theory while the other is based on varying probabilities under dual theory. The authors believe that insurance entities reflect different levels of risk aversion based on their sizes. In fact, there is one type of risk aversion under both expected utility theory and dual theory. The difference is presented in their actual expressions. Puppe (1991, p. 67) argues this point:

> *"Two concepts of risk aversion will be considered here. The first concept defines an individual to be risk averse if the sure gain E(F) of the expectation of a distribution F is always preferred to the distribution itself. An alternative definition of risk aversion, suggested by Rothschild and Stiglitz (1970), requires a risk averse individual to prefer a distribution F to any mean preserving spread of F."*

The two definitions of risk aversion are equivalent only under expected utility theory. In regards to insurance pricing theory, an insurer's pricing principle reflects its attitude towards risks. Insurers who are risk averse expect their pricing principles to preserve stop-loss order, which is consistent with the second definition of risk aversion particularly under non-expected utility theory. To avoid any confusion, the second definition is preferred. It is also known that risk aversion is equivalent to the convexity of $v(x)$ under expected utility theory, to the concavity of $g(x)$ under dual theory, and to both of them under anticipated utility theory.

Dual theory parallels expected utility theory from the standpoint of utilizing probabilities versus wealth. Even so, risk aversion based on expected utility theory and risk aversion based on dual theory cannot be compared. This result can be seen by the characterization theorem of comparative risk aversion discussed by Puppe (1991, p. 71). Hence considering only the size of an insurer is insufficient in determining which pricing principle an insurer should utilize. The following theorems clarify the aforementioned risk aversion comparisons.

Lemma 3.2: Let $\bar{V}$ and $\bar{V}^*$ be rank-dependent utility functionals with corresponding generalized utility functions $\bar{v}$ and $\bar{v}^*$, respectively. Assume $\bar{v}_{12}$ and $\bar{v}^*_{12}$ exist everywhere and are differentiable with respect to both arguments. Then, $\bar{V}$ is more risk averse than $\bar{V}^*$ if and only if for all $(x,p) \in \Pi \times [0,1]$ the following two relations hold.

$$\frac{\bar{v}_{121}(x,p)}{\bar{v}_{12}(x,p)} \leq \frac{\bar{v}^*_{121}(x,p)}{\bar{v}^*_{12}(x,p)} \text{ and } \frac{\bar{v}_{122}(x,p)}{\bar{v}_{12}(x,p)} \geq \frac{\bar{v}^*_{122}(x,p)}{\bar{v}^*_{12}(x,p)}.$$

(Note the prior assumptions regarding $v(x)$ and $g(x)$ are implied in the following theorems.)

Theorem 3.3: An insurer is more risk averse under expected utility theory than under dual theory if and only if $\frac{v''(x)}{v'(x)} \leq 0$ and $\frac{g''(p)}{g'(p)} \leq 0$ for all $x \in \Pi$ and $0 \leq p \leq 1$.

Theorem 3.4: An insurer is more risk averse under expected utility theory and dual theory than under anticipated utility theory.

Within the same theoretical system, it is true that the degree of risk aversion is closely related to the size of an insurer, i.e., its wealth. This is the concept of decreasing risk aversion. In general, decreasing risk aversion implies *"an individual with utility function $u(x)$ is more risk averse than another one with utility function $u(w+x)$, $w>0$, under the standard of maximizing expected utility functions"*. The Arrow-Pratt measure of risk aversion, $r(x)$, is a decreasing function of x applicable to expected utility theory. However, this measure does not make any sense in dual theory because here the Arrow-Pratt measure is zero. According to lemma 3.2, under anticipated utility theory *"the characterization of decreasing risk averse is exactly the same as in the expected utility model."*

The above discussion of risk aversion stems from an insurer's point of view. However, from an insured's perspective, results will be perfectly opposite. Arguably, insurance is the outcome of high-speed development of an economy. The result is the existence of a luxury commodity, insurance, which allows an individual to exchange uncertain outcomes for a certain one after having certain wealth accumulation. In addition, the more wealth an individual has, the more care they are likely to place in the insurance market.

3.3. Optimal Reinsurance

From an insurance company's perspective, the optimization criterion of a reinsurance contract is to minimize the insurance premium of retained risks. A reinsurance contract $I^*(X) \in I$ is said to be an optimal reinsurance contract with respect to the pricing principle π if $\pi[X-I^*(X)] < \pi[X-I(X)]$ for all $I(X) \in I$, where $I = \{I(x) : I(0) = 0, 0 \le I' \le 1\}$ is a set of reinsurance contracts. The most useful two subsets of I are $I_{\pi,P}$, I_μ, where $I_{\pi,P}$, $= \{I(x) : I(0) = 0, 0 \le I' \le 1, \pi[I(X)] = P\}$, $I_\mu = \{I(x) : I(0) = 0, 0 \le I' \le 1, E[I(X)] = \mu\}$, and P, μ are fixed. Goovaerts et al. (1990) gives an informative exposition regarding optimal reinsurance in the case of I_μ. Wang (1998) and Young (1999) study this problem with respect to the distortion pricing principle.

Lemma 3.2 (Goovaerts et al., 1990): For any optimization criterion preserving stop-loss order, the optimal reinsurance contract over set I_μ is of the form $I^*(X) = (X-d)_+$ and is called the stop-loss contract.

Theorem 3.5: According to the mean-value-distortion pricing principle, the stop-loss contract is the optimal reinsurance contract for I_μ and $I_{g,\mu}$, where $I_{g,\mu} = \{I : I(0) = 0, 0 \le I' \le 1, E_g[I(X)] = \mu\}$.

Proof: Applying theorem 2.8 and lemma 3.2, it is easy to prove the result for I_μ. Further, according to equation (9), $v(\pi(X)) = E_g[v(X)]$ and since $v'' \ge 0$,

$$v(t) - v(z) \ge v'(z)(t-z) \text{ for all } t,\, z \in R.$$

Therefore, $v[x-I(x)]-v[x-I^*(x)] \geq v'[x-I^*(x)][I^*(x)-I(x)]$. If less $I^*(x)-I(x) > 0$, then $x-I^*(x) = d$ and

$$v[x-I(x)]-v[x-I^*(x)] \geq v'(d)[I^*(x)-I(x)]. \tag{13}$$

Otherwise, $x-I^*(x) \leq d$, $v'[x-I^*(x)] \leq v'(d)$ and inequality (13) also exists. Substituting X for x in inequality (13) and integrating both sides with respect to $1-g(S_X(x))$ yields

$$v[\pi(X-I(X))]-v[\pi(X-I^*(X))] \geq v'(d)E_g[I^*(X)-I(X)] = 0$$

This implies $I^*(X) = (X-d)_+$ is the optimal reinsurance contract. □

Corollary 3.1: If $\pi(X) = \pi_g(X) = \int_0^\infty g(S_X(t))dt$, then $I_{\pi,P} = I_{g,P}$. In this case the stop-loss contract is the optimal reinsurance contract for I_μ and $I_{\pi,P}$ according to the distortion pricing principle.

For the mean-value-distortion pricing principle, the problem of an extreme value with respect to $\pi[X-I(X)]$ is identical to $v[\pi(X-I(X))]$. In this case a larger set $I^{\tilde{v}}_{g,P} = \{I(x){:}I(0) = 0, 0 \leq I' \leq 1, E_g[\tilde{v}(I(X))] = P\}$ is considered, where $\tilde{v}(x)$ make the integral exist. Obviously I_μ, $I_{\pi,P}$ and $I_{g,\mu}$ are all subsets of $I^{\tilde{v}}_{g,P}$. Moreover, since $v[\pi(x-I(x))]$ is Gâteaux differentiable with respect to $I(x)$, the method of resolving a constrained extreme value of a functional to find an optimal reinsurance contract is applied.

Theorem 3.6: In the set $I^{\tilde{v}}_{g,P}$, an optimal reinsurance contract $I^*(x)$ is determined by the equation $v'[x-I(x)] = \lambda\tilde{v}'[I(x)]$, where λ satisfies the constraint $E_g[\tilde{v}(I^*(X))] = P$.

Proof: Let $\varphi(I) = E_g[v(X-I(X))]$ and $\Phi(I) = E_g[\tilde{v}(I(X))]$. The aim is to minimize $\varphi(I)$ under the constraint $\Phi(I) = P$. Let $f(I, \lambda) = \varphi(I)-\lambda\Phi(I)$. For all real t and functional $h(x)$, optimal reinsurance $I^*(x)$ satisfies $\lim_{t\to 0}\frac{f(I+th,\lambda)-f(I,\lambda)}{t} = 0$. That is,

$$\lim_{t\to 0}\frac{\varphi(I+th)-\varphi(I)-\lambda[\Phi(I+th)-\Phi(I)]}{t} = 0. \tag{14}$$

Substituting for $\varphi(I)$ and $\Phi(I)$ in equation (14), we have

$$E_g\{[v'(X-I(X))-\lambda\tilde{v}'(I(X))]h(X)\} = 0 \quad \text{for all functional } h(x).$$

Thus, $v'[x-I(x)] = \lambda\tilde{v}'[I(x)]$ for all $x \geq 0$, where λ satisfies $E_g[\tilde{v}(I^*(X))] = P$. □

Corollary 3.2: The optimal reinsurance contract $I^*(x)$ for $I_{\pi,P}$ is determined by $v'[x-I(x)] = \lambda v'[I(x)]$, where λ satisfies the constraint $\pi[I^*(X)] = P$.

Corollary 3.3: If $v(x) = x^2$, the optimal reinsurance contract for $I_{\pi,P}$ is a quota share contract.

4. Conclusion

This paper discusses the insurance or risk premium calculation known as the mean-value-distortion pricing principle in the general framework of anticipated utility theory. Essential properties such as non-negative loading, non-excessive loading, scale and translation invariant, stop-loss order preservation, and sub-additivity are preserved in the analysis of the pricing principle. It is also shown that for non-degenerate risks, independence and comonotonicity do not exist simultaneously. Risk aversion is not comparable under expected utility theory and Yaari's dual theory. This fact suggests consideration of insurance problems in a larger theoretical frame. Finally, optimal reinsurance contracts are derived by different computational methods.

Acknowledgements

The author gratefully acknowledges anonymous referees for their helpful comments and suggestions. The author is also grateful to Professor Jean Lemaire and Mr. Martin Halek from the Wharton School of the University of Pennsylvania for their work regarding the English language.

References

Bühlmann, H. (1970) *Mathematical Methods in Risk Theory*. Springer-Verlag.

Chateauneuf, A., Cohen, M. and Meilijson I. (1997) Comonotonicity, rank-dependent utilities and a search problem. In: Benes, V. and Stepan J. (Ed.). *Distributions with given marginals and moment problems*. Kluwer Academic Publishers, Amsterdam.

Denneberg, D. (1994) *Non-Additive Measure and Integral*. Kluwer Academic Publishers, Boston.

Denuit, M., Dhaene, J. and Van Wouve, M. (1999) The Economics of Insurance: a review and some recent developments. *Bulletin of the Swiss Association of Actuarie*s, 137-175.

Goovaerts, M.J., De Vylder, F. and Haezendonck J. (1984) *Insurance Premiums: Theory and Applications.* North-Holland, Amsterdam.

Goovaerts, M.J., Kaas, R., Van Heerwaarden, A.E. and Bavwelinck, X.T. (1990) *Effective Actuarial Methods*. North-Holland, Amsterdam.

Hürlimann, W. (1997) On quasi-mean value principles. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* XXIII, 1-16.

Hürlimann, W. (1998) On stop-loss order and the distortion pricing principles. *ASTIN Bulletin* **28(2)**, 119-134.

Puppe, C. (1991) *Distorted Probabilities and Choice under Risk*. Springer-Verlag.

Quiggin, J. (1982) A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3**, 323-343.

Rothschild, M. and Stiglitz J.E. (1970) Increasing risk: I. A Definition. *Journal of Economic Theory* **2**, 225-243.

Segal, U. (1989) Anticipated utility theory: a measure representation approach. *Annals of Operations Research* **19**, 359-373.

Schmeidler, D. (1986) Integral representation without additivity. *Proceedings of American Mathematical Society* **97**, 255-261.

Wang, S. (1996) Premium calculation by transforming the layer premium density. *ASTIN Bulletin* **26(2)**, 71-92.

WANG, S., YOUNG, V.R. and PANJER H.H. (1997) Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics* **21**,173-183.

WANG, S. (1998) An actuarial index of the right-tail risk. *North American Actuarial Journal* **2(2)**, 89-101.

WANG, S. and YOUNG V.R. (1998) Ordering of risks: utility theory verse Yaari's dual theory of risk. *Insurance: Mathematics and Economics* **22**, 145-161.

YAARI, M.E. (1987) The dual theory of choice under risk. *Econometrica* **55**, 95-115.

YOUNG,V.R. (1998) Families of updated rules for non-additive measures: Applications in pricing risks. *Insurance: Mathematics and Economics* **23**, 1-14.

YOUNG, V.R. (1999) Optimal insurance under Wang's premium principle. *Insurance: Mathematics and Economics* **25**, 109-122.

CUNCUN LUAN
Department of Mathematics
Nanjing University
Nanjing 210093
P.R. China
Email: Luancc@263.net

HEAVY-TAILED DISTRIBUTIONS AND RATING

By

J. Beirlant, G. Matthys and G. Dierckx

University Center of Statistics, Katholieke Universiteit Leuven

Abstract

In this paper we consider the problem raised in the Astin Bulletin (1999) by Prof. Benktander at the occasion of his 80th birthday concerning the choice of an appropriate claim size distribution in connection with reinsurance rating problems. Appropriate models for large claim distributions play a central role in this matter. We review the literature on extreme value methodology and consider its use in reinsurance. Whereas the models in extreme-value methods are non-parametric or semi-parametric of nature, practitioners often need a fully parametric model for assessing a portfolio risk both in the tails and in more central portions of the claim distribution. To this end we propose a parametric model, termed the generalised Burr-gamma distribution, which possesses such flexibility. Throughout we consider a Norwegian fire insurance portfolio data set in order to illustrate the concepts. A small sample simulation study is performed to validate the different methods for estimating excess-of-loss reinsurance premiums.

1. Introduction

The topic raised by Professor Benktander on the occasion of his 80th birthday concerning the choice of an appropriate claim size distribution in connection with a (multi-layer) rating problem is indeed a very fundamental area of discussion, both in the academic as in the practical (re-)insurance world.

On the one hand, modelling extreme events through Pareto-type and other heavy-tailed distributions attracts more and more attention. The number of statisticians working in extreme value methodology and the number of publications in this area is systematically growing; see the reference list for some recent books and papers with special emphasis on actuarial applications. Several important methods in this area were influenced by methods developed in the actuarial literature, not in the least by the paper by Benktander and Segerdahl *On the analytical representation of claim distributions with special reference to excess-of-loss distributions* (the XVIth International Congress of Actuaries, Brussels, 1960). Indeed, in that contribution the concept of the mean excess (or mean residual life) function was illuminated, which turned out to be quite a useful tool in extreme value statistics. Professor Benktander was also one of the first to introduce the concept of probability and quantile plotting in actuarial statistical practice, which, in our opinion, is the right way to view the data with the aim of tail modelling.

On the other hand, actuaries working in a reinsurance context are sometimes feeling uneasy with this material. One of the main problems is that the statistical extreme value models concern only the *ultimate tail* section of the distribution while a practitioner faced with reinsurance rating will need to model also more *central* areas of the distribution in order to handle the different layers in a flexible way. This, we believe, leads to another important merit of the abovementioned paper by Professor Benktander: the Benktander I and II distributions offer a nice compromise between statistical flexibility and efficiency, and computational simplicity with regard to premium rates. These classes contain all popular heavy-tailed models ranging from the Pareto distributions, over lognormal-type models to Weibull-type tails. At the same time the elegant expressions of their mean excess functions makes them especially attractive for the actuarial practitioner.

In this text we present a personal view on the link between statistical extreme value methods and the selection of appropriate statistical claim size models on the one side, and actuarial concepts, in particular the mean excess function, on the other. Proposals for statistical models that are able to capture both central and tail characteristics of the distribution will be presented. Finally, recent new directions in extreme value statistics, again motivated mainly by actuarial applications, will also be discussed. In Section 2 the relation between quantile plotting and the mean excess function is explained. In Section 3 we add the connection with extreme value methods. We order the presentation of the different approaches from non-parametric techniques over semi-parametric ones to a final fully parametric model in Section 4. The implications to premium rating are clarified along the way.

Throughout the text we use the fire claim data from a Norwegian portfolio in 1990 (taken from [1]) to illustrate the different methods and to give an idea of the typical problems with claim data modelling.

2. Quantile plotting and the mean excess function

Let $x_1, x_2, ..., x_n$ be claim data that come from a random sample $X_1, X_2, ..., X_n$ with distribution function F and survival function $\bar{F}(x) = P(X > x)$, denoting the probability to obtain a claim larger than x. The ordered data will be denoted by

$$x_{1,n} \leq x_{2,n} \leq ... \leq x_{n,n},$$

which are the sample values of the order statistics $X_{1,n} \leq ... \leq X_{n,n}$.
In case the expected value of X exists, i.e. $E(X) < \infty$, the *mean excess function* is given by

$$m(x) = E(X - x \mid X > x),$$

the expected excess claim size.

This function plays a central role in the rating of an excess-of-loss reinsurance in excess of a retention or priority level R, as the corresponding risk premium $\Pi(R)$ for the layer from R to infinity is given by (a multiple of)

$$\Pi(R) = \bar{F}(R)m(R) = E((X-R)_+).$$

It is a well-known fact that the most efficient way to derive m from $\bar{F}$ is by using the expression

$$m(x) = \frac{\int_x^\infty \bar{F}(u)du}{\bar{F}(x)},$$

while the inverse operation is given by

$$\bar{F}(x) = \frac{m(B)}{m(x)} \exp\left(-\int_B^x \frac{du}{m(u)}\right), \text{ for } x > B,$$

where B denotes the left limit of the support of F.

In practice, the mean excess function m is easily estimated at $x = X_{n-k,n}$ for some $k = 1, ..., n-1$ by the (empirical) average excess of the k data points higher than $X_{n-k,n}$:

$$\hat{m}_{k,n} = \frac{1}{k}\sum_{j=1}^{k} X_{n-j+1,n} - X_{n-k,n}.$$

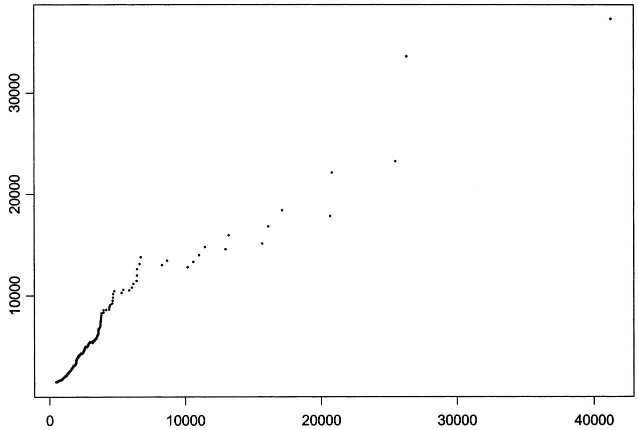

FIGURE 1: Plot of $\hat{m}_{k,n}$ as a function of $x_{n-k,n}$ for the Norwegian fire insurance data.

The *hazard rate* $\mu(x)$ defined by

$$\mu(x) = \frac{F'(x)}{\bar{F}(x)},$$

is closely linked to the mean excess function through the expression

$$\mu(x) = \frac{1 + m'(x)}{m(x)}.$$

So far for the recapitulation of the basic notions from (re)insurance mathematics. On the statistical side the potential of *quantile plotting* through quantile-quantile or QQ plots (or, alternatively, of probability plotting) for the graphical description and for the analysis of claim data has been stressed by several authors considering extreme value methods, see for instance [1], [17]. What may look innocuous or only somewhat suspect in a density comparison may become quite glaring in a QQ plot. Starting from the point of view that a heavy-tailed distribution is a distribution for which the tail is heavier than any exponential tail, i.e.

$$\lim_{x\to\infty} \frac{\exp(-\lambda x)}{\bar{F}(x)} = 0, \text{ for any } \lambda > 0,$$

the degree of deviation can be depicted through visual inspection of an *exponential quantile plot* of points with coordinates

$$\left(-\log\left(\frac{j}{n+1}\right), X_{n-j+1,n}\right).$$

Here the empirical quantiles $X_{n-j+1,n}$ appear as estimates of the unknown quantiles $Q(1-\frac{j}{n+1})$, defined as the claim levels that are surpassed in $\frac{j}{n+1}$ 100% of the cases. Hence, a straight line pattern in the exponential quantile plot will direct the practitioner to a model of the type

$$Q(1-p) = a + \frac{1}{\lambda}(-\log p)$$

for some a and $\lambda > 0$, and hence to

$$\bar{F}(x) = \exp(-\lambda(x-a)), x > a$$

i.e. an exponential model, perhaps shifted over a distance a.

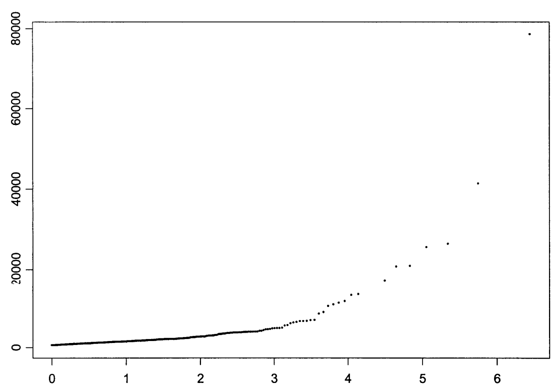

FIGURE 2: Exponential quantile plot for the Norwegian fire insurance data.

By the definition of the exponential quantile plot itself, namely that the vertical coordinates of the plotted positions are given by the data themselves, it follows that in case of a distribution with a tail heavier than any exponential, the plot will bend upwards away from a linear fit which is 'in line' with the exponential model. Rephrasing the expression 'bend upwards' more rigorously, we are led to stating that for such 'sub-exponential'[1] distributions the slope or the derivative of the exponential quantile plot increases as we increase the claim level.

One very naive way to estimate the slope of the exponential plot to the right of a point, say the position $\left(-\log(\frac{k+1}{n+1}), X_{n-k,n}\right)$, is to use the quotient of the average vertical and horizontal excesses over this position:

$$\frac{\frac{1}{k}\sum_{j=1}^{k} X_{n-j+1,n} - X_{n-k,n}}{\frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{n+1}{j}\right) - \log\left(\frac{n+1}{k+1}\right)},$$

or, even simpler,

$$\frac{1}{k}\sum_{j=1}^{k} X_{n-j+1,n} - X_{n-k,n},$$

since the denominator of the first expression is very closely approximated by 1 (as it is an approximation of the mean excess function of the unit exponential distribution, which is constantly equal to 1).

Hence we conclude that the empirical mean excess function $\hat{m}$ defined above is a naive derivative function for the exponential quantile plot. It also follows easily that the mean excess function of distributions with tails heavier than the exponential model all have an increasing mean excess function. The strongest increase is found for Pareto distributions for which the increase is linear.

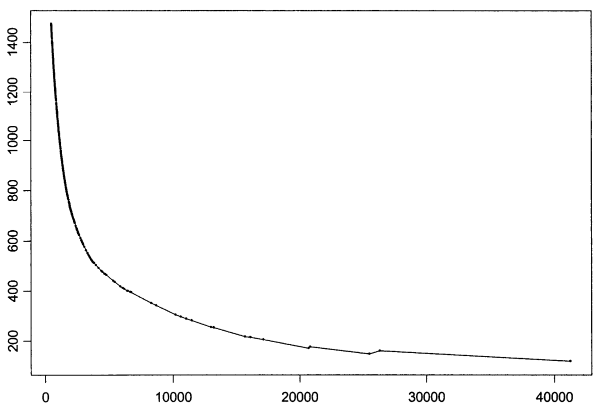

FIGURE 3: Non-parametric estimator of the excess-of-loss net premium as a function of the retention level R for the Norwegian fire insurance data.

[1] The concept of sub-exponentiality is only used here in an intuitive way, for a rigorous treatment see [13]

Further, this relationship shows that the exponential quantile plot is not only useful in the statistical validation of a claim model, but also in the calculation of a risk premium for a layer from $R = x_{n-k,n}$ to ∞, which can be estimated by $\frac{k+1}{n+1}\hat{m}_{k,n}$ in a purely non-parametric way, i.e. without assuming any parametric part in the statistical claim model.

Finally, to clarify the relationship between quantile plotting and the hazard rate μ, observe that the exponential quantile plot is the graph of the function $Q(1-e^{-x})$, which has derivative $e^{-x}/F'(Q(1-e^{-x}))$ or $(1/\mu)\,(Q(1-e^{-x}))$. Hence the reciprocal of the hazard rate follows exactly from the derivative of the exponential quantile plot at a plotting position.

3. Quantile plotting, mean excess and extreme value methods

In contrast to the previous fully *non-parametric* approach for premium calculation for an upper layer, extreme value methods typically use a *semi-parametric* approach, containing one or two parameters next to a functional part which is not specified. This seems reasonable from the fact that these methods are designed to make extrapolations outside the sample, for instance to estimate an extremely large quantile $Q(1-p)$ with $p < \frac{1}{n}$. Using a fully parametric model would then induce a second extrapolation from the sample towards the statistical population, and hence bias risk would only become larger.

3.1. Pareto-type distributions

The most famous example of such a semi-parametric extreme value model is *the Pareto-type model*, which is deduced from limit theory for the maximum $X_{n,n}$ of a sample:

$$\bar{F}(x) = x^{-a}\ell(x),$$

where ℓ is a slowly-varying function (at infinity), i.e. which satisfies,

$$\frac{\ell(tx)}{\ell(x)} \to 1, \text{ as } x \to \infty, \text{ for every } t > 0.$$

Here the tail index a is the important, decisive parameter, while ℓ is a nuisance function. Working under this model amounts to assuming that the survival function behaves in first order as a power law. Examples of popular claim size models which belong to this class are, of course, the (strict) Pareto model itself (and hence the Benktander distribution with the parameter B equal to 0), next to the Burr, the generalised Pareto, the loggamma, the log-logistic and the Fréchet distribution, among others.

The estimator of a which has received by far the most attention (and still does attract a lot of research) was proposed by Hill (1975) [15] and was shown by Mason (1982) [18] to be consistent under the complete Pareto-type model, and in that sense appears to be a perfect semi-parametric estimator at first sight:

$$1/\hat{a}_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n}.$$

The Hill statistic is nothing else than the mean excess estimate of the log-transformed data at $X_{n-k,n}$, and hence can be deduced from a Pareto quantile plot, which is the exponential quantile plot of the log-transformed data. Indeed, under the Pareto-type model such a Pareto quantile plot can be shown to be ultimately linear with slope approaching $1/a$ above some high threshold $X_{n-k,n}$, i.e. for small enough k and large enough n.

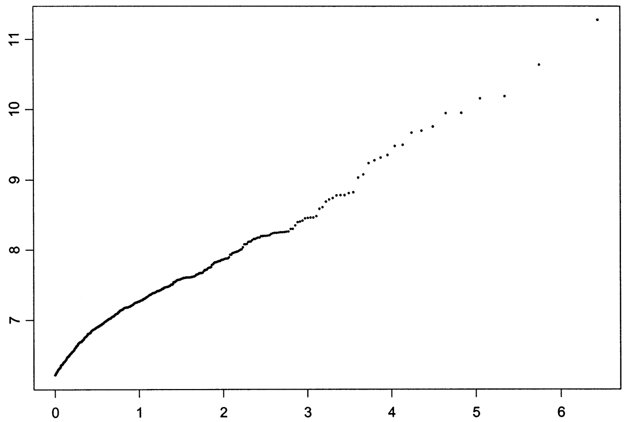

FIGURE 4: Pareto quantile plot for the Norwegian fire insurance data.

Remark that we have in fact as many estimates of a as we have data points; for each value of k we obtain a new estimate of a. Plots of $\hat{a}_{k,n}$ as a function of k are often quite volatile. In [20] it is mentioned that it is helpful to plot the Hill estimates as a function of $\log k$ (in fact, this is equivalent to using the same horizontal scale as in the Pareto quantile plot). For the Norwegian fire claim data, however, there is no apparent gain with this approach.

Several authors have tried to guide the practitioner in choosing k, leading to an adaptive choice $\hat{k}$ such that an estimate of the mean squared error of the Hill estimator is minimised at $\hat{k}$. This was done by bootstrap methods (see, for instance, [9]) or by regression diagnostics on a Pareto quantile plot in [2]. A somewhat different solution was proposed in [12]. In case of the Norwegian fire insurance portfolio the method indicated in [3] yields the value $k = 290$, which results in the estimate $1/\hat{a} = 0.62$.

Other problems are for instance the non-invariance of the Hill estimator with respect to shifts that could be applied to the data, and most importantly, the bias that the Hill estimator exhibits in certain cases. This can be understood from the fact that for certain Pareto-type distributions (as it is the case for the loggamma distribution, for instance) the influence of the slowly-varying

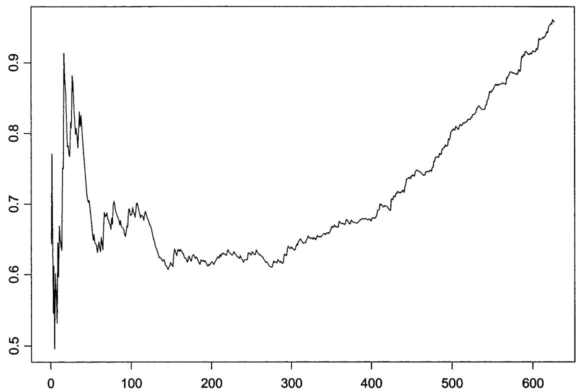

FIGURE 5: Plot of $1/\hat{a}_{k,n}$ as a function of k for the Norwegian fire insurance data.

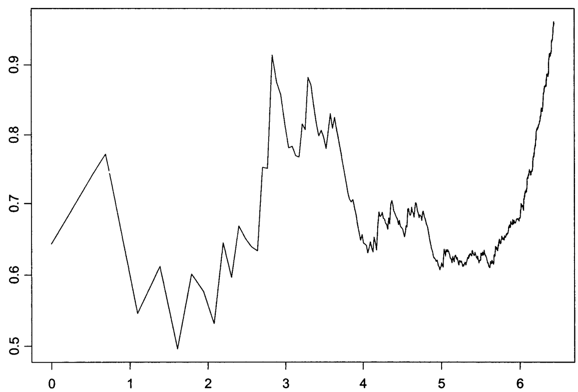

FIGURE 6: Plot of $1/\hat{a}_{k,n}$ as a function of $\log k$ for the Norwegian fire insurance data.

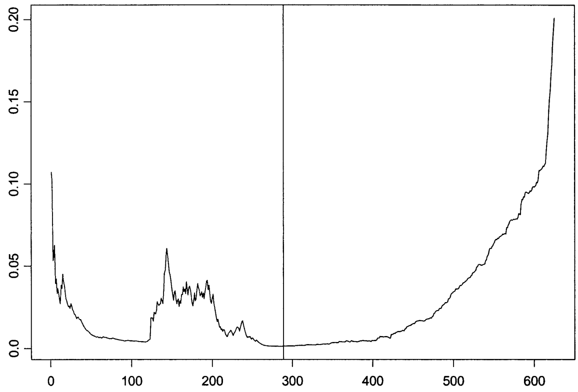

FIGURE 7: Plot of an estimator (see [3]) of the asymptotic mean squared error of the Hill estimator as a function of k for the Norwegian fire insurance data. A minimum is found at $k = 290$.

part ℓ is still imminent near the top end of the Pareto quantile plot. As a consequence, confidence intervals for α will not show the required coverage probability in such cases. This puts a serious restriction on the reliability of these methods.

Next to methods based on high order statistics, such as the Hill estimator, an alternative is offered by the *peaks-over-threshold approach* (POT). This method consists of fitting the *generalised Pareto distribution* (GPD) to the distribution of the excesses $Y = X - u$ (if $X > u$) over a high threshold u, for instance by maximum likelihood methods [23], the method of moments [16], or modern Bayesian estimation methods [8]. By its nature this approach has a natural link with excess-of-loss reinsurance replacing the retention level R by the statistical threshold u ; for a discussion, see [19], [22]. This approach is based on a limit result of Pickands (1975) [20] stating that as $u \to \infty$, the survival function of the excesses tends to the survival function of the GPD given by $(1 + \frac{x}{\alpha\sigma})$ with the scale parameter $\sigma = \sigma_u$ depending on u. Again, every choice of u leads to another estimator of α and of course σ_u. Smith has advised to choose $u = X_{n-k,n}$ at the smallest value of $X_{n-k,n}$ to the right of which the mean excess plot remains approximately linear as a function of the ordered data. Pure adaptive algorithmic choices have not yet been explored systematically, however. The POT method possesses some advantages over the methods based on extreme order statistics such as the ones derived from the Hill estimates: it is invariant with respect to shifts and the plots of the estimates of α as a function of k are often more stable, apparently because of the use of the second parameter σ. However, also here the asymptotic result of Pickands can set in really slowly, leading to biased estimates of α for this method too. In case of the Norwegian fire insurance data, the POT method does not lead to a more stable graph when the estimates are plotted as a function of k; see Figure 8.

Let us now consider again the estimation of the risk premium for a layer from R to ∞ with the semi-parametric approach. Using the concept of the

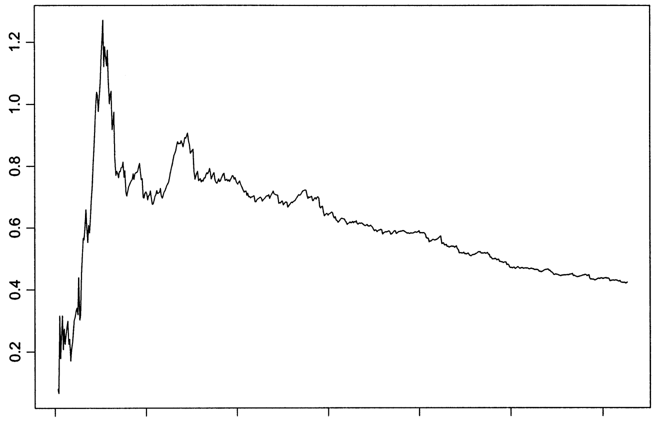

FIGURE 8: Plot of the POT estimates (ML) as a function of k for the Norwegian fire insurance data.

Hill estimator, we arrive at the following approximation based on the famous Karamata theorem (see for instance [7]) for $\alpha > 1$:

$$\begin{aligned} \Pi(R) &= \int_R^\infty u^{-\alpha} \ell(u) du \\ &\sim \frac{1}{\alpha-1} R^{1-\alpha} \ell(R) \\ &= \frac{1}{\alpha-1} R\bar{F}(R). \end{aligned}$$

When the priority R is situated within the sample, i.e. when claims at the magnitude of R have previously been observed and R is taken equal to $x_{n-k,n}$ for some k, this leads to the estimate $\hat{\Pi}(x_{n-k,n}) = \frac{1}{\hat{\alpha}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$. Figure 9 presents these estimates for the fire insurance data.

If R is not fixed at one of the sample points, extreme value formulas for estimation of $\bar{F}(R)$ in the expression for $\Pi(R)$ can be applied (see for instance [1] or [12]): $\hat{\bar{F}}(R) = \left(\frac{\hat{k}+1}{n+1}\right)\left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{\alpha}}$ with $\hat{k}$ denoting an appropriate adaptive choice for the number of extreme order statistics used in the procedure, which can be obtained with the methods mentioned above. This is shown in Figure 10 for our example.

Alternatively, the POT approach suggests substituting the conditional expected value of the GPD for the mean excess function at a *high* priority R (for $R > u$) : $m(R) = \frac{\alpha_u}{\alpha_u - 1} \sigma_u \left(1 - \frac{R-u}{\alpha_u \sigma_u}\right)$, while $\bar{F}(R)$ will be estimated with the formula $\frac{k+1}{n+1}\left(1 - \frac{R-u}{\alpha \sigma_R}\right)^{-\alpha}$ when k observations exceed the threshold u. Replacing α_u and σ_u by their estimates then leads to an estimate of the risk premium as in Figure 11.

When estimating the premium at a retention level within the sample, i.e. $R = x_{n-k,n}$, one can fix the threshold u at R and then the POT approach leads to an estimate $\frac{\hat{\alpha}_R}{\hat{\alpha}_R - 1} \hat{\sigma}_R \frac{k+1}{n+1}$. This is of the same form as the first estimation

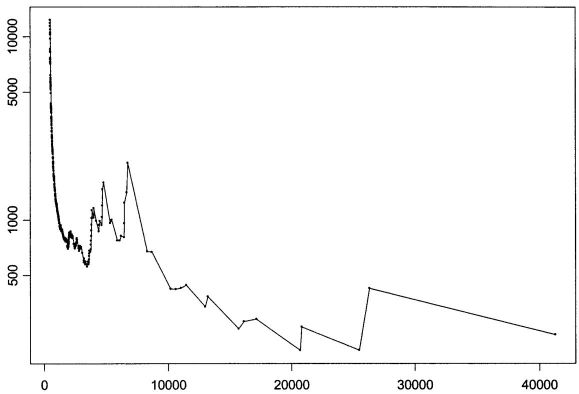

FIGURE 9: Plot of $\frac{1}{\hat{\alpha}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ as a function of $R = x_{n-k,n}$ for the Norwegian fire insurance data. (log-scale on Y-axis).

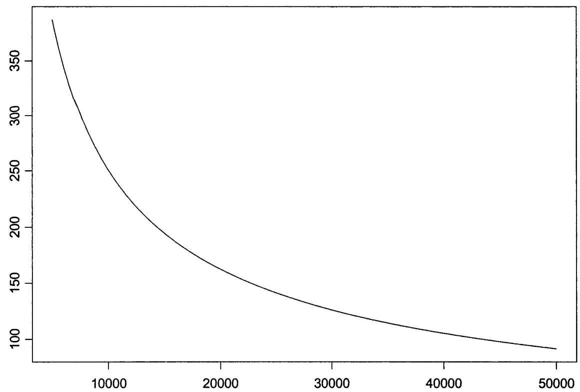

FIGURE 10: Plot of $\frac{1}{\hat{a}_{\hat{k}}-1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ as a function of R for the Norwegian fire insurance data.

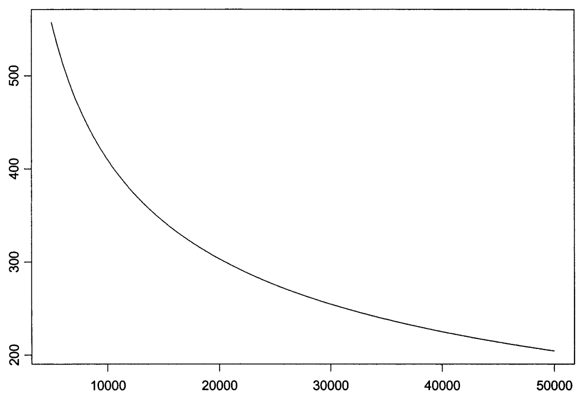

FIGURE 11: Plot of POT-based premium estimates as a function of retentions R situated beyond the threshold $u = x_{n-290,n}$ for the Norwegian fire insurance data.

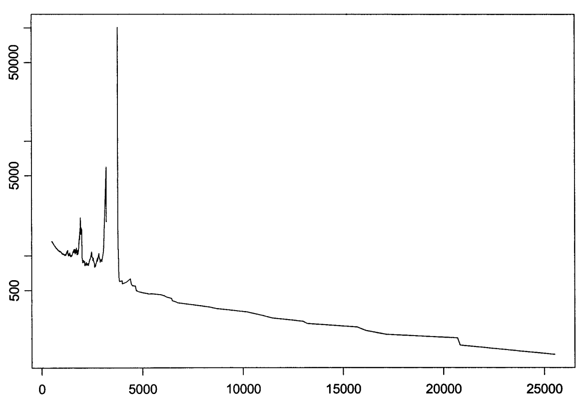

FIGURE 12: Plot of the POT based premium estimates $\frac{\hat{a}_R}{\hat{a}_R-1} \hat{\sigma}_R \frac{k+1}{n+1}$ as a function of $R = x_{n-k,n}$ for the Norwegian fire insurance data. (log-scale on Y-axis).

method based on the Hill estimator, replacing $x_{n-k,n}$ by $\hat{a}_R\hat{\sigma}_R$. A plot of these estimates for the Norwegian fire insurance data is shown in Figure 12.
A summary of all above estimation methods can be found in Table 1.

3.2. Bias reduction in estimating the Pareto index

The abovementioned problems with systematic biases appearing in the 'classical' extreme value methods have only recently led some authors [3], [14] to look in more detail at important (parametrised) subclasses of the set of all slowly-varying functions. The following class was first indicated by Hall (1984):

$$\ell(x) = C\left(1 + Dx^{-\beta}(1 + o(1))\right),$$

(with C, D and β denoting positive constants) to which belong for instance the Burr, the generalised Pareto and the Fréchet distribution. Another helpful subclass is given by

$$\ell(x) = C(\log x)^{\beta}(1 + o(1))$$

to which belongs for instance the loggamma distribution.

It is then shown that for k not too large the scaled logarithmic spacings $Z_j := j(\log X_{n-j+1,n} - \log X_{n-j,n})$, $j = 1, \ldots, k$, can be modelled by the following generalised regression models:

a power regression model

$$Z_j = \left(\frac{1}{\alpha} + b_{n,k}\left(\frac{j}{k+1}\right)^{\rho}\right) f_j, 1 \le j \le k, k < n$$

with $b_{n,k}$ and $\rho(> 0)$ depending on C, D and β, and $f_1, f_2, \ldots$ denoting independent and identically distributed unit exponential random variables; respectively,

a logarithmic regression model

$$Z_j = \frac{1}{\alpha}\left(j\log\left(\frac{j+1}{j}\right) + \beta j \log\frac{\log\frac{n+1}{j}}{\log\frac{n+1}{j+1}}\right) + \varepsilon_j,\ 1 \le j \le k, k < n,$$

where ε_j denote centered exchangeable error random variables. The latter model is to be used when the parameter ρ in the power regression model is close to 0.

Again for every k another estimate of α is obtained, e.g. by joint maximum likelihood estimation of α, ρ and $b_{n,k}$, or β, but typically the plots of the estimates as a function of k are much more stable. Also, the covariate terms $b_{n,k}\left(\frac{j}{k+1}\right)^{\rho}$, respectively $\beta j \log\frac{\log\frac{n+1}{j}}{\log\frac{n+1}{j+1}}$, remove the bias of the original Hill-type estimators to a high extent. Finally, the problem concerning the non-invariance of the original estimators with respect to shifts has also been lifted up, i.e. one

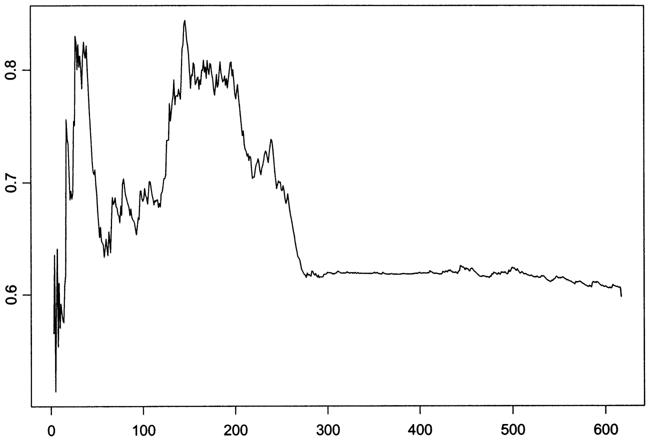

FIGURE 13: Plot of estimates of $\frac{1}{a}$ based on the power regression model as a function of k for the Norwegian fire insurance data.

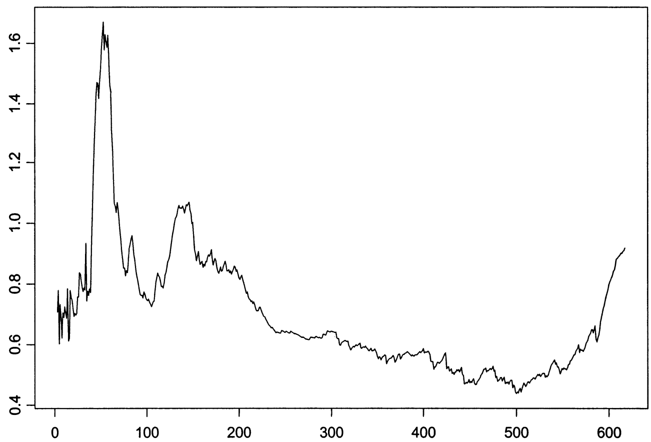

FIGURE 14: Plot of estimates of $\frac{1}{a}$ based on the logarithmic regression model as a function of k for the Norwegian fire insurance data.

can add or subtract values up to the third quartile of the underlying distribution while the bias-corrected estimates remain stable. On the other hand the standard deviation has inflated in comparison with the simpler estimators but it stays of order $1/\sqrt{k}$.

Of course, a practitioner has to choose between the two estimates of a obtained by each of these two generalised regression models. In the Norwegian fire insurance example, the estimates obtained from the power regression model seem to be more stable than those from the logarithmic model. Here again the value around 0.6 appears as an estimate of $1/a$. The estimates corresponding to $k < 290$ indicate the possibility of a mixture with even a heavier tail at the extreme right end of the distribution. In the whole, the logarithmic model does not appear to fit well in this case, which gives rise to a larger variability in the estimates over the range of k-values.

The different ways to estimate a premium for an excess-of loss reinsurance contract with retention R covered above, namely $\hat{\Pi}(x_{n-k,n}) = \frac{1}{\hat{a}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ for a retention $R = x_{n-k,n}$, respectively $\frac{1}{\hat{a}_{\hat{k}}-1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ when $R > x_{n-\hat{k},n}$, can now be recomputed replacing the Hill estimate $\hat{a}$ of a by the new estimates based on the power or logarithmic regression model. The results for the Norwegian fire claim data are given in Figures 15 through 18.

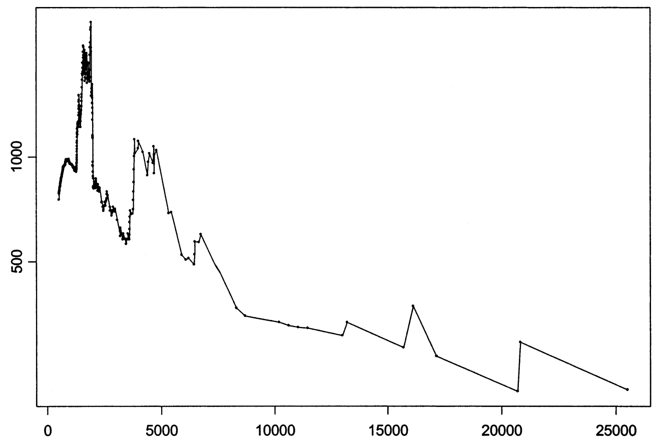

FIGURE 15: Plot of $\frac{1}{\hat{a}_{\hat{k}}-1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ as a function of $R = x_{n-k,n}$ with $\hat{a}$ the ML estimator from the power regression model for the Norwegian fire insurance data. (log-scale on Y-axis).

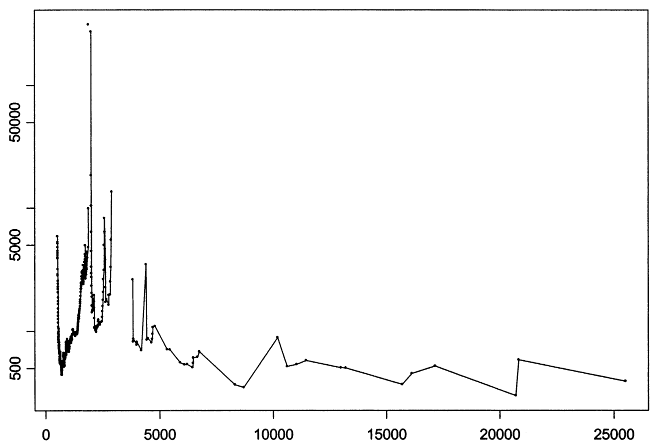

FIGURE 16: Plot of $\frac{1}{\hat{a}_k-1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ as a function of $R = x_{n-k,n}$ with $\hat{a}$ the ML estimator from the logarithmic regression model for the Norwegian fire insurance data. (log-scale on Y-axis).

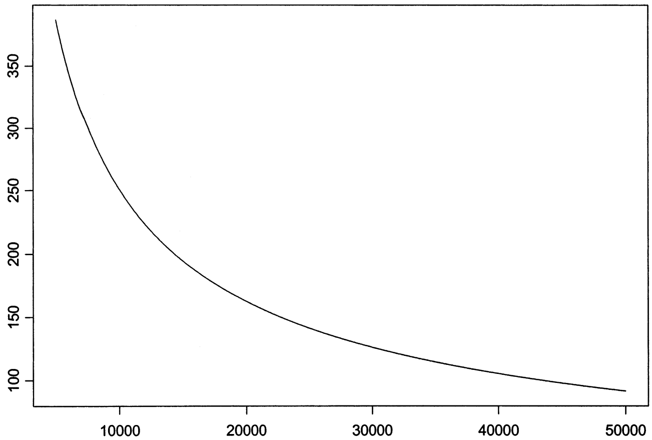

FIGURE 17: The Norwegian fire insurance data: plot of $\frac{1}{\hat{a}_{\hat{k}}-1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ as a function of R with $\hat{a}_{\hat{k}}$ obtained from the power regression model.

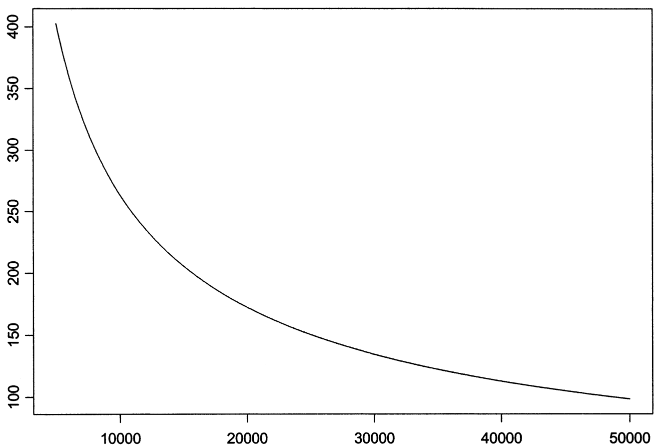

FIGURE 18: The Norwegian fire insurance data: plot of $\frac{1}{\hat{a}_{\hat{k}}-1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ as a function of R with $\hat{a}_{\hat{k}}$ obtained from the logarithmic regression model.

3.3. The Gumbel maximum domain of attraction

Next to the Pareto-type models, important claim distributions such as the lognormal and the Weibull distributions (which are included in the framework of the Benktander I and II classes of distributions) have to be available in a practitioner's toolbox. Formally, this class is defined as the set of distributions for which maxima are attracted in distribution to the Gumbel distribution with distribution function $\exp(-\exp(-x))$ for large sample sizes. In extreme value methodology this group of distributions is modelled with an extension of the Pareto-type distributions through the *extreme value index* $\gamma = 1/a$, defining the extreme value index γ to be 0 for this large class of distributions with exponentially fast decreasing tails. Remark that the lognormal distribution is then really on the borderline between the Pareto-type distributions and the $\gamma = 0$ class, as the first order approximation (for $x \to \infty$) of the survival function of the lognormal distribution is given by $\bar{F}(x) \sim C_1 \exp(-C_2(\log x)^2)$ for some positive constants C_1, C_2.

The difficulties encountered by the extreme value methods can be illustrated by the POT approach, for which the Gumbel class approximation is obtained formally by letting $a \to \infty$ in the definition of the GPD, leading to an exponential fit $\exp(-x/\sigma_u)$. In general the goodness of fit of an exponential distribution to the excess distribution over a high threshold u will only appear to be accurate for extremely large thresholds u, which are only useful in practice for very high sample sizes.

Extensions of the Hill estimator are also available. Here we mention the *moment estimator* [10] of φ given by

$$\hat{\gamma}^{M}_{k,n} = M^{(1)}_{k,n} + 1 - \frac{1}{2}\left(1 - \frac{\left(M^{(1)}_{k,n}\right)^2}{M^{(2)}_{k,n}}\right)^{-1},$$

where

$$M^{(i)}_{k,n} = \frac{1}{k}\sum_{j=1}^{k}(\log X_{n-j+1,n} - \log X_{n-k,n})^i, \quad i = 1, 2$$

with $M^{(1)}_{k,n}$ being the Hill statistic.

An extension to the case $\gamma \geq 0$ of the graphical support that was offered by the Pareto quantile plot for the Pareto-type distributions appears to be a natural question. In [4] it was shown that in this general case, the mean residual life function m satisfies $m(Q(1-p)) = p^{-\gamma}\ell(1/p)$ for some slowly-varying function ℓ (in case $0 \leq \gamma < 1$). Hence the *the quantile – mean excess plot*, or QM plot,

$$\left(-\log\frac{k+1}{n+1}, \log\hat{m}_{k,n}\right), \quad 2 \leq k \leq n$$

will be ultimately linear with slope γ. This then leads to an estimator of γ as it was done on the basis of the Pareto quantile plot in case of $\gamma > 0$ which entailed the Hill estimator and other bias reduced estimators. So, here the message is to plot the log-transformed empirical mean excess values $\log\hat{m}_{k,n}$ against the log-scale k in order to estimate the value of γ and to capture the Pareto versus non-Pareto behaviour of the tail of the distribution: ultimately horizontal QM plots point in the direction of an exponentially decreasing tail.

This technique can be adapted so as to work without the restriction $\gamma < 1$ by replacing $\log\hat{m}_{k,n}$ in the QM plot by $\log\left(X_{n-k,n}\frac{1}{k}\sum_{j=1}^{k}(\log X_{n-j+1,n} - \log X_{n-k,n})\right)$.

3.4. Comparing the different premium calculation techniques

The different semi-parametric ways to estimate excess-of-loss premiums that are covered above are summarised in Table 1. They do yield quite different results in our case study. In order to inspect this in more detail, a small sample simulation study was performed using a Burr distribution with

$$F(x) = 1 - \left(\frac{1}{1+\sqrt{x}}\right)^2.$$

We focus on the methods developed for estimating $\Pi(R)$ when the retention satisfies $R > X_{n-\hat{k},n}$ with $\hat{k}$ chosen to minimise the mean squared error of the Hill estimator, i.e. (2), (4) and (6). The results based on the POT method (4)

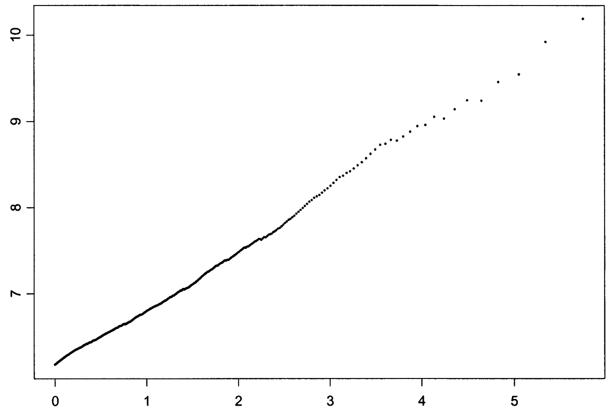

FIGURE 19: Adapted QM plot for the Norwegian fire insurance data.

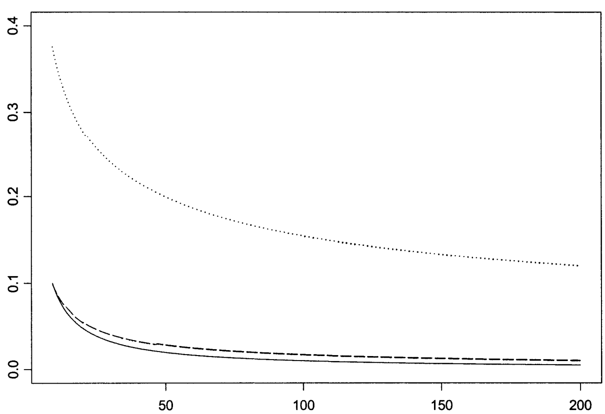

FIGURE 20: Simulation study based on 100 simulated data sets of size 500 from a Burr distribution: exact $\Pi(R)$ (solid line); median $\hat{\Pi}$(R) for methods (4) and (6) (dashed line); median $\hat{\Pi}$(R) for method (2) (dotted line).

TABLE 1

OVERVIEW OF THE DIFFERENT NON- AND SEMI-PARAMETRIC ESTIMATION METHODS FOR AN EXCESS-OF-LOSS REINSURANCE

Estimation method $\hat{a}$	Retention within sample $R = x_{n-k,n} (= u)$	Retention beyond threshold $R > x_{n-\hat{k},n} (= u)$ with $\hat{k}$ obtained by minimizing $AMSE(1/\hat{a})$
Hill estimator	$\hat{\Pi} = \frac{1}{\hat{a}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ (1)	$\hat{\Pi} = \frac{1}{\hat{a}_{\hat{k}} - 1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ (2)
POT ML	$\hat{\Pi} = \frac{\hat{a}_R}{\hat{a}_R - 1} \hat{\sigma}_R \left(\frac{k+1}{n+1}\right)$ (3)	$\hat{\Pi} = \frac{\hat{a}_u}{\hat{a}_u - 1} \hat{\sigma}_u \left(\frac{\hat{k}+1}{n+1}\right) \left(1 + \frac{R-u}{\hat{a}_u \hat{\sigma}_u}\right)^{1-\hat{a}_u}$ (4)
Regression model ML	$\hat{\Pi} = \frac{1}{\hat{a}_k - 1} x_{n-k,n} \left(\frac{k+1}{n+1}\right)$ (5)	$\hat{\Pi} = \frac{1}{\hat{a}_{\hat{k}} - 1} R \left(\frac{\hat{k}+1}{n+1}\right) \left(\frac{R}{X_{n-\hat{k},n}}\right)^{-\hat{a}_{\hat{k}}}$ (6)

and the regression model estimates (6) are almost identical and are in fact quite satisfactory. The simplest method based on the Hill estimator typically overestimates the correct Π function and entails a strong positive bias. In Figure 20 the median curves $\hat{\Pi}$ as a function of R are given, based on 100 simulated data sets of size $n = 500$.

4. CAPTURING CENTRAL ÁND TAIL CHARACTERISTICS

Having explained the difficulties and merits with nowadays' methods from extreme value statistics, we clearly recognise the need for completely *parametric claim models* that are capable to fit well both the tail and more central parts of the claim domain. However, fitting any such model, if existing, cannot be performed in a classical statistical way, e.g. by the use of χ^2 goodness-of-fit techniques. The parameters linked with the tail behaviour need to be estimated by methods from extreme value statistics as described above.

One such class of distributions was recently proposed in [5], termed the *generalised Burr-gamma distribution*. The distribution function is given by

$$F(x) = \frac{1}{\Gamma(p)} \int_0^{u_\xi(x)} e^{-u} u^{p-1} du$$

where

$$u_\xi(x) = \frac{1}{\xi} \log(1 + \xi u(x))$$

with $u(x) = x^{\frac{1}{b}} / \tau$.[2]

[2] In fact, in [5] $u(x)$ is modelled by $\exp\left\{\psi(p) + \frac{\log(x) + \mu}{\sigma} \sqrt{\psi'(p)}\right\}$, with ψ, resp. ψ', denoting the digamma, resp. the trigamma function. For simplicity we introduce the parameters b and τ here.

It can be seen that the parameter $b\xi$ equals the extreme value index for this parametric model. Several sub-models have appeared in the discussion above and show the flexibility of this model:

- If $\xi = 0$ then X is distributed as a generalised gamma distribution. Remark that in this case u_ξ is to be read as u and hence this model provides a generalization of the Weibull distribution

$$\bar{F}(x) = \frac{1}{\Gamma(p)} \int_{\frac{x^{1/b}}{\tau}}^{\infty} e^{-u} u^{p-1} du.$$

  The Weibull distribution is obtained choosing $p = 1$.

- If $\xi = 0$ and $p \to \infty$ this model approximates a lognormal distribution (see [5]).
- In case $\xi > 0$ we find that for p a positive integer

$$\bar{F}(x) = \left(1 + \xi \frac{x^{1/b}}{\tau}\right)^{-1/\xi} \left(\sum_{j=0}^{p-1} \frac{1}{j!} \frac{1}{\xi^j} \log^j \left(1 + \xi \frac{x^{1/b}}{\tau}\right)\right).$$

  Hence important actuarial claim models such as the Burr model (which includes the GPD) and the loggamma distribution are special cases of, or can be mimicked by this model.

How can one proceed to estimate the different parameters p, ξ, b and τ in this model? First, as $b\xi$ is the extreme value index for this model, it can be estimated with the methods discussed in the preceding section. This part of the estimation procedure is then based on a number k of extreme order statistics, i.e. the number k of highest claims in the sample, which is to be chosen adaptively as discussed above. In fact, supposing for instance that $\gamma > 0$, one finds that for this model the extreme value regression model

$$Z_j = \frac{1}{a}\left(j \log\left(\frac{j+1}{j}\right) + \beta j \log \frac{\log \frac{n+1}{j}}{\log \frac{n+1}{j+1}} \right) + \varepsilon_j, \quad 1 \le j \le k$$

holds for $k/n \to 0$ with $a = 1/(b\xi)$ and $\beta = p - 1$. This allows for estimation of $b\xi$ and p, for instance by a least-squares method based on the k highest claim data.

In Figure 21 we show the result for $\hat{p}$ for the fire claim data, which indicates the choice $p = 1$ and confirms the validity of a model without logarithmic factors. Hence, in this case the generalised Burr-gamma model reduces to

$$\bar{F}(x) = \left(1 + \xi \frac{x^{1/b}}{\tau}\right)^{-1/\xi}$$

which is in fact a Burr model. The method of moments yields the following estimates for b and τ

$$\hat{b} = 0.195,$$
$$\hat{\tau} = 8.94 \cdot 10^{14},$$

leading to an estimate $\hat{\xi} = 3.197$ for ξ.

The goodness of fit of this model is analyzed in Figure 22 using a QQ plot that shows the empirical quantiles versus the corresponding theoretical quantiles from the fitted Burr distribution. A point of inflection appears, which confirms our previous supposition of a mixture of distributions in the tail. This of course complicates the analysis.

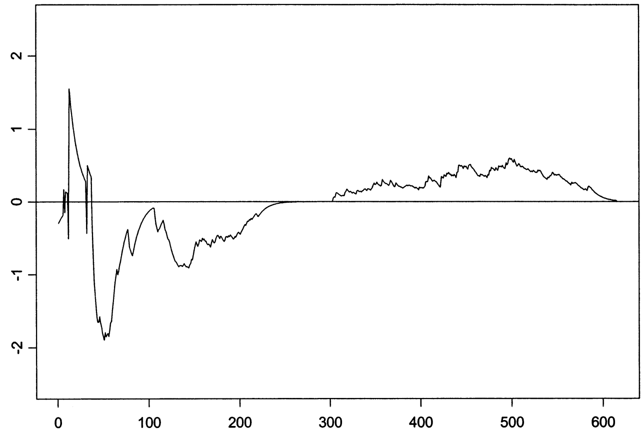

FIGURE 21: The Norwegian fire insurance data: plot of $\hat{\rho}-1$ as a function of k.

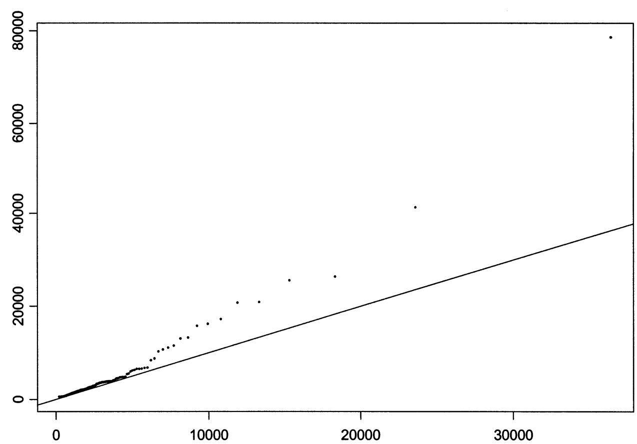

FIGURE 22: The Norwegian fire insurance data: QQ plot of empirical quantiles versus fitted Burr quantiles.

Finally, the premium $\Pi(R)$ for the fitted Burr model is easily computed numerically for different values of R. The result, given in Figure 23, is situated a bit lower than the results obtained in Figures 17 and 18. This can be understood from the fact that – partly due to the complication of the tail mixture – less weight is given to the tail section in this fully parametric analysis.

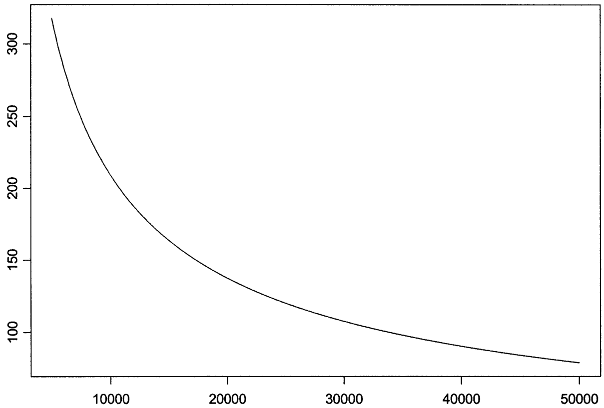

FIGURE 23: The Norwegian fire insurance data: plot of Π as a function of R based on the fitted Burr model.

5. CONCLUSION

In this paper we have tried to overview the different stages in a claim modelling process and risk premium calculation, starting with a completely non-parametric, over a semi-parametric, towards a completely parametric approach. A constant theme throughout this approach is the inspection of the tail behaviour, which is a prerequisite for accurate premium calculations, especially with reinsurance layers which cover the highest risks. Of course this discussion is certainly not the final answer but a description of the state-of-the-art in an active field of research.

REFERENCES

1. BEIRLANT, J., TEUGELS, J.L. and VYNCKIER, P. (1996) *Practical Analysis of Extreme Values*. Leuven University Press.
2. BEIRLANT, J., TEUGELS, J.L. and VYNCKIER, P. (1996) Tail index estimation, Pareto quantile plots and regression diagnostics. *J. Am. Statist. Assoc.*, **91**, 1659-1667.
3. BEIRLANT, J., DIERCKX, G., GOEGEBEUR, Y. and MATTHYS, G. (1999) Tail index estimation and an exponential regression model. *Extremes*, **2**, 177-200.
4. BEIRLANT, J., TEUGELS, J.L. and VYNCKIER, P. (1996) Excess functions and the estimation of the extreme value index. *Bernoulli*, **2**, 293-318.
5. BEIRLANT, J., DE WAAL, D.J. and TEUGELS, J.L. (2000) The generalized Burr-gamma family of distributions with applications in extreme value analysis. To appear in: *Proceedings of the 4th Conference on Limit Theorems in Probability and Statistics of the J. Bolyai Soc.*
6. BENKTANDER, G. and SEGERDAHL, C.O. (1960) On the analytical representation of claim distributions with special reference to excess-of-loss reinsurance. *Trans. 16-th Intern. Congress Actuaries*, 626-646.
7. BINGHAM, N.H., GOLDIE, C.M. and TEUGELS, J.L. (1987) *Regular Variation*. Cambridge University Press.
8. COLES, S.G. and POWELL, E.A. (1996) Bayesian methods in extreme value modelling: a review and new developments. *Int. Statist. Rev.*, **64**, 119-136.

9. Danielsson, J., de Haan, L., Peng, L. and De Vries, C.G. (1997) Using a bootstrap method to choose the sample fraction in tail index estimation. Technical report, Erasmus University, Rotterdam.
10. Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989) A moment estimator for the index of an extreme value distribution. *Ann. Statist.*, **17**, 1833-1855.
11. Dierckx, G. (2000) *Estimation of the extreme value index*. Ph.D. Thesis, Dept. of Mathematics, K.U. Leuven.
12. Drees, H. and Kaufmann, E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Proc. Appl.*, **75**, 149-172.
13. Embrechts, P., Klüppelberg, C. and Mikosch, Th. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin Heidelberg New York.
14. Feuerverger, A. and Hall, P. (1999) Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, **27**, 760-781.
15. Hill, B.M. (1975) A simple approach to inference about the tail of a dsitribution. *Ann. Statist.*, **3**, 1163-1174.
16. Hosking, J.R.M. and Wallis, J.R. (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339-349.
17. Kratz, M. and Resnick, S. (1996) The qq-estimator of the index of regular variation. *Commun. Statist., Stoch. Models*, **12**, 699-724.
18. Mason, D.M. (1982) Laws of large numbers for sums of extreme values. *Ann. Prob.*, 10, 756-764.
19. McNeil, A. (1997) Estimating the tails of severity distributions using extreme value theory. *Astin Bulletin*.
20. Pickands, J. III (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119-131.
21. Resnick, S.I. and Starica, C. (1997) Smoothing the Hill estimator. *Adv. Appl. Probab.*, **29**, 271-293.
22. Rootzén, H. and Tajvidi, N. (1997) Extreme value statistics and wind storm losses: a case study. *Scand. Act. J.*, 70-94.
23. Smith, R.L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174-1207.

ULTIMATE RUIN PROBABILITIES FOR GENERALIZED GAMMA-CONVOLUTIONS CLAIM SIZES

By

M. Usábel

Universidad Carlos III de Madrid

Abstract

A method of inverting the Laplace transform based on the integration between zeros technique and a simple acceleration algorithm is presented. This approach was designed to approximate ultimate ruin probabilities for Γ-*convolutions* claim sizes, but it can be also used with other distributions. The stable algorithm obtained yields interval approximations (lower and upper bounds) to any desired degree of accuracy even for very large values of u (1,000,000), initial reserves, without increasing the number of computations. This last fact can be considered an interesting property compared with other recursive methods previously used in actuarial literature or other methods of inverting Laplace transforms.

Keywords and phrases

Ultimate ruin probability, upper and lower bounds, stable recursive algorithms, numerical inversion of the Laplace transform.

1. Introduction

Let us consider the Classical risk process in continuous time $\{Z_t\}_{t \geq 0}$ with U_k claim sizes and premium ρ per time unit,

$$Z_t = u + \rho\, t - \sum_{k=1}^{N_t} U_k$$

where u are the initial reserves and N_t the total number of claims up to time t following a homogeneous Poisson process of parameter $\lambda > 0$. Let $F(x)$ denote the distribution function of claim sizes U_k with mean p_1 and $p = \lambda\, p_1(1 + \theta)$, where θ is the premium loading factor. We will also assume that $\theta > 0$ and $F(x) = 0$.

Using a renewal argument and Theorem 13.5.1 of the text-book by Bowers et al. (1997), the ultimate non-ruin probability can be expressed using the following integral equation (Volterra integral equation of the second kind)

$$\Phi(u) = \frac{\theta}{(1+\theta)} + \frac{1}{(1+\theta)p_1}\int_0^u \Phi(u-w)(1-F(w))dw \tag{1.1}$$

or in the case of ruin probability, see Gerber (1979, p. 115, equation (3.7)

$$\Psi(u) = \frac{1}{(1+\theta)p_1}A(u) + \frac{1}{(1+\theta)p_1}\int_0^u \Phi(u-w)(1-F(w))dw \tag{1.2}$$

where

$$A(w) = \int_w^\infty (1-F(z))dz \tag{1.3}$$

Since the early 1980s, many methods have been developed in order to approximate $\Psi(u)$. They were based on a discretization of some aspect of the risk process and derived recursive expressions; see for example Goovaerts and De Vylder (1984), Panjer (1986), Dickson (1989), Dickson and Waters (1991), Ramsay (1992b), Dickson, Egidio dos Reis and Waters (1995). Panjer and Wang (1993) describe the conditions under which these recursions are stable.

Although some of these recursive approaches may be able to determine $\Psi(u)$ to any desired degree of accuracy, they may not be suitable for heavy-tailed distributions, such as the Pareto or lognormal distribution for two reasons, citing Ramsay and Usábel (1997):

1. To achieve a reasonable degree of accuracy, the interval of discretization must be at most one unit of the mean of length. If we standardize the unit of currency such that $p_1 = 1$, then to obtain $\Psi(10)$ we must recursively estimate every intermediate unit point $\Psi(u)$ for $k = 0, 1, 2, \ldots, 9, 10$. This may be acceptable if we need only small values of u; however, for large values of u, say $u = 500$ units, this method can be slow. For the Pareto, $\Psi(500)$ is not insignifcant.
2. The quadrature rules inherent in the recursive schemes are usually of low order. This further reduces its accuracy and its rate of convergence. To improve accuracy, the intervals of discretization are made even smaller. This substantially increases the number of intermediate calculations required, making the process of finding $\Psi(u)$ slower.

The above presented problem was partially solved using product integration by Ramsay and Usábel (1997); where it was proved that the convergence of the method was significantly faster than former methods of actuarial literature. Nevertheless, in this method, accuracy is also eventually menaced by increasing values of the initial reserves and the convergence is of order $O(h^2)$ and the true errors are not easily estimated.

Before the shift to recursive methods explained in the last paragraph, the problem of ruin in the Collective Risk Theory had been extensively treated in actuarial literature using integral transforms.

Since the paper by Sparre Andersen in 1955 many authors developed approximations for the ruin probability using Laplace-Stieltjes transforms.

Cramér (1955) used the Winer-Höpf method for the classical case and Thorin (1970, 71, 77) introduced the generalization when epochs of claims form a renewal process. Thorin and Wikstad (1971, 73, 77) used Piessens (1969) inversion method of the Laplace transforms and Bohman (1971, 74, 75), focussed on inversions of Fourier transforms and Seal (1971, 74) dealt with both Laplace and Fourier numerical inversions. Seal (1977) obtained an interesting result for the classical case and exponential claim size distribution using the Bromwich-Mellin inversion formula for Laplace transforms.

Numerical illustrations obtained using this methodology were based on Laplace transforms inversion techniques due to Piessens (1969, 71). These methods can be considered very accurate in the cases contemplated but the theoretical error is not easy to control. In this context, we should mention the very much cited approach, and commonly used in non-life insurance practice in North America, presented by Heckman and Meyers (1983). Other works on ruin probability approximations are Gerber, Goovaerts and Kaas (1987); Ramsay (1992a); Cai and Garrido (1998) and Usábel (1999).

The study of the tail probabilities of the stationary waiting times is the counterpart concept in the context of queueing systems. Many interesting works were presented in this field dealing with long-tailed distributions such as Choudhury, Gupta and Agarwal (1992); Abate, Choudhury and Whitt (1994); Abate, Choudhury and Whitt (1995); Glynn and Whitt (1995). Some of them were specially focussed on the use of Laplace of Fourier transforms as Abate and Whitt (1992, 95, 96) and Choudhury and Whitt (1997). Finally, some interesting asymptotic approximations were produced by Embrechts and Veraverbeke (1982); Willekens and Teugels (1992) and Abate, Choudhury and Whitt (1994).

The study of the Laplace transform of the ultimate ruin probability when claim sizes follow a generalized Γ-*convolution* function is contained in section 2. Some of the most frequently used heavy-tailed distributions in actuarial science belongs to this family. Thorin (1977a) or Berg (1981) proved that Pareto distributions are members of this family; so Thorin (1977 b) did with Log-normal distributions. Other outstanding works related with Γ-*convolution* functions are Thorin (1978) and Goovaerts, D'Hooge and De Pril (1977).

In the present work, a method of inverting Laplace transforms based on the integration between zeros method (along with an acceleration algorithm: a generalization of the Euler method) is introduced when approximating ultimate ruin probabilities in the Classical case of risk theory. We will show that it is specially recommended for large values of u and heavy-tailed distributions.

In sections 3, 4 and 5 we will lay the theoretical foundations to consider the integration between zeros technique an interesting approach when solving integral (3.2) and, subsequently, obtaining interval approximations (lower and upper bounds) for the ultimate ruin probability function when claim sizes are Γ-*convolution* functions. In section 6, the use of the mid-point integration technique or three point trapezoidal rule will be proved to reduce drastically the number of calculations involved. Section 7 is devoted to asymptotic results for large values of the initial reserves u. Finally, numerical examples are presented in section 8, confirming the promised efficiency explained in the theoretical results.

2. The Laplace transform of the ultimate ruin probability for generalized Γ-convolution claim size

Using the Laplace transform on expression (1.1)

$$\Phi^*(s) = \frac{\frac{\theta}{1+\theta}\left(\frac{1}{s}\right)}{1-\frac{1}{(1+\theta)p_1}\left(\frac{1}{s}-F^*(s)\right)}$$

$$= \frac{\frac{\theta}{1+\theta}}{s-\left(\frac{1}{(1+\theta)p_1}\right)+\left(\frac{1}{(1+\theta)p_1}\right)f^*(s)} \tag{2.1}$$

where $f(x)$ is the d.f. of the claim size distribution and,

$$f^*(s) = \int_0^{+\infty} f(x)\, e^{-sx}\, dx = sF^*(s) \tag{2.2}$$

It is obvious that the ruin probability

$$\Psi(u) = 1-\Phi(u)$$

has the following Laplace transform

$$\Psi^*(s) = \frac{1}{s} - \Phi^*(s) \tag{2.3}$$

A distribution function defined on the non-negative real axis is a generalized Γ-*convolution* if its Laplace transform can be written

$$f^*(s) = \int_0^{+\infty} f(x)e^{-sx}\,dx = e^{-as}\, e^{\int_0^{\infty}\ln\left(\frac{1}{1+\left(\frac{s}{y}\right)}\right)dU(y)} \quad Re(s) \geq 0 \tag{2.4}$$

where $a \geq 0$ and $U(y)$ is *nondecreasing* and such that

$$U(0) = 0$$

$$\int_0^1 |\ln(y)|\, dU(y) < \infty$$

$$\int_0^{\infty} \frac{dU(y)}{y} < \infty$$

Some of the most frequently used heavy-tailed distributions in actuarial science belongs to this family. Thorin (1977) or Berg (1981) proved that Pareto distributions are members of this family with parameters $a = 0$ and

$$U'(y) = -\frac{1}{\pi}\frac{q'(y)}{1+(q(y))^2} \quad y > 0$$

and

$$q'(y) = \frac{\Gamma(1+h)}{\pi} e^{gy} g(gx)^{-h-1}$$

where the Pareto c.d.f. with parameters g and h was defined as

$$F(x) = 1 - \left(\frac{x}{g}\right)^{-h} \quad x \geq g$$

Thorin (1977) also proved the same condition for Log-normal c.d.fs. with parameters α and β

$$\Lambda(x) = N\left(\frac{\log(\alpha x)}{\beta}\right) \quad x > 0$$

obtaining that, again in this case, $a = 0$ and

$$U(y) = \frac{1}{\pi} arctg \frac{\operatorname{Im} \lambda^{+}(y)}{\operatorname{Re} \lambda^{+}(y)}$$

where

$$\lambda^{+}(y) = \frac{e^{\pi^2/(2/\beta^2)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x}{\alpha} e^{-\beta u} - \frac{1}{2} u^2 + \frac{I \pi u}{\beta}\right) du$$

and

$$I = \sqrt{-1}$$

The following theorem (proved in the appendix) will be most interesting for future developments,

Theorem 1. For $c > 0$ and $z \geq 0$, the real part of the Laplace transform of the ruin probability for Γ-convolution function claim size, $\operatorname{Re}(\Psi^*(c + zI))$

1. is asymptotically close to 0 with increasing values of z.

$$\lim_{z \to \infty} \operatorname{Re}(\Psi^*(c + zI)) = 0$$

2. is always bound and smooth.

3. The integration between zeros approach

In order to obtain the inverse Laplace transform we can use the Bromwich-Mellin inversion formula,

$$\Psi(u) = \frac{1}{2\pi I} \int_{c-\infty i}^{c+\infty i} e^{su} \Psi^*(s) ds \tag{3.1.}$$

where c is a positive real constant that exceeds the real part of all singularities of $\Psi^*(s)$.

Unfortunately, when $f(x)$ is a generalized Γ-*convolution* function, $\Psi^*(s)$ has no isolated singular points because $f^*(s)$ is not defined for $\mathrm{Re}(s) < 0$ and we proved in Lemma 1 a) that $\Psi^*(s)$ is always bounded for $\mathrm{Re}(s) \geq 0$. This last fact means that we cannot benefit, for instance, from Residues Theorem in order to approximate $\Psi(u)$ using its Laplace transform.

Within the actuarial literature, Seal (1977) proposed the following expression

$$\Psi(u) = \frac{2e^{cu}}{\pi} \int_0^{\infty} \mathrm{Re}(\Psi^*(c + zI))\cos(uz)\,dz \tag{3.2.}$$

and Heckman and Myers (1983) used an alternative formula based on the results by Kendall and Stuart (1977)

$$\Psi(u) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \frac{\mathrm{mod}(\Psi^*(z/\sigma))}{z} \sin\left(\frac{zu}{s} - \arg(\Psi^*(z/\sigma))\right) dz \tag{3.3}$$

where mod and arg are the modulus and argument functions respectively.

It is not an easy task solving (3.2) or (3.3) numerically because the integrand is a rapidly oscillating function. As $z \to \infty$ we will face plus areas and minus areas of nearly equal size and the resulting cancellation of area is attended by a loss of significance, specially when $\lim_{z \to \infty} \mathrm{Re}(\Psi^*(c + zI)) = 0$ (see figure 1).

Heckman and Meyers (1983) already used the integration between zeros in order to approximate integral (3.3) to assess the total claims distributions. Davies and Rabinowitz (1984) also cited the integration between zeros as a valuable method in approximating these integrals. They all argued that it is advantageous to use a rule that employs the values of the integrand at the endpoints of the integration intervals. Since the integrand is zero at these points, more accuracy is obtained without additional computation.

We should, nevertheless, point out that the formula (3.3) used by Heckman and Myers (1983) is not very suitable for fully exploiting the advantages of the integration between zeros to numerically approximate the ultimate ruin probability for Γ-convolutions claim size. The main reasons argued are

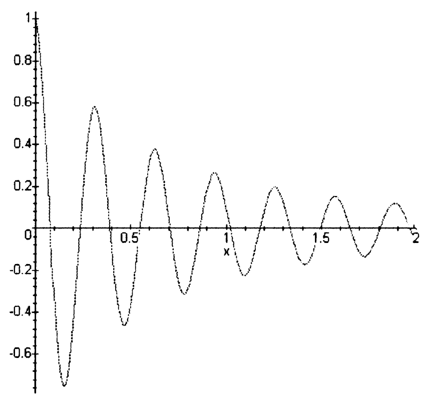

FIGURE 1

1. The formula uses the parameter σ that must be obtained from the second order moment of the claim size distribution. The heaviest-tailed members of the Γ-convolutions claims size family may have no such moment.
2. The basic interval length, $h = 2\pi\sigma\,/\,$(maximum claim amount), cannot be defined for claim size distributions with support in all the positive real axis (see p. 40 of the original work for further details).

Let us now review again this methodology on solving (3.2). It is very easy to prove that

$$\begin{aligned}\Psi(u) &= \frac{2e^{uc}}{\pi}\int_0^{\frac{\pi}{2u}}\mathrm{Re}(\Psi^*(c+zI))\cos(uz)\,dz \\ &\quad+\frac{2e^{uc}}{\pi}\sum_{i=1}^{\infty}\int_{\frac{\pi}{2u}+\frac{(i-1)\pi}{u}}^{\frac{\pi}{2u}+\frac{i\pi}{u}}\mathrm{Re}(\Psi^*(c+zI))\cos(uz)\,dz \\ &= \sum_{i=0}^{\infty}a_i(u)\end{aligned} \tag{3.4}$$

where

$$\begin{aligned}a_0(u) &= \frac{2e^{uc}}{\pi}\int_0^{\frac{\pi}{2u}}\mathrm{Re}(\Psi^*(c+zI))\cos(uz)\,dz \\ a_i(u) &= \frac{2e^{uc}}{\pi}\int_{\frac{\pi}{2u}+\frac{(i-1)\pi}{u}}^{\frac{\pi}{2u}+\frac{i\pi}{u}}\mathrm{Re}(\Psi^*(c+zI))\cos(uz)\,dz \\ i &= 1,2,\ldots\end{aligned} \tag{3.5}$$

Approximating integral (3.2) using the series formula above, as many other numerical techniques, generates two types of errors

1. Discretization error ($\varepsilon^d(u)$).
2. Truncation error ($\varepsilon_n(u)$).

4. THE DISCRETIZATION ERROR ($\varepsilon^d(u)$)

It is obvious that the family of integrals $\{a_i(u)\}_{i=0}^{\infty}$ in most cases, shall be evaluated numerically and the errors must be considered. The discretization total error then will be the sum of all of them. Obviously, the stability of the method fully relies on the numerical technique used when approximating the mentioned integrals and not on other more complicated considerations. Numerical analysis libraries can certainly provide us with stable, fast and accurate approximations algorithms in order to obtain these integrals.

It is very plain to see that the reduction of the dicretization total error will be paid in terms of more evaluations of the function $\mathrm{Re}(\Psi^*(c + zI))$ within the intervals $z \in \left(\frac{\pi}{2u} + \frac{(i-1)\pi}{u}, \frac{\pi}{2u} + \frac{i\pi}{u}\right)$ and the subsequent increase on the computation times.

We should be fully aware that large values of the initial reserves, u, will generate tighter intervals with a positive effect on the discretization error.

5. The truncation error ($\varepsilon_n(u)$)

Using (3.4), it is clear that the sequence $\{a_i(u)\}_{i=0}^{\infty}$ generates the approximations $\{S_n(u)\}$

$$S_n(u) = \sum_{i=0}^{n} a_i(u)$$

It is obvious that the larger the n considered the more accuracy and the more computations required.

Nevertheless, the order of convergence of the initial sequence ought not to be necessarily large and, subsequently, this approach would not be very efficient when seeking for a high precision approximation. As it is suggested by Davies and Rabinowitz (1984), many extrapolations techniques were proved to work effectively in accelerating the convergence of sequences of partial sums of series resulting from the integration of oscillatory integrals. These authors cited the Euler transformation, or the variation introduced by Longman (1956), Richardson's extrapolation, the ε-transformation, iterations of Aitken's Δ^2 method, Levin's V-transformation (Levin (1973) or the work by van de Vooren and van Linde (1966). For a survey of extrapolation processes in numerical analysis see Joyce (1971).

However, it is not proved that the mentioned standard extrapolation techniques applied to the former sequences will yield upper and lower limits. Actually, most of them cannot guarantee this fact. Moreover, these techniques, although useful in some cases, are not proved to accelerate the convergence in many others.

We will now show how the convergence can be substantially accelerated using a simple generalization of the Euler extrapolation technique obtaining upper and lower bounds sequences and decreasing errors.

With the following theorem (proved in the appendix), we will now show why integration between zeros, used on formula (3.2), is most interesting when approximating ultimate ruin probabilities for generalized Γ-*convolution* claim sizes.

Theorem 2. When the ultimate ruin probability can be expressed as the alternating series

$$\Psi(u) = \sum_{i=0}^{\infty} (-1)^i a_i(u)$$

where $\{a_i(u)\}_{i=0}^{\infty}$ is a bound and smooth sequence for which $\lim_{i\to\infty} a_i(u) = 0$, the following equality holds

$$\Psi(u) = A^j(u) + \left(\frac{(-1)^j}{2^j}\right) \sum_{i=0}^{\infty} (-1)^i \Delta_i^j(u) \quad j = 1, 2, \ldots \tag{5.1}$$

where

$$A^j(u) = \sum_{i=0}^{j-1} (-1)^i \frac{\Delta_0^i(u)}{2^{i+1}}$$

expressing the ruin probability as an alternating series of forward-differences of arbitrary order j, $\left\{\Delta_i^j(u)\right\}_{i=0}^{\infty}$ also bound and smooth and $\lim_{i\to\infty}\Delta_i^j(u)=0$

Remark 1. The forward-differences are defined as

$$\Delta_i^j(u)=\sum_{k=0}^{j}(-1)^k\binom{j}{k}a_{i+j-k}=\Delta_{i+1}^{j-1}(u)-\Delta_i^{j-1}(u)\quad j>0$$
$$\Delta_i^0(u)=a_i(u),$$

Remark 2. The reader has probably realized that the terms $A^j(u)$ presented in the former Theorem are the successive approximations of the Euler acceleration technique, see for instance p. 131 of Press et al. (1986); just the very first term in the context of the approximations presented in this Theorem.

As a corollary of the former Theorem, we can consider the following family of approximations truncating (5.1), with explicit error terms

$$\begin{aligned}\Psi(u) &= S_n^j(u)+\varepsilon_n^j(u)\\ \Psi(u) &\simeq S_n^j(u)=A^j(u)+\left(\frac{(-1)^j}{2^j}\right)\sum_{i=0}^{n}(-1)^i\Delta_i^j(u)\\ n &= 0,1,2,\ldots\quad j=1,2,\ldots\end{aligned}$$

It is easy to prove that this family of approximations can be obtained using the following alternative recursive formula

$$S_n^j(u)=\frac{S_n^{j-1}(u)+S_{n+1}^{j-1}(u)}{2};\quad S_n^0(u)\equiv S_n(u)=\sum_{i=0}^{n}a_i(u) \tag{5.2}$$

with a very positive effect on the loss of significant digits due to the substraction of small quantities compared with the former expression in which the forward-differences are directly involved. The formula used by Abate and Whitt (1995)

$$E(m,n,u)=\sum_{k=0}^{m}\binom{m}{k}2^{-m}S_{m+k}(u)$$

can be considered less stable due to the increasing magnitude of the factorials involved in the calculations.

It is then obvious that the error can be expressed as an alternating series with bound and smooth terms and $\lim_{i\to\infty}\Delta_i^j(u)=0$

$$\varepsilon_n^j(u)=\left(\frac{(-1)^j}{2^j}\right)\sum_{i=n+1}^{\infty}(-1)^i\Delta_i^j(u)$$

The error magnitude will be decreasing with j because

$$\Delta_i^j(u) = h^j f^{j)}(x_i) + O(h_j) \tag{5.3}$$

where $f(x)$ is the original function which the original terms $a_i(u)$ come from and $x_i = x_0 + hi$. We should remember that in the context of integration between zeros, see the former section,

$$h \backsimeq \frac{\pi}{u}$$

reducing significantly the magnitude of the forward-differences, and subsequently the error terms as j increases, when u is large.

Remark 3. The value of the forward-differences, using the alternative formula (5.2), is obtained with the expression

$$\left(\frac{(-1)^j}{2^j}\right)(-1)^n \Delta_n^j(u) = S_n^j(u) - S_{n-1}^j(u)$$

In order to offer an estimation of the error, we can apply the same acceleration technique in the error term series and obtain

$$\varepsilon_n^j(u) = \left(\frac{(-1)^{j+n+1}}{2^j}\right)\sum_{i=0}^{\infty}(-1)^i \frac{\Delta_{n+1}^{j+i}(u)}{2^{i+1}} \simeq \left(\frac{(-1)^{j+n+1}}{2^j}\right)\sum_{i=0}^{m}(-1)^i \frac{\Delta_{n+1}^{j+i}(u)}{2^{i+1}} \tag{5.4}$$

Remark 4. This last formula offers a substantial improvement in the error estimation stated by Abate and Whitt (1995) in reference of Hosono (1984), where the following expression was used to assess the error

$$\varepsilon_n^j(u) \simeq S_{n+1}^j(u) - S_n^j(u) = \frac{(-1)^{j+n+1}\Delta_{n+1}^j(u)}{2^j},$$

the simplest case in our approach, when $m = 1$.
Moreover, because of (5.3)

$$\left|\Delta_{n+1}^j(u)\right| > \left|\Delta_{n+1}^{j+1}(u)\right|$$

$$sign\left(\varepsilon_n^j(u)\right) \simeq (-1)^{j+n+1}\Delta_{n+1}^j(u)$$

and it is easy to deduce that the sign of the error terms, $\varepsilon_n^j(u)$ will be alternating for two consecutive values $(n, n+1)$ when

$$sign\left(\Delta_{n+1}^j(u)\right) = sign\left(\Delta_{n+2}^j(u)\right)$$

producing a sequence of upper and lower bounds for the ultimate ruin probability. One should bear in mind that this last condition is observed in most cases.

6. The mid-point integration algorithm

Let us now present a way to control the discretization total error avoiding a massive number of evaluations of the function $\mathrm{Re}(\Psi^*(c+zi))\cos(uz)$ within the intervals $z \in \left(\frac{\pi}{2u} + \frac{(k-1)\pi}{u}, \frac{\pi}{2u} + \frac{\kappa\pi}{u}\right)$ for $k = 1, 2, \ldots$

The use of the mid-point integration rule with step size $h = \pi / 2u$ and parameter $c = A / 2u$ on the family of integrals $\{a_k(u)\}_{k=0}^{\infty}$ yields the expression

$$\Psi(u) = \frac{e^{A/2}}{2u}\mathrm{Re}\left(\Psi^*\left(\frac{A}{2u}\right)\right) \tag{6.1}$$

$$+\frac{e^{A/2}}{u}\sum_{j=1}^{\infty}(-1)^j\mathrm{Re}\left(\Psi^*\left(\frac{A}{2u}+\left(\frac{j\pi}{u}\right)i\right)\right) \tag{6.2}$$

Remark 5. The reader can easily realize that in this context the mid-point integretion is exactly the same as the three-point trapezoidal rule,
and assign

$$a_0(u) = \frac{e^{A/2}}{2u}\mathrm{Re}\left(\Psi^*\left(\frac{A}{2u}\right)\right)$$

$$a_i(u) = \frac{e^{A/2}}{u}\mathrm{Re}\left(\Psi^*\left(\frac{A}{2u}+\left(\frac{j\pi}{u}\right)i\right)\right) \quad i = 1,2,\ldots$$

This approach is similar to the one stated in Abate and Whitt (1995), in reference of Dubner and Abate (1968), but obtained in the context of the Fourier-series method of Laplace transform inversion. Using this integration method, we only need the mid-point of each and every interval to approximate the family $\{a_i(u)\}_{i=0}^{\infty}$ with a dramatic reduction in the number of evaluations of the Laplace transform.

Moreover, Abate and Whitt (1995) also proved that the discretization error of the former expression can be controlled with the parameter A and the formula

$$\left|e^d(u)\right| \le \frac{e^{-A}}{1-e^{-A}} \simeq e^{-A} \quad (\text{for } e^{-A} \text{ small})$$

This last fact means, as stated by the mentioned authors, that in order to have at most $10^{-\gamma}$ discretization error $A = \gamma \log(10)$.

One should consider that the larger the value of A (and smaller truncation error) the more accuracy is needed in the calculations of $\left\{\mathrm{Re}\left(\Psi^*\left(\frac{A+2i\pi I}{2t}\right)\right)\right\}_{i=0}^{\infty}$ because, to obtain the final approximation, the common factor $\frac{e^{A/2}}{u}$ will be used in (6.1). So the trade-off between significant digits used and discretization error becomes clear.

Unfortunately, the mentioned trade-off is not a new fact at all in numerical approximations; for a very clear example of this fact on approximating ruin probabilities see Usábel (1999). As a consequence, although much faster, when we use the mid-point integration, the algorithm can become unstable when searching for a high degree of accuracy and the significant digits are not upgraded.

Once more, dealing with large initial reserves, *u*, will be positive in the context of this methodology because of the above cited factor $\frac{e^{A/2}}{u}$, avoiding the need of increasing the significant figures to perform the calculations.

7. Asymptotic Formulas

One of the main advantages of the method presented in this work is that it is not negatively affected by the size of the initial reserves considered as happened to be with other methods previously used in actuarial literature (see the introduction). For large values of u most of the methods became either unstable or of a very slow convergence.

We cannot conclude this work without a mention to the main asymptotic approximations for ultimate ruin probability or the tail probabilities of the stationary waiting times, as its counterpart in queueing theory.

Considering a heavy-tailed service time or claim size distribution leads to the search of special formulae designed for these subexponential distributions. In the context of risk theory, Embrechts and Veraverbeke (1982) produced the formula

$$\Psi(u) \sim \frac{\int_u^\infty (1-F(y))dy}{\theta p_1} \quad u \to \infty$$

Later Willekens and Teugels (1992) found a generalization of the above result including more terms in the final formula in the context of M/ G/1 queues

$$\Psi(u) \sim \frac{\int_u^\infty (1-F(y))dy}{\theta p_1} + \left(\frac{(p_1)^2(1-\theta)}{p_2\theta^2}\right)(1-F(u))$$
$$+ \left(\frac{3\,(p_1)^3(1-\theta)^2}{4\,(p_2)^2\theta^3} + \frac{(p_1)^2(1-q)}{3p_3\theta^2}\right) f(u) \tag{7.1}$$

where

$$p_k = \int_0^\infty x^k f(x)dx$$

Unfortunately, the three terms approximations cannot be used for the heaviest-tailed members of the family of the Γ-convolution claim sizes because the moments are not defined (see table 1 and 2) and the approximations in the numerical illustrations considered later are not so good as expected for very large initial reserves u.

Abate, Choudhury and Whitt (1994) also obtained an special formula for Pareto mixture of exponentials service time distributions.

8. Numerical Illustration

We will consider as an illustration, Pareto claim sizes with c.d.f.

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\lambda+1} \quad x \geq 0 \quad \lambda > 0 \text{ (integer)}$$

as one of the heaviest tailed members of the family of Γ-convolutions claim size. The Laplace transforms of the example considered can be expressed in terms of the exponential integral (see for example Gradshteyn and Ryzhik (1994) formula 3.353.2).

We will find approximations for the rapidly oscillatory integral (3.2), the Bromwich-Mellin inversion formula, using the simple recursive expression (5.2) and the mid-point integration for the integrals (3.5)

$$\begin{aligned} S_n^j(u) &= \frac{S_n^{j-1}(u) + S_{n+1}^{j-1}(u)}{2}; \quad S_n^0(u) \equiv S_n(u) = \sum_{i=0}^{n} a_i(u) \\ a_0(u) &= \frac{e^{A/2}}{2u} \mathrm{Re}\left(\Psi^*\left(\frac{A}{2u}\right)\right) \\ a_i(u) &= \frac{e^{A/2}}{u} \mathrm{Re}\left(\Psi^*\left(\frac{A}{2u} + \left(\frac{j\pi}{u}\right) I\right)\right) \quad i = 1, 2, \ldots \end{aligned}$$

The value of the parameter A was set so that the approximations had a discretization error of at most 15 significant digits (A = 15 log(10)), see section 6

TABLE 1

Pareto claim size λ = 1
Ultimate ruin probability intervals A = 15 log(10)

θ	u	$\left(S_9^{20}(u), S_{10}^{20}(u), S_{11}^{20}(u)\right)$	Asymptotic (one term)
0.1	1	(8.50144942, 8.50144943) 10^{-1}	
	10	(6.271279490, 6.27179501) 10^{-1}	
	100	(1.64859138, 1.64859141) 10^{-1}	
	1,000	(1.13443368, 1.13443373) 10^{-2}	
	10,000	(1.016661353, 1.016661386) 10^{-3}	9.9990001 10^{-4}
	100,000	(1.00209834, 1.00209837) 10^{-4}	9.9999000 10^{-5}
	1,000,000	(1.0002553, 1.0002559) 10^{-5}	9.9999900 10^{-6}
0.25	1	(6.909906847, 6.909906853) 10^{-1}	
	10	(3.726769676, 3.726769680) 10^{-1}	
	100	(5.22265530, 5.22265551) 10^{-2}	
	1,000	(4.1948538, 4.1948539) 10^{-3}	
	10,000	(4.0260816, 4.0260817) 10^{-4}	3.999600 10^{-4}
	100,000	(4.00332776, 4.00332778) 10–5	3.999960 10^{-5}
	1,000,000	(4.00040606, 4.00040606) 10^{-6}	3.999996 10^{-6}

for details. Note that in Table 2 this value is upgraded to 201n(10) because the probability is very small.

The lower and upper bounds for the ultimate ruin probability will be spotted using the simple rule stated in the last paragraph of section 5. The approximations considered were $\left(S_9^{20}(u), S_{10}^{20}(u), S_{11}^{20}(u)\right)$ so that the maximum number of evaluations of the function $\mathrm{Re}(\Psi^*(c + zI))$ is just 31)!

The asymptotic approximations are based on the formula by Willekens and Teugels (1992), expression (7.1), considering the maximum number of terms possible depending on the value of λ.

TABLE 2

PARETO CLAIM SIZE $\lambda = 2$
ULTIMATE RUIN PROBABILITY INTERVALS A = 15 LOG(10)

θ	u	$\left(S_9^{20}(u), S_{10}^{20}(u), S_{11}^{20}(u)\right)$	Asymptotic (two terms)
0.1	1	(8.41831695, 8.41831696) 10^{-1}	
	10	(5.22719526, 5.22719527) 10^{-1}	
	100	(1.8279697, 1.8279700) 10^{-2}	
	1,000	(4.3448088, 4.3448093) 10^{-5}	
	10,000	(4.0308031, 4.0308034) 10^{-7}	4.00019940 10^{-7}
	100,000	(4.0030442, 4.0030445) 10^{-9}	4.00001999 10^{-9}
	1,000,000	*(4.00030, 4.00036) 10^{-11}	4.00000200 10^{-11}
0.25	1	(6.760398370, 6.760398375) 10^{-1}	
	10	(2.522264643, 2.522264644) 10^{-1}	
	100	(2.4590058, 2.4590063) 10^{-3}	
	1,000	(1.6478781, 1.6478783) 10^{-5}	
	10,000	(1.645162, 1.6045163) 10^{-7}	1.5996000 10^{-7}
	100,000	(1.6004484, 1.6004485) 10^{-9}	1.5999600 10^{-9}
	1,000,000	*(1.600035, 1.600060) 10^{-11}	1.5999960 10^{-11}

Remark 6. The intervals with the asterisk (*) were produced using $A = 20\ln(10)$ because of their small magnitude. An increase in the relative amplitude of the intervals in then observed.

It is very important to mention that the relative amplitude of the intervals (the relative error) is not quite affected when considering very large figures for the initial reserves, except in the cases mentioned in the remark above due to the change in the parameter A. All calculations were programmed in Maple V, release 4 using 22 significant digits.

9. CONCLUDING COMMENTS

The method presented in this work to approximate the ultimate ruin probability for Γ-convolutions claims size is specially recommended for large values of the initial reserves, u. As it is highlighted in sections 4, 5 and 6, considering even huge values for the initial reserves will not endanger either the accuracy or efficiency of this algorithm.

The general method of inverting Laplace transforms of tail probability distributions presented by Abate and Whitt (1995) is revisited in the context of the integration between zeros with three added improvements (see section 5 for details):

1. A better error estimation
2. The generation of upper and lower bounds.
3. A more stable recursive formula.

Consequently, the approach contemplated in this work is granted with the main advantages of the resursive methods based on discretization (see section 1) but not with their main drawback: the accuracy and efficiency is menaced considering large initial reserves. The reader should consider the accuracy obtained with just 31 evaluations of the Laplace transform and initial reserves u = 1,000,000 and calculations performed with 22 significant figures.

On the other side, the classical works of actuarial literature, mainly developed by the Scandinavian School in the 70's (see section 1), based their numerical illustrations on the Piessens' algorithms of inverting Laplace transforms (Piessens (1969, 71). In the survey of numerical methods for inverting the Laplace transform by Davis and Martin (1979), a clear conclusion can be reached: the error of the approximations was not easy to control.

Methods considered very good in this study, for instance Piessens and Branders (1971), can lead to poor results when dealing with increasing values of the initial reserves u. Seal (1975) already claimed that Laguerre series cannot be recommended as a practical method of numerical inversion of Laplace transform and the result can also affect to any other orthogonal polynomials. Piessens (1971, section 7, comment 2) also cited the limitation of Gaussian quadrature based methods for increasing accuracy demands. The method presented is a good alternative to the Gaver-Stehfest algorithm of inverting Laplace transforms, see Usábel (1999), because the demands of significant figures in the calculations are, by far, less restrictive. However, complex numbers are involved in the calculations.

Finally, the asymptotic formulae presented in section 8 did not yield, in the examples considered (see Table 1 and 2), so good approximations as expected for values of the initial reseves of magnitude u = 1,000,000.

The algorithm was originally designed for Γ-*convolution* functions claim sizes because this family includes some of the most famous heavy-tailed distributions used in actuarial works, i.e. Pareto and log-normal. Nevertheless, the approach can be extended to any other claim size d.f. when

$$\lim_{z\to\infty} \mathrm{Re}(f^*(c+zi)) = 0 \quad c \geq 0$$

that is usually observed in density functions with not very restrictive smoothness conditions as it is proved in Abate and Whitt (1995).

APPENDIX

Proof of Theorem 1

Let us start proving that the limit for the ruin probability function for generalized Γ-convolutions claim size gets asymptotically close to 0

$$\lim_{z\to\infty} \mathrm{Re}(\Psi^*(c+zI)) = 0$$

Abate and Whitt (1995) already showed that, under some restrictive smoothness conditions, a complementary c.d.f. fulfills

$$\mathrm{Re}(\Psi^*(c+zI)) = \frac{c-\Psi'(0)}{z^2} + o(z^{-2})$$

and obviously the asymptotic limit of the expression above.

However, we wanted to prove the same statement straight from the definition of Γ-convolutions claim size in order to offer a more complete proof for this type of distributions regardless extra considerations on smoothness. The following lemma shows the asymptotic behaviour for the Laplace transform of the d.f. of the Γ-*convolution* claim size distributions

Lemma 3. The Laplace transform of the density function of Γ-convolution function can be expressed

$$f^*(c+zI) = R(c,z)\cos(\Theta(c,z)) + IR(c,z)\sin(\Theta(c,z)) \tag{9.1}$$

where

$$\begin{aligned} R(c,z) &= e^{-ac} e^{-\frac{1}{2}\int_0^\infty \ln\left(1+2\left(\frac{c}{y}\right)+\left(\frac{c}{y}\right)^2+\left(\frac{z}{y}\right)^2\right) dU(y)} \\ \Theta(c,z) &= -az - \int_0^\infty \arctan\left(\frac{z}{y+c}\right) dU(y) \end{aligned} \tag{9.2}$$

modulus and argument respectively, and

$$\lim_{z\to\infty} f^*(c+zi) = 0 \quad c \geq 0 \tag{9.3}$$

Proof. Let us expand the following natural logarithm

$$\begin{aligned} &\ln\left(\frac{y}{y+c+zI}\right) = \ln(y) - \ln((y+c)+zI) \\ &= \ln(y) - \ln\left(\sqrt{(y+c)^2+z^2}\right) - \arctan\left(\frac{z}{y+c}\right) I \end{aligned}$$

and

$$\begin{aligned} e^{\int_0^\infty \ln\left(\frac{1}{1+\frac{s}{y}}\right) dU(y)} &= e^{\int_0^\infty \ln\left(\frac{y}{y+s}\right) dU(y)} \\ &= \left(e^{-\frac{1}{2}\int_0^\infty \ln(1+2(\frac{c}{y})+(\frac{c}{y})^2+(\frac{z}{y})^2)\, dU(y)}\right) \\ &\quad e^{I\left(-\int_0^\infty \arctan\left(\frac{z}{y+c}\right) dU(y)\right)} \end{aligned}$$

Using the definition (2.4)

$$\begin{aligned} f^*(s) &= e^{-as} e^{\int_0^\infty \ln\left(\frac{1}{1+\left(\frac{s}{y}\right)}\right) dU(y)} \\ &= \left(e^{-ac-\frac{1}{2}\int_0^\infty \ln(1+2(\frac{c}{y})+(\frac{c}{y})^2+(\frac{z}{y})^2)\, dU(y)}\right) \\ &\quad e^{I\left(-az-\int_0^\infty \arctan\left(\frac{z}{y+c}\right) dU(y)\right)} \\ &= R(c,z)\cos(\Theta(c,z)) + IR(c,z)\sin(\Theta(c,z)) \end{aligned}$$

The last statement (9.3) follows from the limit

$$\lim_{z\to\infty} \ln\left(1+2\left(\frac{c}{y}\right)+\left(\frac{c}{y}\right)^2+\left(\frac{z}{y}\right)^2\right) = \infty$$

and

$$\lim_{z\to\infty}\int_0^\infty \ln\left(1+2\left(\frac{c}{y}\right)+\left(\frac{c}{y}\right)^2+\left(\frac{z}{y}\right)^2\right) dU(y) = \infty$$

where $U(y)$ is nondecreasing ($U'(y) \geq 0$) using the definition of generalized Γ-*convolution* functions. Finally, using (9.2) and (9.1)

$$\lim_{z\to\infty} R(c,z) = 0 \Rightarrow \lim_{z\to\infty} f^*(c+zI) = 0 \qquad \blacksquare$$

Let us now finally prove that, for Γ-*convolution* function claim size, the Laplace transform of the ruin probability function will be asymptotically close to zero as z increases in any event.

If we expand expression (2.3) using the results previously obtained from the former lemma (9.3)

$$\Psi^*(c+zI) = \frac{1}{c+zI} - \frac{\left(\frac{\theta}{1+\theta}\right)}{(c+zI)-\left(\frac{1}{(1+\theta)p_1}\right)}$$

and just focusing on the real part

$$\begin{aligned} &\lim_{z\to\infty} \mathrm{Re}(\Psi^*(c+zI)) \\ &= \lim_{z\to\infty} \frac{c}{c^2+z^2} - \frac{\theta((1+\theta)p_1 c-1)}{((1+\theta)p_1 c-1)^2+((1+\theta)p_1 z)^2} = 0 \end{aligned}$$

proving statement 1 of Theorem 1.

The real part of the Laplace transform of the ruin probability function

$$\begin{aligned}\Psi^*(c+zI) &= \int_0^\infty e^{-(c+zI)x}\Psi(x)dx \\ &= \int_0^\infty e^{-cx}\cos(-zx)\Psi(x)dx \\ &\quad + \left(\int_0^\infty e^{-cx}\sin(-zx)\Psi(x)dx\right)I\end{aligned}$$

can be bound using the absolute value convergence Theorem for integrals

$$\int_0^\infty \left|e^{-cx}\cos(-zx)\Psi(x)\right|dx \leq \int_0^\infty e^{-cx}\Psi(x)dx$$

The last integral will be always convergent for non-negative c because $\Psi(x)$ is a decreasing function and

$$\lim_{x\to\infty}\Psi(x) = 0$$

and it is clear that we cannot find any singularities of $\Psi^*(s)$ when c is non-negative, proving that it is bound and smooth.

Proof of Theorem 2

If the initial sequence $\{a_i(u)\}_{i=0}^{\infty}$ is bound and smooth and $\lim_{i\to\infty} a_i = 0$, the forward-differences defined as

$$\begin{aligned}\Delta_i^j(u) &= \sum_{k=0}^{j}(-1)^k\binom{j}{k}a_{i+j-k}(u) = \Delta_{i+1}^{j-1}(u) - \Delta_i^{j-1}(u) \quad j>0 \\ \Delta_i^0(u) &= a_i(u)\end{aligned}$$

will be also bound and smooth and decreasing with i, $\lim_{i\to\infty}\Delta_i^j(u) = 0$.

Remark 7. the simplified notation a_i and Δ_i^j will be used henceforth. The alternating initial series can be easily modified into

$$\begin{aligned}\Psi(u) &= \sum_{i=0}^{\infty}(-1)^i a_i \\ &= \frac{a_0}{2} - \frac{(a_1 - a_0)}{2} + \frac{(a_2 - a_1)}{2} - \dots \\ &= \frac{a_0}{2} - \frac{\Delta_0^1}{2} + \frac{\Delta_1^1}{2} - \dots \\ &= \frac{a_0}{2} - \sum_{i=0}^{\infty}(-1)^i\frac{\Delta_i^1}{2}\end{aligned}$$

and the same procedure can be applied to the new alternating series obtained $\sum_{i=0}^{\infty}(-1)^i \frac{\Delta_i^1}{2}$. If we continue applying the same easy transform recursively to the new alternating series in terms of the forward-differences the expression (5.1) results.

Acknowledgments

I am indebted to the anonymous referees of this paper for their useful comments, specially on bibliography, that lead to a substantial improvement of the previous version of this work.

References

Abate, J. and Whitt, W. (1992) The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems* **10**, 5-88.

Abate, and Whitt, W. (1995) Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing* **7**, No. 1, 36-43.

Abate, and Whitt, W. (1995) An operational calculus for probability distributions via Laplace Transforms. *Advances in Applied Probability,* **28**, 75-113.

Abate, J., Choudhury, G. and Whitt, W. (1994) Waiting-time tail probabilities in queues with long-tail service-time distribution. *Queueing Systems*, **16**, 311-338.

Abate, J., Choudhury, G. and Whitt, W. (1995) Calculating the M/G/1 busy period density and LIFO waiting-time distribution by direct numerical transform inversion. *Operations Research Letters* **18**, 113-119.

Abramowitz, M. and Stegun, I.A. (1972) *Handbook of Mathematical Functions*. Ninth Edition. New York, N.Y. Dover Publications.

Berg, C. (1981) The Pareto distribution is a generalized Γ-convolution – a new proof. *Scan. Act. Journal*, 117-119.

Boas, R. (1987) *Invitation to complex analysis.* The Random House/Birkhäuser mathematics series. New York.

Bohman, H. (1971) Ruin probabilities. *Skandinavisk Aktuarietidskrift*, 159-163.

Bohman, H. (1974) Fourier Inversion-Distribution functions-Long tails. *Scand. Actuarial Journal*, 43-45.

Bohman, H. (1975) Numerical inversion of characteristic functions. *Scan. Act. Journal*, 121-124.

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbith, C.J. (1997) Actuarial Mathematics. *Society of Actuaries*.

Cai, J. and Garrido, J. (1998) Aging properties and bounds for ruin probabilities and stop-loss premiums. *Insurance: Mathematics and Economics* **23**, Number 1, 33-44.

Choudhury, M.L., Gupt, U.C., Agarwal, M. (1992) Exact and approximate numerical solutions to steady-state single-server queues: M/G/1 – a unified approach. *Queueing Systems* **10**, 351-380.

Coudhury, M.L. and Whitt, W. (1997) Scaling for Numerical Transform Inversion to analyze Teletraffic Models. *ITC* **15**, 933-942.

Conway, J. (1978) *Functions of one complex variable I.* Second edition. Springer-Verlag. New York.

Cramèr, H. (1955) *Collective Risk Theory*. Jubille Volume of F. Skandia.

Davies, B. and Martin, B. (1979) Numerical inversion of the Laplace transform: a survey and comparison of methods. *Journal of computational physics* **33**, 1-32.

Davies, J.D. and Rabinowitz, P. (1984) *Methods of numerical Integration*. Second edition. Academic Press Inc. Boston.

DICKSON, D.C.M. (1989) Recursive calculation of the probability and severity of ruin. *Insurance: Mathematics and Economics* **8**, 145-148.

DICKSON, D.C.M., EGIDIO DOS REIS, A.D., WATERS, H.R. (1995) Some Stable Algorithms in Ruin Theory and their Application. *ASTIN Bulletin* **25**, 153-175.

DICKSON, D.C.M. and WATERS, H.R. (1991) Recursive calculation of survival probabilities. *ASTIN Bulletin* **21**, 199-221.

EMBRECHTS, P. and VERAVERBEKE, N. (1982) Estimates for Probability of Ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics* **1**, 55-72.

GERBER, H.U. (1979) *An Introduction to Mathematical Risk Theory*. Huebner Foundation Monograph 8. Philadelphia, Pa., University of Pennsylvania.

GERBER, H. GOOVAERTS, M., KAAS, R. (1987) On the probability and severity of ruin. *ASTIN Bulletin* **17**, n. 2, 151-163.

GLYNN, P. and WHITT, W. (1995) Heavy-traffic extreme-value limits for queues. *Operations Research Letters* **18**, 107-111.

GOOVAERTS, M. D'HOOGE, L., DE PRIL, N. (1977) On a class of generalized Γ-convolutions. *Scan. Act. Journal*, 21-30.

GOOVAERTS, M. and DE VYLDER, F. (1984) A Stable Recursive Algorithm for Evaluation of Ultimate Ruin Probabilities. *ASTIN Bulletin* **14**, 53-59.

GRADSHTEYN, I.S. and RYZHIK, I.M. (1994) *Table of Integrals, Series and Products*. Fifth edition. Academic Press.

HECKMAN, P. and MEYERS, G. (1983) The calculation of aggregate loss distributions from claim severity and claim count distributions. *Proceedings of the Casuality Actuarial Society* **LXX**, 22-61.

HOSONO, T. (1984) *Fast Inversion of the Laplace Transform by Basic.* Kyoritsu Publishers, Japan.

JOYCE, D.C. (1971) Survey of extraolation processes in numerical analysis. *SI AM Rev.* **13**, 435-490.

KENDALL, M. and STUART, A. (1977) *The Advanced Theory of Statistics.* Vol. I, 4th Edition. MacMillan.

LEVIN, D. (1973) Development of non-linear transformations for improving convergence of sequences. *Internat. J. Comput. Math.* **3**, 371-388.

LONGMAN, I.M. (1956) Note on a method for computing infinite integrals of oscillatory functions. *Proc. Cambridge Philos. Soc.* **52**, 764-768.

PANJER, H.H. (1986) Direct Calculation of Ruin Probabilities. *Journal of Risk and Insurance* **53**, 521-529.

PANJER, H.H., WANG, S. (1993) On the Stability of recursive Formulas. *ASTIN Bulletin* **23**, 227-258.

PANJER, H.H., WILLMOT, G.E. (1992) *Insurance risk Models*. Society of Actuaries, Schaumburg.

PIESSENS, R. (1969) New quadrature formulas for the numerical inversion of Laplace transforms. *BIT 9*, 351-361.

PIESSENS, R. (1971) Gaussian Quadrature Formulas for the Numerical Integration of Bromwich's integral and the Inversion of the Laplace Transform. *Journal of Engineering Math.* **5**, 1-5.

PIESSENS R. and BRANDERS, M. (1971) Numerical Inversion of the Laplace Transform using generalised Laguerre Polynomials. *Proceedings of the IEE* **118**, NO. 10.

RAMSAY, C.M. (1992a) A Practical Algorithm for Approximating the Probability of Ruin. *Transactions of the Society of Actuaries* **XLIV**, 443-59.

RAMSAY, C.M. (1992b) Improving Goovaerts' and De Vylder's Stable Recursive Algorithm. *ASTIN Bulletin* **22**, 51-59.

RAMSAY, C.M. and USÁBEL, M.A. (1997). Calculating Ruin probabilities via Product integration. *ASTIN Bulletin* **27**, n. 2, 263-271.

SEAL, H. (1971) Numerical calculation of the Bohman-Esscher family of convolution-mixed negative binomial distribution functions. *Mitt. Verein. schweiz. Versich.-Mathr.* **71**, 71-94.

SEAL, H. (1974) The numerical calculation of $U(w,t)$, the probability of non-ruin in an interval $(0,t)$. *Scan. Act. Journal*, 121-139.

SEAL H.L. (1975) A note on the use of Laguerre polynomials in the inversion of Laplace transforms. *Blätter der DAVM* **12**, 131-134.

SEAL, H. (1977) Numerical inversion of characteristic functions. *Scan. Act. Journal*, 48-53.
THORIN, O. (1970) Some remarks on the ruin problem in case the epochs of claims form a renewal process. *Skandinavisk Aktuarietidskrift*, 29-50.
THORIN, O. (1971) Further remarks on the ruin problem in case the epochs of claims form a renewal process. *Skandinavisk Aktuarietidskrift*, **14-38**, 121-142.
THORIN, O. (1973) The ruin problem in case the tail of a distribution is completely monotone. *Skandinavisk Aktuarietidskrift*, 100-119.
THORIN, O. (1977) Ruin probabilities prepared for numerical calculation. *Scandinavian Actuarial Journal*.
THORIN, O. (1977) On the infinite divisibility of the Pareto distribution. *Scan. Act. Journal*, 31-40.
THORIN, O. (1977) On the infinite divisibility of the Lognormal distribution. *Scan. Act. Journal*, 121-148.
THORIN, O. (1978) An extension of the notion of a generalized Γ-convolution. *Scan. Act. Journal*, 141-149.
THORIN, O. and WIKSTAD, N. (1973) Numerical evaluation of ruin probabilities for a finite period. *ASTIN Bulletin* **VII:2**, 138-153.
USÁBEL, M. (1999) Calculating multivariate ruin probabilities via Gaver-Stehfest inversion technique. *IME* **25**, 133-142.
VAN DE VOOREN, A.I. and VAN LINDE, H.J. (1966) Numerical calculation of integrals with strongly oscillating integrand. *Math. Comp.*, 232-245.
WIKSTAD, N. (1971) Exemplifications of ruin probabilities. *ASTIN Bulletin* **6**, part 2.
WIKSTAD, N. (1977) How to calculate Ruin probabilities according to the classical Risk Theory. *Scand. Actuarial Journal*.
WILLEKENS, E. and TEUGELS, J.L. (1992) Asymptotic expansions for waiting time probabilities in a M/G/1 with long-tailed service time. *Queueing Systems* **10**, 295-312.

Miguel A. USÁBEL
Universidad Carlos III de Madrid
Avda. Universidad Carlos, 22
Colmenarejo 28270 (Madrid)
Spain
E-mail: usabel@emp.uc3m.es

EXPERIENCE RATING SCHEMES FOR FLEETS OF VEHICLES*

By

Denise Desjardins[1], Georges Dionne[1,2], Jean Pinquet[3]

Abstract

This paper proposes bonus-malus systems for fleets of vehicles, by using the individual characteristics of both the vehicles and the carriers. Bonus-malus coefficients are computed from the history of claims or from the history of safety offences of the carriers and the drivers. The empirical results are derived from a data set obtained from the Société de l'Assurance Automobile du Québec, the public insurer for bodily injuries and the regulator of road safety.

Keywords

Stratified portfolios, credibility, vehicle, fleet, accidents, safety offences.

1. Introduction

This paper stems from a study carried out for the Société de l'Assurance Automobile du Québec, later referred as the SAAQ (see also Dionne, Desjardins, Pinquet (1999, 2000a)). Its objective is to provide Bonus-Malus Systems (later referred to as BMS) for fleets of vehicles from the history of claims or from that of safety offences.

Fleets of vehicles are owned by firms, which are commercial motor carriers in the SAAQ portfolio. A portfolio of insurance contracts subscribed by firms has a stratified structure, and the size of the stratum (the set of policies held by

* The paper benefited from useful comments of two anonymous referees. This research was funded by the programme de recherche universitaire en sécurité routière of the Ministère des Transports du Québec (MTQ) and the Société de l'Assurance Automobile du Québec. The authors also acknowledge financial support from the Fédération Française des Sociétés d'Assurances (FFSA) and the FCAR in Quebec. They remain responsible for the errors, if any. A first version was presented at two research meetings of the FFSA and at the Risk Theory Seminar of the American Risk and Insurance Association.

[1] Centre for Research on Transportation, University of Montreal.
[2] Risk Management Chair, HEC-Montreal.
[3] (Corresponding Author) University Paris X.

a given firm) is a key variable in risk analysis. The propensity to self-insurance increases with the size of the stratum. Insurance contracts for fleets of vehicles often use stop-loss risk sharing schemes (see Marie-Jeanne (1994) for their properties as a function of the fleet size, and Teugels, Sundt (1991) for experience rating schemes on the aggregate loss). These rating structures are designed for large fleets, which is not the case on average for the portfolio analyzed in this article. Notice that, in general, fleet insurance business is offered mostly for fleets with little or medium size.

In our data set, the characteristics of each fleet are recorded by the SAAQ in real-time (see Section 2), and the tariff structures proposed in this article use the individual characteristics of both the vehicles and the carriers. The history of a vehicle should have a greater ability to predict the risk level of this vehicle than that of the other vehicles in the fleet. The basic issue in the statistical analysis of the portfolio is the assessment of these predictive abilities. Information on the drivers is not available in the data set, so a new vehicle can only be related to the fleet to which it belongs. Bonus-malus coefficients for the next period will then depend on an expected turnover for the vehicles of the fleet. Since the insurance premium is paid at the firm level, the bonus-malus coefficients computed in the paper depend on the history of claims or safety violations at the fleet level. However, an experience rating scheme using full information on the claims history is designed in Section 3.5.

The experience rating schemes are based on models with hierarchical random effects (see Jewell (1975)). Two types of BMS are analyzed. BMS designed from the number of claims are presented in Section 3, and another one obtained from the history of safety offences is given in Section 4. We explain the number of claims for bodily injuries. Bonus-malus coefficients are obtained from vehicle-specific and fleet-specific credibilities. They take into account an expected turnover for the vehicles within the fleets.

Compensations for bodily injuries are performed in Quebec within a pure no-fault framework (Devlin (1992); Boyer and Dionne (1987)), so it is difficult to use the history of claims in the rating structure, because standard BMS always have a "crime and punishment" flavour. Since 1992, the history of safety offences is used in the tariff structure of the SAAQ for pleasure vehicles (see Dionne and Vanasse (1997a) and Dionne, Maurice and Pinquet (2000b) for a related study).

The BMS designed in Section 3 is consistent with respect to the fleet-specific components, which is not the case when claims are replaced by safety offences as in Section 4. However, the BMS based on safety offences outperforms the one based on accidents after a year of experience with our data. The explanation of this somewhat surprising finding is the following. The frequency of offences is fourteen times higher than that of claims with bodily injuries. Even if the BMS based on safety offences is less efficient than the one based on accidents in the long run, the former system is closer to its limit in the short run, due to the higher frequency of safety offences.

A short conclusion summarizes the main results and proposes some extensions to the models presented in this article.

2. Economic environment and data set

Let us precise first the context of the study. The Province of Quebec introduced a new *Automobile Insurance Act* in March, 1978 to govern accident compensations. The Government had two goals in mind in tabling this legislation – to provide a rapid and reliable method for compensating all victims of bodily injuries, and to ensure better control of the cost of car repairs and faster compensation for property damage.

Fault has been entirely eliminated for bodily injuries. Compensation is provided by a compulsory and universal public plan. This plan is administered by a public corporation, the SAAQ. There is a maximum indemnity (which was estimated to compensate the total loss of income of 85 per cent of the population in 1978) for disability and death benefits. The indemnities for bodily injuries are in lieu of all rights to sue for bodily injuries or death, and no action is admitted before any court of justice.

The pricing procedure is very simple. The main sources of financing are from drivers' permits and automobile registration fees. Weight and type of vehicle driven are taken into consideration for vehicles other than pleasure vehicles. Past driving experience is taken into account since 1992 by using demerit points of the drivers.

So the SAAQ is a state insurer which provides motor insurance for bodily injuries in a monopolistic situation. As a state company, the SAAQ is also involved in road safety regulation. Consequently, it has a direct access to the information on individual safety offences. It was decided in 1992 to use such information for the pricing of private cars insurance. Besides their ability of screening risks, experience rating schemes provide incentives to careful driving. Indeed, the frequency of claims decreased by at least five per cent since the new regulation (see Dionne, Maurice and Pinquet (2000b) for more details).

The SAAQ also provides insurance for bodily injuries for fleets of vehicles. This insurance is also compulsory. Information is brought in real time for each vehicle, a situation which is not often encountered in this market. In order to create road safety incentives, the introduction of an experience rating scheme (as well as an a priori rating structure) is under consideration, which motivated the present study. This type of insurance rating would be easy to implement for the SAAQ since it has a direct access to all the necessary data.

Since January 1991, the SAAQ has been mandated to verify that commercial vehicles respect the laws and regulations governing, for example, the vehicle load and size limits, etc. In addition, the SAAQ was also given the mandate to verify the mechanical conformity of the vehicles.

In our working sample, the vehicles were observed during the years 1995 and 1996. The duration of observation of a vehicle is the validity duration of its licence plate. The weight of the vehicles has to be greater than 3,000 kgs, hence fleets of cars do not belong to this sample. The portfolio contains 50,746 fleets and 124,629 vehicles, and fleets are of small size on average. The size of the fleet is measured in vehicle-years, which is the sum of the validity durations. The other fleet-specific rating factors are the age of the firm and its activity sector. The vehicle-specific rating factors are the weight, the type of use, the type of fuel, the number of cylinders and the number of axles.

The initial file is the file of all registered motor carriers as of July 23, 1997. To be in that file a motor carrier must own or lease (long term) one or more vehicles.

We matched the information concerning the vehicles and the firms with the characteristics of safety violations committed at the carrier or at the vehicle level. The characteristics concerning mechanical conformity of the vehicles which had a recent mechanical check-up were linked to the data set already obtained.

The unit of observation in the working sample is a vehicle with at least one day with a valid license plate in 1996. In considering the safety offences committed in 1995 in the analyses, 24,581 trucks with no day with a valid license plate in 1995 have been dropped from the data set.

3. Bonus-malus systems from the number of claims

3.1. Bonus-malus coefficients as functions of the size of the fleet: Two limit examples

On a stratified portfolio, fixed and random effects introduced to design an optimal BMS must have a hierarchical structure (Jewell (1975)). The risk distribution of each vehicle includes then a vehicle-specific effect and a fleet-specific effect. Let us compute bonus-malus coefficients in two limit situations:

- Only the vehicle-specific effect is retained. The history of a vehicle cannot be used to predict the risk levels of the other vehicles in the fleet. If all the vehicles have the same a priori frequency risk, the credibility computed at the fleet level is the one given to each vehicle. As the variance of the ratio between the number of claims and the frequency premium decreases towards 0 when the size of the fleet goes to infinity, the same result holds for the variance of the bonus-malus coefficient.

- Only the fleet-specific effect is included in the number of claims distribution. Denote m as the number of vehicles in a given fleet, n_i as the number of claims reported by the vehicle i and λ as the a priori frequency risk for all the vehicles. We then have

$$N_i \sim P(\lambda u)(\forall i = 1,\dots,m) \Rightarrow \sum_{i=1}^{m} N_i \sim P(m\lambda u),$$

  if the N_i are independent in the fixed effects model (the fixed effect common to the vehicles in the fleet is denoted as u). If we write $E(U) = 1$; $V(U) = \sigma^2$ in the random effects model, the credibility granted to the fleet in the prediction is equal to

$$a = \frac{m\widehat{\lambda}\widehat{\sigma^2}}{1 + m\widehat{\lambda}\widehat{\sigma^2}}$$

This credibility increases towards one when the size m goes to infinity, and the bonus-malus coefficient converges towards the fleet-specific fixed effect u. The variance of the bonus-malus coefficient increases with the size of the fleet in the random effects model.

If the two random effects are included in a hierarchical model, the credibility granted to the history of the fleet will increase with its size if the estimated variance of the fleet-specific random effect is greater than zero. On the other hand, the variance of the bonus-malus coefficients is not a monotonic function of the size of the fleets. The increase of risk revelation with the size of the fleet is balanced by risk compensation between the vehicles.

3.2. Estimation of a model with random effects on a stratified portfolio

The hierarchical nature of the portfolio is taken into account by a double indexation. The fleets are indexed by $f = 1, \ldots, F$, and the vehicles are indexed by $i = 1, \ldots, m_f$, where m_f is the size of the fleet f. If N_{fi} is the number of claims reported by the vehicle i in the fleet f, we write

$$N_{fi} \sim P(\lambda_{fi} u_{fi});\ f = 1, \ldots, F;\ i = 1, \ldots, m_f$$

in the fixed effects model. The number of claims N_{fi} follows a Poisson distribution in the fixed effects model. The parameter λ_{fi} is a function of rating factors observed at the fleet level or at the vehicle level. The fixed effect u_{fi} represents the residual heterogeneity in the number of claims distribution. We distinguish firm-specific and vehicle specific effects in the regression and heterogeneity components, and write

$$\lambda_{fi} = d_{fi} \exp(x_f \gamma + z_{fi} \delta); u_{fi} = r_f s_{fi}.$$

The parameter λ_{fi} is proportional to the duration of observation of the vehicle d_{fi}. The line-vectors x_f and z_{fi} are the regression components connected to the fleet and to the vehicle. The related parameters are represented by the column-vectors γ and δ. The fixed effect u_{fi} splits into a fleet-specific effect r_f and a vehicle-specific effect s_{fi}. Vehicle-specific heterogeneity components could reflect the behaviour of the drivers, if a given vehicle is used by few drivers. This heterogeneity component can also reflect hidden features which are only related to the vehicle. You might think of annual mileage, which depends on the missions assigned to a given truck, but not on the drivers. The behaviour of the firms will influence the fleet-specific heterogeneity components. Fleet owners may obey (or not) to safety rules related to the mechanical check-up of vehicles, bulk trucking regulation, driving and work hour rules, etc. The financial structure of the carrier (which is not recorded by the SAAQ) probably influences safety activities, and hence the risk level. Economic and empirical results on the relationship between the financial structure of air carriers and safety are given by Dionne et al. (1997b).

The preceding distributions hold for real individuals, and the variables $(N_{fi})_{f=1,\dots,F;\, i=1,\dots,m_f}$ are supposed to be independent in the fixed effects model. This is the usual assumption in actuarial models (observed contagion on risk variables is supposed to be only apparent). The random effects $(R_f)_{f=1,\dots,F}$ and $(S_{fi})_{f=1,\dots,F;\, i=1,\dots,m_f}$ are i.i.d. in each family and mutually independent. Distributions in the model with random effects are mixtures of Poisson distributions, and they refer to generic individuals, who represent a class of real individuals with the same observable characteristics (see Pinquet (2000) for instance). The independence between the $(R_f)_{f=1,\dots,F}$ and $(S_{fi})_{f=1,\dots,F;\, i=1,\dots,m_f}$ can be assumed without loss of generality since the decomposition $U_{fi} = R_f S_{fi}$ is not unique. The random effect S_{fi} reflects a residual heterogeneity in the risk distribution of the vehicle. If R and S are random variables with these distributions, we suppose that

$$E(R) = E(S) = 1;\; V(R) = V_{RR};\; V(S) = V_{SS}.$$

Within a semiparametric approach, the distributions on the random effects will only be specified by the variances. If $U=RS$, we have

$$E(U) = E(R)E(S) = 1;\; V(U) = V_{UU} = E(R^2)E(S^2) - 1 = V_{RR} + V_{SS} + V_{RR}V_{SS}.$$

With the total variance and covariance formula and the independence assumed in the model with fixed effects, we obtain

$$V(N_{fi}) = \lambda_{fi} + \lambda_{fi}^2 V(U_{fi}) = \lambda_{fi} + \lambda_{fi}^2 V_{UU};$$

$$Cov(N_{fi}, N_{fi'}) = \lambda_{fi}\lambda_{fi'}\, Cov(U_{fi}, U_{fi'}) = \lambda_{fi}\lambda_{fi'}\, V_{RR}\,(i \neq i') \tag{2}$$

in the random effects model. As the size of the portfolio is large, we will use a frequentist approach, and will describe the data by consistent estimators.

The a priori rating model is a Poisson model without fixed or random effects, i.e. $N_{fi} \sim P(\lambda_{fi})\, \forall f, i$. Let $\widehat{\lambda_{fi}} = d_{fi} \exp(x_f \widehat{\gamma} + z_{fi} \widehat{\delta})$ be the frequency premium computed in the a priori rating model, where $\widehat{\gamma}$ and $\widehat{d}$ are the maximum likelihood estimators. The likelihood equations in this model are

$$\sum_{f,i} (n_{fi} - \widehat{\lambda_{fi}})^t x_f = 0;\; \sum_{f,i} (n_{fi} - \widehat{\lambda_{fi}})^t z_{fi} = 0. \tag{3}$$

They reflect an orthogonality relationship between number-residuals and the regression components. Since $E(N_{fi}) = \lambda_{fi} E(U) = \lambda_{fi}$ in the model with random effects, the m.l.e. in the Poisson model without fixed and random effects are consistent estimators of the corresponding parameters in the model with random effects (see Gouriéroux et al. (1984)). Hence, a frequency premium computed for an individual in the a priori rating model converges towards the frequency risk of the related generic individual in the model with random effects.

From the moments computed in (2), we obtain the following limits

$$\widehat{V_{RR}} = \frac{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} (n_{fi} - \widehat{\lambda_{fi}})(n_{fi'} - \widehat{\lambda_{fi'}})}{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} \widehat{\lambda_{fi}} \widehat{\lambda_{fi'}}} \to V_{RR};$$

$$\widehat{V_{UU}} = \frac{\sum_{f,i} \left[(n_{f,i} - \widehat{\lambda_{fi}})^2 - n_{fi} \right]}{\sum_{f,i} \widehat{\lambda_{fi}}^2} \to V_{UU}. \tag{4}$$

Thus consistent estimators of $V(U)$ and $V(R)$ are obtained from the estimators derived in the a priori model. Since $V_{UU} = V_{RR} + V_{SS} + V_{RR} V_{SS}$,

$$\widehat{V_{SS}} = \frac{\widehat{V_{UU}} - \widehat{V_{RR}}}{1 + \widehat{V_{RR}}}$$

is a consistent estimator of V_{SS}.

Let us interpret these results. The estimator $\widehat{V_{RR}}$ assesses observed contagion between the claims histories connected to different vehicles within the same fleet. If $\widehat{V_{RR}}$ is greater than zero, the positive observed contagion means that the history of a vehicle can reveal hidden features in the risk distributions of every vehicle in the same fleet. The numerator of the ratio which defines the estimator $\widehat{V_{RR}}$ is easily derived from

$$\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} (n_{fi} - \widehat{\lambda_{fi}})(n_{fi'} - \widehat{\lambda_{fi'}}) = \sum_f (n_f - \widehat{\lambda_f})^2 - \sum_{f,i} (n_{fi} - \widehat{\lambda_{fi}})^2,$$

if we write $\sum_{1 \le i \le m_f} n_{fi}, \widehat{\lambda_f} = \sum_{1 \le i \le m_f} \widehat{\lambda_{fi}}$. We then have

$$\widehat{V_{SS}} > 0 \Leftrightarrow \widehat{V_{UU}} > \widehat{V_{RR}} \tag{5}$$

$$\Leftrightarrow \frac{\sum_{f,i} \left[(n_{f,i} - \widehat{\lambda_{fi}})^2 - n_{fi} \right]}{\sum_{f,i} \widehat{\lambda_{fi}}^2} > \frac{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} (n_{fi} - \widehat{\lambda_{fi}})(n_{fi'} - \widehat{\lambda_{fi'}})}{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} \widehat{\lambda_{fi}} \widehat{\lambda_{fi'}}}$$

$$\Leftrightarrow \frac{\sum_{f,i} \left[(n_{f,i} - \widehat{\lambda_{fi}})^2 - n_{fi} \right]}{\sum_{f,i} \widehat{\lambda_{fi}}^2} > \frac{\sum_f \left[(n_f - \widehat{\lambda_f})^2 - n_f \right]}{\sum_f \widehat{\lambda_f}^2}$$

The estimated variance of the vehicle-specific random effect is greater than zero if the relative overdispersion derived at the vehicle level is greater than its counterpart computed at the fleet level.

These moment-based estimators are unconstrained (i.e. estimated variances are not bound to be positive). Suppose for instance that $\widehat{V_{RR}} < 0$ on a sample. Such an estimation would be related to a null estimator of V_{RR} within a constrained approach (for instance m.l.e., which is costly to perform if the likelihood does not admit a closed form). In this case, the fleet-specific random effect must be abandoned whatever is the estimation strategy. Hence, the unconstrained nature of the estimators retained in the paper is not a drawback.

These estimators are consistent and asymptotically normal in the model with random effects. Their asymptotic variance can be reduced if weights related to overdispersion are introduced in the regression (see Liang, Zeger (1986)).

Let us precise this point. Denote the parameters of the a priori rating model and of the mixing distribution as

$$\eta = \begin{pmatrix} \gamma \\ \delta \end{pmatrix}; \; \theta = \begin{pmatrix} V_{RR} \\ V_{SS} \end{pmatrix}.$$

If we stack the numbers of claims reported on a given fleet in a vector $sn_f = vec_{1 \le i \le m_f}(n_{fi})$ the m.l.e. of the Poisson model (3) can be expressed as the solution in η of the equation

$$\sum_f^2 \left(\frac{\partial}{\partial \eta} E(SN_f | \eta) \right) \left[V(SN_f | \eta) \right]^{-1} \left(sn_f - E(SN_f | \eta) \right) = 0, \qquad (6)$$

where the moments are computed in the Poisson model without fixed or random effects. Let $E(SN_f | \eta, \theta)$ and $V(SN_f | \eta, \theta)$ be the expectation and variance derived in the random effects model (we have $E(SN_f | \eta, \theta) = E(SN_f | \eta) \, \forall \eta, \theta$). The moment-based estimators of V_{RR} and V_{SS} given in this section from the regression provide a function $\eta \to \widehat{\theta}(\eta)$. A "generalized estimating equation" includes the estimated moments of the random effects in equation (6). The corresponding estimator is the solution in $\widehat{\eta}$ of equation (6), where $V(SN_f | \eta)$ is replaced by $V(SN_f | \eta, \widehat{\theta}(\eta))$. This estimator η can be shown to have optimal properties in terms of asymptotic variance. Then a new estimator $\widehat{\theta}(\widehat{\eta})$ is obtained for the parameters of the mixing distribution. The random effects model retained in this section is usually referred to as an "exchangeable correlations" model.

3.3. Linear credibility predictors

In this section, we compute linear credibility predictors (Bühlmann (1967)) for each vehicle. They are derived from the history of claims observed at the fleet level, whereas the credibility coefficient depends on the vehicle. Let i_0 be a vehicle which belongs to the fleet f_0. The portfolio is observed during one period, and a bonus-malus coefficient is computed for the next one. In order to allow for a turnover in the portfolio, this vehicle may appear at the second period. Predictors are obtained separately for each fleet, and the fleet index is suppressed in order to simplify the notations. The fleet is supposed to contain m vehicles during the first period.

The bonus-malus coefficient for the vehicle i_0 is supposed to depend only on the number of claims reported on the whole fleet. It is written as $\widehat{a}_{i_0}+\widehat{b}_{i_0}\left(\sum_{i=1}^{m} n_i\right)$, with

$$(\widehat{a}_{i_0},\widehat{b}_{i_0}) = \arg\min_{a,b} \widehat{E}\left[\left(U_{i_0}-a-b\left(\sum_{i=1}^{m} N_i\right)\right)^2\right].$$

The estimated expectation is derived in the random effects model. Notice that no specific weight is given to the history of the vehicle. As $E(U_{i_0}) = 1$, we have

$$\widehat{a}_{i_0}+\widehat{b}_{i_0}\left(\sum_{i=1}^{m} n_i\right) = 1+\widehat{b}_{i_0}\left(\sum_{i=1}^{m}(n_i-\widehat{\lambda}_i)\right) = (1-cred_{i_0})+cred_{i_0}\frac{\sum_{i=1}^{m} n_i}{\sum_{i=1}^{m}\widehat{\lambda}_i},$$

with

$$cred_{i_0} = \widehat{b}_{i_0}\left(\sum_{i=1}^{m}\widehat{\lambda}_i\right) = \frac{\widehat{Cov}(U_{i_0},\sum_{i=1}^{m} N_i)}{\widehat{V}(\sum_{i=1}^{m} N_i)}\left(\sum_{i=1}^{m}\widehat{\lambda}_i\right).$$

Consistent estimators for the individual moments are

$$\begin{aligned}
&\widehat{Cov}(U_{i_0},N_i) = \widehat{\lambda}_i\widehat{Cov}(U_{i_0},U_i) = \begin{cases}\lambda_i\widehat{V_{RR}}\,(i_0\neq i)\\ \lambda_{i_0}\widehat{V_{UU}}\,(i_0=i)\end{cases};\\
&\widehat{V}(N_i) = \widehat{\lambda}_i+\widehat{\lambda}_i^{\,2}\,\widehat{V_{UU}};\ \widehat{Cov}(N_i,N_{i'}) = \widehat{\lambda}_i\widehat{\lambda}_{i'}\widehat{V_{RR}}(i\neq i'),
\end{aligned}\tag{7}$$

with the estimators obtained in the preceding section. In the computation of the credibility coefficient, two situations may happen:

- Either the vehicle was not observed during the first period, which means that it joined the fleet during the forecast period ($i_0 \neq i\ \forall i = 1,\dots, m$). From the estimations obtained in (7), we have

$$cred_{i_0} = \alpha = \frac{\widehat{V_{RR}}\left(\sum_{i=1}^{m}\widehat{\lambda}_i\right)}{1+\left(\widehat{V_{RR}}\left(\sum_{i=1}^{m}\widehat{\lambda}_i\right)\right)+\left((\widehat{V_{UU}}-\widehat{V_{RR}})\dfrac{\sum_{i=1}^{m}\widehat{\lambda}_i^2}{\sum_{i=1}^{m}\widehat{\lambda}_i}\right)}$$

This fleet-specific credibility coefficient roughly increases with the estimated variance of the fleet-specific random effect and with the frequency-premium computed at the fleet level.

- Or the vehicle was observed during the first period ($1 \le i_0 \le m$). Then

$$cred_{i_0} = \alpha+\beta_{i_0};\ \beta_{i_0} = \frac{(\widehat{V_{UU}}-\widehat{V_{RR}})\widehat{\lambda}_{i_0}}{1+\left(\widehat{V_{RR}}\left(\sum_{i=1}^{m}\widehat{\lambda}_i\right)\right)+\left((\widehat{V_{UU}}-\widehat{V_{RR}})\dfrac{\sum_{i=1}^{m}\widehat{\lambda}_i^2}{\sum_{i=1}^{m}\widehat{\lambda}_i}\right)}.\tag{8}$$

The credibility coefficient is the sum of the fleet-specific coefficient and of a vehicle-specific coefficient. It can be computed only if the estimated variance of the vehicle-specific effect $\widehat{V_{SS}}$ is greater than zero (which amounts to $\widehat{V_{UU}} > \widehat{V_{RR}}$ from (5)), a condition fulfilled in our data. This coefficient is the bonus granted to the firm if no claim is recorded on its vehicles.

Fleets are open in most cases, which means that an endorsement is not brought to the insurance policy after each arrival or departure of a vehicle in the fleet. In this context, bonus-malus coefficients computed at the vehicle level may appear unrealistic. If ρ is the expected turnover for the vehicles of the fleet, a credibility equal to $\alpha + ((1-\rho)\overline{\beta})$ can be retained at the fleet level, where $\overline{\beta}$ is the average of the β_i.

3.4. Empirical results

Table 1 presents the results of a Poisson model which explains the number of claims reported in 1996 by regression components derived from rating factors. The only continuous rating factor is the age of the firm. We observe that the frequency of claims decreases – ceteris paribus – by 3.4% with a supplementary year of age. The other rating factors have a finite number of categories.

In Table 1, the vehicles are weighted by the risk exposure measured by the number of days the vehicle is authorized to circulate. The estimated exponential of the coefficients (written in a multiplicative way) related to the different levels of each rating factor are averaged to one (column ST. COFF., for standardized coefficient). Two advantages are obtained.

- The coefficients do not depend on the category that must be omitted in the regression for each rating factor in order to avoid colinearity. This is due to the fact that the vector of frequency-premiums derived from a Poisson model with regression components depends only on the linear space spanned by the covariates. Hence, the multiplicative coefficients derived from the Poisson model are defined up to a multiplicative constant for each rating factor, whatever are the omitted levels.
- These coefficients can be compared to the relative frequency of each category, which is the frequency of claims for one category divided by the global frequency, column REL. FRE. in Table 1. Consider for instance the category “bulk transport” of the rating factor “firm’s activity sector”. The relative frequency is 1.617, whereas the standardized coefficient derived from the Poisson model equals 1.146. From the likelihood equations of the Poisson model (see (3)), the number of claims equals the sum of the frequency premiums for each level. The ratio 1.617/1.146=1.411 means that the vehicles belonging to this type of fleet have, with respect to other rating factors, a frequency risk level which is 41% higher than the average.

Table 1 also provides levels of significance for the coefficients estimated in the regression. The P-VALUE column is obtained from a studentized statistic (i.e. the ratio between the estimated coefficient and its estimated standard

TABLE 1

RATING SCORE FOR THE FREQUENCY OF CLAIMS WITH BODILY INJURIES

	WEIGHT (%)	REL.FRE.	ST.COFF.	P-VALUE
VARIABLE: FIRM'S ACTIVITY SECTOR				
general merchandise transport	13.7	1.508	1.233	0.011
bulk transport	10.9	1.617	1.146	0.079
short term rental	2.5	0.959	0.840	0.501
independent trucker, other sector	72.9	0.813	0.940	ref. group
VARIABLE: VEHICLES-YEARS				
0 or 1 vehicle-year	31.8	0.758	0.803	ref. group
2 vehicle-years	11.9	0.887	0.920	0.145
3 vehicle-years	7.2	1.032	1.055	0.010
4 to 9 vehicle-years	17.1	1.111	1.083	<0.001
10 to 20 vehicle-years	9.6	1.292	1.177	<0.001
more than 20 vehicle-years	22.4	1.183	1.164	<0.001
VARIABLE: TYPE OF FUEL				
gasoline	20.4	0.430	0.597	<0.001
fuel oil	79.6	1.147	1.104	ref. group
VARIABLE: WEIGHT OF THE VEHICLE				
from 3,000 to 3,870 kgs	20	0.624	0.718	0.014
from 3,871 to 6,220 kgs	20	0.674	0.888	0.025
from 6,221 to 7,620 kgs	20	1.174	1.108	0.982
from 7,621 to 8,850 kgs	20	1.428	1.174	0.479
more than 8,850 kgs	20	1.099	1.110	ref. group
VARIABLE: TYPE OF USE				
commercial use	75.8	0.809	0.969	0.508
bulk transport	10.4	1.724	1.351	0.005
other types of transport	13.8	1.501	0.904	ref. group
VARIABLE: NUMBER OF AXLES				
unknown	1.3	5.706	6.835	<0.001
2 axles, less than 4,000 kgs	21.2	0.573	1.174	ref. group
2 axles, more than 4,000 kgs	26.9	0.694	0.797	0.023
3 axles	18.0	0.917	0.781	0.022
4 axles	5.4	0.908	0.760	0.028
5 axles	8.8	0.876	0.635	0.001
6 axles and more	18.4	1.775	1.141	0.869
VARIABLE: NUMBER OF CYLINDERS				
1 to 5 cylinders	1.4	0.840	0.982	0.410
6 to 7 cylinders	59.9	1.261	1.122	<0.001
8 cylinders and more	38.7	0.600	0.812	ref. group
Number of vehicles		124,629		

deviation). For each rating factor, the reference group is related to the level which was suppressed in order to avoid colinearity.

The frequency of claims increases with the size of the fleet. This result could be explained by a greater exposure to risk (as measured by annual mileage) for the vehicles belonging to large fleets. The same reason probably also explains why gasoline-powered vehicles are much less risky than fuel-powered ones.

Annual mileage was estimated for the vehicles which had a recent mechanical check-up (54,699 vehicles). The estimation of the rating model with this supplementary variable leads to the following results, with a level of significance equal to 10%.

- The fuel effect disappears.
- The size effect decreases, but remains significant.
- The firm activity sectors are not significant.
- The number of cylinders effect disappears.

Detailed results can be obtained in Dionne, Desjardins, Pinquet (1999).

On the sample, the estimators given in the preceding section are equal to

$$\widehat{V_{RR}} = 0.153;\ \widehat{V_{UU}} = 1.121 \Rightarrow \widehat{V_{SS}} = \frac{\widehat{V_{UU}} - \widehat{V_{RR}}}{1 + \widehat{V_{RR}}} = 0.840. \tag{9}$$

The estimated variances of random effects are close to the malus applied to the a priori frequency premium after one claim if this premium is close to zero. This is the case for most of the fleets in the portfolio because of their small size on average, and because of the low frequency of claims for bodily injuries, which is equal to 1.6% per year on average. Hence, one claim reported on such a fleet would entail a malus close to 15% for a new vehicle.

The estimated variance $\widehat{V_{SS}}$ of the vehicle-specific random effect is important. The history of a vehicle will have much more ability to predict the risk level of this vehicle than that of the other vehicles in the fleet.

The preceding estimators are not really modified by a "generalized estimating equation" (see the end of Section 3.2). The frequency premiums are very close, and estimated variances of the random effects are

$$\widehat{V_{RR}} = 0.161;\ \widehat{V_{UU}} = 1.110.$$

We use the estimators obtained in equation (9) at the end of the section.

Bonus-malus coefficients are computed at the fleet level in Table 2, for the two limit values of the turnover. Credibilities of the histories and standard deviations of the bonus-malus coefficients are given for each size level retained in the regression components (see Table 1).

Since the bonus-malus coefficients are computed at the fleet level, all the averages computed in Table 2 are weighted by the frequency premiums of the fleets. Due to the important value of the variance of the vehicle-specific random effect, the credibility strongly depends on the turnover for fleets with little or

TABLE 2

AVERAGE CREDIBILITIES FOR FLEETS AND VEHICLES
STANDARD DEVIATIONS OF BONUS-MALUS COEFFICIENTS AT THE FLEET LEVEL

Fleet size	$\bar{\alpha}$	$\bar{\alpha}+\bar{\beta}$	$\sigma_{bonmal_{\alpha}}$ **(turnover = 100%)**	$\sigma_{bonmal_{\alpha+\beta}}$ **(turnover = 0%)**
0 or 1 vehicle-year	0.003	0.019	0.020	0.136
2 vehicle-years	0.006	0.026	0.030	0.126
3 vehicle-years	0.009	0.030	0.037	0.122
from 4 to 9 vehicle-years	0.019	0.041	0.053	0.116
from 10 to 20 vehicle-years	0.048	0.072	0.083	0.129
more than 20 vehicle-years	0.245	0.262	0.189	0.203

medium size. The same result holds for the dispersion of the bonus-malus coefficients. As expected from the conclusion of Section 3.1., the standard deviation of the bonus-malus coefficients is not a monotonic function of the size of the fleet when the turnover is equal to zero.

3.5. An experience rating scheme using full information on the claims history

Since the drivers do not pay insurance premiums of firm-owned vehicles, the computation of premiums at the vehicle level may appear irrelevant. However, disaggregated information on the premium may be of interest for the firm. In this context, you can think of using full information on the claims history. Different weights can be given to the histories of the vehicles in the derivation of the bonus-malus coefficient for a given vehicle.

Bonus-malus coefficients obtained at the vehicle level from the approach retained in Section 3.3 have a very low within fleets dispersion. This is due to the fact that the credibility granted to the history of the vehicle is applied to a ratio computed at the fleet level. The within fleets dispersion of the bonus-malus coefficients, as measured by the standard deviation, is at most equal to three per cent of the total dispersion for the different size levels.

In this section, we compute linear credibility predictors which give specific weights to the history of each vehicle in the prediction of the risk frequency for a given vehicle. The intuition is that the predictor should overweight the history of this vehicle, as compared to the one obtained in Section 3.3. As a result, the within fleets dispersion of the bonus-malus coefficients should increase.

As in Section 3.3, we consider a fleet with m vehicles during the first period, and a vehicle i_0 which belongs to the fleet during the forecast period. We suppress the fleet index, and write the bonus-malus coefficient as $\widehat{a}_{i_0} + {}^t\widehat{b}_{i_0} n$ with

$$n = \underset{1 \le i \le m}{vec}(n_i); b_{i_0} = \underset{1 \le i \le m}{vec}(b_{i_0,i}); (\widehat{a}_{i_0}, \widehat{b}_{i_0}) = \arg \min_{a \in \mathbb{R}, b \in \mathbb{R}} \widehat{E}\left[(U_{i_0} - a - {}^t bN)^2\right].$$

The estimated expectation is derived in the model with random effects. Since $E(U_{i_0}) = 1$ the bonus-malus coefficient is equal to

$$\widehat{a}_{i_0}, {}^t\widehat{b}_{i_0} n = 1 + {}^t\widehat{b}_{i_0}(n - \widehat{E}(N)) = 1 + {}^t\widehat{b}_{i_0}(n - \widehat{\lambda}).$$

The vector of frequency-premiums $\widehat{\lambda} = vec_{1 \le i \le m}(\widehat{\lambda_i})$ is derived from m.l.e. in the a priori rating model. It is a consistent estimator for the frequency risks in the model with random effects. From the consistent estimators of individual variances and covariances derived in Section 3.3, we infer

$$\widehat{b}_{i_0} = [\widehat{V}(N)]^{-1}\widehat{Cov}(N, U_{i_0}),$$

with

$$\widehat{V}(N) = D + \left(\widehat{V_{RR}}\widehat{\lambda}\, {}^t\widehat{\lambda}\right), D = \underset{1 \le i \le m}{diag}\left(\widehat{\lambda_i} + \left[\widehat{\lambda_i}^2 (\widehat{V_{UU}} - \widehat{V_{RR}})\right]\right);$$
$$\widehat{Cov}(N, U_{i_0}) = \widehat{V_{RR}}\widehat{\lambda} + (\widehat{V_{UU}} - \widehat{V_{RR}})\,\widehat{\lambda}_{i_0} e_{i_0}.$$

The last term exists if $i_0 \le m$ (i.e. the vehicle was observed during the first period). The vector e_{i_0} belongs to the canonical basis of $\mathbb{R}^m$, with the corresponding index.

Let us compute $[\widehat{V}(N)]^{-1}$: From $\widehat{V}(N) = D\left[I_m + \left(\widehat{V_{RR}} D^{-1}\widehat{\lambda}\, {}^t\widehat{\lambda}\right)\right]$ and $\left(D^{-1}\widehat{\lambda}\, {}^t\widehat{\lambda}\right)^2 =$

$$\|\widehat{\lambda}\|^2_{D^{-1}} D^{-1}\widehat{\lambda}\, {}^t\widehat{\lambda}, \|\widehat{\lambda}\|^2_{D^{-1}} = {}^t\widehat{\lambda} D^{-1}\widehat{\lambda} = \sum_{i=1}^{m} \frac{\widehat{\lambda_i}}{1 + [\lambda_i (\widehat{V_{UU}} - \widehat{V_{RR}})]},$$

we obtain

$$[\widehat{V}(N)]^{-1} = \left[I_m + \left(\widehat{V_{RR}} D^{-1}\widehat{\lambda}\, {}^t\widehat{\lambda}\right)\right]^{-1} D^{-1} = \left[I_m - \left(\frac{\widehat{V_{RR}}}{1 + \|\widehat{\lambda}\|^2_{D^{-1}}\widehat{V_{RR}}} D^{-1}\widehat{\lambda}\, {}^t\widehat{\lambda}\right)\right] D^{-1}.$$

A first expression of $\widehat{b}_{i_0} = [\widehat{V}(N)]^{-1}\widehat{Cov}(N, U_{i_0})$ is

$$\widehat{b}_{i_0} = \left[I_m - \left(\frac{\widehat{V_{RR}}}{1 + \|\widehat{\lambda}\|^2_{D^{-1}}\widehat{V_{RR}}} D^{-1}\widehat{\lambda}\, {}^t\widehat{\lambda}\right)\right] D^{-1}\left[\widehat{V_{RR}}\widehat{\lambda} + (\widehat{V_{UU}} - \widehat{V_{RR}})\,\widehat{\lambda e_{i_0}}\right].$$

Since

$$\left[I_m - \left(\frac{\widehat{V_{RR}}}{1 + \|\widehat{\lambda}\|^2_{D^{-1}}\widehat{V_{RR}}} D^{-1}\widehat{\lambda}\, {}^t\widehat{\lambda}\right)\right] D^{-1}\widehat{\lambda} = \frac{1}{1 + \|\widehat{\lambda}\|^2_{D^{-1}}\widehat{V_{RR}}} D^{-1}\widehat{\lambda},$$

we have

$$\widehat{b}_{i_0} = \frac{1}{1 + \|\widehat{\lambda}\|^2_{D^{-1}}\widehat{V_{RR}}}\left[\left(\widehat{V_{RR}}D^{-1}\widehat{\lambda}\right) + \widehat{\lambda_{i_0}}(\widehat{V_{UU}} - \widehat{V_{RR}})\,D^{-1}e_{i_0}\right]. \tag{10}$$

The bonus-malus coefficient for the vehicle i_0 is obtained from the credibility formula

$$1 + {}^t\widehat{b}_{i_0}(n - \widehat{\lambda}) = \left[1 + \sum_{i=1}^{m} a_i\left(\frac{n_i}{\widehat{\lambda}_i} - 1\right)\right] + \widehat{\beta}_{i_0}\left(\frac{n_{i_0}}{\widehat{\lambda}_{i_0}} - 1\right),$$

if we write

$$a_i = \widehat{\lambda}_i\widehat{b}_{i_0,i}(i \neq i_0); a_{i_0} + \beta_{i_0} = \widehat{\lambda}_{i_0}\widehat{b}_{i_0,i_0}, \tag{11}$$

with a_{i_0} expressed as the a_i, $i \neq i_0$. The bonus-malus coefficient is a sum of two terms:

- The first one does not depend on the vehicles within the fleet, and is applied to the new vehicles.
- The second one exists only if the vehicle was observed in the past ($1 \leq i_0 \leq m$). The credibility coefficient β_{i_0} is applied to the individual history. The corresponding coefficient was applied to the history at the fleet level in Section 3.3, and this explains the more important within fleet dispersion of bonus-malus coefficients which use all the information.

From equations (10) and (11), the credibility coefficients are respectively equal to

$$a_i = \frac{\dfrac{\widehat{\lambda}_i\widehat{V_{RR}}}{1 + [\widehat{\lambda}_i(\widehat{V_{UU}} - \widehat{V_{RR}})]}}{1 + \sum_{i'=1}^{m}\dfrac{\widehat{\lambda}_{i'}\widehat{V_{RR}}}{1 + [\widehat{\lambda}_{i'}(\widehat{V_{UU}} - \widehat{V_{RR}})]}}; \quad \beta_{i_0} = \frac{\dfrac{\widehat{\lambda}_{i_0}(\widehat{V_{UU}} - \widehat{V_{RR}})}{1 + [\widehat{\lambda}_{i_0}(\widehat{V_{UU}} - \widehat{V_{RR}})]}}{1 + \sum_{i'=1}^{m}\dfrac{\widehat{\lambda}_{i'}\widehat{V_{RR}}}{1 + [\widehat{\lambda}_{i_0}(\widehat{V_{UU}} - \widehat{V_{RR}})]}}$$

$\forall i$, $i_0 = 1, \ldots, m$. As in Section 3.3, this credibility system makes sense only if $\widehat{V_{UU}} > \widehat{V_{RR}}$ which means that the estimated variance $\widehat{V_{SS}}$ of the vehicle-specific effect is greater than 0 (see equation (5)).

Let us compare results obtained in this section and in Section 3.3 with an example. We use the estimations given in equation (9). Consider a fleet of five vehicles observed during a period, with frequency-premiums equal to 0.02 for each vehicle. Suppose that one claim was reported during the first period. The bonus-malus coefficients for the next period are given in the following table.

TABLE 3

BONUS-MALUS COEFFICIENTS WITH LIMITED AND FULL INFORMATION (EXAMPLE)

Fleet size	History at the fleet level	Full information
new vehicle	1.133	1.135
claimless vehicle	1.301	1.116
vehicle with one claim reported	1.301	2.063

Let us comment the coefficients given by the BMS with full information. The four vehicles without claim reported are penalized because the malus at the fleet level outweighs the bonus generated by the individual history. The bonus-malus coefficient of the vehicle with one claim reported is much more important than that of the four other ones because of the differences between the individual credibilities.

The within fleets dispersion of bonus-malus coefficients will be more important with this BMS. This is shown in the following table which provides between fleets and total standard deviation of bonus-malus coefficients computed on the portfolio with the present BMS.

TABLE 4

TOTAL AND BETWEEN FLEETS STANDARD DEVIATIONS OF BONUS-MALUS COEFFICIENTS LINEAR CREDIBILITY WITH FULL INFORMATION

Fleet size	turnover = 100% $\sigma_{between}$	turnover = 0% $\sigma_{between}$	turnover = 0% σ_{total}
0 or 1 vehicle-year	0.020	0.136	0.138
2 vehicle-years	0.030	0.123	0.151
3 vehicle-years	0.037	0.123	0.158
from 4 to 9 vehicle-years	0.053	0.116	0.169
from 10 to 20 vehicle-years	0.083	0.128	0.187
more than 20 vehicle-years	0.189	0.203	0.231

The between fleets dispersions of the bonus-malus coefficients are very close to those obtained in Table 2 for the same value of the turnover. This means that using only the history of the fleet in the prediction did not entail a loss of efficiency for bonus-malus coefficients computed at the fleet level.

Optimal BMS using all the information on the claim history can also be derived with an expected value principle (Lemaire (1985), Dionne et al. (1989), Pinquet (1997)). The negative binomial model with random effects (Hausman, Hall, Griliches (1984)) can be used for that purpose. Initially designed for longitudinal count data, it can be used in our context, due to the analogy between panel data and stratified samples. For example, consider an individual as a stratum and a period as an individual within a stratum. This model is developed in Dionne, Desjardins, Pinquet (2000a).

4. BONUS-MALUS SYSTEMS FROM THE NUMBER OF SAFETY OFFENCES

4.1. Safety offences used as regression components

Owing to the no-fault setting, the history of claims cannot be included in the tariff structure of the SAAQ. However, safety offences can be used to perform experience rating. In our data base, safety offences of different types were recorded at the carrier level and at the driver level. Those which were recorded in 1995 are added here as regression components in the Poisson model estimated in Table 1. Hence the number of claims reported in 1996 is explained by rating factors and by the safety offences recorded the year before. Each estimated coefficient related to a given type of safety offence leads to a relative malus, if this coefficient is positive. The safety offences which did entail a malus are presented in Table 5.

TABLE 5

RELATIVE MALI DERIVED FROM SAFETY OFFENCES

Type of safety offence recorded in 1995	**related to**	**relative malus (%)**	**P-value**
exceeding speed limits	vehicles	42	<0.001
not wearing the seat belt	vehicles	93	<0.001
not respecting hazardous goods rules	carriers	105	0.008
excess load	carriers	12	0.089
not stopping at an agent's signal	vehicles	38	0.091
not respecting driving hours rules	carriers	72	0.013
number of vehicles		100,048	

We retained the vehicles with a positive duration of authorization for the licence plate during 1995 and 1996. Other safety offences which were not retained by the model are the following: Exceeding size limits, not respecting bulk trucking regulation, not respecting mechanical check-up rules, driving with a sanction, not stopping at a red light. Many of them are significant when we consider all types of road accidents (property damages and bodily injuries). See Dionne et al. (1999) for more details, including regression results related to the rating factors. An optimal BMS is designed in the next section from a model with random effects on two types of events, namely the claims for bodily injury and all the safety offences.

4.2. The model with random effects

Let INF_{fi} be the number of safety offences (whatever their type) recorded on the vehicle i belonging to the fleet f. We write

$$INF_{fi} \sim P(\tau_{fi} t_{fi}),$$

where $\tau_{fi} = d_{fi} \exp(x_f \zeta + z_{fi} \eta)$ is the component of $E(INF_{fi})$ which is explained by the duration of exposure to safety offences and by both fleet-specific and vehicle-specific regression components, and where t_{fi} is the fixed effect. The hierarchical structure of the portfolio is taken into account by writing $t_{fi} = p_f q_{fi}$ where p_f and q_{fi} are the fleet-specific and vehicle-specific fixed effects. All the number variables are supposed independent in the fixed effects model. Let U, R, T and P be random variables with the same joint distribution as any random vector such as $(U_{fi}, R_f, T_{fi}, P_f)$ (we use the notations of Section 3.2). The assumption $E(U) = 1$ made in Section 3.2 is relaxed now. A natural multivariate distribution family with explicit moments for non-negative random effects is the log-normal distribution family, and the expectation depends then on the variance. Let $\widehat{\tau_{fi}} = d_{fi} \exp(x_f \widehat{\zeta} + z_{fi} \widehat{\eta})$ be the estimation of $E(INF_{fi})$ derived from likelihood maximization in the Poisson model without fixed or random effects. If data are generated in the random effects model, we have

$$\widehat{\lambda_{fi}} \to E(N_{fi}) = \lambda_{fi} E(U); \widehat{\tau_{fi}} \to E(INF_{fi}) = \tau_{fi} E(T). \tag{12}$$

The expectation is computed in the random effects model. From (12) and results similar to those given in (2), we obtain the following limits

$$\widehat{V_{UT}}^1 = \frac{\sum_{f,i} (N_{fi} - \widehat{\lambda_{fi}})(INF_{fi} - \widehat{\tau_{fi}})}{\sum_{f,i} \widehat{\lambda_{fi}} \widehat{\tau_{fi}}} \to \frac{Cov(U,T)}{E(U)E(T)};$$

$$\widehat{V_{RP}}^1 = \frac{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} (N_{fi} - \widehat{\lambda_{fi}})(INF_{fi'} - \widehat{\tau_{fi'}})}{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} \widehat{\lambda_{fi}} \widehat{\tau_{fi'}}} \to \frac{Cov(R,P)}{E(R)E(P)};$$

$$\widehat{V_{PP}}^1 = \frac{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} (INF_{fi} - \widehat{\tau_{fi}})(INF_{fi'} - \widehat{\tau_{fi'}})}{\sum_f \sum_{1 \le i, i' \le m_f; i \ne i'} \widehat{\tau_{fi}} \widehat{\tau_{fi'}}} \to CV^2(P) = \frac{V(P)}{E^2(P)};$$

$$\widehat{V_{TT}}^1 = \frac{\sum_{f,i} \left[(INF_{fi} - \widehat{\tau_{fi}})^2 - \widehat{\tau_{fi}} \right]}{\sum_{f,i} \widehat{\tau_{fi}}^2} \to CV^2(T) = \frac{V(T)}{E^2(T)}. \tag{13}$$

The superscript “1” is used for the preceding estimators because they are obtained at the first step of the Newton-Raphson algorithm of likelihood maximization, where the initial value is the m.l.e. for the a priori rating model. For instance, the estimator $\widehat{V_{RP}}^1$ reflects the predictive power that safety offences recorded on a given vehicle have on the risk level of every other vehicle in the same fleet. Not surprisingly, the fleet-specific credibility obtained in the next section will depend on this estimator.

4.3. Linear credibility predictors

An optimal BMS using both claims and safety offences would be more efficient than those designed in the preceding sections (see Pinquet (1998) for a comparison of short-term effects). We now consider the case where claims cannot be used and the frequency of claims is predicted from the history of safety violations only. Notice that an optimal BMS using both claims and safety offences could easily be obtained from the preceding estimators. It would be enough to adapt the linear credibility system given in the aforementioned paper to a stratified portfolio.

Let us compute the bonus-malus coefficient for the frequency of claims reported by the vehicle i_0 belonging to the fleet f_0. The fleet index is suppressed in order to simplify the expressions. The bonus-malus coefficient is written a $\widehat{a_{i_0}} + \widehat{b_{i_0}}\left(\sum_{i=1}^{m} \mathrm{inf}_i\right)$, with

$$(\widehat{a_{i_0}}, \widehat{b_{i_0}}) = \arg\min_{a,b} \widehat{E}\left[\left(\frac{U_{i_0}}{E(U_{i_0})} - a - b\left(\sum_{i=1}^{m} INF_i\right)\right)^2\right].$$

From computations similar to those performed in Section 3.3, we obtain the following bonus-malus coefficient

$$bonmal_{i_0} = 1 + \widehat{b_{i_0}}\left(\sum_{i=1}^{m} (\mathrm{inf}_i - \widehat{\tau_i})\right);\ \widehat{b_{i_0}} = \frac{\widehat{Cov}\left(\frac{U_{i_0}}{E(U_{i_0})}, \sum_{i=1}^{m} INF_i\right)}{\widehat{V}\left(\sum_{i=1}^{m} INF_i\right)},$$

with

$$\widehat{Cov}\left(\frac{U_{i_0}}{E(U_{i_0})}, \sum_{i=1}^{m} INF_i\right) = \widehat{V_{RP}}^1\left(\sum_{i=1}^{m}\widehat{\tau_i}\right) + \left((\widehat{V_{UT}}^1 - \widehat{V_{RP}}^1)\widehat{\tau_{i_0}}\right).$$

The last term must be suppressed if the vehicle i_0 is not observed during the first period. Following the computations of Section 3.3, we obtain then

$$\widehat{V}\left(\sum_{i=1}^{m} INF_i\right) = \left(\sum_{i=1}^{m}\widehat{\tau_i}\right) + \left[\widehat{V_{PP}}^1\left(\sum_{i=1}^{m}\widehat{\tau_i}\right)^2\right] + \left[\left(\widehat{V_{TT}}^1 - \widehat{V_{PP}}^1\right)\left(\sum_{i=1}^{m}\widehat{\tau_i}^2\right)\right];$$

$$bonmal_{i_0} = (1 - cred_{i_0}) + cred_{i_0}\frac{\sum_{i=1}^{m}\text{inf}_i}{\sum_{i=1}^{m}\widehat{\tau_i}};$$

$$cred_{i_0} = \alpha\,(i_0 \notin \{1,\ldots,m\});\; cred_{i_0} = \alpha + \beta_{i_0}(i_0 \in \{1,\ldots,m\});$$

$$\alpha = \frac{\widehat{V_{RP}}^1\left(\sum_{i=1}^{m}\widehat{\tau_i}\right)}{1 + \left(\widehat{V_{PP}}^1\left(\sum_{i=1}^{m}\widehat{\tau_i}\right)\right) + \left[\left(\widehat{V_{TT}}^1 - \widehat{V_{PP}}^1\right)\left(\frac{\sum_{i=1}^{m}\widehat{\tau_i}^2}{\sum_{i=1}^{m}\widehat{\tau_i}}\right)\right]};$$

$$\beta_{i_0} = \frac{\left(\widehat{V_{UT}}^1 - \widehat{V_{RP}}^1\right)\widehat{\tau_{i_0}}}{1 + \left(\widehat{V_{PP}}^1\left(\sum_{i=1}^{m}\widehat{\tau_i}\right)\right) + \left[\left(\widehat{V_{TT}}^1 - \widehat{V_{PP}}^1\right)\left(\frac{\sum_{i=1}^{m}\widehat{\tau_i}^2}{\sum_{i=1}^{m}\widehat{\tau_i}}\right)\right]}. \tag{14}$$

The fleet-specific credibility coefficient α increases with $\widehat{V_{RP}}^1$ a term related to the covariance between the two fleet-specific random effects. The coefficient β_{i_0} is the vehicle-specific credibility. It makes sense only if $\widehat{V_{UT}}^1 > \widehat{V_{RP}}^1$, a condition fulfilled in our data.

4.4. Empirical results

The frequency of claims with bodily injury reported in 1996 is predicted from the number of safety offences recorded in 1995, and we retained the vehicles with a positive duration of authorization for the licence plate during 1995 and 1996. The detailed results of the regression explaining the number of safety offences recorded in 1995 are presented in Table 6. Let us emphasize two points:

- The annual frequency of recorded offences is equal to 22.2%. It is much superior to that of the claims with bodily injury liability. This will explain later the better short term performance of the prediction designed in this section.
- The frequency of offences increases with the size of the fleet, but decreases for fleets with more than 20 vehicle-years.

TABLE 6

RATING SCORE FOR THE FREQUENCY OF SAFETY OFFENCES

	WEIGHT (%)	REL.FRE.	ST.COFF.	P-VALUE
VARIABLE: FIRM'S ACTIVITY SECTOR				
general merchandise transport	13.3	1.269	1.048	0.013
bulk transport	10.7	2.045	0.997	0.314
short term rental	2.5	1.297	1.742	<0.001
independent trucker, other sector	73.5	0.789	0.967	ref. group
VARIABLE: VEHICLES-YEARS				
0 or 1 vehicle-year	32.7	0.953	1.055	ref. group
2 vehicle-years	11.4	1.022	1.119	0.008
3 vehicle-years	7.1	1.147	1.174	<0.001
4 to 9 vehicle-years	17.4	1.262	1.210	<0.001
10 to 20 vehicle-years	9.7	1.256	1.056	0.725
more than 20 vehicle-years	21.7	0.686	0.606	<0.001
VARIABLE: TYPE OF FUEL				
gasoline	22.2	0.392	0.582	<0.001
fuel oil	77.8	1.174	1.119	ref. group
VARIABLE: WEIGHT OF THE VEHICLE				
from 3,000 to 3,870 kgs	20.0	0.653	0.991	0.371
from 3,871 to 6,220 kgs	20.0	0.654	0.939	<0.001
from 6,221 to 7,620 kgs	20.5	1.112	0.960	<0.001
from 7,621 to 8,850 kgs	19.2	1.517	1.061	0.473
more than 8,850 kgs	20.3	1.082	1.050	ref. group
VARIABLE: TYPE OF USE				
commercial use	76.0	0.776	0.927	0.644
bulk transport	10.2	2.356	1.667	<0.001
VARIABLE: NUMBER OF AXLES				
unknown	2.3	0.998	1.119	0.047
2 axles, less than 4,000 kgs	20.6	0.651	0.984	ref. group
2 axles, more than 4,000 kgs	27.6	0.600	0.751	<0.001
3 axles	18.3	0.802	0.696	<0.001
4 axles	5.7	0.893	0.856	0.017
5 axles	8.2	0.987	0.865	0.044
6 axles and more	17.3	2.305	1.836	<0.001
VARIABLE: NUMBER OF CYLINDERS				
1 to 5 cylinders	1.4	0.714	0.834	0.699
6 to 7 cylinders	59.0	1.294	1.130	<0.001
8 cylinders and more	39.6	0.572	0.812	ref. group
Number of vehicles		100,048		

As for the random effects, the numerical values of the estimators are

$$\widehat{V_{UT}}^1 = 0.519;\ \widehat{V_{PP}}^1 = 0.465;\ \widehat{V_{RP}}^1 = 0.141;\ \widehat{V_{TT}}^1 = 1.263.$$

These moment-based estimators can be connected to explicit distributions. If log-normal distributions are retained for the random effects, we can write

$$R = \exp(a_1 G_1);\ U_1 = \exp(a_1 G_1 + a_2 G_2^i) \Rightarrow U_i = RS_i, S_i = \exp(a_2 G_2^i).$$

The fleet index is suppressed, and the random variables $G_1, (G_2^i)_{i=1,\ldots,m}$ follow independent standard normal distributions. In the same way, we can write

$$P = \exp(a_3 G_1 + a_4 G_3); T_i = \exp(a_3 G_1 + a_4 G_3 + a_5 G_2^i + a_6 G_4^i) \Rightarrow T_i = PQ_i,$$
$$Q_i = \exp(a_5 G_2^i + a_6 G_4^i),$$

with similar assumptions on the random variables $G_3, (G_4^i)_{i=1,\ldots,m}$. It is easily seen that

$$G \sim N(0, I_q) \Rightarrow \frac{Cov({}^t aG, {}^t bG)}{E({}^t aG)E({}^t bG)} = \exp({}^t ab) - 1$$

$\forall$a, b $\in \mathbb{R}^q$. The moment-based estimators are then connected with the following values

$$a_1 = 0.381;\ a_2 = 0.828;\ a_3 = 0.346;\ a_4 = 0.512;\ a_5 = 0.346;\ a_6 = 0.562.$$

The predictor computed in this section cannot be consistent with respect to the fleet specific component, since the event for which the frequency is predicted is not retained in the history. When the size of the fleet m converges towards infinity, we obtain from equation (14)

$$\lim_{m \to +\infty} cred_{i_0} = \frac{\widehat{V_{RP}}^1}{\widehat{V_{PP}}^1} = 0.303 \forall i_0.$$

The credibility coefficient $cred_{i_0}$ is defined in (14). As we have the following limit

$$\lim_{m \to +\infty} \frac{\sum_{i=1}^m INF_i}{\sum_{i=1}^m \widehat{\tau_i}} = \frac{P}{E(P)}$$

in the random effects model, the limit of the bonus-malus coefficient is

$$\lim_{m \to +\infty} bonmal_{i_0} = \left(1 - \frac{\widehat{V_{RP}}^1}{\widehat{V_{PP}}^1}\right) + \left(\frac{\widehat{V_{RP}}^1}{\widehat{V_{PP}}^1} \frac{P}{E(P)}\right).$$

Hence, the BMS is not consistent (the limit should be $R/E(R)$ for a consistent predictor). The limit is the estimated affine regression of $R/E(R)$ with respect to $P/E(P)$.

Although this BMS is less efficient in the long run than the one based on the number of claims, it is more efficient after one year, as shown in Table 7.

TABLE 7

AVERAGE CREDIBILITIES FOR FLEETS AND VEHICLES
STANDARD DEVIATIONS OF BONUS-MALUS COEFFICIENTS AT THE FLEET LEVEL

Fleet size	$\bar{\alpha}$	$\bar{\alpha}+\bar{\beta}$	$\sigma_{bonmal_{\alpha}}$	$\sigma_{bonmal_{\alpha+\beta}}$
0 or 1 vehicle-year	0.027	0.094	0.064	0.216
2 vehicle-years	0.049	0.113	0.087	0.198
3 vehicle-years	0.070	0.132	0.107	0.196
from 4 to 9 vehicle-years	0.114	0.168	0.136	0.197
from 10 to 20 vehicle-years	0.175	0.209	0.186	0.222
more than 20 vehicle-years	0.242	0.252	0.220	0.235

Table 7 was obtained in the same way as Table 2. Standard deviations of bonus-malus coefficients are more important in this table for fleets with little or medium size. This BMS is less efficient in the long run than the one presented in Section 3, but it is closer to its limit, due to the higher frequency of safety offences.

5. Conclusion

The objective of this paper was to propose BMS for fleets of vehicles. The models were applied to fleets of trucks, but they could be used for other stratified portfolios if individual information on the insurance contracts was available.

Two systems were presented: one based on past accidents and the other based on past safety offenses. It was shown that the former system is more efficient in the long run, while it is outperformed by the latter BMS after one year, a result explained by the higher frequency of safety offences.

Many extensions of this article can be done. One is to use information on many periods in order to build up a panel. The corresponding panel would be very useful to analyze the stability of the BMS over time. It would also permit to verify for how long period the system based on safety offences dominates the one based on accidents. However such extensions will not be straightforward since we would have to introduce dynamic random effects along with the fleet effects in order to take into account the serial correlations.

References

1. Boyer, M. and Dionne, G. (1987) Description and analysis of the quebec automobile insurance plan. *Canadian Public Policy* **13**, 181-195.
2. Bühlmann, H. (1967) Experience rating and credibility. *ASTIN Bulletin* **4**, 199-207.
3. Devlin, R.A. (1992) Liability versus no-fault automobile insurance regimes: An analysis of the experience in Quebec. *Contributions to Insurance Economics* (Editor: Georges Dionne), Kluwer Academic Publishers. Huebner International Series on Risk, Insurance, and Economic Security, Boston.
4. 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.
5. Dionne, G., and Vanasse, C. (1997a) Une evaluation empirique de la nouvelle tarification de l'assurance automobile au Québec. *Actualité Economique* **73**, 47-80.
6. Dionne, G., Gagné, R., Gagnon, F. and Vanasse, C. (1997b) Debt, moral hazard and airline safety: An empirical evidence. *Journal of Econometrics* **79**, 379-402 (Special Issue on Duration, Transition and Count Data Models. Editors: Christian Gouriéroux and Thierry Magnac).
7. Dionne, G., Desjardins, D. and Pinquet, J. (1999) *L'évaluation du risque d'accident des transporteurs en fonction de leur secteur d'activité, de la taille de leur flotte et de leur dossier d'infractions.* Research Report 99-28, Centre de Recherches sur les Transports, Université de Montréal.
8. Dionne, G., Desjardins, D. and Pinquet, J. (2000a) Optimal insurance for fleets of vehicles under asymmetric information. Mimeo, Risk Management Chair, HEC Montreal.
9. Dionne, G., Maurice, M. and Pinquet, J. (2000b) The role of memory and saving in long-term contracting with moral hazard: An empirical evidence in automobile insurance. Mimeo, Risk Management Chair, HEC-Montreal.
10. Gouriéroux, C., Monfort, A. and Trognon, A. (1984) Pseudo likelihood methods: Applications to Poisson models. *Econometrica* **52**, 701-720.
11. Hausman, J.A., Hall, B.H. and Griliches, Z. (1984) Econometric models for count data with an application to the patents – R&D relationship. *Econometrica* **52**, 909-938.
12. Jewell, W.S. (1975) The use of collateral data in credibility theory: A hierarchical model. *Giornale dell'Instituto degli Attuari* **38**, 1-16 (quoted in "Advanced Risk Theory," F. Etienne de Vylder, Edition de l'Université de Bruxelles et Swiss Association of Actuaries, (1996)).
13. Lemaire, J. (1985) *Automobile Insurance: Actuarial Models.* Huebner International Series on Risk, Insurance and Economic Security, Kluwer Academic Publishers, Boston.
14. Liang, K.Y., and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.
15. Marie-Jeanne, P. (1994) *Problèmes spécifiques des flottes automobiles.* Proceedings of the ISUP conference "Cours Avancé sur l'Assurance Automobile".
16. Pinquet, J. (1997) Allowance for cost of claims in bonus-malus systems. *ASTIN Bulletin* **27**, No. 1, 33-57.
17. Pinquet, J. (1998) Designing optimal bonus-malus systems from different types of claims. *ASTIN Bulletin* **28**, No. 2, 205-220.
18. Pinquet, J. (2000) Experience rating through heterogeneous models. *Handbook of Insurance*, chapter 14 (Editor: Georges Dionne), Kluwer Academic Publishers. Huebner International Series on Risk, Insurance and Economic Security, 459-500.
19. Teugels, J.L. and Sundt, B. (1991). "A stop-loss experience rating scheme for fleets of cars". *Insurance: Mathematics and Economics*, North-Holland, 173-179.

Denise Desjardins
Centre de Recherche sur les Transports
Université de Montréal
C.P. 6128, Succursale Centre-ville

Montréal (Québec)
Canada H3C 3J7
email: denise@crt.umontreal.ca

GEORGES DIONNE
Ecole des Hautes Etudes Commerciales
Chaire de Gestion des Risques
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec)
Canada H3T 2A7
email: georges.dionne@hec.ca

JEAN PINQUET
U.F.R. de Sciences Economiques
Université de Paris X
200, avenue de la République
92001 Nanterre Cedex
France
email: pinquet@u-paris10.fr

ANALYTICAL EVALUATION OF ECONOMIC RISK CAPITAL FOR PORTFOLIOS OF GAMMA RISKS

BY

WERNER HÜRLIMANN

ABSTRACT

Based on the notions of value-at-risk and expected shortfall, we consider two functionals, abbreviated VaR and RaC, which represent the economic risk capital of a risky business over some time period required to cover losses with a high probability. These functionals are consistent with the risk preferences of profit-seeking (and risk averse) decision makers and preserve the stochastic dominance order (and the stop-loss order). Quantitatively, RaC is equal to VaR plus an additional stop-loss dependent term, which takes into account the average amount at loss. Furthermore, RaC is additive for comonotonic risks, which is an important extremal situation encountered in the modeling of dependencies in multivariate risk portfolios. Numerical illustrations for portfolios of gamma distributed risks follow. As a result of independent interest, new analytical expressions for the exact probability density of sums of independent gamma random variables are included, which are similar but different to previous expressions by Provost (1989) and Sim (1992).

KEYWORDS

Value-at-risk, expected shortfall, risk-adjusted capital, comonotonicity, additivity, supermodular order, stop-loss order, gamma convolutions.

1. ECONOMIC RISK CAPITAL USING VAR AND RAC

Suppose a firm is confronted with a risky business over some time period, and let the random variable X represent the potential *loss* or risk the firm incurs at the end of the period. To be able to cover any loss with a high probability, the firm borrows at the beginning of the time period on the capital market the amount ERC_0, called *economic risk capital*. At the end of the period, the firm has to pay interest on this at the interest rate i_R. To guarantee with certainty the value of the borrowed capital at the end of the period, the firm invests ERC_0 at the risk-free interest rate $i_f < i_R$. The value of the economic risk capital at the end of the period is thus $ERC = ERC_0 \cdot (1 + i_f - i_R)$. The risky business will be successful at the end of the period provided the event $\{X > ERC\}$ occurs only with a small tolerance probability.

There exist several risk management principles applied to evaluate *ERC*. Two simple methods that have been considered so far are the value-at-risk and the expected shortfall approach (e.g. Arztner et al. (1997a/b), Arztner (1999), Embrechts (1995), Hürlimann (1998a), Schröder (1996), Wirch (1999)). According to the value-at-risk method one identifies the economic risk capital with the *value-at-risk* of the loss setting

$$ERC = VAR_{\alpha}[X] := Q_X(\alpha), \tag{1.1}$$

where $Q_X(u) = \inf\{x \mid F_X(x) \geq u\}$ is a quantile function of X, with $F_X(x) = \Pr(X \leq x)$ the distribution of X. This quantile represents the maximum possible loss, which is not exceeded with the (high) probability α (called security level). According to the expected shortfall method one identifies the economic risk capital with the *risk-adjusted capital* of the loss setting

$$ERC = RaC_{\alpha}[X] := E[X \mid X > VaR_{\alpha}[X]]. \tag{1.2}$$

This value represents the conditional expected loss given the loss exceeds its value-at-risk. Clearly one has

$$RaC_{\alpha}[X] = Q_X(\alpha) + m_X[Q_X(\alpha)] = Q_X(\alpha) + \frac{1}{\varepsilon}\pi_X[Q_X(\alpha)], \tag{1.3}$$

where $m_X(x) = E[X - x \mid X > x]$ is the mean excess function, $\pi_X(x) = (1 - F_X(x)) \cdot m_X(x)$ is the stop-loss transform, and $\varepsilon = 1 - \alpha$ is interpreted as loss probability (called loss tolerance level). In Arztner (1999) the expression (4.3) is called *tail conditional expectation* and abbreviated TailVaR there (for *tail value-at-risk*). Mathematically, VaR and RaC, which have been defined as functions of random variables, may be viewed as functionals defined on the space of probability distributions associated with these random variables. By abuse of language, we will use the terminology functionals when appropriate.

It is important to observe that both *ERC* functionals satisfy two important risk-preference criteria in the economics of insurance (see Denuit et al. (1999) for a recent review). They are consistent with the risk preferences of profit-seeking decision makers respectively profit-seeking risk averse decision makers. To see this let us first recall two partial orders of riskiness.

Definitions 1.1. A risk X is less dangerous than a risk Y in the *stochastic order*, written $X \leq_{st} Y$, if $Q_X(u) \leq Q_Y(u)$ for all $u \in [0,1]$. A risk X is less dangerous than a risk Y in the *stop-loss order*, written $X \leq_{sl} Y$, if $\pi_X(x) \leq \pi_Y(x)$ for all x.

To compare economic risk capitals using criteria, which do not depend on the choice of the loss tolerance level, let us introduce two further partial orders of riskiness.

Definitions 1.2. A loss X is less dangerous than a loss Y in the *VaR order*, written $X \leq_{VaR} Y$, if the value-at-risk quantities satisfy $VaR_{\alpha}[X] \leq VaR_{\alpha}[Y]$, for all $a \in [0,1]$. A loss X is less dangerous than a loss Y in the *RaC order*, written $X \leq_{RaC} Y$, if the risk-adjusted capital quantities satisfy $RaC_a[X] \leq RaC_a[Y]$, for all $a \in [0,1]$.

The value-at-risk and expected shortfall methods are consistent with ordering of risks in the sense that profit-seeking (risk averse) decision makers require higher VaR (RaC) by increasing risk, where risk is compared using the stochastic order $\leq_{st}$ (stop-loss order $\leq_{sl}$). Reciprocally, increasing VaR (RaC) is always coupled with higher risk. The following result expresses these ordering properties mathematically.

Theorem 1.1. If X and Y are two loss random variables, then $X \leq_{VaR} Y \Leftrightarrow X \leq_{st} Y$ and $X \leq_{RaC} Y \Leftrightarrow X \leq_{sl} Y$.

Proof. Since $X \leq_{st} Y \Leftrightarrow Q_X(u) \leq Q_Y(u)$ for all $a \in [0,1]$, the first property is immediate by (1.1). Consider the Hardy-Littlewood transform defined by

$$HL_X(u) = \begin{cases} \frac{1}{1-u} \cdot \int_u^1 Q_X(t)dt, & u < 1 \\ Q_X(1), & u = 1 \end{cases} \tag{1.4}$$

Its name stems from the Hardy-Littlewood (1930) maximal function and has been extensively used in both theoretical and applied mathematics (e.g. Blackwell and Dubins (1963), Dubins and Gilat (1978), Meilijson and Nàdas (1979), Kertz and Rösler (1990/92/93), Rüschendorf (1991), Hürlimann (1998b/c/d)). One knows that there exists a random variable X^H associated to X such that (e.g. Hürlimann (1998b), Theorem 2.1)

$$HL_X(u) = Q_{X^H}(u) = Q_X(u) + m_X[Q_X(u)], \tag{1.5}$$

hence $RAC_a[X] = Q_{X^H}(a)$ by (1.3). The result follows from the fact that $X \leq_{sl} Y$ if and only if $X^H \leq_{st} Y^H$, where $\leq_{st}$ denotes the usual stochastic dominance order (e.g Kertz and Rösler (1992), Lemma 1.8, or Hürlimann (1998c), Theorem 2.3). For the convenience of the reader, an alternative perhaps more accessible proof should also be pointed out. Consider the so-called distortion function $g_a(x) = \min\left\{\frac{x}{1-a}, 1\right\}$. It is easy to show that

$$RaC_a[X] = \int_0^\infty g_a[1 - F_X(x)]\,dx \tag{1.6}$$

identifies the RaC functional with a member of the class of distortion pricing principles in Wang (1996). The result follows by Dhaene et al. (2000), Theorem 3, which contains a proof of the stated equivalence. ◊

Finally, it is important to observe that, except for a world of elliptical linear portfolio losses (Embrechts et al. (1998), Fundamental Theorem of Risk Management), the VaR functional has several shortcomings. It is not subadditive and not scalar multiplicative, and it cannot discriminate between risk-averse and risk-taking portfolios (examples 1 to 3 in Wirch (1999)). Some more details for the practitioner are in order. Recall that a risk measure $R[\cdot]$ acting on the set of all risks is *subadditive* provided $R[X+Y] \le R[X] + R[Y]$ for all X, Y, that is merging two risks does not create extra risk. If a firm must meet a requirement of extra economic risk capital that did not satisfy this property, the firm might separate in two subunits requiring less capital, a matter of concern for the supervising authority. A risk measure is *scalar multiplicative* provided $R[cX] = cR[X]$ for all X, all constants $c \ge 0$. In situations where no diversification occurs capital requirement depends on the size of the risk. In contrast to this, the RaC functional, which is subadditve and scalar multiplicative, is a coherent risk measure in the sense of Arztner et al. (1997) and appears thus more suitable in general applications. A recent work devoted to the evaluation of economic risk capital in life-insurance using the VaR and RaC approaches is Ballmann and Hürlimann (2000).

2. The maximum RaC for the aggregate risk of portfolios

An important but complex problem is the evaluation of RaC for the aggregate risk of portfolios. Let $\boldsymbol{X} = (X_1, ..., X_n)$ be a portfolio of multivariate risks, where the marginal risks X_i have distributions $F_i(x)$, $i = 1, ..., n$. In a first step, one is interested in the maximum RaC for the aggregate risk $S(X) = X_1 + ... + X_n$ whenever $\boldsymbol{X} \in D(F_1, ..., F_n)$, the set of all multivariate risks with given marginals $F_i(x)$. It will be shown below that the maximum RaC is attained when the margins X_i show the strongest possible dependence structure, an extremal situation for which one says that $X_1, ..., X_n$ are mutually comonotonic.

A multivariate loss $(X_1, ..., X_n)$ is called *comonotonic* whenever an increase of a single loss $X_i(\omega_1) < X_i(\omega_2)$ for two events ω_1, ω_2 implies a nondecrease of all other losses $X_j(\omega_1) \le X_j(\omega_2)$, $j \ne i$ (Schmeidler (1986), Yaari (1987)). For $\boldsymbol{X} \in D(F_1, ..., F_n)$ this is exactly the case when $X = \left(F_1^{-1}(U), ..., F_n^{-1}(U)\right)$ with U a uniform (0,1) random variable noting that $F_i^{-1}(U)$ has distribution F_i and F_i^{-1} is increasing for all i. The distribution F of a comonotonic random vector is determined by its marginal distributions F_i through the relationship $F(x_1, ..., x_n) = \min_{1 \le i \le n} \{F_i(x_i)\}$. Mathematically, four equivalent defining conditions of comonotonicity can be given.

Definition 2.1. (Bäuerle and Müller (1998)) The components of a random vector $\boldsymbol{X} = (X_1, ..., X_n) \in D(F_1, ..., F_n)$ are called mutually *comonotonic* if any of the following equivalent conditions hold:

(C1) The multivariate distribution $F(x_1, ..., x_n)$ of $(X_1, ..., X_n)$ identifies with the so-called Fréchet upper bound $F(x_1, ..., x_n) = \min_{1 \le i \le n} \{F_i(x_i)\}$.

(C2) There exists a random variable Z and non-decreasing real functions u_1, ..., u_n such that $(u_1(Z), ..., u_n(Z))$ has the distribution F.

(C3) The random vector $\left(F_1^{-1}(U),\dots,F_n^{-1}(U)\right)$, where U is uniformly distributed on [0,1], has distribution F.

(C4) There is a random vector $\boldsymbol{X}$, distributed as F, such that $X_i(\omega_1) < X_i(\omega_2)$ implies $X_j(\omega_1) \le X_j(\omega_2)$ for all $j \ne i$.

We need further the notion of supermodular order.

Definition 2.2. A random vector **X** precedes **Y** in the *supermodular order*, written $\mathbf{X} \le_{sm} \mathbf{Y}$, if $E[f(X)] \le E[f(Y)]$ for all supermodular functions f such that the expectations exist, where f is called *supermodular* if

$$f(x \wedge y) + f(x \vee y) \ge f(x) + f(y) \text{ for all } x, y, \in R^n, \tag{2.1}$$

with the notation $(x_1,\dots,x_n) \wedge (y_1,\dots,y_n) = (x_1 \wedge y_1,\dots,x_n \wedge y_n)$, $\wedge$ the minimum operator, and $(x_1,\dots,x_n) \vee (y_1,\dots,y_n) = (x_1 \vee y_1,\dots,x_n \vee y_n)$, $\vee$ the maximum operator.

Intuitively the notion of supermodular function can be grasped as follows. Let $x_1, \dots, x_n$ be n individual losses in a portfolio, and let $f(x_1, \dots, x_n)$ be the aggregate loss caused by these losses. Then supermodularity of the function f means that the influence on the aggregate loss of an increase of a single loss is greater, the higher the other losses are. In the literature supermodular functions are also called superadditive, and have been originally studied in applied mathematics and operations research (e.g. Marshall and Olkin (1979)). They have been extensively applied in economics (e.g. Topkis (1998)). The related supermodular order allows for a comparison of the strength of dependence between random vectors. Its origin in the statistical literature can be traced back to Block and Sampson (1988), Joe (1990), Meester and Shanthikumar (1993), Szekli et al. (1994), Shaked and Shanthikumar (1997). Actuarial applications of this order are discussed in Müller (1997), Bäuerle and Müller (1998), Goovaerts and Dhaene (1999).

To compare the riskiness of portfolios, one says that a portfolio $\boldsymbol{X} = (X_1, \dots, X_n)$ is less risky than a portfolio $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ if the corresponding aggregate risks $S(X) = X_1 + \dots + X_n$ and $S(Y) = Y_1 + \dots + Y_n$ are stop-loss ordered, that is $S(X) \le_{sl} S(Y)$. A sufficient condition for this is the supermodular order.

Theorem 2.1. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ be random vectors in $D(F_1, \dots, F_n)$ such that $\mathbf{X} \le_{sm} \mathbf{Y}$, then one has $S(X) \le_{sl} S(Y)$.

Proof. This is shown in Müller (1997), Theorem 3.1. ◊

The significance of the supermodular order for economic risk capital calculations is now immediate. Given two portfolios $\boldsymbol{X}, \boldsymbol{Y} \in D(F_1, \dots, F_n)$ such that $\mathbf{X} \le_{sm} \mathbf{Y}$, it is possible to compare the RaC of the aggregate risk $S(X)$ with the RaC of the aggregate risk $S(Y)$.

Corollary 2.1. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ be random vectors in $D(F_1, \dots, F_n)$ such that $\mathbf{X} \le_{sm} \mathbf{Y}$, then one has $RaC_a[S(X)] \le RaC_a[S(Y)]$ for all $a \in [0,1]$.

Proof. This is an immediate consequence of Theorem 2.1 and Theorem 1.1. ◊

Even more, one obtains that the portfolio $\boldsymbol{X}^c = \left(F_1^{-1}(U),...,F_1^{-1}(U)\right) \in D(F_1,..., F_n)$ with mutually comonotonic margins yields the maximum RaC.

Theorem 2.2. The maximum RaC for the aggregate risk of a portfolio with fixed marginal risks is attained at the portfolio with mutually comonotonic components, that is one has

$$\max_{X \in D(F_1,..., F_n)} \{RaC_a[S(X)]\} = RaC_a\left[S(X^c)\right]. \tag{2.2}$$

Proof. By the inequality of Lorentz (1953) (e.g. Theorem 5 in Tchen (1980)), one knows that $\mathbf{X} \leq_{\text{sm}} \mathbf{X^c}$ for all $\boldsymbol{X} \in D(F_1, ..., F_n)$. The result follows by Corollary 2.1. Alternatively, it is possible to prove directly that $\mathbf{X} \leq_{\text{sl}} \mathbf{X^c}$ for all $\boldsymbol{X} \in D(F_1, ..., F_n)$ as shown by Goovaerts et al. (2000) (see also Dhaene et al. (2000), Corollary 6). Then Theorem 1.1 implies the result. ◊

This result means that comonotonicity, which displays the strongest possible dependence structure, corresponds to the riskiest portfolio under all portfolios with the same marginal risks and requires the maximum RaC under all these portfolios. It is further remarkable that under a simple regularity condition the maximum RaC is an additive functional.

Theorem 2.3. Let $\boldsymbol{X}^c = (X_1, ..., X_n)$ be a portfolio of mutually comonotonic risks with absolutely continuous marginal distributions $F_i(x)$, $i = 1, ..., n$. Then the RaC functional satisfies the additive property

$$RaC_a\left[S(X^c)\right] = \sum_{i=1}^{n} RaC_a[X_i]. \tag{2.3}$$

Proof. Denote by $F_s(x)$ the distribution of $S(X^c)$. Consider the quantiles $d = Q_S(a)$, $d_i = Q_{x_i}(a)$, $i = 1, ..., n$, and the stop-loss transforms $\pi_s(x)$, $\pi_i(x) := \pi_{x_i}(x)$, $i = 1, ..., n$. By the comonotonic assumption, the quantiles and stop-loss transforms behave additively, that is one has $d = \sum_{i=1}^{n} d_i$ (e.g. Landsberger and Meilijson (1994), Denneberg (1994), Kaas et al. (2000)) and $\pi_s(d) = \sum_{i=1}^{n} \pi_i(d_i)$ (Dhaene et al. (2000), Theorem 8, special case of absolutely continuous distributions, or Kaas et al. (2000)). The assertion follows from (1.3) using the relationship $(1-F_X(x)) \cdot m_X(x) = \pi_X(x)$ between the mean excess function and the stop-loss transform by means of the equalities

$$RaC_a\left[S(X^c)\right] = d + \tfrac{1}{\varepsilon} \cdot \pi_s(d) = \sum_{i=1}^{n} \left\{d_i + \tfrac{1}{\varepsilon} \cdot \pi_i(d_i)\right\} = \sum_{i=1}^{n} RaC_a[X_i]. \ \diamond.$$

Remark 2.1.

As pointed out by a referee, the additive relation (2.3) is a special case of a more general result due to Dellacherie (1970) and quoted in Schmeidler

(1986). Let A be a σ-algebra of subsets of a set S, and χ the set of all bounded real-valued A-measurable functions on S. For a monotone set function v on S such that $v(\emptyset) = 0$, $v(S) = 1$, and a non-negative real valued function $X \in \chi$, consider the Choquet integral $H_v[X] = \int_S X dv = \int_0^\infty v(X > x) dx$. Dellacherie's result states that if $X, Y \in \chi$ are comonotonic, then $H_v[X + Y] = H_v[X] + H_v[Y]$. In the special case of a probability space (Ω, P, A), consider the distortion function $g_a(x) = \min\left\{\frac{x}{1-a}, 1\right\}$ and the set function $v = g_a \circ P$. With (1.6) one obtains $RaC_a[X] = \int_0^\infty g_a[P(X > x)] dx = H_v[X]$. The additivity (2.3) for comonotonic risks follows from Dellacherie's result. However, note that our Theorem 2.3 is not restricted to bounded random variables, an essential assumption in Schmeidler's paper.

An interesting problem concerns the impact of various "positive" dependence structures between risks $X_1, \ldots, X_n$ on the evaluation of RaC for the aggregate risk $S(X) = X_1 + \ldots + X_n$. Independent risks with an aggregate denoted by $S^i = X_1 \oplus \ldots \oplus X_n$ and comonotonic risks with an aggregate $S^c = X_1 +_c \ldots +_c X_n$ are two extreme cases of primary importance. Let us motivate this assertion. In virtue of Corollary 2.1 and Theorem 2.2 it seems reasonable to restrict the attention to positive dependent portfolios $X \in D(F_1, \ldots, F_n)$ satisfying the supermodular inequality $\mathbf{X^i} \leq_{\mathrm{sm}} X \leq_{\mathrm{sm}} \mathbf{X^c}$, which implies $S^i \leq_{sl} S(X) \leq_{sl} S^c$ and $S^i \leq_{RaC} S(X) \leq_{RaC} S^c$. As an example, the family of multivariate elliptically contoured distributions is increasing in the supermodular order as the correlation increases (Block and Sampson (1988)). Portfolios satisfying only the stop-loss inequality $S^i \leq_{sl} S(X) \leq_{sl} S^c$, which by Theorem 1.1 is sufficient to imply $S^i \leq_{RaC} S(X) \leq_{RaC} S^c$, might also be of interest (e.g Bäuerle and Müller (1998), Section 4).

It is well-known that the stop-loss order relation $S^i \leq_{sl} S^c$ implies a considerable difference between the corresponding stop-loss premiums. However, the quantitative impact of this relation on the evaluation of RaC has not yet been examined. The additive property of Theorem 2.3 is of evident help for the quantitative analysis of the property $RaC_a[S^i] \leq RaC_a[S^c]$. Since insurance risks are often quite well approximated by gamma distributed risks or translations thereof (e.g. Seal (1977), Dufresne et al. (1991), Dickson and Waters (1993)), we will restrict ourselves in the present paper to a quantitative evaluation of this inequality for gamma risks. Since the exact distribution of sums of independent gamma random variables is not very well-known among actuaries, the next Section is of additional independent interest.

3. Sums of independent gamma random variables

Gamma distributions, which include the exponential, Erlang and chi-square distributions, are among the most important distributions widely used in applications. They are also of great importance in theoretical work. Thorin

(1977) introduced the class of generalized gamma convolutions, defined as the smallest class of distributions on the positive real line that contains the gamma distributions and is closed with respect to convolution and weak limits, to prove the infinite divisibility of many distributions. The class of generalized gamma convolutions is surprisingly rich and has a remarkable structure. It has been extensively studied in the last century by Bondesson (1992).

Though not noticed in actuarial science (e.g. one misses them in Panjer and Willmot (1992)), expressions for the exact probability density of sums of *independent* gamma random variables are known from the statistical literature. For example, Johnson et al. (1994), pp. 384-85, refers to Mathai (1982), Moschopoulos (1985) and Sim (1992). One can add Provost (1989), which determines the exact density applying the inverse Mellin transform. The result by Sim (1992) uses the following direct elementary approach. Let $X_i \sim \Gamma(a_i, \beta_i)$, $i = 1, ..., n$, be n independent gamma random variables with densities

$$(3.1) \qquad f_i(x) = g(\beta_i x; a_i) := \frac{(\beta_i x)^{a_i}}{\Gamma(a_i)} \cdot \frac{e^{-\beta_i x}}{x}, \quad x > 0,\ a_i,\ \beta_i > 0.$$

The special case of identical scale parameters being well-known, one assumes that $\beta_1 > \beta_2 > ... > \beta_n$. The density of the independent sum $S_n = X_1 \oplus ... \oplus X_n$ can be obtained from the convolution formula

$$(3.2) \qquad f_{S_n}(s) = \int_0^s f_{S_{n-1}}(t) \cdot f_n(s-t)\,dt$$

applying mathematical induction. A calculation yields the result by Sim (1992) (see also Johnson et al. (1994), formula (17.110)):

$$(3.3) \quad f_{S_n}(s) = \frac{1}{\Gamma(a^{(n)})} \cdot \left[\prod_{i=1}^{n} \beta_i^{a_i}\right] \cdot s^{a^{(n)}-1} \cdot e^{-\beta_1 s} \cdot \sum_{k=0}^{\infty} C_k^n \cdot \frac{\left(a^{(n-1)}\right)_k}{k! \cdot \left(a^{(n)}\right)_k} \cdot [(\beta_1 - \beta_2)s]^k,$$

where $a^{(k)} = a_1 + ... + a_k$, $(c)_k = \dfrac{\Gamma(c+k)}{\Gamma(c)}$, and

$$(3.4) \qquad C_k^i = \begin{cases} 1, & i = 2 \\ \displaystyle\sum_{j=0}^{k} C_j^{i-1} \cdot \binom{k}{j} \cdot \frac{\left(a^{(i-2)}\right)_j}{\left(a^{(i-1)}\right)_j} \cdot \left[\frac{\beta_{n-i+2} - \beta_{n-i+3}}{\beta_{n-i+1} - \beta_{n-i+2}}\right]^j, & i = 3, ..., n. \end{cases}$$

A rearrangement shows that (3.3) is an infinite linear combination of gamma densities with the same scale parameter β_1, a property already observed by Provost (1989). Applying another elementary approach, we obtain below a new similar representation of the exact probability density, which differs from the results by Provost (1989) and Sim (1992).

Theorem 3.1. Let $X_i \sim \Gamma(\alpha_i, \beta_i)$ be n independent gamma random variables such that $\beta_1 > \beta_2 > \ldots > \beta_n$. Then the density of the independent sum $S_n = X_1 \oplus \ldots \oplus X_n$ is analytically described by the infinite series

$$f_{S_n}(s) = A_n \cdot \sum_{k=0}^{\infty} C_k^n \cdot g(\beta_1 s; \alpha^{(n)} + k), \text{ with} \tag{3.5}$$

$$\alpha^{(n)} = \alpha_1 + \ldots + \alpha_n, \quad A_n = \prod_{j=2}^{n} \left[\frac{\beta_j}{\beta_1}\right]^{\alpha_j}, \quad n \geq 2, \quad A_1 = 1,$$

$$C_k^i = \begin{cases} \left[1 - \dfrac{\beta_2}{\beta_1}\right]^k \cdot \dfrac{(\alpha_2)_k}{k!}, & i = 2 \\ \displaystyle\sum_{j=0}^{k} C_j^{i-1} \cdot \left[1 - \frac{\beta_i}{\beta_1}\right]^{k-j} \cdot \frac{(\alpha_i)_{k-j}}{(k-j)!}, & i = 3, \ldots, n \end{cases} \tag{3.6}$$

$$C_k^1 = 0 \quad k = 1, 2, \ldots, \quad C_0^i = 1 \quad i = 1, \ldots, n.$$

Proof. This is shown through induction. Clearly, the series representation holds for n=1. By induction, assume the representation holds for the index n and show it for the index n+1. For convenience set

$$S_n = X_n \oplus S_{n-1} \in (0, \infty), \quad R_n = \frac{X_n}{S_n} \in (0, 1). \tag{3.7}$$

Applying the standard method of transformation of random variables based on Jacobians (e.g. Fisz (1973), p. 77), the density of the sum S_n is determined recursively by the formulas

$$f_{S_n}(s) = \int_0^1 f_{(S_n, R_n)}(s, r)\, dr, \tag{3.8}$$

$$f_{(S_n, R_n)}(s, r) = s \cdot f_{X_n}(rs) \cdot f_{S_{n-1}}((1-r)s).$$

Using this and the induction assumption, one obtains

$$f_{S_{n+1}}(s) = A_n \cdot \sum_{k=0}^{\infty} C_k^n \cdot s^{\alpha^{(n+1)} + k - 1} \cdot e^{-\beta_1 s} \cdot \frac{\beta_1^{\alpha^{(n)} + k}}{\Gamma(\alpha^{(n)} + k)} \cdot I(n, k), \tag{3.9}$$

with

$$I(n,k) = \int_0^1 r^{\alpha_{n+1} - 1} \cdot (1-r)^{\alpha^{(n)} + k - 1} \cdot e^{-(\beta_{n+1} - \beta_1)^{rs}}\, dr \tag{3.10}$$

$$= \sum_{j=0}^{\infty} \frac{(\beta_1 - \beta_{n+1})^j}{j!} \cdot s^j \cdot \frac{\Gamma(\alpha^{(n)} + k)\Gamma(\alpha_{n+1} + j)}{\Gamma(\alpha^{(n+1)} + k + j)}.$$

Through rearrangement it follows that

$$(3.11)\quad \begin{aligned} f_{s_{n+1}}(s) &= A_n \cdot \left[\frac{\beta_{n+1}}{\beta_1}\right]^{a_{n+1}} \cdot \sum_{k=0}^{\infty} C_k^n \cdot \left\{ \sum_{j=0}^{\infty} \frac{\left(1-\frac{\beta_{n+1}}{\beta_1}\right)^j \Gamma(a_{n+1}+j)}{j!\Gamma(a_{n+1})} \cdot g(\beta_1 s; a^{(n+1)}+k+j) \right\} \\ &= A_{n+1} \cdot \sum_{k=0}^{\infty} C_k^{n+1} \cdot g(\beta_1 s; a^{(n+1)}+k). \end{aligned}$$

The analytical formula (3.5) is shown. ◊

Remark 3.1.

Though the coefficients of $g(\beta_1 s; a^{(n)}+k)$ in (3.3) and (3.5) are evaluated using different expressions, they are identical. However, the formulas (3.6) are more symmetric and simpler, and for this reason they should be preferred.

Using the incomplete gamma function defined by

$$(3.12)\qquad G(\beta x; a) = \frac{1}{\Gamma(a)} \cdot \int_0^{\beta x} t^{a-1} e^{-t}\, dt,$$

the distribution function of an independent gamma sum is through integration of (3.5) equal to

$$(3.13)\qquad F_{s_n}(s) = A_n \cdot \sum_{k=0}^{\infty} C_k^n \cdot G(\beta_1 s; a^{(n)}+k).$$

The evaluation of RaC for portfolios of independent gamma risks requires an analytical expression for the stop-loss transform of S_n.

Corollary 3.1. The stop-loss transform $\pi_{s_n}(d) = E[(S_n - d)_+]$ of a sum $S_n = X_1 \oplus \ldots \oplus X_n$ of n independent gamma random variables $X_i \sim \Gamma(a_i, \beta_i)$, $i = 1, \ldots, n$, such that $\beta_1 > \beta_2 > \ldots > \beta_n$, is determined by the analytical formula

$$(3.14)\quad \pi_{s_n}(d) = E[S_n] - d \cdot \overline{F}_{s_n}(d) - \frac{1}{\beta_1} \cdot A_n \cdot \sum_{k=0}^{\infty} (a^{(n)}+k) \cdot C_k^n \cdot G(\beta_1 d; a^{(n)}+k+1)$$

Proof. This follows without difficulty noting that $\pi_{s_n}(d) = E[S_n] - d \cdot \overline{F}_{s_n}(d) - \int_0^d s f_{s_n}(s) ds$, and $\int_0^d s g(\beta s; a) ds = \frac{a}{\beta} G(\beta d; a+1)$. ◊

Remark 3.2.

The special case $\beta_1 = ... = \beta_n = \beta$ of identical scale parameters is well-known. In this situation, the above formulas are replaced by the very simple ones

$$f_{s_n}(s) = g(\beta s; \alpha^{(n)}), \quad F_{s_n}(s) = G(\beta s; \alpha^{(n)}),$$
$$\pi_{s_n}(d) = \left(\frac{\alpha^{(n)}}{\beta} - d\right) \cdot \overline{F}_{s_n}(d) + \frac{d}{\beta} \cdot f_{s_n}(s), \tag{3.15}$$

where the last one is obtained through partial integration.

4. Numerical examples

First, let us calculate RaC for portfolios of independent gamma risks. Given the loss tolerance level ε, first determine using (3.13) the solution d_ε of the equation $F_{s_n}(d_\varepsilon) = 1 - \varepsilon$. Inserting the obtained value in (3.14), one obtains the formula

$$RaC_{1-\varepsilon}\left[S_n^i\right] = d_\varepsilon + \frac{1}{\varepsilon} \cdot \pi_{s_n}(d_\varepsilon)$$
$$= \frac{1}{\varepsilon} \cdot \left[E[S_n] - \frac{1}{\beta_1} \cdot A_n \cdot \sum_{k=0}^{\infty} (\alpha^{(n)} + k) \cdot C_k^n \cdot G(\beta_1 d_\varepsilon; \alpha^{(n)} + k + 1)\right]. \tag{4.1}$$

In the special case $\beta_1 = ... = \beta_n = \beta$ of Remark 3.2, the quantile d_ε is solution of $G(\beta d_\varepsilon; \alpha^{(n)}) = 1 - \varepsilon$ and (4.1) simplifies to

$$RaC_{1-\varepsilon}\left[S_n^i\right] = E[S_n] \cdot \left\{1 + \frac{1}{\varepsilon} \cdot \frac{d_\varepsilon}{\alpha^{(n)}} \cdot g(\beta d_\varepsilon; \alpha^{(n)})\right\}. \tag{4.2}$$

Second, let us calculate RaC for portfolios of comonotonic gamma risks. The evaluation uses the additive property of Theorem 2.2. For $i = 1, ..., n$, determine the solution $d_{i,\varepsilon}$ of $G(\beta_i d_{i,\varepsilon}; \alpha_i) = 1 - \varepsilon$. Replacing $\alpha^{(n)}$ by α_i in (3.15) one gets

$$\pi_i(d_{i,\varepsilon}) = \varepsilon \cdot \left(E[X_i] - d_{i,\varepsilon}\right) + \frac{d_{i,\varepsilon}}{\beta_i} \cdot g(\beta_i d_{i,\varepsilon}; \alpha_i).$$

It follows that

$$RaC_{1-\varepsilon}\left[S_n^c\right] = \sum_{i=1}^{n} RaC_{1-\varepsilon}[X_i] = \sum_{i=1}^{n} \left\{d_{i,\varepsilon} + \frac{1}{\varepsilon} \cdot \pi_i(d_{i,\varepsilon})\right\}$$
$$= E[S_n] + \frac{1}{\varepsilon} \cdot \sum_{i=1}^{n} \frac{d_{i,\varepsilon}}{\beta_i} \cdot g(\beta_i d_{i,\varepsilon}; \alpha_i). \tag{4.3}$$

Example 4.1: independence versus comonotonic assumption

In the special case $\beta_1 = ... = \beta_n = \beta$ of Remark 3.2, a comparison of (4.2) and (4.3) yields the difference formula for RaC:

$$(4.4)\quad \beta\varepsilon \cdot \left(RaC_{1-\varepsilon}\left[S_n^c\right] - RaC_{1-\varepsilon}\left[S_n^i\right]\right) = \sum_{i=1}^{n} \frac{d_{i,\varepsilon}}{\beta_i} \cdot g(\beta_i d_{i,\varepsilon}; \alpha_i) - d_\varepsilon \cdot g(\beta d_\varepsilon; \alpha^{(n)}),$$

where $G(\beta d_\varepsilon; \alpha^{(n)}) = 1-\varepsilon$ and $G(\beta_i d_{i,\varepsilon}; \alpha_i) = 1-\varepsilon$, $i = 1,..., n$. A numerical illustration for the exponential case $\beta = 1$, $\alpha_1 = ... = \alpha_n = 1$, is summarized in Table 4.1 below. In this situation $RaC_{1-\varepsilon}\left[S_n^c\right] = n \cdot \{1-\ln(\varepsilon)\}$ depends *linearly* on the number of exponential risks. The difference increases non-linearly according to the formula

$$(4.5)\qquad RaC_{1-\varepsilon}\left[S_n^c\right] - RaC_{1-\varepsilon}\left[S_n^i\right] = \frac{1}{\varepsilon} \cdot \left\{ n \cdot \varepsilon \cdot [-\ln(\varepsilon)] - \frac{d_\varepsilon^n e^{-d_\varepsilon}}{(n-1)!} \right\},$$

where $G(d_\varepsilon; n) = 1-\varepsilon$. As an interesting observation, one notes for increasing n a *decreasing percentage increase* of $RaC_{1-\varepsilon}\left[S_n^i\right]$ over the ε-range between 0.05 and 0.001.

TABLE 4.1
RaC FOR EXPONENTIAL RISK PORTFOLIOS UNDER INDEPENDENCE (I) AND COMONOTONIC (C) ASSUMPTION

n	ε = 0.05		ε = 0.01		ε = 0.001	
	(i)	(c)	(i)	(c)	(i)	(c)
1	4	4	5.6	5.6	7.9	7.9
2	5.9	8	7.8	11.2	10.3	15.8
3	7.6	12	9.6	16.8	12.4	23.7
4	9.2	16	11.4	22.4	14.3	31.6
5	10.7	20	13	28	16.1	39.5
10	17.6	40	20.5	56.1	24.2	79.1
20	30.3	79.9	34	112.1	38.6	158.2
50	65.7	199.8	70.9	280.3	77.3	395.4
100	121.7	399.6	128.7	560.5	137.2	790.8

Example 4.2: sums of independent gamma risks versus gamma and normal approximations

Suppose an insurer desires to calculate VaR and RaC for a portfolio of n independent risks, which follow a classical risk model. Each risk $X_i = Y_{i,1} + ... + Y_{i,N_i}$ has a compound Poisson distribution, where N_i is Poisson distributed and the $Y_{i,j}$'s are the individual claims. Assume that one knows the expected number of claims $\lambda_i = E[N_i]$, as well as the first and second moments $v_i = E[Y_{i,1}]$, $m_{i,2} = E\left[Y_{i,1}^2\right]$ of the severity distributions, $i = 1, ..., n$. Let $k_i = c_i \cdot \lambda_i^{-\frac{1}{2}}$, with $c_i = \sqrt{m_{2,i}} \cdot v_i^{-1}$, be the coefficient of variation of X_i. As mentioned previously, it is often possible to assume that X_i is gamma distributed with parameters

$$\alpha_i = \frac{1}{k_i^2} = \frac{\lambda_i}{c_i^2}, \quad \beta_i = \frac{1}{c_i^2} \cdot \frac{1}{v_i}, \quad i = 1,..., n. \tag{4.6}$$

The risk $S_n = X_1 \oplus ... \oplus X_n$ of the portfolio is again compound Poisson distributed with corresponding parameters

$$\lambda = \sum_{i=1}^{n} \lambda_i, \quad v = \sum_{i=1}^{n} \left[\frac{\lambda_i}{\lambda}\right] \cdot v_i, \quad k = c \cdot \lambda^{-\frac{1}{2}}, \quad c^2 = \sum_{i=1}^{n} \left[\frac{\lambda_i}{\lambda}\right] \cdot \left[\frac{v_i}{v}\right]^2 \cdot c_i^2. \tag{4.7}$$

Now, it is possible to approximate S_n either by a sum of n independent gamma risks with the parameters α_i, β_i in (4.6) or by a gamma risk with parameters $\alpha = \frac{\lambda}{c^2}$, $\beta = \frac{1}{c^2} \cdot \frac{1}{v}$ as defined in (4.7). To illustrate, we compare the VaR and RaC values of these approximations with the values obtained from a normal approximation for portfolios of 5 risks with parameters (typical for the aggregate claims of collectives of life insurance policies):

$$\begin{aligned} &(\lambda_1,..., \lambda_5) = m \cdot (1,1,1,1,1), \quad m = 1,2,5,10,20,50, \\ &(v_1,..., v_5) = (2,2,1,3,2), \quad (c_i,..., c_5) = (1.25,1.75,2.5,1.5,2). \end{aligned} \tag{4.8}$$

The parameters for the overall gamma approximation are by (4.7) equal to $\lambda = 5m$, $v = 2$, $c = 1.74642$. Table 4.2 shows that the VaR and RaC values of both gamma approximations differ only slightly, but the normal approximation underestimates systematically these values, especially for small λ_i's and more considerably for RaC than for VaR.

TABLE 4.2

VaR AND RaC COMPARISONS BY FIXED $\varepsilon = 0.05$

m	VaR			RaC		
	normal approx.	**gamma approx.**	**sum of ind. gamma**	**normal approx.**	**gamma approx.**	**sum of ind. gamma**
1	22.8	25.3	25.3	26.1	32.1	32.4
2	38.2	40.9	41.0	42.8	49.1	49.5
5	78.7	81.8	81.9	86.0	92.6	93.0
10	140.6	143.8	144.0	150.9	157.6	158.1
20	257.5	260.7	260.9	272.0	278.8	279.3
50	590.8	594.2	594.4	613.9	620.7	621.2

REFERENCES

Artzner, P. (1999) Application of coherent risk measures to capital requirements in insurance. *North American Actuarial Journal* **3(2)**, 11-25.

ARTZNER, P., DELBAEN, F., EBER, J.-M. and HEATH D. (1997a) *Definition of coherent measures of risk.* Risk Management Symposium of the European Finance Association Meeting, Vienna, August 27-30, 1997.

ARTZNER, P., DELBAEN, F., EBER, J.-M. and HEATH D. (1997b) Thinking coherently. *RISK* 10, no. 11 (November 1997), 68-70.

BALLMANN, M., and HÜRLIMANN W. (2000) *Zur Modellierung des ökonomischen Risikokapitals in der Lebensversicherung*. Manuscript.

BÄUERLE, N., and MÜLLER A. (1998) Modeling and comparing dependencies in multivariate risk portfolios. *ASTIN Bulletin* **28(1)**, 59-76.

BLACKWELL, D. and DUBINS L.E. (1963) A converse to the dominated convergence theorem. *Illinois Journal of Mathematics* **7**, 508-514.

BLOCK, H.W. and SAMPSON A.R. (1988) Conditionally ordered distributions. *Journal of Multivariate Analysis* **27**, 91-104.

BONDESSON, L. (1992) *Generalized gamma convolutions and related classes of distributions*. Lecture Notes in Statistics 76. Springer-Verlag.

DELLACHERIE, C. (1970) *Quelques commentaires sur les prolongements de capacités*. Séminaire de Probabilités V, Strasbourg. Lecture Notes in Mathematics 191. Springer-Verlag, Berlin and New York.

DENNEBERG, D. (1994) Non-Additive Measure and Integral. Theory and Decision Library, Series B, vol. 27. Kluwer Academic Publishers.

DENUIT, M., DHAENE, J. and VAN WOUVE M. (1999) The Economics of Insurance: a review and some recent developments. *Bulletin of the Swiss Association of Actuaries*, 137-175.

DHAENE, J., WANG, S., YOUNG, V. and GOOVAERTS M.J. (2000) Comonotonicity and maximal stop-loss premiums. *Bulletin of the Swiss Association of Actuaries*, 99-113.

DICKSON, D.C.M., and WATERS H.R. (1993) Gamma processes and finite time survival probabilities. *ASTIN Bulletin* **23(1)**, 259-272.

DUBINS, L.E. and GILAT D. (1978) On the distribution of the maxima of martingales. *Transactions of the American Mathematical Society* **68**, 337-38.

DUFRESNE, F., GERBER, H.U. and SHIU E.S.W. (1991) Risk theory with the Gamma process. *ASTIN Bulletin* **21(2)**, 177-192.

EMBRECHTS, P. (1995) *A survival kit to quantile estimation*. UBS Quant Workshop, Zürich.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH Th. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

EMBRECHTS, P., MCNEIL, A. and STRAUMANN D. (1998) *Correlation and dependency in Risk Management: properties and pitfalls*. Preprint, ETHZ.

FISZ, M. (1973) *Wahrscheinlichkeitrechnung und mathematische Statistik*. VEB Deutscher Verlag der Wissenschaften. Berlin.

GOOVAERTS, M.J. and DHAENE J. (1999) Supermodular ordering and stochastic annuities. *Insurance: Mathematics and Economics* **24**, 281-290.

GOOVAERTS, M.J., DHAENE, J. and DE SCHEPPER A. (2000) Stochastic upper bounds for present value functions. *Journal of Risk and Insurance Theory* **67**, 1-14.

HARDY, G.H. and LITTLEWOOD J.E. (1930) A maximal theorem with function-theoretic applications. *Acta Mathematica* **54**, 81-116.

HÜRLIMANN, W. (1998a) *Risk measurement and financial gain: a paradox and a reconciliation*. Proceedings 2nd International Congress on Insurance: Mathematics and Economics, Lausanne, July 20-22, 1998. Unpublished improved manuscript.

HÜRLIMANN, W. (1998b) On distribution-free safe layer-additive pricing. *Insurance: Mathematics and Economics* **22**, 277-85.

HÜRLIMANN, W. (1998c) On stop-loss order and the distortion pricing principle. *ASTIN Bulletin* **28(1)**, 119-34.

HÜRLIMANN, W. (1998d) *Inequalities for lookback option strategies and exchange risk modelling*. Proceedings of the 1st Euro-Japanese Workshop on Stochastic Modelling for Finance, Insurance, Production and Reliability, Brussels, 1998.

JOE, H. (1990) Multivariate concordance. *Journal of Multivariate Analysis* **35**, 12-30.

JOHNSON, N.L., KOTZ, S. and BALAKRISHNAN N. (1994) *Continuous Univariate Distributions* (2nd ed.), vol. 1. J. Wiley, New York.

KAAS, R., DHAENE, J. and GOOVAERTS M.J. (2000) *Upper and lower bounds for sums of random variables*. Paper presented at the 4th International Congress Insurance: Mathematics and Economics. Barcelona, July 24-26, 2000.

KERTZ, R.P. and RÖSLER U. (1990) Martingales with given maxima and terminal distributions. *Israel Journal of Mathematics* **69**, 173-192.

KERTZ, R.P. and RÖSLER U. (1992) Stochastic and convex orders and lattices of probability measures, with a martingale interpretation. *Israel Journal of Mathematics* **77**, 129-164.

KERTZ, R.P. and RÖSLER U. (1993) *Hyperbolic-concave functions and Hardy-Littlewood maximal functions*. In: Stochastic Inequalities. IMS Lecture Notes – Monograph Series, vol. 22.

LANDSBERGER, M. and MEILIJSON I. (1994) Co-monotone allocations, Bickel-Lehman dispersion and the Arrow-Pratt measure of risk aversion. *Annals Operations Research* **52**, 97-106.

LORENTZ, G.G. (1953) An inequality for rearrangements. *American Mathematical Monthly* **60**, 178-79.

MARSHALL, A.W. and OLKIN I. (1979) *Inequalities, Theory of Majorization and Its Applications*. Academic Press, New York.

MATHAI, A.M. (1982) Storage capacity of a dam with gamma type inputs. *Annals of the Institute of Statistical Mathematics* **34**, 591-97.

MEESTER, L.E. and SHANTHIKUMAR J.G. (1993) Regularity of stochastic processes: a theory based on directional convexity. *Probability in the Engineering and Informational Sciences* **7**, 343-360.

MEILIJSON, I. and NÀDAS A. (1979) Convex majorization with an application to the length of critical paths. *Journal of Applied Probability* **16**, 671-77.

MOSCHOPOULOS, P.G. (1985) The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics* **37**, 541-44.

MÜLLER, A. (1997) Stop-loss order for portfolios of dependent risks. *Insurance: Mathematics and Economics* **21**, 219-223.

PANJER, H.H. and WILLMOT G.E. (1992) *Insurance Risk Models. Society of Actuaries*, Schaumburg, Illinois.

PROVOST, S.B. (1989) On sums of independent gamma random variables. *Statistics* **20**, 583-91.

RÜSCHENDORF, L. (1991) On conditional stochastic ordering of distributions. *Advances in Applied Probability* **23**, 46-63.

SCHMEIDLER, D. (1986) Integral representation without additivity. *Proceedings American Mathematical Society* **97**, 255-261.

SCHRÖDER, M. (1996) T*he Value at Risk Approach – Proposals on a Generalization*. Aktuarielle Ansätze für Finanz-Risiken, AFIR 1996, vol. 1, 151-170. Verlag Versicherungswirtschaft, Karlsruhe.

SEAL, H.L. (1977) Approximations to risk theory's F(x,t) by means of gamma distributions. *ASTIN Bulletin* **9**, 213-18.

SHAKED, M. and SHANTHIKUMAR J.G. (1997) Supermodular stochastic orders and positive dependence of random vectors. *Journal of Multivariate Analysis* **61**, 86-101.

SIM, C.H. (1992) Point processes with correlated gamma interarrival times. *Statistics and Probability Letters* **15**, 135-141.

SZEKLI, R., DISNEY, R.L. and HUR S. (1994) MR/G1/1 queues with positively correlated arrival stream. *Journal of Applied Probability* **31**, 497-514.

TCHEN, A.H. (1980) Inequalities for distributions with given marginals. *The Annals of Probability* **8(4)**, 814-27.

THORIN, O. (1977) On the infinite divisibility of the lognormal distribution. *Scandinavian Actuarial Journal*, 121-148.

TOPKIS, D.M. (1998) *Supermodularity and Complementarity*. Frontiers of Economic Research. Princeton University Press. Princeton, New Jersey.

WANG, S. (1996) Premium calculation by transforming the layer premium density. *ASTIN Bulletin* **26**, 71-92.

WIRCH, J.L. (1999) Raising value at risk. *North American Actuarial Journal* **3(2)**, 106-115.

YAARI, M.E. (1987) Dual theory of choice under uncertainty. *Econometrica* **55**, 95-115.

WERNER HÜRLIMANN
Value and Risk Management
Winterthur Life and Pensions
Postfach 300
CH-8401 Winterthur
Switzerland

THE MIXED BIVARIATE HOFMANN DISTRIBUTION

By

J.F. Walhin*• and J. Paris*

* *Université Catholique de Louvain*
• *Secura Belgian Re*

Abstract

In this paper we study a class of Mixed Bivariate Poisson Distributions by extending the Hofmann Distribution from the univariate case to the bivariate case.

We show how to evaluate the bivariate aggregate claims distribution and we fit some insurance portfolios given in the literature.

This study typically extends the use of the Bivariate Independent Poisson Distribution, the Mixed Bivariate Negative Binomial and the Mixed Bivariate Poisson Inverse Gaussian Distribution.

Keywords

Mixed Bivariate Independent Poisson Distributions, Hofmann Distribution, maximum likelihood, aggregate claims distribution, recursive algorithm, stable algorithm.

1. Introduction

In this paper we study a family of bivariate counting distributions. These distributions are of interest in actuarial sciences when one wants to work with frequencies of dependent variables such as material damage and bodily injury claims in third party liability automobile insurance.

The general family of bivariate distributions we present in this paper has the following particular cases: the Mixed Bivariate Negative Binomial Distribution (MBNBD) and the Mixed Bivariate Poisson Inverse Gaussian Distribution (MBPIGD). These particular cases have already been discussed in Besson and Partrat (1992) and in Partrat (1994).

By extending the univariate Hofmann Distribution described in Walhin and Paris (2000b), we give a general setting for studying Mixed Bivariate Independent Poisson Distributions.

Note that we use the term Mixed Bivariate Independent Poisson Distribution in order to stress on the fact that it is a Bivariate Distribution obtained by mixing the Bivariate Independent Poisson Distribution.

The bivariate version of the Hofmann Distribution obtained by mixing the Bivariate Independent Poisson Distribution will be called the Mixed Bivariate Hofmann Distribution. It remains important to stress on the "Mixed" because Bivariate Hofmann Distributions can also be constructed via the trivariate reduction method (see Walhin and Paris (2000c)).

The rest of the paper is organized as follows. Section 2 reviews the concept of bivariate ordinary generating functions. Section 3 describes the Mixed Bivariate Independent Poisson Distribution. Section 4 extends the univariate Hofmann Distribution to the bivariate case. Section 5 addresses the problem of estimating the parameters of the distribution. Section 6 gives a stable recursion for the aggregate claims distribution which is based on a two-stage algorithm. Section 7 gives the fits for the data sets given in Besson and Partrat (1992) and in Partrat (1994). The two-stage algorithm is also applied. Section 8 gives the conclusion.

2. Bivariate ordinary generating functions

In the following sections we will use the concept of bivariate ordinary generating functions. This is a generalization of the ordinary generating functions (see Panjer and Willmot (1992) for an application in actuarial sciences).

Let us assume a sequence $\{a_{n,m},\ n \in \mathbb{Z},\ m \in \mathbb{Z}\}$ of real numbers.

The ordinary generating function of this sequence is defined as

$$T_{a_{n,m}}(u,v) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{n,m} u^n v^m.$$

Obviously u and v must be chosen in such a way that the sum exists.

Ordinary generating functions have the following nice properties:

- There is a one-to-one correspondence between $\{a_{n,m},\ n \in \mathbb{Z},\ m \in \mathbb{Z}\}$ and $T_{a_{n,m}}(u,v)$.
- $a_{n,m} = \dfrac{1}{n!m!} \left. \dfrac{\partial^n \partial^m T_{a_{n,m}}(u,v)}{\partial u^n \partial v^m} \right|_{u=0,\, v=0}.$
- $c_{n,m} = \alpha a_{n,m} + \beta b_{n,m} \Leftrightarrow T_{c_{n,m}}(u,v) = \alpha T_{a_{n,m}}(u,v) + \beta T_{b_{n,m}}(u,v).$
- $c_{n,m} = \sum_{k=0}^{n} \sum_{l=0}^{m} a_{k,l} b_{n-k,m-l} \Leftrightarrow T_{c_{n,m}}(u,v) = T_{a_{n,m}}(u,v)\, T_{b_{n,m}}(u,v).$
- $T_{na_{n,m}}(u,v) = u \dfrac{\partial}{\partial u} T_{a_{n,m}}(u,v).$

The philosophy behind using ordinary generating functions is the following:

- look for a relation between some sequences $a_{n,m}$, $b_{n,m}$, $c_{n,m}$, ...
- go in the (u,v) map if the calculations become easier (think of the convolution that becomes a product).

– go back in the initial map by inverting the expression in (u,v) thanks to the properties.

In this paper, the sequences $a_{n,m}$ will be probability functions and so, ordinary generating functions are just probability generating functions. In this case we have the convergence of the bivariate sums at least if $|u| < 1$ and $|v| < 1$.

3. The model

We are going to study the random vector (N,M) of counting variables. We will obtain the distribution of (N,M) by mixing the conditional distribution of (N,M) with a random variable Λ with distribution function $U(\lambda)$:

$$\mathbb{P}(N=n,\ M=m) = \int_0^\infty \mathbb{P}(N=n,\ M=m | \Lambda = \lambda) dU(\lambda). \tag{1}$$

Furthermore we assume that

– Conditionally on Λ the random variables N and M are independent.
– The conditional distributions of N and M given that $\Lambda = \lambda$ are univariate Poisson with parameter respectively λ and $\beta\lambda$.

The probability generating function of (N,M), $\psi_{N,M}(u, v) = \mathbb{E}[u^N v^M]$, writes

$$\psi_{N,M}(u, v) = \int_0^\infty e^{\lambda(u-1)+\beta\lambda(v-1)} dU(\lambda).$$

Kemp (1981) introduced the notion of Homogeneous Bivariate Distribution:

Definition: A bivariate probability generating function $\psi(u, v)$ is said to be of the homogeneous type if

$$\psi(u,v) = H(\sigma_1 u + \sigma_2 v),$$

with

$$H(\sigma_1 + \sigma_2) = 1$$ □

If one chooses H such that

$$H(x) = \int_0^\infty e^{-\lambda(1+\beta)} e^{\lambda x} dU(\lambda),$$

we immediately get that (N,M) is a Bivariate Homogeneous Distribution with $\sigma_1 = 1$ and $\sigma_2 = \beta$.
Kocherlakota and Kocherlakota (1992) have given the following characterization theorem:

Theorem: The probability generating function $\psi_{N,M}(u,v)$ is of the homogeneous type if and only if the conditional distribution of N given $N + M = z$ is Binomial distributed: $Bi\left(z, \frac{\sigma_1}{\sigma_1 + \sigma_2}\right)$. □

In our case we have that

$$N \mid N+M = z \sim Bi\left(z, \frac{1}{1+\beta}\right),$$

a result that also has been obtained by Partrat (1994) and Hesselager (1996). From Hesselager (1996) it is possible to extend the result of Kocherlakota and Kocherlakota (1992) by

$$\rho_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2},$$

$$\rho_2 = \frac{\sigma_2}{\sigma_1 + \sigma_2},$$

$$\psi_{N,M}(u,v) = \psi_{N+M}(\rho_1 u + \rho_2 v),$$

where $\psi_{N+M}(u) = \sum_{i=0}^{\infty} \mathbb{P}[N+M=i]\, u^i$ is the probability generating function of $N+M$.

4. The Mixed Bivariate Hofmann Distribution

Walhin and Paris (2000b) described the Hofmann Distribution. Let us recall some concepts.

Let $N(t)$ be the number of claims occurring in the time interval $(0,t]$ with $N(0) = 0$. Assume $N(t)$ is an infinitely divisible Mixed Poisson process (see Grandell (1997)) for which

$$\Pi(n,t) = \mathbb{P}[N(t)=n] = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!}\, dU(\lambda), \tag{2}$$

and

$$\Pi(0,t) = \mathbb{P}[N(t)=0] = e^{-\theta(t)},$$

where $\theta(t)$ is a Bernstein function:

$$\theta(t) \geq 0,$$

$$\theta(0) = 0,$$

$$\frac{d}{dt}\theta(t) \text{ completely monotone.}$$

The probability generating function of $N(t)$, $\psi_{N(t)}(u) = \mathbb{E}\left[u^{N(t)}\right]$, writes

$$\psi_{N(t)}(u) = \Pi(0, t - tu).$$

With the particular choice

$$\frac{d}{dt}\theta(t) = \frac{p}{(1+ct)^{\alpha}}, \quad p > 0, \ c > 0, \ \alpha \geq 0, \tag{3}$$

we have a Hofmann process (see Hofmann (1955) or Kestemont and Paris (1985)).
By integration, one has

$$\theta(t) = \frac{p}{c(1-a)}\left[(1+ct)^{1-a}-1\right] \text{ if } a \neq 1,$$

$$\theta(t) = \frac{p}{c}\ln(1+ct) \text{ by continuity for } a = 1.$$

Particular cases of interest are: $a = 0$ (Poisson), $a = 0.5$ (Poisson Inverse Gaussian), $a = 1$ (Negative Binomial), $a = 2$ (Polya-Aeppli), $a \to \infty$, $c \to 0$, $ac \to b$ (Neymann Type A).

For fixed t, it is possible to express the random variable $N(t)$ in the form of a compound Poisson distribution:

$$N(t) = \sum_{i=1}^{L(t)} \Xi_i,$$

where $L(t)$ is Poisson distributed with mean $\theta(t)$, independent of the Ξ_i which are independent and identically distributed. Moreover as

$$\frac{\mathbb{P}[\Xi = \xi]}{\mathbb{P}[\Xi = \xi - 1]} = r + \frac{s}{\xi}, \quad \xi > 1,$$

the probability distribution of Ξ is a member of the $(r, s, 1)$ class of counting distributions. The $(r, s, 1)$ class is just a reparametrization of the classical $(a, b, 1)$ class in order to avoid confusion with the a of the Hofmann Distribution.

Thanks to this property, it is possible to use the Panjer algorithm in order to evaluate the aggregate distribution of S:

$$S = \sum_{i=1}^{N(t)} X_i.$$

It is however necessary to apply the algorithm two times: we first introduce the random variable

$$U = \sum_{i=1}^{\Xi} X_i,$$

and then

$$S = \sum_{i=1}^{L(t)} U_i.$$

The distribution of U can be evaluated by the extended Panjer algorithm (see Sundt and Jewell (1981)) whereas the distribution of S can be evaluated by the Panjer algorithm (see Panjer (1981)).

For fixed t, in the sequel we will write $\Pi_{p,c,a}(n,t)$ for $\Pi(n,t)$ in order to specify the use of the Hofmann Distribution $Ho(p,c,a)$.

From now on let us fix $t = 1$ and let us use the Hofmann Distribution in our bivariate case. Let us assume that Λ is the mixing variable leading to the Hofmann Distribution.
From (1) and (2) we immediately get

$$\mathbb{P}(N = n, M = m) = \frac{(n+m)!}{n!m!} \frac{\beta^m}{(1+\beta)^{n+m}} \Pi(n+m, 1+\beta), \tag{4}$$

where it is easy to see that

$$\Pi_{p,c,a}(n+m, 1+\beta) = \Pi_{(1+\beta)p,(1+\beta)c,a}(n+m, 1).$$

In fact our model introduces dependency such that:

$$N \sim Ho(p, c, a),$$
$$M \sim Ho(p\beta, c\beta, a),$$
$$N + M \sim Ho(p(1+\beta), c(1+\beta), a).$$

This clearly generalizes the reasoning of Partrat (1994) where only Λ Gamma or Inverse Gaussian distributed are considered.

5. Estimation of the parameters

Let us assume that we have observed a sample (n_i, m_i), $1 \leq i \leq q$ of (N, M). The log-likelihood writes

$$\begin{aligned} l(\beta, p, c, a) &= \ln \prod_{i=1}^{q} \mathbb{P}(N = n_i, M = m_i) \\ &= C + \ln\beta \sum_{i=1}^{q} m_i - \ln(1+\beta) \sum_{i=1}^{q} (n_i + m_i) + l_{n+m}(\beta, p, c, a), \end{aligned}$$

where $l_{n+m}(\beta, p, c, a)$ is the log-likelihood for the univariate Hofmann Distribution $Ho((1+\beta)p, (1+\beta)c, a)$ with the sample $(n_i + m_i)$, $1 \leq i \leq q$ and C is a term that does not depend on the unknown parameters.

As shown in Besson and Partrat (1992), $\frac{\partial}{\partial\beta} l_{n+m}(\beta, p, c, a) = 0$ at the maximum likelihood estimate. So we immediately get

$$\hat{\beta} = \frac{\bar{m}}{\bar{n}},$$

where $\bar{n}$ (resp. $\bar{m}$) is the experimental mean of N (resp. M).

By standard results on the univariate Hofmann Distribution (see Hürlimann (1990)), we know that the maximum likelihood estimate of the mean is the observed frequency. Therefore maximizing $l_{n+m}(\beta, p, c, a)$ implies that

$$\hat{p}(1 + \hat{\beta}) = \bar{n} + \bar{m}.$$

So the estimates $\hat{p}$ and $\hat{\beta}$ are derived analytically. The other two estimates $\hat{c}$ and $\hat{a}$ have to be found by standard numerical maximization techniques. Note that one has to be careful because the likelihood may be very flat and local extrema are not excluded.

6. A STABLE TWO-STAGE RECURSION FOR THE AGGREGATE CLAIMS DISTRIBUTION

In the conclusion of his paper, Partrat (1994) addresses the problem of finding recursions like Panjer's recursion (see Panjer (1981)) in order to obtain the distribution of the aggregate claims.

Let X_i be the random variable representing the ith claim amount of type N and Y_i the random variable representing the ith claim amount of type M. We will assume, as usual, that the X_i and Y_i are mutually independent random variables. They are also arithmetic. The X_i are identically distributed. The Y_i are identically distributed. We also assume that the X_i and Y_i are independent of N and M. We are interested in the distribution of

$$(S,T)=\left(\sum_{i=1}^{N}X_i,\sum_{i=1}^{M}Y_i\right).$$

In the case of the Mixed Bivariate Negative Binomial Distribution the answer to Partrat's question was given by Hesselager (1996). In his paper, Hesselager (1996) gives a stable algorithm for the evaluation of the joint probability function of (S,T) for the particular case of the Mixed Bivariate Negative Binomial Distribution, i.e. when Λ is Gamma distributed.

In this section we will use the same methodology as in Hesselager (1996) in order to derive stable algorithms for the distribution of (S,T). As we know that U is infinitely divisible, it follows from Maceda (1948) that the distribution of (N,M) is also infinitely divisible.

Then we know from Sundt (1999a) that (N,M) can be interpreted as a bivariate Compound Poisson distribution:

$$(N,M)=\left(\sum_{i=1}^{L}\Xi_i,\sum_{i=1}^{L}\Omega_i\right),$$

where the Ξ_i and Ω_i are not independent and L is Poisson distributed independently of the (Ξ_i, Ω_i).

To be able to use the bivariate Panjer algorithm, we introduce the auxiliary vectors

$$(U,V)=\left(\sum_{i=1}^{\Xi}X_i,\sum_{i=1}^{\Omega}Y_i\right),$$

and then

$$(S,T)=\left(\sum_{i=1}^{L}U_i,\sum_{i=1}^{L}V_i\right).$$

We will use the following notation:

Probability functions

$$\begin{aligned}
\mathbb{P}(N=n, M=m) &= p(n,m),\\
\mathbb{P}(K=k) &= f_K(k), \ (K=N+M),\\
\mathbb{P}(X=x) &= f_X(x),\\
\mathbb{P}(Y=y) &= f_Y(y),\\
\mathbb{P}(S=s, T=t) &= f_{S,T}(s,t),\\
\mathbb{P}(\Xi=n, \Omega=m) &= f_{\Xi,\Omega}(n,m).
\end{aligned}$$

Probability generating functions

$$\begin{aligned}
\psi_X(u) &= \sum_{x=0}^{\infty} f_X(x)u^x,\\
\psi_Y(v) &= \sum_{y=0}^{\infty} f_Y(y)v^y,\\
\psi_K(u) &= \sum_{r=0}^{\infty} f_K(k)u^k,\\
\psi_{N,M}(u,v) &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} p(n,m)u^n v^m,\\
\psi_{S,T}(u,v) &= \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f_{S,T}(x,y)u^x v^y,\\
\psi_{\Xi,\Omega}(u,v) &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} f_{\Xi,\Omega}(n,m)u^n v^m.
\end{aligned}$$

The bivariate Panjer's algorithm described in Walhin and Paris (2000a) as well as in Sundt (1999b) and Ambagaspitya (1999) will be used in order to find the distribution of (S,T) knowing the distribution of (U,V).
In a first time we are interested in deriving the distribution of (U,V). Therefore we first need to derive the distribution of (Ξ, Ω).
Remember that we have

$$\psi_{N,M}(u,v) = \psi_{N+M}(\rho_1 u + \rho_2 v).$$

We find

$$\begin{aligned}
\psi_{N,M}(u,v) &= e^{-\theta(1+\beta-(1+\beta)(\rho_1 u+\rho_2 v))},\\
&= e^{-\theta(1+\beta)(1-\psi_{\Xi,\Omega}(u,v))},
\end{aligned}$$

where

$$\begin{aligned}
\psi_{\Xi,\Omega}(0,0) &= 0,\\
\psi_{\Xi,\Omega}(u,v) &= 1-\frac{\theta((1+\beta)(1-(\rho_1 u+\rho_2 v)))}{\theta(1+\beta)}. \qquad (5)
\end{aligned}$$

(Ξ, Ω) also has a bivariate homogeneous distribution. Indeed

$$\begin{aligned}\psi_{\Xi,\Omega}(u,v) &= G(\rho_1 u+\rho_2 v),\\ G(x) &= 1-\frac{\theta((1+\beta)(1-x))}{\theta(1+\beta)},\\ G(\rho_1+\rho_2) &= 1.\end{aligned}$$

The Taylor expansion around $(1+\beta)$ of (5) gives after a few calculations

$$f_{\Xi,\Omega}(n,m)=\frac{\theta^{(m+n)}(1+\beta)}{\theta(1+\beta)}\frac{(-1)^{m+n-1}}{(m+n)!}\beta^m\frac{(n+m)!}{n!m!}.$$

As we have

$$\begin{aligned}\theta(1+\beta) &= \frac{p}{c(1-a)}\left[(1+c(1+\beta))^{1-a}-1\right],\\ \theta^{(n+m)}(1+\beta) &= (-1)^{n+m-1}pc^{n+m-1}\frac{\Gamma(a+m+n-1)}{\Gamma(a)}(1+c(1+\beta))^{1-a-n-m},\end{aligned}$$

we immediately get

$$\begin{aligned}f_{\Xi,\Omega}(n,m) &= \frac{(n+m)!}{n!m!}\frac{c^{n+m}}{(n+m)!}\frac{1-a}{\Gamma(a)}\beta^m\Gamma(a+m+n-1)\frac{[1+c(1+\beta)]^{1-a-n-m}}{[1+c(1+\beta)]^{1-a}-1}\\ &= \frac{(n+m)!}{n!m!}\rho_1^n\rho_2^m\mathbb{P}(W=n+m),\end{aligned}$$

with

$$W=\Xi+\Omega,$$

and

$$\rho_1=\frac{1}{1+\beta},$$
$$\rho_2=\frac{\beta}{1+\beta}.$$

Indeed

$$\begin{aligned}\psi_W(u) &= \psi_{\Xi,\Omega}(u,u)\\ &= 1-\frac{\theta((1+\beta)(1-u))}{\theta(1+\beta)}\end{aligned}\tag{6}$$

A Taylor expansion around $(1+\beta)$ of (6) immediately shows that

$$\mathbb{P}(W=w)=\frac{c^w}{w!}\frac{1-a}{\Gamma(a)}\Gamma(a+w-1)(1+\beta)^w\frac{[1+c(1+\beta)]^{1-a-w}}{[1+c(1+\beta)]^{1-a}-1}.$$

With the particular form of the distribution of W, we immediately find, with $f_W(w) = \mathbb{P}(W = w)$,

$$\frac{f_W(w)}{f_W(w-1)} = \frac{c(1+\beta)}{1+c(1+\beta)} + \frac{c(1+\beta)(a-2)}{1+c(1+\beta)}\frac{1}{w}, \quad w > 1,$$

which is well in the form of the $(r, s, 1)$ class with

$$\begin{aligned} r &= \frac{c(1+\beta)}{1+c(1+\beta)}, \\ s &= \frac{c(1+\beta)(a-2)}{1+c(1+\beta)}, \\ f_W(0) &= 0, \\ f_W(1) &= c(1+\beta)(1-a)\frac{[1+c(1+\beta)]^{-a}}{[1+c(1+\beta)]^{1-a}-1}. \end{aligned}$$

Now let us study the distribution of

$$(U, V) = (X_1 + \ldots + X_{\Xi}, Y_1 + \ldots + Y_{\Omega}).$$

As W is a member of the $(r,s,1)$ class we have:

$$[1 - ru]\frac{d}{du}\psi_W(u) = f_W(1) + (r+s)\psi_W(u). \tag{7}$$

We are now able to extend Hesselager's methodology to find the aggregate claims distribution.

Theorem 1

We have

$$\begin{aligned} f_{\Xi,\Omega}(1,0) &= \rho_1 f_W(1), && (8)\\ f_{\Xi,\Omega}(n,m) &= \rho_1\left(r + \frac{s}{n}\right) f_{\Xi,\Omega}(n-1,m) \\ &\quad + r\rho_2 f_{\Xi,\Omega}(n,m-1), \ \ n \geq 1 \quad \textit{unless if } (n,m) = (1,0), && (9)\\ f_{\Xi,\Omega}(0,1) &= \rho_2 f_W(1), && (10)\\ f_{\Xi,\Omega}(n,m) &= \rho_2\left(r + \frac{s}{n}\right) f_{\Xi,\Omega}(n,m-1) \\ &\quad + r\rho_1 f_{\Xi,\Omega}(n-1,m), \ \ m \geq 1 \quad \textit{unless if } (n,m) = (0,1), && (11) \end{aligned}$$

Proof

We already noticed that

$$\psi_{\Xi,\Omega}(u, v) = \psi_W(\rho_1 u + \rho_2 v). \tag{12}$$

By differentiating (12) with respect to u, and using (7), we get

$$(1-r\rho_1 u - r\rho_2 v)\frac{\partial}{\partial u}\psi_{\Xi,\Omega}(u,v)=\rho_1(f_W(1)+(r+s)\psi_{\Xi,\Omega}(u,v)).$$

Inverting this expression we immediately get (8) and (9).
(10) and (11) are derived similarly. □

Theorem 2

We have

$$f_{U,V}(0,0)= 1-\frac{\theta((1+\beta)(1-(\rho_1 f_X(0)+\rho_2 f_Y(0))))}{\theta(1+\beta)}, \tag{13}$$

$$\begin{aligned} f_{U,V}(x,0)= {} & \frac{1}{1-r\rho_1 f_X(0)}\Bigg(\rho_1 f_W(1) f_X(x) \\ & +\rho_1\sum_{i=1}^{x}(r+\frac{si}{x})f_X(i)f_{U,V}(x-i,0)\Bigg), \quad x>0, \end{aligned} \tag{14}$$

$$\begin{aligned} f_{U,V}(0,y)= {} & \frac{1}{1-r\rho_2 f_Y(0)}\Bigg(\rho_2 f_W(1) f_Y(y) \\ & +\rho_2\sum_{i=1}^{y}(r+\frac{si}{y})f_Y(i)f_{U,V}(0,y-i)\Bigg), \quad y>0, \end{aligned} \tag{15}$$

$$\begin{aligned} f_{U,V}(x,y)= {} & \frac{1}{1-r\rho_1 f_X(0)-r\rho_2 f_Y(0)}\Bigg(\rho_1\sum_{i=1}^{x}(r+\frac{si}{x})f_X(i)f_{U,V}(x-i,y) \\ & +\rho_2 r\sum_{j=1}^{y}f_Y(j)f_{U,V}(x,y-j)\Bigg), \quad x>0, y>0, \end{aligned} \tag{16}$$

$$\begin{aligned} f_{U,V}(x,y)= {} & \frac{1}{1-r\rho_2 f_Y(0)-r\rho_1 f_X(0)}\Bigg(\rho_2\sum_{i=1}^{y}(r+\frac{si}{y})f_Y(i)f_{U,V}(x,y-i) \\ & +\rho_1 r\sum_{j=1}^{x}f_X(j)f_{U,V}(x-j,y)\Bigg), \quad y>0, x>0. \end{aligned} \tag{17}$$

Proof

As $f_{U,V}(0,0)=\psi_{\Xi,\Omega}(f_X(0),f_Y(0))$ we immediately find (13) by using equation (5). From (9) we get

$$f_{\Xi,\Omega}(n,m) = \rho_1(r+\frac{s}{n})f_{\Xi,\Omega}(n-1,m)+r\rho_2 f_{\Xi,\Omega}(n,m-1) \tag{18}$$

$$\Leftrightarrow$$

$$\begin{aligned} nf_{\Xi,\Omega}(n,m) = {} & r\rho_1(n-1)f_{\Xi,\Omega}(n-1,m) \\ & +\rho_1(r+s)f_{\Xi,\Omega}(n-1,m)+r\rho_2 nf_{\Xi,\Omega}(n,m-1). \end{aligned} \tag{19}$$

Multiplying both sides of (19) by $u\psi_X^{n-1}(u)\frac{d}{du}\psi_X(u)\psi_Y^m(v)$ and summing on $n = 1 \to \infty$, $m = 1 \to \infty$ gives

$$\begin{aligned} u\frac{\partial}{\partial u}\psi_{U,V}(u,v) &= r\rho_1 u\frac{\partial}{\partial u}\psi_{U,V}(u,v)\psi_X(u) \\ &\quad + (r+s)\rho_1 u\frac{d}{du}\psi_X(u)\psi_{U,V}(u,v) + r\rho_2\psi_{U,V}(u,v)\psi_Y(v). \end{aligned}$$

Inverting and rearranging this expression gives (16).
Multiplying both sides of (19) by $u\psi_X^{n-1}(u)\frac{d}{du}\psi_X(u)\psi_Y^m(0)$, summing on $n = 2 \to \infty$, $m = 0 \to \infty$ and adding $f_{\Xi,\Omega}(1,0)u\frac{d}{du}\psi_X(u)$ on both sides gives

$$\begin{aligned} u\frac{\partial}{\partial u}\psi_{U,V}(u,v) &= \rho_1 f_W(1)\psi_X(u) + r\rho_1 u\frac{\partial}{\partial u}\psi_{U,V}(u,v)\psi_X(u) \\ &\quad + (r+s)\rho_1 u\frac{d}{du}\psi_X(u)\psi_{U,V}(u,v). \end{aligned}$$

Inverting and rearranging this expression gives (14).
(15) and (17) are derived similarly.

Knowing the distribution of (U,V) it remains to evaluate the distribution of

$$(S,T) = \left(\sum_{i=1}^{L} U_i,\ \sum_{i=1}^{L} V_i\right).$$

This is easily done with the bivariate Panjer's algorithm.
We have

$$\begin{aligned} f_{S,T}(0,0) &= e^{-\theta(1+\beta)(1-f_{U,V}(0,0))}, \\ f_{S,T}(s,t) &= \sum_x^s \sum_y^t \left[\theta(1+\beta)\frac{x}{s}\right] f_{S,T}(s-x,t-y) f_{U,V}(x,y),\ \ s \geq 1, \\ f_{S,T}(s,t) &= \sum_x^s \sum_y^t \left[\theta(1+\beta)\frac{y}{t}\right] f_{S,T}(s-x,t-y) f_{U,V}(x,y),\ \ t \geq 1, \end{aligned}$$

where we use the following notation:

$$\sum_{x,y}^{s,t} g(x,y) = \sum_{x=0}^{s}\sum_{y=0}^{t} g(x,y) - g(0,0).$$

7. Numerical applications

In this section we fit two data sets given in Partrat (1994).

Data set 1 gives the yearly frequencies of hurricanes affecting two zones (zone 1 and zone 3) of the United States.

Data set 2 gives the yearly frequencies of an automobile third party liability portfolio, divided in material damage (type 1) and bodily injury (type 2) claims.

TABLE 1

Data set 1

		zone3			
zone1		0	1	2	3
0	*obs*	27	9	3	2
	$a = 0$	27.67	13.04	3.08	0.49
	$a = 0.0057$	27.59	13.05	3.09	0.49
1	*obs*	24	13	1	0
	$a = 0$	20.45	9.66	2.29	0.36
	$a = 0.0057$	20.47	9.69	2.29	0.36
2	*obs*	8	2	1	0
	$a = 0$	7.57	3.59	0.85	0.14
	$a = 0.0057$	7.59	3.59	0.85	0.13
3	*obs*	1	0	2	0
	$a = 0$	1.88	0.89	0.21	0.03
	$a = 0.0057$	1.88	0.89	0.21	0.03

For this data set, the maximum likelihood procedure for the Mixed Bivariate Hofmann Distribution ($a = 0.0057$) gives almost the Bivariate Independent Poisson Distribution ($a = 0$).

The characteristics of the fit are given in the next table. In order to compute the χ^2 statistic, some cells have been grouped in order that the theoretical frequencies are all larger than 1 and about 80% of the theoretical frequencies are larger than 5.

In the present case we work with 8 classes: (0,0), (0,1), (1,0), (1,1), (2,0), (0,2+) and (1,2+), (2,1+), (3+,0+).

TABLE 2

Parameter estimates, loglikelihood and χ^2 test – Data set 1

	BIPD	**MBHD**
β	0.6377	0.6377
p	0.7419	0.7419
c		0.5137
a	0	0.0058
λ	–187.9615	–187.9607
χ^2	3.73	3.77
df	5	3
p-value	0.589	0.287

A likelihood ratio test does not reject the null hypothesis that Bivariate Poisson Distribution is adequate against the more general model Mixed Bivariate Hofmann Distribution. Within the latter model we cannot reject the fact that both random variables are independent. In this case the principle of parsimony indicates that the Bivariate Independent Poisson Distribution (BIPD) should be retained. Now let us study our second data set:

TABLE 3

DATA SET 2

		Bodily injury		
Material damage		0	1	2
0	*obs*	171345	918	2
	$a = 0$	171086.9	946.0	2.6
	$a = 1$	171348.8	897.1	4.7
	$a = 0.5$	171348.7	897.5	4.6
	$a = 0.2982$	171345.8	898.6	4.5
1	*obs*	8273	73	0
	$a = 0$	8726.4	48.2	0.1
	$a = 1$	8275.5	86.3	0.7
	$a = 0.5$	8279.5	84.9	0.8
	$a = 0.2982$	8289.4	82.8	0.8
2	*obs*	389	5	0
	$a = 0$	222.5	1.2	0.0
	$a = 1$	398.2	6.2	0.1
	$a = 0.5$	391.5	6.9	0.1
	$a = 0.2982$	381.9	7.6	0.1
3	*obs*	31	1	0
	$a = 0$	3.8	0.0	0.0
	$a = 1$	19.1	0.4	0
	$a = 0.5$	21.3	0.6	0
	$a = 0.2982$	23.5	0.8	0
4	*obs*	1	0	0
	$a = 0$	0.0	0.0	0.0
	$a = 1$	0.9	0.0	0.0
	$a = 0.5$	1.3	0.0	0.0
	$a = 0.2982$	1.9	0.1	0.0

TABLE 4

PARAMETER ESTIMATES, LOGLIKELIHOOD AND χ^2 TEST – DATA SET 2

	MBPD	**MBNBD**	**MBPIGD**	**MBHD**
β	0.1084	0.1084	0.1084	0.1084
p	0.0510	0.0510	0.0510	0.0510
c		0.8934	1.8235	3.0695
a	0	1	0.5	0.3006
λ	–43251.57	–43143.09	–43141.79	–41141.27
χ^2	369.76	11.54	8.72	7.44
df	5	4	4	3
p-value	0	0.021	0.068	0.059

The MBPD and MBNBD are rejected at the 5% level. The MBPIGB and MBHD are not rejected at the 5% level. The grouped cells are: (0,0), (0,1), (1,0) , (1,1), (2,0), (2,1), (3,0), the rest.

Let us work with the portfolio given in table 3 and the Mixed Bivariate Hofmann fit of this portfolio.

Let us assume that the distributions X (material damage) and Y (bodily injury) are given by:

TABLE 5

CLAIMS DISTRIBUTIONS

X	$f_X(x)$	Y	$f_Y(y)$
1	0.2	5	0.2
2	0.2	10	0.36
3	0.2	20	0.22
4	0.1	50	0.11
5	0.1	100	0.11
10	0.1		
20	0.1		

We find the aggregate claims distribution:

TABLE 6

AGGREGATE CLAIMS DISTRIBUTION

	$S = 0$	$S = 1$	$S = 2$	$S = 3$	…
$T = 0$	0.9410275	0.0010876	0.0019589	0.0000042	
$T = 5$	0.0100336	0.0000219	0.0000395	0.0000001	
$T = 10$	0.0101349	0.0000223	0.0000403	0.0000001	
$T = 15$	0.0102375	0.0000228	0.0000411	0.0000001	
…					

Obviously, as the span of the Y claims is 5, it is most convenient (and less time consuming) to rescale the Y claims by division by 5 and then to revert the scaling for the final bivariate aggregate claims distribution.

8. CONCLUSION

In this paper we have extended the use of traditional Mixed Bivariate Independent Poisson Distributions into a general family of bivariate counting distributions. This family has interesting properties. On the one hand it authorizes a maximum likelihood estimation in a univariate setting. On the other hand it gives stable algorithms for the evaluation of the bivariate aggregate claims distribution.

The fits of some insurance portfolios are improved thanks to the use of the Mixed Bivariate Hofmann Distribution.

Remark

A first version of this paper has been presented at the Royal Statistical Society Conference 1999, Warwick.

Three anonymous referees are greatly acknowledged for very detailed and constructive comments on earlier versions of this paper.

References

AMBAGASPITYA, R.S. (1999) On the Distributions of Two Classes of Correlated Aggregate Claims. *Insurance: Mathematics and Economics*, **24**, 301-308.

BESSON, J.L. and PARTRAT, C. (1992) Trend et Systèmes de Bonus-Malus. *Astin Bulletin*, **22**, 11-31.

GRANDELL, J. (1997) *Mixed Poisson Processes*. Chapman and Hall.

HESSELAGER, O. (1996) Recursions for Certain Bivariate Counting Distributions and their Compound Distributions. *Astin Bulletin*, **26**, 35-52.

HOFMANN, M. (1955) Uber zusammengesetzte Poisson-Prozesse und ihre Anwendungen in der Unfallversicherung. *Bulletin of the Swiss Actuaries*, **55**, 499-575.

HÜRLIMANN, W. (1990) On Maximum Likelihood Estimation for Count Data Models. *Insurance: Mathematics and Economics*, **9**, 39-49.

KEMP, A.W. (1981) Computer Sampling from Homogeneous Bivariate Discrete Distributions. *ASA Proceedings of the Statistical Computing Section*, pages 173-175.

KESTEMONT, R.M. and PARIS J. (1985) Sur l'Ajustement du Nombre de Sinistres. *Bulletin of the Swiss Actuaries*, **85**, 157-164.

KOCHERLAKOTA, S. and KOCHERLAKOTA, K. (1992) *Bivariate Discrete Distributions*. Marcel Dekker, New-York.

MACEDA, E. (1948) On the Compound and Generalized Poisson Distributions. *Annals of Mathematical Statistics*, **19**, 414-416.

PANJER, H.H. (1981) Recursive Evaluation of a Family of Compound Distributions. *Astin Bulletin*, **12**, 22-26.

PANJER H.H. and WILLMOT, G.E. (1992) *Insurance Risk Models*. Society of Actuaries.

PARTRAT, C. (1994) Compound Model for Two Dependent Kinds of Claims. *Insurance: Mathematics and Economics*, **15**, 219-231.

SUNDT, B. (1999a) Multivariate Compound Poisson Distributions and Infinite Divisibility. Statistical report 33, Department of Mathematics, University of Bergen.

SUNDT, B. (1999b) On Multivariate Panjer Recursions. *Astin Bulletin*, **29**, 29-45.

SUNDT, B. and JEWELL, W.S. (1981) Further Results on Recursive Evaluation of Compound Distributions. *Astin Bulletin*, **12**, 27-39.

WALHIN, J.F. and PARIS, J. (2000a) The Effect of Excess of Loss Reinsurance with Reinstatements on the Cedent's Portfolio. *Blatter Deutsche Gesellschaft für Versicherungsmathematik*, **24**, 615-627.

WALHIN, J.F. and PARIS, J. (2000b) A Large Family of Discrete and Overdispersed Probability Laws. *submitted*.

WALHIN, J.F. and PARIS, J. (2000c) Recursive Formulae for Some Bivariate Counting Distributions Obtained by the Trivariate Reduction Method. *Astin Bulletin*, **30**, 141-155.

JEAN-FRANÇOIS WALHIN
SECURA Belgian Re
Rue Montoyer 12
1000 Bruxelles
E-mail: jfw@secura-re.com

THE NATURAL SETS OF WANG'S PREMIUM PRINCIPLE [1]

BY

XIAN-YI WU

East China Normal University, Shanghai, China

ABSTRACT

Recently, Wang's premium principle (Wang, 1995, 1996) has been discussed by many authors. Considerable attention has been given to the conditions under which Wang's premium principle can be reduced to the standard deviation premium principle. In this paper, we have got two results on this problem. One is that the natural set is a location-scale family if Wang's premium principle can be reduced to the SD premium principle for all surjective distortions. The other is that the natural set is a location-scale family for all power distortions.

KEYWORDS AND PHRASES

Premium principle, distortion function, proportional hazard transformation, natural set, location-scale family, Laplace transformation.

1. INTRODUCTION

Recently, Wang's premium principle (Wang, 1995, 1996) has been discussed by many authors, e.g., Wang, Young and Panjer (1997) and Wang and Young (1998). Christofides (1998) found that for some distribution families, the proportional hazards (PH) premium principle, as a special case of Wang's premium principle when the distortion is a power function, can be reduced to the well known standard deviation (SD) premium principle. He conjectures that for a parametric distribution family with constant skewness, the PH premium principle reduces to the SD principle. Young (1999) showed that Christofides' conjecture is true for location-scale families and for certain other families, but false in general. Wang (2000) introduced the concept of natural set for the distortion, and showed that for a fixed distortion, the natural set on which Wang's premium principle reduces to the SD premium principle is a union of location-scale families which satisfies some condition. Furthermore, he showed

[1] Project 19831020 Supported by National Natural Science Foundation of China.

that the natural set is a location-scale family if Wang's premium principle reduces to the SD premium principle for all distortions, or especially for all two-step-up distortions in which a two-step-up distortion is a distortion that takes only the values 0 and 1. In fact, his results mean that corresponding to each distortion there is a natural set (a union of location-scale families), and the intersection of all these natural sets for all two-set-up distortions is a location-scale family. He proposed a question of whether the natural set is a location-scale family for all power distortions. In this paper, we will provide further discussion on this topic. We have the following two results. One is that the natural set is a location-scale family if Wang's premium principle can be reduced to the SD premium principle for all surjective distortions. Note that any two-step-up distortion is not surjective. The other is that the natural set is a location-scale family for all power distortions.

2. Results

As we all know, the standard deviation premium principle is that

$$\pi(X) = E(X) + \lambda\sqrt{Var(X)}$$

for some $\lambda > 0$, where $\pi(X)$ is the premium charged for insurance risk X. Wang's (1995, 1996) premium principle gives the premium

$$H_g(X) = \int_{-\infty}^{0} \{g[S_X(t)] - 1\}dt + \int_{0}^{\infty} g[S_X(t)]dt$$

where $S_X(t) = P(X > t)$ is the decumulative distribution function (d.d.f., or survival function) of X, and the distortion g is a non-decreasing function from [0, 1] onto itself. If g is a power function $g(w) = w^c$, then $g[S_X(t)]$ is referred as a proportional hazard transformation of $S_X(t)$ (Wang, 1995, 1996), and $H_g(X)$ is called proportional hazard (PH) premium of X. For convenience, denote the PH premium of a random variable X as $H_c(X)$ for any fixed c.

Let g be a fixed distortion, then it is obvious that Wang's premium principle can reduce to the SD principle in a set, F, of distributions, if and only if, for any $X \in F$ and $Y \in F$, the following equation holds,

$$\frac{H_g(X) - E(X)}{\sqrt{Var(X)}} = \frac{H_g(Y) - E(Y)}{\sqrt{Var(Y)}}$$

Now we give the definition of natural set of a distortion function which is introduced by Wang (2000).

Definition 1. *Let Ω be the set of all distributions. The natural set of a distortion g with respect to distribution $X \in \Omega$, denoted as $N_g(X)$, is defined as*

$$N_g(X) = \left\{ Y : \frac{H_g(X) - E(X)}{\sqrt{Var(X)}} = \frac{H_g(Y) - E(Y)}{\sqrt{Var(Y)}},\ Y \in \Omega \right\}$$

If

$$\lambda = \lambda(X) = \frac{H_g(X) - E(X)}{\sqrt{Var(X)}}$$

we also write $N_g(X)$ as $N_g(\lambda)$.

Obviously, a fixed g may have many natural sets with each set corresponding to different parameter λ's. The proposition 1 in Wang (2000) said that $N_g(X)$ is a union of some location-scale families.

Definition 2. *g is called a surjective distortion if g is a non-decreasing surjection from [0, 1] onto itself. We denote the set of all distortion function as G, the set of all two-step-up distortions as T, the set of all surjective distortions as S and the set of all proportional hazard transformations (power distortions) as P.*

Obviously, P, T and S are all subsets of G and a distortion is a surjective distortion if and only if g is a continuous distortion satisfying $g(0) = 0$ and $g(1) = 1$.

By using a two-step-up distortion, Wang (2000) showed that the natural set on which Wang's premium principle can reduce to the standard premium principle for all distortions (or for all two-step-up distortions) is a location-scale family. It means that both $\bigcap_{g \in G} N_g(X)$ and $\bigcap_{g \in T} N_g(X)$ are location-scale families. In fact, the subscript for which the intersection is carried on need not go through the set G of all distortions or the set T of all two-step-up distortions. We will prove that $\bigcap_{g \in S} N_g(X)$ is a location-scale family, and if X has continuous d.d.f. and convex support set, then $\bigcap_{g \in P} N_g(X)$ is also a location-scale family. Before the main results, we present the following two lemmas.

Lemma 1. *Let $S(x)$ be a decumulative distribution function of a random variable X, and*

$$g(w) = \begin{cases} 0, & 0 \le w \le A, \\ 1, & A < w \le 1. \end{cases}$$

be a two-step-up distortion function for $A \in (0,\ 1)$. Let $u = \inf\{t : S(t) \le A\}$, then Wang's premium $H_g(X) = u$.

Proof: Since the decumulative distribution function is right continuous, $S(t)$ is larger than A when $t < u$ and smaller than or equal to A when $t \geq u$. Therefore we have

$$g[S(t)] = \begin{cases} 1, & t < u; \\ 0, & t \geq u. \end{cases}$$

If $u > 0$, we have

$$H_g(X) = \int_{-\infty}^{0} \{g[S_X(t)] - 1\}dt + \int_{0}^{\infty} g[S_X(t)]dt = \int_{0}^{u} dt = u$$

If $u \leq 0$, then

$$H_g(X) = \int_{-\infty}^{0} \{g[S_X(t)] - 1\}dt + \int_{0}^{\infty} g[S_X(t)]dt = \int_{u}^{0} (-1)dt = u$$

The lemma is proved.

Lemma 2. *(Feller, 1971, Chapter 13, Theorem 1a). Let U be a measure on $R^+ = (0, +\infty)$, taking finite values in bounded sets of R^+, and*

$$\phi(\lambda) = \int_{0}^{\infty} e^{-\lambda x} U(dx)$$

be finite for $a < \lambda < \infty$, where $a \geq 0$ is known. Then U is determined by its Laplace transformation $\phi(\lambda)$.

Proposition 1. $\bigcap_{g \in S} N_g(X)$ *is a location-scale family.*

Proof: Let $F = \bigcap_{g \in S} N_g(X)$. For any surjective distortion, Wang's premium principle reduces to the SD premium principle on the set F of distributions. Thus for any $X, Y \in F$ and any surjective distortion g, we have

$$\frac{H_g(X) - E(X)}{\sqrt{Var(X)}} = \frac{H_g(Y) - E(Y)}{\sqrt{Var(Y)}}$$

Let

$$U = \frac{X - E(X)}{\sqrt{Var(X)}}, \quad V = \frac{Y - E(Y)}{\sqrt{Var(Y)}}$$

Hence, we have

$$H_g(U) = H_g(V) \tag{1}$$

because H_g is location and scale equivariant (or linear invariant).

Therefore, in order to prove the proposition it is sufficient to show that if the equation (1) holds for all surjective distortions g, U and V have an identical distribution.

Denote the decumulative distribution functions of U and V as $S_U(t)$ and $S_V(t)$ respectively. Now, using proof by contradiction, we will prove that $S_U(t) = S_V(t)$ for $t \geq 0$. Assume that there exists a $t_0 \geq 0$ such that $S_U(t_0) > S_V(t_0)$. Let $\alpha = S_V(t_0)$, $0 \leq \alpha \leq 1$. Denote $v = \inf\{t : S_V(t) = \alpha\}$, then $v \leq t_0$. Take ϵ such that

$$0 < \epsilon < S_U(t_0) - \alpha$$

and define

$$g(w) = \begin{cases} 1, & \alpha + \epsilon < w \leq 1; \\ (w - \alpha)/\epsilon, & \alpha < w \leq \alpha + \epsilon; \\ 0, & 0 \leq w \leq \alpha. \end{cases}$$

$$h_1(w) = \begin{cases} 1, & \alpha + \epsilon < w \leq 1; \\ 0, & 0 \leq w \leq \alpha + \epsilon. \end{cases}$$

$$h_2(w) = \begin{cases} 1, & \alpha < w \leq 1; \\ 0, & 0 \leq w \leq \alpha. \end{cases}$$

Then it is obvious that $g(w)$ is a surjective distortion and $h_1(w) \leq g(w) \leq h_2(w)$ for all $w \in [0,\ 1]$.

Let $u(\epsilon) = \inf\{t : S_U(t) \leq \alpha + \epsilon\}$, then

$$S_U(u(\epsilon)) \geq \alpha$$

Since $S_U(t_0) > \alpha + \epsilon$, we have

$$u(\epsilon) > t_0$$

Using Lemma 1, we have

$$H_g(U) \geq H_{h_1}(U) = u(\epsilon)$$

and

$$H_g(V) \leq H_{h_2}(V) = v \leq t_0 < u(\epsilon)$$

Hence, $H_g(U) \neq H_g(V)$, contradicting equation (2).

Similar to the proof of proposition 2 in Wang (2000), it can be shown that $S_U(t) = S_V(t)$ for $t < 0$. Hence the proposition is proved.

Proposition 2. *If the d.d.f. of X is a continuous function and if the supporting set of X is a convex set, then* $\bigcap_{g \in P} N_g(X)$ *is a location-scale family.*

Proof: Similar to the above proof, in order to prove this proposition it is sufficient to show that if $H_g(U) = H_g(V)$ for all power distortions $g(p) = p^c$, U and V have an identical distribution where the d.d.f. of both U and V are continuous functions and the supporting sets of both U and V are convex sets.

Let g be a power distortion given by $g(p) = p^c$. Now we represent the PH premium $H_g(X)$ in the form of a Laplace transformation of random variable X with respect to some function.

$$\begin{aligned} H_g(X) &= \int_{-\infty}^{0} [(S_X(t))^c - 1]dt + \int_{0}^{\infty} (S_X(t))^c dt \\ &= -\int_{-\infty}^{\infty} t d(S_X(t))^c \\ &= -c \int_{-\infty}^{\infty} t(S_X(t))^{c-1} dS_X(t) \\ &= -c \int_{-\infty}^{\infty} t(S_X(t))^c d[\ln S_X(t)] \end{aligned}$$

Let $z = -\ln S_X(t)$, then

$$H_g(X) = -c \int_{0}^{\infty} S_X^{-1}(e^{-z}) e^{-cz} dz$$

which means that $H_g(X)$ is the Laplace transformation of $-S_X^{-1}(e^{-z})$, where $S_X^{-1}(\cdot)$ is the inverse function of $S_X(\cdot)$. Because the d.d.f. of X is continuous and the supporting set of X is convex, this Laplace transformation exists. Hence, according to Lemma 2, if $H_g(U) = H_g(V)$, i.e.

$$\int_{0}^{\infty} S_U^{-1}(e^{-z}) e^{-cz} dz = \int_{0}^{\infty} S_V^{-1}(e^{-z}) e^{-cz} dz$$

we have

$$S_U^{-1}(e^{-z}) = S_V^{-1}(e^{-z})$$

which implies that $S_U(t) = S_V(t)$. Hence U and V have an identical distribution. Thus the proof is completed.

To conclude, the PH-transformation principle reduces to the standard deviation principle only on a natural set of location-scale family of distributions. This is merely a result of (and nothing more than) the positive linearty of Wang premium principle.

Acknowledgement

The author thanks the anonymous referees for their suggestions that have highly improved the paper.

References

Christofides, S. (1998) Principle for risk in financial transactions. *Proceedings of the GISG/ASTIN Joint Meeting in Glasgow, Scotland, October, 1998* **2**, 63-109.

Feller, W. (1971) *An introduction to probability and its applications*. John Wiley & Sons, Inc.

Wang, J.L. (2000) A note on Christofides' conjecture regarding Wang's premium principle. *ASTIN Bulletin* **30**, 13-17.

Wang, S.S. (1995) Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics* **17**, 43-54.

Wang, S.S. (1996) Premium calculation by transforming the layer premium density. *ASTIN Bulletin* **26**, 71-92.

Wang, S.S. and Young, V.R. (1998) Ordering risks: Expected utility theory versus Yaari's dual theory of risk. *Insurance: Mathematics and Economics* **22**, 145-161.

Wang, S.S., Young, V.R. and Panjer, H.H. (1997) Axiomatic characterization of insurance price. *Insurance: Mathematics and Economics* **21**, 173-183.

Young, V.R. (1999) Discussion of Christofides' conjecture regarding Wang's premium principle. *ASTIN Bulletin* **29**, 191-195.

Wu Xian-Yi
Department of Statistics
East China Normal University
3663 Zhongshan Road (Northern)
Shanghai 200062
P.R. China

GEOGRAPHIC PREMIUM RATING BY WHITTAKER SPATIAL SMOOTHING

BY

GREG TAYLOR

Taylor Fry Consulting Actuaries, Sydney Australia
and
University of Melbourne, Australia

ABSTRACT

Whittaker graduation is applied to the spatial smoothing of insurance data. Such data (e.g. claim frequency) form a surface over the 2-dimensional geographic domain to which they relate. Observations on this surface are subject to sampling error. They need to be smoothed spatially if a reliable estimate of the underlying surface is to be obtained.

A measure of smoothness of a surface has been defined. This has been incorporated in 2-dimensional Whittaker graduation to effect the necessary smoothing. The details of this are worked out in Section 4. The procedure is illustrated by numerical example in Section 5. The Bayesian interpretation of this form of spatial smoothing is discussed, and used to assist in the selection of the Whittaker relativity constant.

KEYWORDS

Geographic premium rating, spatial smoothing, Whittaker graduation.

1. INTRODUCTION

In certain lines of insurance business the risk varies geographically. This is typical of domestic lines, where the geographic variation may be related to directly geographic factors (e.g. traffic density, proximity to arterial roads in auto insurance) or socio-demographic factors (perhaps affecting theft rates in house insurance).

In such cases it will be desirable to estimate the geographic variation in risk premium and to price accordingly. Usually data will be available by quite fine geographic divisions, e.g. zipcodes in the US, postcodes (or sub-postcodes in the UK). The subdivisions will typically be fine enough that sampling error in each is substantial.

As a result, a mapping of sampled geographic risk takes on a rather patchy appearance. Despite this, general trends in geographic variation will often be visible. It is necessary to find a way of smoothing sampling error from one subdivision to the next in order to estimate the underlying geographic signal.

Taylor (1989) applied 2-dimensional spline functions to this problem. Boskow and Verrall (1994) provided an alternative treatment which made use of the Gibbs sampler to implement a Bayesian revision of the observations on subdivisions. The Bayesian framework recognised the magnitudes of sampling error and also incorporated the concept of smoothness over neighbouring subdivisions.

The present paper takes a rather similar approach, applying an accepted actuarial technique for compromising between smoothness and fit to data. This is Whittaker graduation, which has also been shown (Taylor, 1992; Verrall, 1993) to have a Bayesian interpretation.

2. Model and notation

Consider a random variable X whose mean μ is characterised by n covariates. One covariate comprises a pair of spatial (Cartesian) coordinates.

For example, X might denote claim frequency, and the coordinates might represent the centroid of a postcode region.

Although n may be any natural number, it will suffice here, and maintain brevity of notation, if the concepts are illustrated for the case $n = 3$. The extension to the general case will be obvious.

Thus, let i, j, k represent specific values of the 3 covariates, with i representing the spatial member. These values define a cell of data $\{N_{ijk}, X_{ijk}\}$, where N_{ijk} is a volume measure. In the above example it might be number of years of policy exposure.

Consistent with the notation given above,

$$E[X_{ijk}] = \mu_{ijk}. \tag{2.1}$$

Suppose that the spatial effect is separable as follows:

$$\mu_{ijk} = \nu_i \, \theta_{jk}, \tag{2.2}$$

for suitable parameters ν_i, θ_{jk}.

It will be supposed that the ν_i are to be estimated, but the θ_{jk} are known, perhaps by means of an earlier estimation program.

There is one degree of redundancy among the parameters ν_i, θ_{jk}. It will often be useful to set the scale of the θ_{jk} so that they are scattered about 1. If this is done, then θ_{jk} can be thought of as an adjustment multiplier to correct the quantity ν_i so that it is specific to cell ijk.

Define

$$\mathrm{Y_i} = \sum_{jk} N_{ijk} \left(X_{ijk} / \theta_{jk} \right) \Big/ \sum_{jk} N_{ijk}. \tag{2.3}$$

On this definition, Y_i is the summary of experience in region i but standardized for other covariates, as discussed by Brockman and Wright (1992). This adjustment to data is found in Taylor (1989) and Boskow and Verrall (1994).

By (2.1)-(2.3),

$$E[\mathrm{Y_i}] = \nu_i, \tag{2.4}$$

showing that Y isolates the spatial effect. It will assumed that

$$V[\mathrm{Y_i}] = \sigma^2/N_i \tag{2.5}$$

with

$$N_i = \sum_{jk} N_{ijk}, \tag{2.6}$$

and for some suitable (though perhaps unknown) constant $\sigma^2 > 0$.

This last assumption is convenient but will often involve some degree of approximation. For example, when X denotes claim frequency with N_{ijk} X_{ijk} Poisson, one finds that

$$V[\mathrm{Y_i}] = \phi_i \nu_i / N_i, \tag{2.7}$$

with

$$\phi_i = \sum_{jk} \left(N_{ijk}/N_i\right)\theta_{jk}^{-1}. \tag{2.8}$$

At this point, it is useful to write ν_i in the alternative form:

$$\nu_i = \nu(x_i), \tag{2.9}$$

expressing the fact that $\nu : R^2 \to R$ is a function of the spatial coordinates x.

Similarly, write $\mathrm{Y_i} = \mathrm{Y}(x_i)$.

3. Whittaker smoothing

Whittaker graduation was devised by Whittaker (1923) and introduced into the actuarial literature by Henderson (1932). Since then, it has appeared in a number of standard actuarial texts, e.g. London (1985).

All of these early treatments involved smoothing a 1-dimensional sequence of observations. The generalisation to 2 or more dimensions was begun by McKay and Wilkin (1977), and a number of subsequent papers have published developments (e.g. Lowrie, 1992).

Consider points $x \in R^2$, Euclidean 2-space. The objective is to find values $W(x_i)$ which provide a smooth version of the $\mathrm{Y}(x_i)$ and estimate $\nu(x_i)$. Define

$$D = \sum_i N_i[\mathrm{Y}(x_i) - W(x_i)]^2 \tag{3.1}$$

which is a measure of the deviation, or error, in the observations relative to their smoothed version.

The use of $\{N_i\}$ as the set of weights in (3.1) is justified by assumption (2.5).

Define

$$F = D + kS, \tag{3.2}$$

where S is a suitably chosen measure of smoothness of $W(\cdot)$, and $k(0 < k < \infty)$ is the tuning constant, or relativity constant. This constant is often chosen empirically, although Taylor (1992) and Verrall (1993) have given an analytical basis.

Whittaker smoothing consists of choosing $\{W(x_i)\}$ so as to minimise (3.2), thus achieving a compromise between error and smoothness.

The choice of S can conveniently be based on the theory of **thin plate splines** (see e.g. Green and Silverman, 1994), which uses the following penalty for lack of smoothness:

$$S = \int \left[\left(\frac{\partial^2 W}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 W}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 W}{\partial y^2} \right)^2 \right] dx\, dy \tag{3.3}$$

for $W = W(x, y)$ (using the (x, y) notation just for this equation) and with integration over the entire spatial region of interest.

Approximated in finite form, this is:

$$S = \sum_i S(x_i) \tag{3.4}$$

with

$$S(x) = \left[\Delta^2_{11} W(x)\right]^2 + 2\left[\Delta^2_{12} W(x)\right]^2 + \left[\Delta^2_{22} W(x)\right]^2 \tag{3.5}$$

and

$$\Delta^2_{pq} W(x) = \Delta_p \left[\Delta_q W(x)\right], \quad p, q = 1,\ 2 \tag{3.6}$$

$$\Delta_q W(x) = W(x + e_q) - W(x), \tag{3.7}$$

with e_1, e_2 denoting (1, 0) and (0, 1) respectively. Thus, Δ_q is the difference operator in the direction of the q-th coordinate axis.

4. Application

4.1. Smoothing

The basic procedure of Whittaker smoothing needs to be adapted to the situation in which the arrangement of points x at which $Y(\cdot)$ is sampled is irregular, rather than forming a lattice as in Section 3. In the case of general $\{x_i\}$, it is not clear how the differences $\Delta^2_{pq} W(x_i)$ should be calculated.

Note that, in the lattice case, $W(y)$ is required at 6 distinct values of y to determine the 3 terms $\Delta^2_{pq}W(x)$ at a fixed x. This reflects the fact that a general quadratic defined on 2-space, is defined by 6 parameters.

One possibility therefore would be to select 5 points "close to" a given x and fit a quadratic $Q_x(\cdot)$ to the 6 points. The values $\Delta^2_{pq}W(x)$ for that x could then be read off from the coefficients of $Q_x(\cdot)$.

To the extent that $Q_x(\cdot)$ is merely an approximation for $W(\cdot)$ in the neighbourhood of x, there will be disagreement between the functions $Q_x(\cdot)$ defined by different x. Fitting a quadratic through precisely 6 points will cause that function to be highly sensitive to the values assumed at those points. In the same way, working with a function $f : R \rightarrow R$, fitting a 5th degree polynomial to each set of 6 consecutive values $\{f(x), f(x+1), ..., f(x+5)\}$ would be liable to produce eccentric fitted functions and a high degree of conflict between them.

For this reason, it is doubtful that the most meaningful smoothness measure is obtained by the precise quadratic fitting described above. One alternative, and preferable, procedure is to fit each $Q_x(\cdot)$ by reference to a larger number of points than 6. Let h be the number of such points.

This is done in Appendix B and the smoothness measure $S(x)$ in (3.5) calculated by reference to the fitted $Q_x(\cdot)$. The calculation procedure is as follows.

Suppose that $\mathrm{Y}(x)$ is observed at m points $x_1, ..., x_m$. Let z denote the vector $[W(x_1), ..., W(x_m)]^T$ of smoothed observations. Let $y_x^{(1)}, ..., y_x^{(h)}$ be the subset of $\{x_i\}$ consisting of the h points closest to x, including $y_x^{(1)} = x$ and let $z_x = \left[W(y_x^{(1)}), ..., W(y_x^{(h)})\right]^T$.

For any $y \in R^2$, write $y = (y_1, y_2)$ and

$$y^{\otimes} = \left(\frac{1}{2}y_1^2, y_1y_2, \frac{1}{2}y_2^2, y_1, y_2, 1\right)^T, \tag{4.1}$$

and define X_x as the $h \times 6$ matrix with $\left[y_x^{(i)}\right]^{\otimes}$ in its i-th row.

Define

$$\underset{6\times h}{A_x} = (X_x^T X_x)^{-1} X_x^T, \tag{4.2}$$

and B_x as the $6 \times m$ matrix containing the h columns of A_x placed within B_x in the same positions as the components of z_x occupy in z.

Define $\tilde{B}_x$ as the $3 \times h$ submatrix of B consisting of the latter's first 3 rows. Then

$$S(x) = z^T \tilde{B}_x^T C \tilde{B}_x z, \tag{4.3}$$

$$S = z^T \left[\sum_i \tilde{B}_{x_i}^T C \tilde{B}_{x_i}\right] z, \tag{4.4}$$

with

$$C = \mathrm{diag}(1, 2, 1). \tag{4.5}$$

The Whittaker criterion (3.2) may be written in the matrix notation established above.

$$F = (\mathrm{Y} - z)^T \Lambda (Y - z) + kz^T Mz, \tag{4.6}$$

where Y is the m-vector of observations $\mathrm{Y}(x_i)$,

$$\Lambda = \mathrm{diag}(N_1, \ ..., \ N_m), \tag{4.7}$$

$$M = \sum_i \tilde{B}_{x_i}^T C \tilde{B}_{x_i}. \tag{4.8}$$

Minimisation of F is carried out by differentiating (4.6) with respect to z and setting the result equal to zero. This yields the smoothed vector

$$z = (1 + k\Lambda^{-1}M)^{-1}\mathrm{Y}. \tag{4.9}$$

The smoothed vector is equal to the unsmoothed plus a "smoothing correction" $k\Lambda^{-1}M\mathrm{Y}$. The greater the value chosen for the relativity constant k, the greater the correction. The correction made at any point is inversely proportional to the volume of experience at that point; the greater that volume, the less it requires smoothing.

Note that, because F is quadratic in Y and z, a scale change in Y induces the same scale change in z, provided that k is changed appropriately. This allows the useful device of replacing $\mathrm{Y_i}$, defined in (2.3), by $\mathrm{Y_i}/\bar{\nu}$, where $\bar{\nu}$ is in some sense an overall average of the ν_i.

This converts the observed values of $\mathrm{Y_i}$ to a scatter about 1, and (2.4) and (2.7) are replaced by $E[\mathrm{Y_i}] = \nu_i/\bar{\nu}$ and $V[\mathrm{Y_i}] = (\nu_i/\bar{\nu})\phi_i/N_i\bar{\nu}$.

The Bayesian interpretation of k may be obtained from Taylor (1992). Straightforward extension of the 1-dimensional reasoning given there to 2 dimensions shows that, if all second differences of z in coordinate axis directions are viewed as subject to independent priors each with variance τ^2, then

$$k = \sigma^2/4\tau^2, \tag{4.10}$$

with σ^2 defined by (2.5).

4.2. Zoning

Consider the framework established in Section 2, in which some pricing function, such as claim frequency is being estimated by postcode. The smoothing formula (4.9) produces such estimates. In principle, it is feasible to price accordingly, i.e. postcode by postcode.

Often, however, an insurer will wish to group postcodes into convenient rating zones, or regions. For pricing purposes, the geographic effect will be taken as constant over such a zone.

A procedure for achieving this would be as follows:
1. Map the smoothed vector z to postcodes on a proper geographic map of the whole region being priced.
2. Colour code the postcodes according to the values of z_i. For example, values of z_i might range (mainly, ignoring a scatter of extreme values) from 40% for rural areas to 150% for inner city areas. In this case, different colours might be applied to the ranges $< 70\%$, 70-80%, 80-90%, etc. The colours should be spectrally ordered, e.g. red, pink, orange, yellow, etc. While the constant bandwidth illustrated above is simply to apply, it may be neater to use a multiplicative (logarithmic) scale, e.g. 91-100%, 100-110%, 110-121%, etc.
3. Scan the map to select zones consisting largely of the one colour, or of a small number of colours adjacent in the chromatic scheme. This requirement of chromatic homogeneity will need to be balanced against the desirability of spatially continuous connected zones.
4. Re-fit the whole model with the collection of selected zones introduced as a rating variable.

In Step 4, the model structure (2.2) is still assumed, but the spatial index i now applies to the coarser zoning determined in Step 3 instead of individual postcode.

In addition, θ_{jk} can no longer be assumed known, since these values will have been estimated with geographic effects ignored. Now that the effects are "known", the θ_{jk} need to be reestimated, taking them into account.

For example, suppose that (2.2) may be expanded in the form:

$$\mu_{ijk} = \nu_i \, \theta_j \, \phi_k; \tag{4.11}$$

equivalently,

$$\log \mu_{ijk} = \log \nu_i + \log \theta_j + \log \phi_k. \tag{4.12}$$

Then the usual regression modelling (e.g. generalised linear) may be applied to estimate the parameters. Note that, at this stage, the numerical results obtained from the smoothing (the vector z) are discarded, and the only role played by the smoothing is the determination of zones.

An example is provided in Section 5.

4.3. Further research

The smoothing procedure described in the foregoing sections appears to work effectively most of the time. It seems well adapted to metropolitan areas, usually characterised by densely packed postcodes.

An example of cases in which it tends to break down is illustrated in schematic form in Figure 1. The polygons in the figure are a stylised representation of postcodes in a rural area. The circled numbers label the 10 postcodes. The uncircled numbers represent unsmoothed values of geographic risk, i.e. $Y(x_i)$.

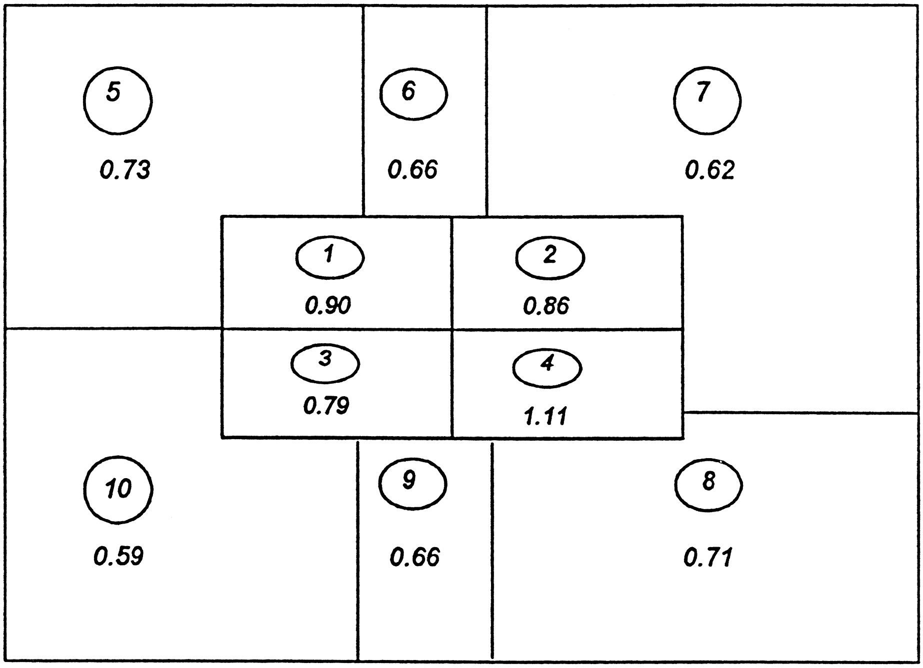

FIGURE 1: Example of geographic variation.

The small postcodes 1-4 represent a regional town, and postcodes 5-10, geographically much larger, represent surrounding rural districts.

The spatial smoothing of Section 4.1 seems poorly adapted to these circumstances. The difficulty relates to the highly localised variation of postcodes 1-4 from the surrounding trend.

According to Section 4.1, the smoothness measure for postcode 1 is calculated by reference to h postcodes. Suppose $h = 10$, and the 10 relevant postcodes are those appearing in Figure 1. In the detail of Appendix B, the smoothness measure is calculated from the curvature of the surface fitted over these 10 postcodes.

The surface needs to be mostly flat, but with a peak concentrated over a small area formed by postcodes 1-4. It is impossible to obtain this by fitting a quadratic surface, which will be much flatter than the experience of the diagram suggests.

This means that the smoothing algorithm sees postcodes 1-10 as relatively smooth before smoothing, and so applies little smoothing to them. The "smoothed" results are likely to exhibit the same "patchiness" over postcodes 1-4 as found in the unsmoothed.

Such failures are less likely to occur in metropolitan areas, since the localised extremes which cause the difficulty are less likely in these cases.

The solution to this difficulty may lie in some form of variation of h with the local topography of the $Y(x)$ surface. Such techniques might be akin to adaptive kernel smoothing of the type discussed by Bailey and Gatrell (1995, p. 87).

5. Example

The smoothing procedure derived in Section 4.1 is applied to a particular claim frequency data set in Figures 2 to 6. These maps show smoothing produced by different choices of the relativity constant k, which increases steadily from Figure 2 to Figure 6.

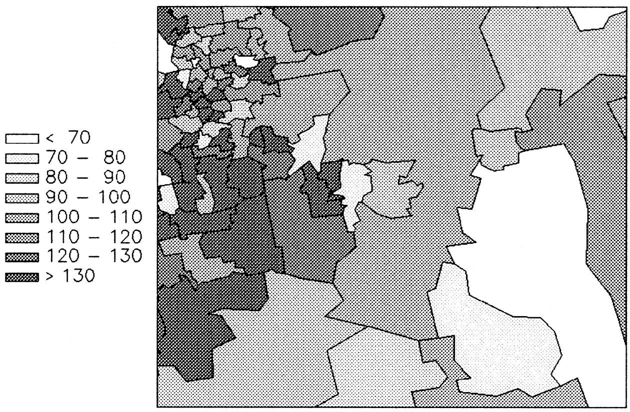

Figure 2: No smoothing ($k = 0$).

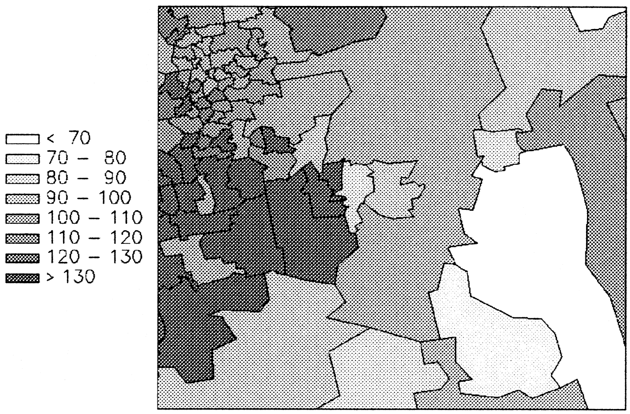

Figure 3: Smoothing with $k = 100$

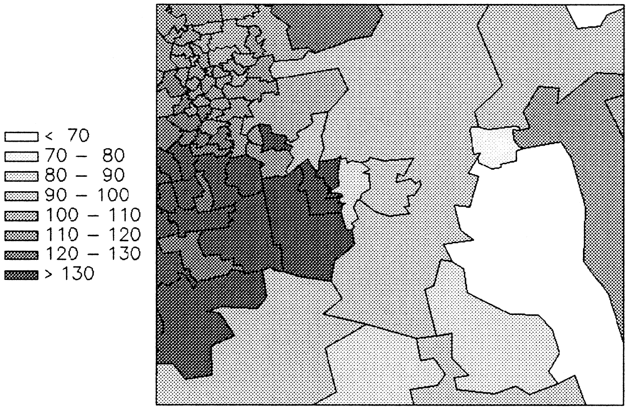

FIGURE 4: Smoothing with $k = 500$.

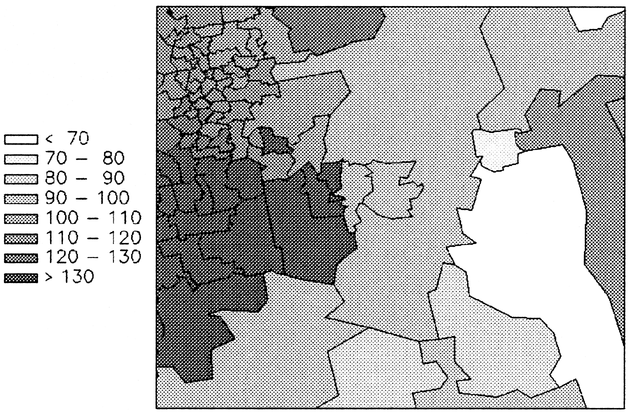

FIGURE 5: Smoothing with $k = 1000$.

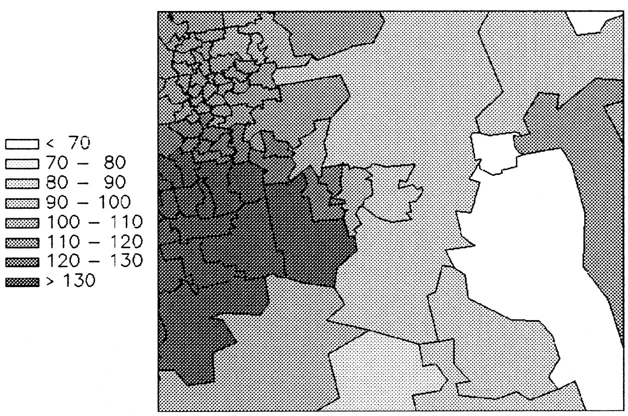

FIGURE 6: Smoothing with $k = 5000$.

Figure 2 uses $k = 0$, which reproduces the unsmoothed statistics (see (4.9)). These statistics are "residual ratios" from a regression which models all effects other than geographic. The residual ratios are defined as follows:

$$\text{Residual ratio for postcode } i = \frac{\text{number of claims observed in postcode}}{\text{model fit of this number}}.$$

This is the case in which X_{ijk} represents claim frequency, but $\mathrm{Y_i}$ has been rescaled by the device mentioned at the end of Section 4.1. The rescaled values of $\mathrm{Y_i}$ are scattered about 1, and the legend in Figures 2 to 6 expresses the ratio as a percentage. For example, the band 70-80 includes residual ratios 70-80%.

The example illustrated here uses $h = 10$. The increased smoothing power resulting from increasing k is evident through the sequence of diagrams.

Equation (4.10) gives the theoretical value of k. The paramater σ^2 is defined by (2.5). For $\mathrm{Y_i}$ representing claim frequency, N_i $\mathrm{Y_i}$ might be assumed Poisson, with N_i denoting exposure. This case, together with the rescaling of ν_i effected by the residual ratios, is dealt with at the end of Section 4.1, where

$$V[\mathrm{Y_i}] = (\nu_i/\bar{\nu})\phi_i/N_i\bar{\nu}, \tag{5.1}$$

giving

$$\sigma^2 = (\nu_i/\bar{\nu})\phi_i/\bar{\nu}. \tag{5.2}$$

Now the values of $\nu_i/\bar{\nu}$ are scattered about 1, and ϕ_i will also be of the order 1 if the θ_{jk} are "centralised" in the manner suggested just prior to (2.3). Thus, a rough approximation is $\sigma^2 = 1/\bar{\nu}$, so that (4.10) gives $k = 1/4\bar{\nu}\tau^2$.

Alternatively, one might take σ^2 to be somewhat greater than this, to allow for some overdispersion relative to Poisson.

The nature of τ^2 is described at the end of Section 4.1. It is clearly more difficult to estimate, but some indication can be obtained.

For reasons of data confidentiality, an estimate of $\bar{\nu}\,\tau^2$ is given here, rather than separate estimates of $\bar{\nu}$ and τ^2. Based on Figure 5, a reasonable estimation of $\bar{\nu}\,\tau^2$ appears to lie in the range 1/8,000 to 1/6,000, giving k in the range 1,500 to 2,000. This just fails to match $k = 1,000$ in Figure 5.

Alternatively, consider Figure 6 which suggests a value of τ^2 smaller by a factor of perhaps 2, giving k in the range 3,000 to 4,000. This fails to match $k = 5,000$ in Figure 6, but the discrepancy lies in the opposite direction from that of Figure 5.

In view of the roughness of the estimated τ^2, it is inappropriate to regard the above calculations too literally. To the extent that they are meaningful, however, they indicate a value of k between 1,000 and 5,000. This suggests that Figure 6 may be over-smoothed, while Figures 3 and 4 are perhaps under-smoothed.

Acknowledgement

The author acknowledges the assistance of Steven Lim, who programmed the smoothing algorithm developed in this paper, and linked it to the mapping software used to produce Figures 2 to 6. The assistance of Dr. Richard Verrall, who provided several references to multi-dimensional smoothing procedures, is also acknowledged.

Appendix A
Fitting a quadratic defined on 2-space to h observations

Consider a quadratic $Q : R^2 \to R$. Let $y^{(1)}, ..., y^{(h)} \in R^2$ and $z^{(1)}, ..., z^{(h)}$ be "observations":

$$z^{(i)} = Q(y^{(i)}) + \text{error}. \tag{A.1}$$

If the quadratic is written out explicitly, it is

$$Q(y) = q_{20}y_1^2 + q_{11}y_1y_2 + q_{02}y_2^2 + q_1y_1 + q_2y_2 + q_0, \tag{A.2}$$

where $y = (y_1, y_2)^T$. Write this as:

$$Q(y) = q^T y^{\otimes}, \tag{A.3}$$

with

$$q = (q_{20}, q_{11}, q_{02}, q_1, q_2, q_0)^T, \tag{A.4}$$

$$y^{\otimes} = (y_1^2, y_1y_2, y_2^2, y_1, y_2, 1)^T. \tag{A.5}$$

By (A.1) and (A.3),

$$z^{(i)} = q^T \left[y^{(i)}\right]^{\otimes} + \text{error}, \quad i = 1, ..., h. \tag{A.6}$$

OLS regression of the $z^{(i)}$ on the $\left[y^{(i)}\right]^{\otimes}$ yields the following estimate $\hat{q}$ of q:

$$\hat{q} = \mathrm{A}z, \tag{A.7}$$

with

$$\mathrm{A} = (X^TX)^{-1}X^T, \tag{A.8}$$

where

$$z = (z^{(1)}, ..., z^{(h)})^T, \tag{A.9}$$

and X is the $h \times 6$ matrix with $[y_i^{\otimes}]^T$ as i-th row.

APPENDIX B
CALCULATION OF THE SMOOTHNESS MEASURE

Section 4.1 requires that a local approximation $Q_x(\cdot)$ be fitted to $W(\cdot)$ by reference to the values of $W(y)$ at h points y.

The fitting can be carried out by OLS regression. Write

$$Q_x(y) = q_x^T y^{\otimes} \tag{B.1}$$

where $y = (y_1, y_2)^T$, and

$$y^{\otimes} = (\frac{1}{2}y_1^2, y_1 y_2, \frac{1}{2}y_2^2, y_1, y_2, 1)^T, \tag{B.2}$$

and q_x is the corresponding vector of coefficients.

Let $y_x^{(1)}$, ..., $y_x^{(h)}$ be the h points closest to x, including $y_x^{(1)} = x$, for which $Y(\cdot)$ is sampled, and let

$$z_x = \left[W(y_x^{(1)}), \ ..., \ W(y_x^{(h)})\right]^T. \tag{B.3}$$

Appendix A shows that

$$\underset{6\times 1}{q_x} = \underset{6\times h}{\mathrm{A}_x} \ \underset{h\times 1}{z_x} \ , \tag{B.4}$$

where

$$\mathrm{A}_x = (X_x^T X_x)^{-1} X_x^T, \tag{B.5}$$

and X_x is the $h \times 6$ matrix with $\left[y_x^{(i)}\right]^{\otimes T}$ as i-th row.

The result (B.4), which is expressed in terms of the "local" set of smoothed values z_x, needs to be expressed in terms of the global set $z = [W(x_1), \ ..., \ W(x_m)]^T$ corresponding to the whole set of observations $Y(x_i)$, $i = 1, \ ..., \ m$.

This is done by rewriting (B.4) as:

$$\underset{6\times 1}{q_x} = \underset{6\times m}{B_x} \ \underset{m\times 1}{z} \ , \tag{B.6}$$

where B_x is the $6 \times m$ matrix containing the h columns of A_x placed so as to reference h components of z_x as components of z, and zeros elsewhere.

The required differences $\Delta_{pq}^2 W(x)$ can now be approximated by the corresponding differences of $Q_x(x)$, which are given by the first 3 components of q_x. The relevant part of (B.6) is therefore

$$\underset{3\times 1}{\tilde{q}_x} = \underset{3\times m}{\tilde{B}_x} \ \underset{m\times 1}{z}, \tag{B.7}$$

where the tilde indicates the operation "take the first 3 rows of".

It is now possible to express $S(x)$ from (3.5) as a quadratic form in $\tilde{q}_x$:

$$S(x) = \tilde{q}_x^T C \tilde{q}_x \tag{B.8}$$

with

$$C = \mathrm{diag}(1,\ 2,\ 1). \tag{B.9}$$

By (B.7) and (B.8),

$$S(x) = z^T \tilde{B}_x^T C \tilde{B}_x z, \tag{B.10}$$

and finally, by (3.4),

$$S = z^T \left[\sum_i \tilde{B}_{x_i} C \tilde{B}_{x_i} \right] z. \tag{B.11}$$

References

Bailey, T.C. and Gatrell, A.C. (1995) *Interactive spatial data analysis*. Longman, England.

Boskow, M. and Verrall, R.J. (1994) Premium rating by geographic area using spatial models. *ASTIN Bulletin* **24**, 131-143.

Brockman, M.J. and Wright, T.S. (1992) Statistical motor rating: making effective use of your data. In *Journal of the Institute of Actuaries* **119**, 457-526.

Green, P.J. and Silverman, B.W. (1994) *Non-parametric regression and generalised linear models: a roughness penalty approach*. Chapman and Hall, London, New York.

Henderson, R. (1924) A new method of graduation. In *Transactions of the Actuarial Society of America* **25**, 29-40.

London, R. (1985) *Graduation: the revision of estimates*. ACTEX, Winsted and Abington, CT.

Lowrie, W.B. (1992) Multidimensional Whittaker-Henderson graduation with constraints and mixed differences. In *Transactions of the Society of Actuaries* **45**, 215-255.

McKay, S.F. and Wilkin, J.C. (1977) Derivation of a two-dimensional Whittaker-Henderson type B graduation formula. Appendix to *Experience of Disabled-Worker Benefits Under OASDI, 1965-74*. Actuarial Study No. 74, U.S. Department of Health, Education, and Welfare.

Taylor, G.C. (1989) Use of spline functions for premium rating by geographic area. *ASTIN Bulletin* **19**, 91-122.

Taylor, G.C. (1992) A Bayesian interpretation of Whittaker-Henderson graduation. In *Insurance: mathematics and economics* **11**, 7-16.

Verrall, R.J. (1993) A state space formulation of Whittaker graduation, with extensions. In *Insurance: mathematics and economics* **13**, 7-14.

Whittaker, E.T. (1923) On a new method of graduation. In *Proceedings of the Edinburgh Mathematical Society* **41**, 63-75.

Greg Taylor
Taylor Fry Consulting Actuaries
Level 4, 5 Elizabeth Street
Sydney NSW 2000
Australia

Professorial Associate, Centre for Actuarial Studies
Faculty of Economics and Commerce
University of Melbourne
Parkville VIC 3052
Australia

OPTIMAL LOSS FINANCING UNDER BONUS-MALUS CONTRACTS

BY

JON HOLTAN

if P&C Insurance Ltd., Oslo, Norway

ABSTRACT

The paper analyses the question: *Should an insurance customer carry an occurred loss himself, or should he make a claim to the insurance company?* This question is important within bonus-malus contracts with individual experience adjustments of the premium. The analysis model includes a bonus hunger strategy where the customers prefer the most profitable financial alternative, that is, the alternative which represents the lowest rate of interest. Hence the loss of bonus after a claim is calculated as a rate of interest paid from the customer to the insurer. Within this model the paper outlines the existence of a true compensation function and a relative cost function for each customer. A set of properties for bonus-malus contracts are presented and discussed. A concrete example of a bonus-malus system and an insurance compensation function illustrates the theoretical framework in a practical manner.

KEYWORDS

Insurance contracts, bonus-malus, bonus hunger, true compensation, true deductible, relative cost function, optimal loss financing.

1. INTRODUCTION

Should an insurance customer carry an occurred loss himself, or should he make a claim to the insurance company? This question is quite fundamental under bonus-malus contracts, that is, under insurance contracts with bonus-malus (individual experience rating or no-claim) adjustments, like e.g. motor insurance contracts. It is the general tendency for insurance customers to carry small losses themselves to avoid increases of future premium costs which explains the relevance of the question. This tendency is called the bonus hunger of the insurance customers. Bonus hunger has been widely discussed and analyzed in actuarial literature, see e.g. Lemaire (1995), chapter 7, where pp. 101-102 contains a partial review of this literature. The bonus hunger question is more relevant to a customer the harder the loss of bonus rules are, the higher the premium is, and the smaller an occurred loss is, and vice versa. The aim of this paper is to outline and describe how this bonus hunger effect may be taken into account within the framework of optimal loss financing under bonus-malus contracts seen from the customer's point of view.

The question of optimal loss financing is important not only at the time of the loss occurrence, but also when the customers purchase their insurance contracts. If it is rarely worth to let the insurance company carry a loss, why then purchase the contract? This question is in fact part of the general problem of purchasing optimal insurance coverage, which has been extensively studied under varying conditions in insurance economics. Holtan (2001) analyses this problem particularly for bonus-malus contracts. But to do so, we first have to outline the necessary concepts of – and insight to – bonus-malus contracts, which is part of the objective of this paper.

The paper is organized as follows: Sections 2 and 3 describe the general insurance contract and the bonus hunger strategy of the customers. Section 4 outlines the general existence of a true compensation function for all insurance contracts with bonus-malus adjustments. Section 5 outlines the existence of relative cost functions for the customer and their general properties. Sections 6 and 7 illustrate some of the ideas in sections 4 and 5 by doing special assumptions on the bonus-malus system and the insurance compensation function. Section 8 gives some concluding remarks.

2. The general insurance contract

Consider an insurance buyer representing a risk of loss X, where X is a stochastic variable with probability density function $f(x)$ where $x \geq 0$. The insurance contract is characterized by a continuous premium process $p(t)$ transferred from the insured to the insurer at time t, and a compensation $c(x)$ transferred the opposite way if loss $X = x$ obtains. The compensation $c(x)$ is hereby called the *contractual compensation*. Any admissible contractual compensation function satisfies $0 \leq c(x) \leq x$ for all $x \geq 0$. This constraint reflects that there is no compensation if there is no loss ($c(0) = 0$) and that the customer cannot make a profit by gambling on his or her risk ($c(x) \ngtr x$).

3. The bonus hunger strategy

Let the premium process $p(t)$ depend on a bonus-malus system. In principle a bonus-malus system gives the insurance customer a future premium increase if a loss occurrence is compensated by the insurer, and a premium reduction if no loss is compensated. The premium increase is called the loss of bonus, and depends usually only on the number of compensated claims, and not on their amounts. Hence, a customer often saves money by self-financing an occurred loss in order to avoid future premium increase, instead of financing the loss by a compensation from the insurer, and thereby accepting a future premium increase. This phenomenon is called *bonus hunger*. After a loss occurrence the customer's decision problem is hereby to choose the most profitable financial alternative. Trivial investment theory solves this problem by using *the rate of interest* as the optimal financial criteria, that is, the customer should prefer the financial alternative which represents the lowest rate of interest.

In order to define a bonus hunger strategy (a financial decision rule) for the insurance customer, we use the following notations and assumptions: The premium paid at time t after a loss occurrence at time s is denoted by $p_1(s+t)$ if the loss is reported to the insurer, and by $p_0(s+t)$ if the loss is not reported. Assume $p_0(s+t)$ and $p_1(s+t)$ to be continuous non-stochastic premium processes for all $t > 0$.

Definition 1: Given a loss occurrence at time s with a fixed loss amount $X = x$, the constant discount rate $\delta(x)$ determined by the net present value equation

$$c(x) = \int_0^\infty e^{-\delta t}(p_1(s+t) - p_0(s+t))dt \tag{1}$$

is a *relative measure* of the loss of bonus following that loss. □

The discount rate $\delta(x) \in \langle -\infty, \infty \rangle$ is by definition the effective rate of interest for the insurance compensation. To choose the optimal financial alternative after a loss occurrence, the rate of interest for the insurance compensation has to be compared to the effective rate of interest for the financial alternatives, that is, self-financing by savings, by borrowing or by a combination of savings and borrowing, according to the liquidity of each customer. Let λ_1 be the borrowing rate and λ_2 be the saving rate (the return rate) at the loss occurrence time s, and assume λ_1 and λ_2 to be positive time-constant parameters. The non-stochastic rate of interest for self-financing, λ, is then defined by (for simplicity disregard taxes):

$$\lambda = \begin{cases} \lambda_1 & \text{if financed by borrowing,} \\ \lambda_2 & \text{if financed by savings,} \\ \beta\lambda_1 + (1-\beta)\lambda_2 & \text{if a share of } \beta \text{ is financed by borrowing and a share of} \\ & (1-\beta) \text{ is financed by savings.} \end{cases}$$

Hence the bonus hunger strategy for the insurance customer is identified by:

If $\delta > \lambda \Rightarrow$ Self-financing.

If $\delta < \lambda \Rightarrow$ Financing by compensation from the insurer.

If $\delta = \lambda \Rightarrow$ Indifference between the two choices of financing.

4. The true compensation function

A loss of bonus after a claim is obviously paid by the customer, not by the insurer. In principle, this fact identifies the loss of bonus as a deductible paid by the customer to the insurer over a period of time. Hence the *true deductible* of the customer is a combination of the *contractual deductible*, $x - c(x)$, and the loss of bonus. Thus we may define the excess point of the true deductible as that loss amount which makes the insurance customer *indifferent* between the two choices of loss financing, that is, when $\delta = \lambda$. The existence of a true deductible obviously generates a corresponding existence of a *true compensation*, which differs from the earlier defined *contractual compensation* $c(x)$. Exact expressions of the true compensation and the true deductible are defined as follows:

From (1) we define the fixed amount z when $\delta = \lambda$:

$$z = \int_0^{\infty} e^{-\lambda t}(p_1(s+t) - p_0(s+t))dt. \tag{2}$$

z is in this context a constant because of the non-stochastic assumptions of λ_1, λ_1, p_0(s + t) and p_1(s + t).

From (1) and (2) we obtain the following modifications of the bonus hunger strategy:

If $\delta > \lambda \Rightarrow c(x) < z \Rightarrow$ Self-financing.

If $\delta < \lambda \Rightarrow c(x) > z \Rightarrow$ Financing by compensation from the insurer.

If $\delta = \lambda \Rightarrow c(x) = z \Rightarrow$ Indifference between the two choices of financing.

Hence, if we assume the customers to follow this optimal bonus hunger strategy, we have:

Definition 2: The true compensation of an occurred loss $X = x$ is defined by:

$$c^*(x) = \begin{cases} c(x) - z & \text{if } c(x) > z \\ 0 & \text{if } c(x) \le z \end{cases}$$ □

Definition 3: The true deductible of an occurred loss $X = x$ is defined by:

$$d^*(x) = x - c^*(x) = \begin{cases} x - c(x) + z & \text{if } c(x) > z \\ x & \text{if } c(x) \le z \end{cases}$$ □

$x - c(x) + z$ is the excess point of the true deductible, where $x - c(x)$ is the excess of the contractual deductible and z is the excess of the deductible generated by the loss of bonus. Explicitly we may define:

Definition 4: A bonus-malus contract has a contractual compensation function $c(x)$ and a true compensation function $c^*(x)$. □

From (2) we observe that the lower λ is, the higher z is, which by definition 2 gives a lower true compensation $c^*(x)$. Hence we introduce the following proposition:

Proposition 1: A decreasing force of interest in the money market generates a decreasing true compensation, and hence a less favorable insurance profitability for the insurance customers, and vice versa. □

Within this framework there exists a lower and an upper limit of $d^*(x)$; both greater than zero. The lower limit is defined when $\lambda \to \infty$, that is, when the relative cost of self-financing goes to infinity, while the upper limit is defined when $\lambda \to 0$, that is, when the relative cost of self-financing goes to zero. Let $z_{\min} = \lim_{\lambda \to \infty} z$ and $z_{max} = \lim_{\lambda \to 0} z$. Hence by definition 3 the lower and upper limit of $d^*(x)$ is defined by:

$$0 < \min\{x, x - c(x) + z_{\min}\} \leq d^*(x) \leq \min\{x, x - c(x) + z_{max}\} \qquad (3)$$

By definition 2 we hereby also define the lower and upper limit of $c^*(x)$:

$$\max\{0, c(x) - z_{max}\} < c^*(x) < \max\{0, c(x) - z_{\min}\} \qquad (4)$$

Hence by (4) and definition 2 we state two important propositions:

Proposition 2: Independent of the contractual compensation function, the true compensation function has always an individual deductible. □

Proposition 3: The compensation function of a bonus-malus contract without a contractual deductible is equivalent to the compensation function of a standard insurance contract with an individual deductible. □

Proposition 3 is based on the fact that the true compensation function *max* $(c(x) - z, 0)$ reduces to *max* $(x - z, 0)$ when the bonus-malus contract has no contractual deductible. Since z in this context is the non-stochastic excess value of the deductible generated by the loss of bonus, the compensation function *max* $(x - z, 0)$ has by definition the same structure as a standard insurance contract with a deductible z.

Note firstly that even if z is paid over a period of time by increased premiums within a bonus-malus contract, z can nevertheless be considered as a fixed deductible at the time of the loss occurrence. And, not to forget, the customers act as if z is a fixed deductible because they have to make a decision at the time of the loss occurrence. Note secondly that z depends on x via the premium process $p(t)$ and the bonus-malus rules. Hence there exists different compensation functions *max* $(x - z, 0)$ for different customers. This existence does not, however, influence the validation of proposition 3 if we allow *individual* deductibles in standard insurance contracts. In general we have:

Proposition 4: There exists different true compensation functions for different customers. □

5. The relative cost function

Definition 1 in section 3 expresses the rate of interest for the insurance compensation on the assumption that the loss amount is already known. If we do not know the size of the loss amount, or more precisely, if the loss amount is a stochastic variable, the rate of interest will also be a stochastic variable. The sample space of this stochastic rate of interest generates something we may call the *relative cost function*. Hence this function is identified by:

Definition 5: The sample space of the stochastic discount rate $\delta(X) \in \langle -\infty, \infty \rangle$ determined by the net present value function

$$c(x) = \int_0^\infty e^{-\delta t}(p_1(s+t) - p_0(s+t))dt$$

is called the *relative cost function* for all possible loss amounts. □

The relative cost function, or *ReCoF* for short, expresses the relationship between all possible loss amounts $x \geq 0$ and their correspondingly rate of interest δ for the insurance compensation. The practical utility of this relationship is obvious: At the beginning of the insurance period the *ReCoF* gives the insurance customers information about their true compensation and their true deductible if a loss occurs during the period. This information is essential within the practical choice of insurance coverage. Figure 1 illustrates the general *ReCoF* and some of the correspondingly vital information.

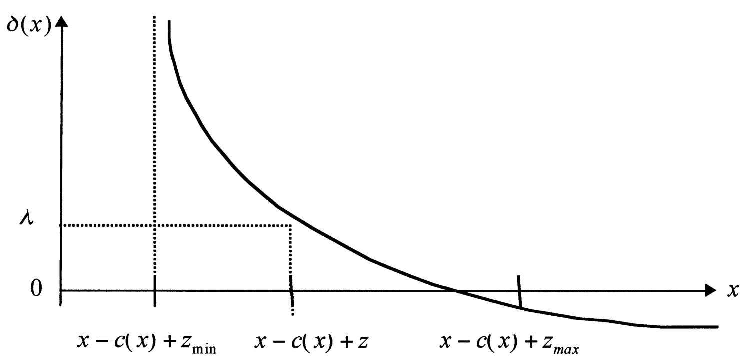

Figure 1: The general ReCoF – The relative cost function.

In figure 1 we observe that the relative cost (the rate of interest δ) for insurance compensation is high for small claims and low for large claims, or more precisely, the *ReCoF* has a decreasing form. The discount component of the present value function in definition 4 makes this characteristic common to all existing bonus-malus systems.

Three essential values are marked out at the horizontal x-axis in figure 1: The left hand value, $x - c(x) + z_{\min}$, is the lower limit of the excess point of the true deductible, while the right hand value, $x - c(x) + z_{max}$, is the upper limit of the excess point of the true deductible. The middle value, $x - c(x) + z$, that is, when $\delta(x) = \lambda$, is the real excess point of the true deductible for all possible loss amounts $X = x$ given a time-constant rate of interest for self-financing λ. As we e.g. observe, an uncritical reporting of losses with amounts close to $x - c(x) + z_{\min}$ may generate astronomical sized rate of interests for insurance compensation for the customers.

The left hand value and the right hand value at the x-axis generate three essential outcomes of an occurred loss $X = x$:

Outcome 1: $x < x - c(x) + z_{\min} \Leftrightarrow 0 < c(x) < z_{\min}$

Outcome 2: $x - c(x) + z_{\min} < x < x - c(x) + z_{max} \Leftrightarrow z_{\min} < c(x) < z_{max}$

Outcome 3: $x > x - c(x) + z_{max} \Leftrightarrow c(x) > z_{max}$

Common to outcome 1 and 3 are their *independence* of the market parameter λ. In other words, the optimal financial choice of outcome 1 is always self-financing, and the choice is hereby independent of λ. In the same way, the optimal financial choice of outcome 3 is always a financing by insurance compensation. Hence this choice is also independent of λ. Note that if a loss within outcome 1 is less that the excess point of the contractual deductible, then the customer cannot demand any insurance compensation, and hence there exists no financial choice at all.

Outcome 2 is more complex: The financial choice is, unlike outcome 1 and 3, directly *dependent* of the market parameter λ. Within our model, where λ is assumed to be a time-constant parameter, there exists two different outcomes for outcome 2:

Outcome 2a: $x - c(x) + z_{\min} < x < x - c(x) + z \Leftrightarrow z_{\min} < c(x) < z$

Outcome 2b: $x - c(x) + z < x < x - c(x) + z_{max} \Leftrightarrow z < c(x) < z_{max}$

Hence, by the bonus hunger strategy in section 3, outcome 2a generates an optimal self-financing, while outcome 2b generates an optimal financing by insurance compensation. It should be noted that the optimal financial choices within outcome 2 are modified if λ is assumed to be a *stochastic variable*; see section 8 for further discussion/comments on this.

From definition 4 we observe that the *ReCoF* also depends on the individual premium processes $p_0(s + t)$ and $p_1(s + t)$. These processes are again dependent on the individual premium tariff criteria of each customer. Hence we admit the existence of different *ReCoF* for different customers, and hence also the existence of different true deductibles and different true compensations for different customers. We observe e.g. that the higher premium costs a customer pays, the higher is his or her rate of interest for insurance compensation, and hence the higher is his or her true deductible. Given identical losses $X = x$, a customer with high premium costs has to pay a higher rate of interest for insurance compensation than a customer with lower premium costs. Hence, a high risk individual is not only punished once by a high premium, but twice by also a high rate of interest of insurance compensation (which is equivalent to a high true deductible).

According to proposition 4 we conclude this section by the following proposition:

Proposition 5: There exists different relative cost functions for different customers, and all functions are decreasing. □

6. Theoretical example

To give further illustrations on the optimal decision problem of the customers, we do the following assumptions of the bonus-malus system and the contractual compensation function:

Bonus-malus system: Let the insurance contract depend on a bonus-malus system which is characterized by a continuous bonus scale where the customer receives a constant premium reduction of percentage (1-k) if no loss is compensated, and a constant premium increase of amount m if a loss is compensated; $0 < k < 1$ and $m > 0$. This system is a modified version of a credibility system with geometric weights described by Sundt (1988), and is chosen because of simple calculating properties. Another modification of this system has been practiced within motor insurance for 10 years (1987-97) by the Norwegian insurance company Storebrand Ltd. (the earlier name of the norwegian part of if P&C Insurance); see a detailed description in Neuhaus (1988).

Let us interpret p as the premium paid by the customer at time s, i.e. at the time of the loss occurrence. Hence we have $p_0(s + t) = pk^t$ and $p_1(s + t) = (p + m)k^t$. From (1) we then find:

$$c(x) = \int_0^\infty e^{-\delta t} mk^t \, dt = \frac{m}{(\delta - \ln k)} \tag{5}$$

$$\Leftrightarrow \delta(x) = \begin{cases} \dfrac{m}{c(x)} + \ln k & \text{if } c(x) > 0 \\ \text{not defined} & \text{if } c(x) = 0 \end{cases} \tag{6}$$

And from (2) we find:

$$z = \int_0^\infty e^{-\lambda t} mk^t \, dt = \frac{m}{(\lambda - \ln k)}. \tag{7}$$

The bonus hunger effect within the credibility system with geometric weights has been studied by Sundt (1989). His bonus hunger strategy is close to our strategy, but unlike us, he does not give attention to the optimal financial choice of the customers, i.e. he does not use relative cost as the sufficient bonus hunger criteria.

Contractual compensation: Let the contractual compensation function follow the ordinary excess of loss function identified by:

$$c(x) = \begin{cases} x - d & \text{if } x > d \\ 0 & \text{if } x \le d \end{cases}$$

where d is a fixed amount called the contractual excess point. Hence the specified bonus-malus system and the contractual compensation function give us specified expressions of $\delta(x)$, $c^*(x)$ and $d^*(x)$ as follows:

$$\delta(x) = \begin{cases} \dfrac{m}{x - d} + \ln k & \text{if } x > 0 \\ \text{not defined} & \text{otherwise} \end{cases} \tag{8}$$

$$c^*(x, k, m, \lambda) = \begin{cases} x - d - \dfrac{m}{\lambda - \ln k} & \text{if } x > d + \dfrac{m}{\lambda - \ln k} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$d^*(x, k, m, \lambda) = \begin{cases} d + \dfrac{m}{\lambda - \ln k} & \text{if } x > d + \dfrac{m}{\lambda - \ln k} \\ x & \text{otherwise} \end{cases} \tag{10}$$

where $d + m/(\lambda - \ln k)$ is called the true excess point.

It should be noted that the expressions (8)-(10) contain an underlying assumption of a fixed dependency between the contractual excess point d and the premiums $p_0(s+t)$ and $p_1(s+t)$. In other words; in (8)-(10) d can not be interpreted as a varying parameter. See section 7 for a wider discussion on this. From (3) and (10) we find the lower and upper limit of $d^*(x)$:

$$z_{\min} = \lim_{\lambda \to \infty} \frac{m}{(\lambda - \ln k)} = 0 \qquad z_{max} = \lim_{\lambda \to 0} \frac{m}{(\lambda - \ln k)} = \frac{m}{-\ln k}.$$

Hence we have:

$$0 < \min(m,d) \leq d^*(x) \leq \min(x, d - m/\ln k). \tag{11}$$

And from (4) and (9) we also find the lower and upper limit of $c^*(x)$:

$$max(0, x-d+m/\ln k) < c^*(x) < max(0, x-d) \tag{12}$$

The assumed bonus-malus system in this example is, as mentioned, chosen because of it's simple analytical calculating properties. Most other bonus systems in force today are quite different, and also much more difficult to calculate correspondingly. Hence, the only practical way to (numerical) calculate the expressions (8)-(12) for most systems is to use data simulation methods.

7. Numerical studies

There are several ways to study the expressions (8)-(10) numerically. Here we briefly present two of them:

Study 1:

Let k = 0.87 (like the old system in Storebrand) and let d = 0. Hence table 1 shows some values of $\delta(x, m)$ when $\$500 \leq x \leq \$5000$ and $\$100 \leq m \leq \$500$.

TABLE 1

	x = 500	x = 1000	x = 2000	x = 3000	x = 4000	x = 5000
m = 100	6,1%	–3,9%	–8,9%	–10,6%	–11,4%	–11,9%
m = 200	26,1%	6,1%	–3,9%	–7,3%	–8,9%	–9,9%
m = 300	46,1%	16,1%	1,1%	–3,9%	–6,4%	–7,9%
m = 400	66,1%	26,1%	6,1%	–0,6%	–3,9%	–5,9%
m = 500	86,1%	36,1%	11,1%	2,7%	–1,4%	–3,9%

As earlier indicated, we observe from the table that the smaller loss amounts x and the higher premium increase m after a loss compensation, the higher rate of interest δ is for the insurance compensation, and vice versa. A combination of a small amount x and a high premium increase m gives very high rate of interest for the insurance compensation, and of course a considerable financial loss for the customer. The shadowed cells indicate when the rate of interest is positive, while the other cells represent negative values. In study 1 the lower limit of the excess point of the true deductible, $x - c(x) + z_{\min}$, is zero (by straightforward calculation), while the upper limit of the excess point of the true deductible, $x - c(x) + z_{max}$, is $m/(-\ln k) = m/0.139$.

Study 2:

Like study 1 let k = 0.87 and d = 0. Let m take three different values: m = \$100, m = \$300 and m = \$500. Hence figure 2 shows a two dimensional figure of the true excess point $d + m / (\lambda - \ln k)$ as a function of the market interest λ (where $0 \le \lambda \le 20\%$) for the three values of m.

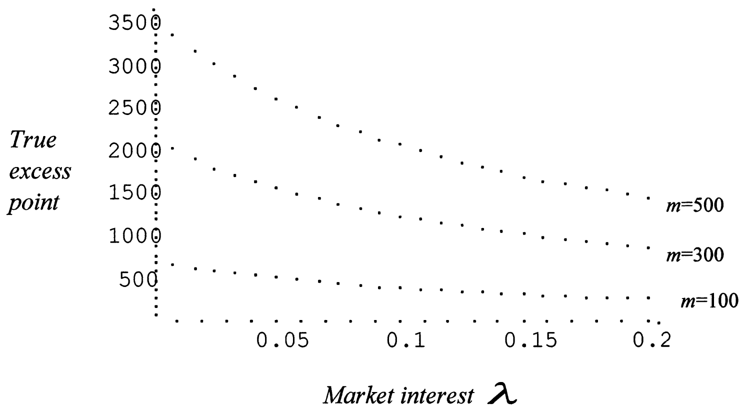

FIGURE 2

In figure 2 we observe – as earlier pointed out generally in proposition 1 – that the higher the market interest λ is, the lower is the true excess point, and hence the more favorable the insurance contract is for the insurance customers. Or in other words; an increasing force of interest in the money market generates a lower true excess point and more reported claims to the insurance company, and vice versa. And as we observe, this effect is stronger the harder the premium increase m is after a claim.

Premium vs. the contractual excess point

Recall the underlying assumption of a fixed dependency between the contractual excess point d and the premiums $p_0(s+t)$ and $p_1(s+t)$ in the expressions (8)-(10) in section 6. In a real world the premium size is obviously influenced by the size of d. Hence, if we want to interpret d as a varying parameter, we have to make concrete assumptions about the dependency between the premium and the customer's choice of d. One very simple method is to let the premium = $p(d) = \omega p$ where $\omega = \exp(-\beta d)$. ω is here interpreted as the per cent discount of the deductible d: $p(0) = p$ and $p(\infty) = 0$. The parameter β has to be determined such that ω generates reasonable values.

The above premium modifications give $p_0(s+t) = \exp(-\beta d)pk^t$ and $p_1(s+t) = \exp(-\beta d)(p+m)k^t$. Hence formula (5) in section 6 is e.g. corrected to $c(x) = m\exp(-\beta d) / (\delta - \ln k)$, which again leads to similarly corrections in the expressions (6)-(12).

8. Concluding remarks

The outline of optimal loss financing under bonus-malus contracts is based on a set of assumptions of the purchasing behavior of the customers. Two assumptions may generate some discussion: 1) The loss of bonus z after a loss occurrence will always be paid by the customer, and 2) the customer will always choose the most profitable financial alternative after a loss occurrence.

An objection to the first assumption may be that the loss of bonus z becomes zero if the customer breaks the insurance contract the year after the claim. This situation is, however, taken care of by definition (2) of z in section 4. Definition (2) gives in this case $z = 0$ since both $p_1(s + t) = 0$ and $p_0(s + t) = 0$ after the contract break. Hence, the value of z is based on the individual (behavior of the) customer, as earlier pointed out in section 4. If the time horizon of the customer is to break the contract the year after a loss, then z becomes zero and the bonus-malus contract becomes a one period standard contract without malus adjustments.

An objection to the second assumption may be that the customers may choose insurance compensation even if it is more optimal to carry occurred losses themselves. Hence $c^*(x)$ in definition 2 in section 4 becomes negative. This will typically happen if the customer is forced to choose insurance compensation because of his or her bad financial position. This situation may be eliminated if the insurer offers a loan facility as a supplement to the bonus-malus insurance contract, and hence takes care of the financial needs of the insurance customer. These needs are probably underestimated by insurance companies as well as by banks. Holtan (1995) gives some ideas of financial services based on these needs. Anyway, because the aim of this paper is to find optimal loss financing properties under bonus-malus contracts, the second assumption seems after all to be a reasonable assumption.

The question of optimal loss financing is directly linked to insurance purchasing questions like: If it is rarely worth to let the insurance company carry a loss, why then purchase the contract? Or in other words: Should – or should not – an individual buy a bonus-malus insurance contract? And if so, what insurance coverage should he or she prefer? These questions lead to a classical field of insurance economics: optimal insurance coverage. Holtan (2001), which is based on – and a direct extension of – this paper, analyses these questions particularly for bonus-malus contracts.

As a concluding remark we may put forward a final question: What does all this mean for the future and the design of insurance contracts with bonus-malus scales? A good prediction seems to be that bonus-malus systems will still exist within products and markets where individual claim experience is a significant risk parameter. However, the ordinary systems seem to be more customer friendly as they will be more and more part of the marketing profiled product advantages within the insurance companies; see e.g. Holtan (1994) and a reply paper by Lemaire & Zi (1994) where such a system is presented, discussed and analyzed. On the other hand, this movement generates an even stronger emergence of individual price adjustments which is not communicated openly to the customers. These price adjustments will take care of the real need for at least limited differentiated experience rating, and hence be based on pure statistical rating techniques, preferably on customer – not product – level. In other words, we may tend to have two bonus-malus systems which influence each customer, one open and communicative system and one "black box" closed and not communicative system. This trend will obviously be a consequence of the increased competition world wide within the non-life insurance markets.

Acknowledgment

This paper was presented at the 30th International ASTIN Colloquium in Tokyo, Japan, 22-25 August 1999. I would like to thank my earlier colleague in if P&C Insurance Ltd., Walther Neuhaus, for some valuable comments to an earlier version of this paper.

References

HOLTAN, J. (1994) Bonus Made Easy. *ASTIN Bulletin* **24**, 61-74.

HOLTAN, J. (1995) Sparing, lån og bilforsikring. (Savings, borrowing and motor insurance). *Nordisk Forsikringstidsskrift* **2**, 181-187. (Nordic insurance journal).

HOLTAN, J. (2001) Optimal Insurance Coverage under Bonus-Malus Contracts. *This volume of ASTIN Bulletin*.

LEMAIRE, J. and ZI, H. (1994) High Deductibles instead of Bonus-Malus: Can it work? *ASTIN Bulletin* **24**, 75-88.

LEMAIRE, J. (1995) *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers, Boston, Dordrecht, London.

NEUHAUS, W. (1988) A bonus-malus system in automobile insurance. *Insurance: Mathematics and Economics* **7**, 103-112.

SUNDT, B. (1988) Credibility estimators with geometric weights. *Insurance: Mathematics and Economics* **7**, 113-122.

SUNDT, B. (1989) Bonus hunger and credibility estimators with geometric weights. *Insurance: Mathematics and Economics* **8**, 119-126.

JON HOLTAN
if P&C Insurance Ltd.
P.O. Box 1340 Vika, N-0114 Oslo, Norway
Telephone: + 47 22 31 52 53
Facsimile: + 47 22 31 24 34
E-mail: jon.holtan@if.no

OPTIMAL INSURANCE COVERAGE UNDER BONUS-MALUS CONTRACTS

BY

JON HOLTAN

if P&C Insurance Ltd., Oslo, Norway

ABSTRACT

The paper analyses the questions: *Should – or should not – an individual buy insurance? And if so, what insurance coverage should he or she prefer?* Unlike classical studies of optimal insurance coverage, this paper analyses these questions from a bonus-malus point of view, that is, for insurance contracts with individual bonus-malus (experience rating or no-claim) adjustments. The paper outlines a set of new statements for bonus-malus contracts and compares them with corresponding classical statements for standard insurance contracts. The theoretical framework is an expected utility model, and both optimal coverage for a fixed premium function and Pareto optimal coverage are analyzed. The paper is an extension of another paper by the author, see Holtan (2001), where the necessary insight to – and concepts of – bonus-malus contracts are outlined.

KEYWORDS

Insurance contracts, bonus-malus, optimal insurance coverage, deductibles, utility theory, Pareto optimality.

1. INTRODUCTION

Should – or should not – an individual buy insurance? And if so, what insurance coverage should he or she prefer? These fundamental questions are of main practical interest within the field of insurance purchasing, and have been extensively studied under varying conditions in insurance economics. Classical references are e.g. Mossin (1968), Arrow (1974) and Raviv (1979). A common factor of all these studies is their straightforward focus on insurance contracts *without* individual bonus-malus (experience rating or no-claim) adjustments. However, both from a customer's point of view and from a theoretical point of view, insurance contracts *with* bonus-malus adjustments, like e.g. motor insurance contracts, are usually much more complex to consider with regard to optimizing the insurance coverage. The increased complexity is caused by the bonus hunger mechanism of the customers; that is, the tendency for insurance customers to carry small losses themselves in order to avoid an increase of future premium costs. The aim of this paper is to take this bonus hunger mechanism into account within the framework of optimal insurance coverage under bonus-malus contracts. The paper is an extension of another paper by the author, Holtan (2001), where the necessary insight to – and concepts of – bonus-malus contracts are outlined.

The paper is organized as follows: Sections 2 and 3 describe the general insurance contract and an expected utility approach to the problem. Sections 4-6 outline some new propositions of the field of optimal insurance coverage particularly for bonus-malus contracts. These propositions are compared to their correspondingly classical propositions for standard insurance contracts. Section 5 treats optimal coverage for a fixed premium function, while section 6 treats pareto optimal coverage. Section 7 gives a summary of the conclusions of the paper.

2. The general insurance contract

We recapitulate briefly the main features of a general bonus-malus insurance contract as outlined in Holtan (2001). Consider an insurance buyer representing a risk of loss X, where X is a stochastic variable with probability density function $f(x)$ and $x \geq 0$. The damage side of the contract is characterized by a contractual compensation $c(x)$ and a true compensation $c^*(x)$ if loss $X = x$ occurs, where

$$c^*(x) = \begin{cases} c(x) - z & \text{if } c(x) > z \\ 0 & \text{if } c(x) \leq z. \end{cases} \tag{1}$$

The true compensation function $c^*(x)$ is the actual compensation function because of its bonus hunger component z, while the contractual compensation function $c(x)$ is no more than the loss amount minus the contractual deductible. The fixed amount z is the excess point of the optimal choice of self-financing generated by the customer's bonus hunger strategy after the loss occurrence, or in other words, the present value of the loss of bonus, and is defined as

$$z = \int_0^\infty e^{-\lambda t}(p_1(s+t) - p_0(s+t))dt, \tag{2}$$

where λ is the non-stochastic market rate of interest of self-financing, $p_1(s+t)$ is the premium paid at time t after a loss occurrence at time s if the loss is reported to the insurer, and $p_0(s+t)$ is the correspondingly premium if the loss is not reported. The premium processes $p_0(s+t)$ and $p_1(s+t)$ are assumed to be continuous non-stochastic for all $t > 0$. See Holtan (2001) for a more detailed and complete description of bonus-malus effects on an insurance contract.

An important statement which partly follows from (1) and (2) is that *independent of the contractual compensation function, the true compensation function has always an individual deductible*; see proposition 2 in Holtan (2001). As we discuss later, this statement explains much of the optimal coverage characteristics of bonus-malus contracts outlined in this paper.

3. An expected utility approach

The existence of a true compensation function obviously influences the individual in his or her choice of insurance coverage within bonus-malus insurance contracts. Recall hereby our introductory questions in section 1: Should – or should not – an individual buy insurance? And if so, what insurance coverage should he or she prefer? Or more precisely: What is the optimal insurance coverage for the individual? As pointed out earlier these questions have traditionally been treated within the framework of insurance economics; in general see e.g. Borch (1990), chapter 2.1, 2.9, 6.3 and 6.4, or a more updated overview in Aase (1993), chapter 8. A brief summary of this classical treatment is as follows:

Consider the insurance customer and the insurance contract described in section 2. Assume w to be the certain initial wealth of the customer. Assume the risk taking preference of the customer to be represented by expected utility $Eu(\cdot)$, that is, facing an uncertain choice the customer is assumed to maximize his expected utility of wealth. Or more precisely: The customer will prefer an uncertain wealth W_1 to another uncertain wealth W_2 if $Eu(W_1) \geq Eu(W_2)$. The preference period of the customer is assumed to be one-period, which is the usual contractual period in non-life insurance. Note that even if the loss of bonus is accumulated over many years, the customers act on the present value of the loss of bonus, and hence the one-period preference period is a consistent assumption in this context.

Classical optimal condition: For the moment consider the classical point of view where the insurance contract *has no* bonus-malus adjustments. Thus the necessary condition for the customer to purchase a coverage $c(\cdot)$ for a premium p is:

$$Eu(w - X + c(X) - p) \geq Eu(w - X). \tag{3}$$

In other words; the customer prefers to buy an insurance coverage $c(\cdot)$ if the expected utility of the coverage is greater than or equivalent to the expected utility of not buying insurance at all. Note that within this framework the random variable X represents the total risk exposure of the customer, which not only includes the uncertain loss amount, but also the uncertain probability of loss occurrence. The probability distribution of X, $f(x)$, is hereby a mixed distribution, containing the probability that no accident occurs at the mass point $x = 0$ and, conditional on one or more accidents, a continuous loss size distribution for $x > 0$.

There may of course exist more than one coverage which satisfies (3). Hence the optimal choice of insurance coverage is the one which maximizes the left hand side of (3) with respect to the function $c(\cdot)$ and the function p, where p in this context obviously must depend on $c(\cdot)$.

Bonus-malus optimal condition: Let us now consider the situation where the insurance contract contains bonus-malus adjustments. Thus (3) is not longer a valid purchasing condition for the customer. The corrected optimal condition is rather influenced by the generalized true compensation function which was defined by (2). More precisely, the necessary condition for the customer to purchase a contractual coverage for a premium p is simply:

$$Eu(w - X + c^*(X) - p) \geq Eu(w - X). \tag{4}$$

In (4) p follows the rules of a general bonus-malus system and is also a function of $c(\cdot)$. If (4) holds for at least one contractual coverage $c(x)$, then the bonus-malus optimal choice of coverage is simply the one which maximizes the left hand side of (4).

Within the framework of bonus-malus insurance contracts condition (4) will obviously influence a wide specter of classical propositions and statements within the theory of optimal insurance coverage. In sections 4-6 some of these classical propositions are presented and thereafter corrected by the effect of the true compensation function within a bonus-malus framework.

4. The indifferent premium

Classical proposition I: Assume the classical framework of a standard insurance contract with no bonus-malus adjustments. The maximum premium the customer will pay for the insurance coverage is the premium $p = p_{max}$ which generates a "=" instead of a "≥" in (3). The premium p_{max} is hence the premium where the customer is indifferent between buying and not buying the insurance coverage, and is therefore also called the *indifferent premium*. The existence of such a premium is actually one of the axioms of the von Neumann-Morgenstern utility theory.

The utility function $u(\cdot)$ is usually assumed to be concave and monotonically increasing, i.e. $u'(\cdot) > 0$ and $u''(\cdot) < 0$, which means that the customer is assumed to be risk averse. Hence, by trivial use of Jensen's inequality, we may find that

$$p_{max} > Ec(X), \tag{5}$$

which is one of the key propositions in insurance economics. A practical interpretation of (5) is that a risk averse customer is willing to participate in an unfair game ($p_{max} = Ec(X)$ is a fair game).

Bonus-malus proposition I: Indifferent premium

Within the framework of a bonus-malus contract the indifferent premium satisfies

$$p_{max} > Ec^*(X), \tag{6}$$

where $c^*(\cdot)$ is defined by (1). □

Proof: From Jensen's inequality it follows that $u(w - EX) > Eu(w - X)$ since $u''(\cdot) < 0$. Hence the equality sign in (4) will hold for some $p_{max} > Ec^*(X)$. □

The practical interpretation of (6) is in fact the same as for (5), that is, a risk averse customer is willing to participate in an unfair game, but the unfair premium limit (the indifferent premium) is different between (5) and (6).

5. Optimal coverage for a fixed premium function

The two introductory questions in section 1 concern the problem of rational insurance purchasing for a fixed set of bonus-malus contracts offered by the insurer. In other words, the terms of the insurance contract are assumed to be exogenously specified and imposed on the insurance customer. This approach reflects a realistic purchasing situation in an insurance mass market, where the customers just within certain limits have possibilities to influence the terms of the insurance contract. The next proposition give attention to a classical statement and to a correspondingly bonus-malus statement within such an exogenous point of view. The contractual compensation assumes to be on excess of loss form, which is probably the most common contractual compensation form in the world wide insurance market.

Classical proposition II:

Assume the classical framework of a standard insurance contract with no bonus-malus adjustments. Assume the contractual compensation to be $c(X) = max[X - d, 0]$, where $d \geq 0$ is the contractual excess point, and the premium to be $p(d) = (1 + \gamma)Ec(X) + k$, where $\gamma \geq 0$ is a safety loading factor and $k \geq 0$ is a flat cost fee. If $\gamma = 0$ (and $w > p(d) + d$) and k is not too high, *it is always optimal to buy maximal contractual coverage*, that is, $d = 0$ is the optimal choice of insurance coverage. If k is too high, the only alternative is not to buy insurance at all. □

This classical statement is e.g. outlined in Borch (1990), pp. 33-34. As we will find, this statement of maximal coverage is also valid under bonus-malus contracts. The point is, however, that the specification of maximal coverage is different under bonus-malus contracts.

Bonus-malus proposition II:

Within the framework of a bonus-malus contract assume the contractual compensation to be $c(X) = max[X-d, 0]$, where $d \geq 0$ is the contractual excess point, and the premium to be $p(d) = (1+\gamma)Ec^*(X) + k$, where $\gamma \geq 0$ is a safety loading factor and $k \geq 0$ is a flat cost fee. If $\gamma = 0$ (and $w > p(d) + z(d) + d$) and k is not too high, *it is always optimal to buy maximal true coverage*, that is, a value of d which gives $z'(d) = -1$ is the optimal choice of insurance coverage. If k is too high, the only alternative is not to buy insurance at all. □

Remark: It is not obvious that insurance companies explicitly calculate $Ec^*(X)$ in the premium expression $p(d) = (1+\gamma)Ec^*(X) + k$. However, implicitly they do because they use the actual reportet claims – which are affected by the bonus hunger of the customers – as data input to the risk premium estimation.

Proof: From (1) the true compensation is $c^*(X) = max[X-d-z(d), 0]$, where the bonus hunger excess point $z(d)$ obviously is a function of d since $p(d)$ depends on d.

The optimal coverage maximizes the left hand side of (4). Hence we have:

$$U(d) = Eu[w - X + c^*(X) - p(d)]$$

$$= \int_0^{d+z(d)} u[w-x-p(d)]\,f(x)dx + u[w-d-z(d)-p(d)] \int_{d+z(d)}^{\infty} f(x)dx. \tag{7}$$

The first order condition for a maximum is $U'(d) = 0$. Hence by straightforward calculus we find:

$$U'(d) = -p'(d) \int_0^{d+z(d)} u'[w-x-p(d)]\,f(x)dx$$

$$-(1+z'(d)+p'(d))u'[w-d-z(d)-p(d)] \int_{d+z(d)}^{\infty} f(x)dx. \tag{8}$$

We have:

$$p(d) = (1+\gamma)Ec^*(X) + k = (1+\gamma) \int_{d+z(d)}^{\infty} (x-d-z(d))f(x)dx + k, \tag{9}$$

which gives:

$$p'(d) = -(1+\gamma)(1+z'(d)) \int_{d+z(d)}^{\infty} f(x)dx, \tag{10}$$

and hereby:

$$1 + z'(d) + p'(d) = (1 + z'(d))\left[\int_0^{d+z(d)} f(x)dx - \gamma \int_{d+z(d)}^{\infty} f(x)dx\right]. \tag{11}$$

Hence, by substituting (10) and (11) into (8), followed by straightforward calculus, we find:

$$U'(d) = [1 + z'(d)] \int_{d+z(d)}^{\infty} f(x)dx \cdot \tag{12}$$

$$\left[(1+\gamma)\int_0^{d+z(d)} [u'(w - p(d) - x) - u'(w - p(d) - d - z(d))]\, f(x)dx + \gamma v'(w - p(d) - d - z(d)) \int_{d+z(d)}^{\infty} f(x)dx\right]$$

Since $u'(\cdot) > 0$ and $u''(\cdot) < 0$, we observe from (12) that if $\gamma = 0$ and $w > p(d) + z(d) + d$, we have

$$U'(d) = 0 \text{ if and only if } z'(d) = -1.$$

From (12) we also have generally

$$U'(d) > 0 \text{ if } z'(d) < -1$$
$$U'(d) < 0 \text{ if } z'(d) > -1,$$

which implies that $z'(d) = -1$ is a maximum point of $U(d)$, as shown illustratively in figure 1:

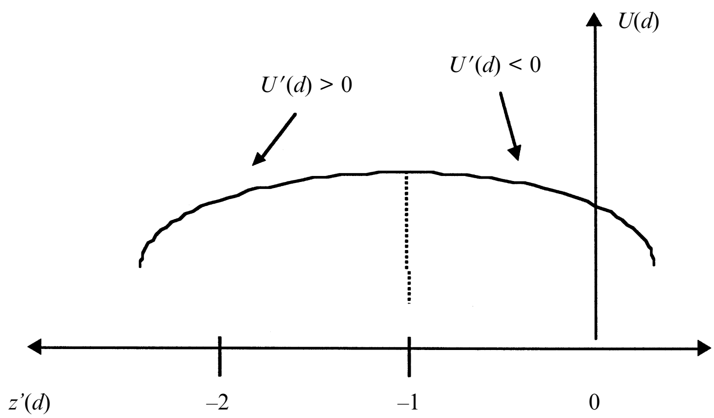

FIGURE 1

Hence, if $\gamma = 0$ and $w > p(d) + z(d) + d$, then a value of d which gives $z'(d) = -1$ generates an optimal coverage solution (maximal expected utility) for the customer. □

If $D(d) = d + z(d)$ = the true deductible, then $D'(d) = 1 + z'(d)$, and hence $D'(d) = 0$ if $z'(d) = -1$. Since $D'(d) < 0$ when $z'(d) < -1$ and $D'(d) > 0$ when $z'(d) > -1$, then $z'(d) = -1$ represents a *minimum point* of $D(d)$. This minimum is greater than zero because $z(d) > 0$ for all $d \geq 0$.

From (10) we have correspondingly $p'(d) = 0$ if and only if $z'(d) = -1$. Since $p'(d) > 0$ when $z'(d) < -1$ and $p'(d) < 0$ when $z'(d) > -1$, then $z'(d) = -1$ represents a *maximum point* of $p(d)$.

Hence we conclude: $z'(d) = -1$ generates a maximum value of the premium $p(d)$ and a minimum value of the true deductible $d + z(d)$, which together gives maximal true coverage. In other words, *maximal true coverage gives maximal expected utility for the customer*, given that $\gamma = 0$ in the assumed premium function.

Note that there may exist more than one value of d satisfying $z'(d) = -1$; call them d_{max}. All other values different from d_{max} give lower expected utility from the customers point of view. Figure 2 gives an illustrative interpretation of this result by illustrating the existence of a tangent line $z'(d) = -1$ touching $z(d)$ in the maximum expected utility point d_{max}. For simplicity the figure assumes the existence of just one maximum point d_{max} satisfying $z'(d) = -1$ and a bonus-malus contract with *decreasingly* premium reduction generated by the deductible d.

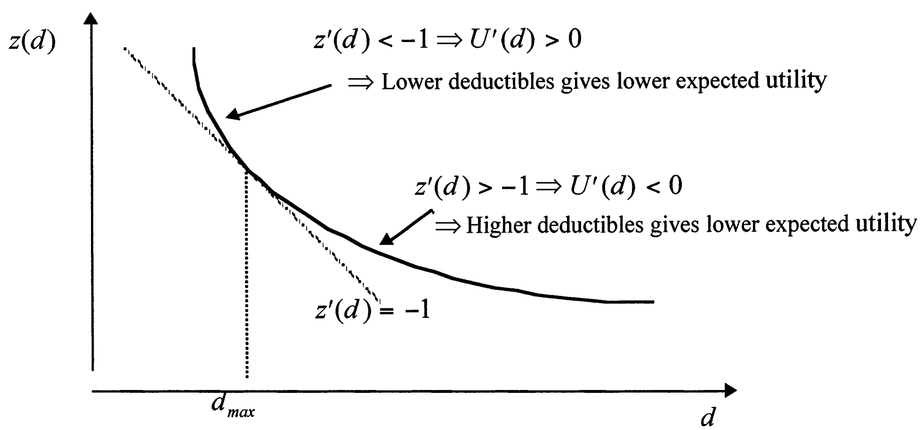

FIGURE 2

The bonus-malus rules of the contract, the market rate λ and the individual premium level at the purchasing time, decide the individual value(s) of d_{max} as well as the decreased expected utility for values of d different from d_{max}. This quite complex and individual dependent conclusion reflects to some extent the practical purchasing situation: Both the insurance company and the insurance customers find it difficult to recommend and choose an individual contractual deductible under bonus-malus contracts. And, if the premium reduction generated by the contractual deductible d has an upper limit (as most insurers have), there may not for some customers exist individual value(s) of d_{max} at all. Hence, these customers should probably not buy the insurance coverage either. These customers are typically customers with low bonus level, high premium level and hard malus rules in an economic market with low market rate λ. On the other hand, customers with high bonus level, low premium level and nice malus rules in an economic market with high market rate λ, should obviously buy the insurance coverage and choose d_{max} as the contractual deductible.

To summarize this section we conclude within our bonus-malus model: Given an excess of loss contractual compensation and a premium function without a safety loading factor, then there exists a specific choice of contractual coverage which gives maximum expected utility compared to other choices of coverage. This optimal contractual coverage is defined when the true insurance coverage is maximal, that is, when $z'(d) = -1$. This conclusion is in accordance with the correspondingly standard insurance contract without bonus-malus adjustments, where maximal (contractual) coverage is optimal for the customers.

Note that even if we in our model have defined the true deductible as a net present value based on an infinite-horizon consideration of the loss of bonus, the above conclusions will also hold for other considerations and assumptions of $z(d)$. The only condition is that $z(d)$ depends on d in some way.

6. Pareto optimal coverage

The conclusion in section 5 leads to a more general approach of deriving the optimal insurance coverage under bonus-malus contracts. A reversed key question is hereby: What is the optimality of a bonus-malus contract in an insurance market? And even more critical: Does there exist such an optimality at all? These problems involve *Pareto optimal* analysis techniques, where both the insurance customer and the insurer is analyzed from a risk-sharing point of view.

Hence consider a general insurance contract with bonus-malus adjustments. The necessary condition for the insurer to offer the true compensation $c^*(X) = max[c(X)\text{-}z(p), 0]$ for a premium p is obviously

$$Eu_0(w_0 - c^*(X) + p) \geq u_0(w_0), \tag{13}$$

where $u_0(\cdot)$ is the utility function of the insurer satisfying $u_0'(\cdot) > 0$ and $u_0''(\cdot) \leq 0$, w_0 is the initial wealth of the insurer and p follows the rules of a general bonus-malus system. In order for a bonus-malus contract to be acceptable to both the insurer and the customer, both (13) and (4) have to be satisfied. If such a contract exists at all, then the Pareto optimal contract is the one which maximizes the total risk-exchange utility for the insurer and the insured, that is, the contract which maximizes the left hand side of (4) and (13). This simple risk-exchange model is hereafter referred to as *the standard risk-exchange model*, which is e.g. part of Borch's classical 1960-theorem of Pareto Optimality. Within this framework Borch's theorem says in fact that a *sufficient* condition that our (bonus-malus) contract is Pareto Optimal is that there exist positive constants k_0 and k such that

$$k_0 u'_0(w_0 + p - c^*(X)) = ku'(w - p - X + c^*(X)),$$

which mathematically expresses a common linear maximizing of the left hand side of both (4) and (13). See Borch (1990), chapter 2.5 or Aase (1993), chapter 3, for a more detailed presentation.

We have:

Bonus-malus proposition III: A bonus-malus contract can not be Pareto Optimal within the standard risk-exchange model. □

Proof: A direct application of Borch's Theorem gives the first order condition for the Pareto optimal sharing rule between the insurer and the customer

$$u'_0(w_0 + p - c^*(X)) = \left[k/_{k_0} \right] u'(w - p - X + c^*(X)), \tag{14}$$

where k and k_0 are arbitrary positive constants. Following Aase (1993), chapter 8, a differentiating of (14) with respect to X leads to

$$\frac{\partial}{\partial X} c^*(X) = \frac{R(w - p - X + c^*(X))}{R_0(w_0 + p - c^*(X)) + R(w - p - X + c^*(X))}, \tag{15}$$

where R and R_0 are the Arrow-Pratt measures of absolute risk aversions for the customer and the insurer. If both the customer and the insurer are risk averse, then directly from (15) we establish the general Pareto optimal criteria

$$0 < \frac{\partial}{\partial X} c^*(X) < 1 \text{ for all X} \geq 0. \tag{16}$$

On the other hand, under bonus-malus contracts we have $c^*(X) = max[c(X) - z(p), 0]$. Hence $\frac{\partial}{\partial X} c^*(X) = 0$ for $c(X) \leq z(p)$, and hence quite generally the Pareto optimal criteria (16) will not hold for all $X \geq 0$. □

Under standard insurance contracts without bonus-malus adjustments the corresponding proposition is as follows; see Aase (1993), chapter 8, for a general proof which follows the same lines as the above proof:

Classical proposition III: The Pareto optimal sharing rule of a standard insurance contract without bonus-malus adjustments involves a positive amount of coinsurance within the standard risk-exchange model. A contractual compensation with a deductible can, however, not be Pareto optimal within the standard risk-exchange model. □

Proposition 2 in Holtan (2001) states that independent of the contractual compensation function, the true compensation function has always an individual deductible under bonus-malus contracts. Hence, given the standard risk-exchange model, it is intuitively correct that the Pareto optimal statement for standard contracts with a deductible is valid *in general* for bonus-malus contracts.

As concluded in Aase (1993), chapter 8, standard insurance contracts "with a deductible can only be Pareto optimal in models where one or more of the following are included; costs, moral hazard, asymmetric information, non-observability or alternative preferences (e.g. star-shaped utility)". Standard references within this context are: Arrow (1974), who included a fixed percentage (cost)loading to show optimality of deductibles, Raviv (1979), who found that a deductible is Pareto optimal if and only if the insurance costs depends on the insurance coverage, Rothschild & Stiglitz (1976), who included asymmetric information and found that low-risk individuals would choose high deductibles, and Holmstrøm (1979), who found that moral hazard gives rise to deductibles.

These expanded model assumptions are in general in accordance with the main intentions of a bonus-malus system in an insurance market:

1) *Adverse selection*: Measure and smooth out asymmetric information by individual a posteriori tariffication.
2) *Moral hazard*: Reduce the claim probability by economic punishment.
3) *Costs*: Reduce the administrative costs generated by claims handling.

Hence, since no one of these intentions was included in the model in this paper, we put forward the following conjecture:

Conjecture: A bonus-malus contract can only be Pareto optimal if the risk-exchange model includes one or more of the bonus-malus intentions 1-3. □

Proposition 3 and 4 in Holtan (2001) state that the compensation function of a bonus-malus contract without a contractual deductible is equivalent to the compensation function of a standard insurance contract with an individual deductible. Hence it should be easy to formally prove the existence of the conjecture for bonus-malus contracts without a contractual deductible.

On the other hand, if we do not restrict a bonus-malus contract in this way, then the size of the loss of bonus deductible depends on the individual choice of the contractual deductible, cf. the discussion in section 5. This dependency complicates the Pareto optimal analysis, and hence also the proof of the above conjecture.

As a concluding remark to the above discussion, we may point out that ordinary deductibles are usually used in the insurance market as the main instrument to reduce the claim probability (moral hazard) and to reduce the costs generated by claims handling. Therefore, the main intention of a bonus-malus system is to handle the problem of adverse selection generated by individual asymmetric information (even if Holtan (1994) outlines a model with high deductibles financed over a period of time as an adverse selection alternative to bonus-malus systems). Hence, as a general rule bonus-malus systems should only be used if individual loss experience is a significant risk parameter within the insurance market.

7. Summary

The paper outlines some new statements of optimal insurance coverage under bonus-malus contracts and compares them with corresponding classical statements under standard insurance contracts. The theoretical framework is an expected utility model, but neither adverse selection, moral hazard nor costs are part of the model. Under the assumption of an excess of loss contractual compensation and a premium function without a safety loading factor, it is outlined that maximal true coverage gives maximal expected utility for the customers. This result is in accordance with classical theory of standard contracts without bonus-malus adjustments. On the other hand and within the same expected utility model, it is outlined that bonus-malus contracts are not optimal to both the customers and the insurer at the same time, that is, Pareto optimal. The conjecture in section 6, which is not formally proved, states as a natural consequence that bonus-malus contracts can only be Pareto optimal if adverse selection, moral hazard and/or costs are included in the analysis model.

Acknowledgment

This paper was presented at the 30th International ASTIN Colloquium Tokyo, Japan, 22-25 August 1999.

References

Aase, K.K. (1993) Equilibrium in a reinsurance syndicate; Existence, uniqueness and characterization. *ASTIN Bulletin* **23**, 185-211.

Arrow, K.J. (1974) Optimal insurance and generalized deductibles. *Skandinavisk Aktuarietidsskrift*, 1-42.

Borch, K.H. (1990) *Economics of Insurance*. Advanced Textbooks in Economics Series. (Editors: Knut K. Aase and Agnar Sandmo), North Holland; Amsterdam, New York, Oxford, Tokyo.

Holmström, B. (1979) Moral hazard and observability. *The Bell Journal of Economics* **10**, 74-91.

Holtan, J. (1994) Bonus Made Easy. *ASTIN Bulletin* **24**, 61-74.

Holtan, J. (2001) Optimal Loss Financing under Bonus-Malus Contracts. *This volume of ASTIN Bulletin.*

Mossin, J. (1968) Aspects of rational insurance purchasing. *Journal of Political Economy* **76**, 553-568.

Raviv, A. (1979) The design of an optimal insurance policy. *American Economic Review* **69**, 84-96.

Rothschild, M. and Stiglitz, J. (1976) Equilibrium in competitive insurance markets. An essay on the economics of imperfect information. *Quarterly Journal of Economics* **90**, 629-650.

Jon Holtan
if P&C Insurance Ltd.
P.O. Box 1340 Vika
N-0114 Oslo, Norway

FINANCIAL DATA ANALYSIS WITH TWO SYMMETRIC DISTRIBUTIONS

BY

WERNER HÜRLIMANN

ABSTRACT

The normal inverted gamma mixture or generalized Student t and the symmetric double Weibull, as well as their logarithmic counterparts, are proposed for modeling some loss distributions in non-life insurance and daily index return distributions in financial markets. For three specific data sets, the overall goodness-of-fit from these models, as measured simultaneously by the negative log-likelihood, chi-square and minimum distance statistics, is found to be superior to that of various "good" competitive models including the log-normal, the Burr, and the symmetric α-stable distribution. Furthermore, the study justifies on a statistical basis different important models of financial returns like the model of Black-Scholes (1973), the log-Laplace model of Hürlimann (1995), the normal mixture by Praetz (1972), the symmetric α-stable model by Mandelbrot (1963) and Fama (1965), and the recent double Weibull as limiting geometric-multiplication stable scheme in Mittnik and Rachev (1993). As an application, the prediction of one-year index returns from daily index returns is discussed.

KEYWORDS

Claim size data, financial market data, index return, normal inverted gamma mixture, generalized Student t, symmetric double Weibull, goodness-of-fit.

1. INTRODUCTION

The fitting of probability distributions to financial data is a statistical subject with a long tradition in both actuarial and financial literature. The detailed analysis of the available models leads to many unsolved problems of theoretical and practical importance, and this field of research always generates new challenges. The present contribution is a further piece of this big puzzle.

The proposed models belong, after appropriate transformation, to the class of symmetric distributions. Let us argue in favor of such a seemingly severe restriction. First of all, applying an adequate transform $T(X)$ to a random variable X often reveals approximate symmetry in the sense that

$$T(X) = \mu + c \cdot Z, \tag{1.1}$$

where (μ, c) are *location* and *scale* parameters, and Z is a *symmetric* random variable with mean zero. The ubiquitous transform in this respect is the logarithmic transform $T(X) = \ln(X)$. Another motivation for considering symmetric distributions for Z in (1.1) is the desire to measure the departure from a normal random variable. Besides its practical appeal, the latter working hypothesis finds some theoretical foundation (e.g. Efron (1982)). Empirical arguments are also available. Important financial data for which this approach has been considered adequate include in particular daily returns in equity markets (e.g. Taylor (1992), p. 45). Furthermore, the logarithm of non-life claim sizes has often a low skewness, and can therefore be modeled using the device (1.1). A short outline of our study follows.

Section 2 presents the method applied to determine the unknown parameters and the goodness-of-fit statistics used to assess the overall fit of an estimated distribution. Sections 3 and 4 introduce the proposed symmetric distributions. The required formulas to do all calculations for the comparative distributions used in our study are summarized in the Appendix. The results of our extensive data analysis are exposed in Sections 5 and 6. Finally, to illustrate the potential use of the proposed models, we show in Section 7 how one-year index returns can be predicted from the distributions of daily index returns.

2. Estimation method and goodness-of-fit statistics

Given a restriction to two and three parameter distributions, the distribution of Z in (1.1) is either parameter-free or contains one *shape* parameter. The location and scale parameters μ and c in (1.1) are throughout estimated with the maximum likelihood method. The theoretical justification of this procedure lies in asymptotic statistics, and is explained in many of the modern statistical textbooks. A recent unification result about the maximum likelihood estimation of location and scale parameters is presented in Hürlimann (1998a). A remaining shape parameter a is either included in a three parameter maximum likelihood estimation or it is treated as nuisance parameter. In the latter case, it is chosen to minimize individually or simultaneously some of the goodness-of-fit statistics presented below. Maximum likelihood estimators are denoted $\hat{\mu}, \hat{c}, \hat{a}$. The value of a nuisance parameter is simply denoted by a (without a "hat"). The estimation procedure for the shape parameter is motivated as follows. As our experience has shown, a simultaneous three parameter maximum likelihood estimation often causes numerical difficulties, and does not always lead to an overall best fit. The latter point is illustrated in the text with the NIG ranked 5 in Table 6.3.

In the practical analysis, it is assumed that the data sets consisting of n observations are grouped into m classes with boundaries $\xi_0, \xi_1, \ldots, \xi_m$. The only available information are the frequencies $\lambda_1, \lambda_2, \ldots, \lambda_m$ of the corresponding classes $(\xi_0, \xi_1], (\xi_1, \xi_2], \ldots, (\xi_{m-1}, \xi_m]$. Financial results are supposed to occur at the average values in $(\xi_{i-1}, \xi_i]$, say at $m\xi_i$, i=1, …, m. Often, in particular in case the average value is not known, we set by convention $m\xi_i = \frac{1}{2}(\xi_{i-1} + \xi_i)$, $i = 1, \ldots, m$. Suppose that the data are observations from a random variable X with a survival distribution $S(x) = S_x(x;\theta)$, $\theta = (\theta_1, \ldots, \theta_p)$ an unknown parameter vector, and that the data are truncated at ξ_0. Of interest is thus the truncated random variable $X_0 = (X|X > \xi_0)$ with survival distribution

$$S_0(x) = \begin{cases} 1, & x \le \xi_0, \\ \dfrac{S(x)}{S(\xi_0)}, & x > \xi_0. \end{cases} \tag{2.1}$$

In case the data is not truncated at ξ_0, we assume that $S(\xi_0) = 1$ and our subsequent analysis remains valid with $S_0(x) = S(x)$.

The quality of fit of the various models will be measured using 5 goodness-of-fit statistics, the first 3 of which have a well-known theoretical justification. The other 2 ad hoc statistics are based on reliability measures, and have been used and motivated by some actuaries (see e.g. Hogg and Klugman (1984), pp. 108-111). To assess the *overall goodness-of-fit*, a decision under multiple criteria is necessary. Our simple overall rank is based on the first 3 theoretical criteria. A fitted distribution is ranked before another one if two of the negative log-likelihood, chi-square and minimum distance statistics have a smaller value. One should emphasize that the defined criterion is merely another decision rule, which helps to select good models. It cannot replace a *formal statistical test* like the chi-square goodness-of-fit test or the Kolmogorov-Smirnov test for the ultimate validation or rejection of a model (consult Klugman et al. (1998), Section 2.9, for further discussion on this). In particular, any informal decision rule is necessarily a subjective judgement, which may lead to inappropriate conclusions. A significant illustration of this phenomenon is provided in Section 6. We do not include the other empirical measures in our overall goodness-of-fit criterion. There are two reasons for this. First, our examples show that the LE- and ME-statistics defined below are quite sensitive to changes in parameter values. Second, it is possible to find low LE- and ME-values even if the 3 theoretical criteria do not attain at all their minimal values (e.g. the lnNIG ranked 5 in Table 5.3, the NIG ranked 5 and the log-normal in Table 6.3). A decision including the LE- and ME-statistics appears thus *inconsistent* with our estimation method, at least with respect to the negative log-likelihood and chi-square criteria.

2.1. The negative log-likelihood statistic

The *negative log-likelihood* of X_0 reads

$$-\ln L = n \cdot \ln S(\xi_0) - \sum_{k=1}^{m} \lambda_k \cdot \ln\left[S(\xi_{k-1}) - S(\xi_k)\right], \tag{2.2}$$

and the goal is to minimize this quantity. This is achieved through application of the *scoring method* (e.g. Hogg and Klugman (1984), chap. 3.7 and 4.3, or Klugman et al. (1998)). Define

$$P_i = P_i(\theta) = \frac{S(\xi_{i-1}) - S(\xi_i)}{S(\xi_0)}, \quad i = 1,\ldots,m, \tag{2.3}$$

and consider the information matrix $A = A(\theta)$ with elements

$$a_{rs} = n \cdot \sum_{i=1}^{m} \frac{\partial P_i}{\partial \theta_r} \cdot \frac{\partial P_i}{\partial \theta_s} \cdot \frac{1}{P_i}, \quad r,s = 1,\ldots,p, \tag{2.4}$$

and the score vector $S = S(\theta)$ with elements

$$S_r = \frac{\partial \ln L}{\partial \theta_r} = \sum_{i=1}^{m} \lambda_i \cdot \frac{\partial P_i}{\partial \theta_r} \cdot \frac{1}{P_i}, \quad r = 1,\ldots,p. \tag{2.5}$$

Given a preliminary estimate θ_0, then an iterative method to get the maximum likelihood estimate $\hat{\theta}$ of θ is described by the recursion

$$\theta_k = \theta_{k-1} + A(\theta_{k-1})^{-1} \cdot S(\theta_{k-1}), \quad k = 1,2,\ldots \tag{2.6}$$

In case this sequence converges to $\hat{\theta}$, insertion in (2.2) will yield a numerical approximation to the desired minimum value of –ln$L$. Even if the sequence does not converge, it is possible to obtain with this method estimates $\hat{\theta}$ with a comparatively small practical value of –ln$L$.

2.2. The chi-square statistic

With grouped data the quality of fit is often measured using *Pearson's goodness-of-fit statistic*

$$\chi^2 = \chi^2(\theta) = \sum_{i=1}^{m} \frac{\left(\lambda_i - nP_i(\theta)\right)^2}{nP_i(\theta)}, \text{ with} \tag{2.7}$$

A comparatively small value χ^2 of is an indicator of an acceptable fit according to the following elegant theory (e.g. Hogg and Klugman (1984), p. 107). Suppose $\theta_{\min}$ solves the minimization problem

$$\chi^2_{\min} = \chi^2(\theta_{\min}) = \min_{\theta}\left[\chi^2(\theta)\right]. \tag{2.8}$$

Then the statistic $\chi^2_{\min}$ has an approximate chi-square distribution with $m-1-p$ degrees of freedom (e.g. Cramér (1946), Fisz (1973), p. 512-513). If $\chi^2_{\min}$ is sufficiently small, then one accepts $S_0(x;\theta_{\min})$ as a reasonable model. However, if $\chi^2_{\min} \geq z_\alpha$, where $\Pr(\chi^2(m-1-p) \geq z_\alpha) = \alpha$, then one rejects the model $S_0(x; \theta_{\min})$ at the α significance level. Even if the minimum chi-square estimate $\theta_{\min}$ is actually too tough to be calculated, the χ^2-statistic is very useful. For example, if the maximum likelihood estimate or another estimate $\hat{\theta}$ is substituted into $\chi^2(\theta)$ instead of $\theta_{\min}$, then $\chi^2(\hat{\theta}) \geq \chi^2_{\min}$. Therefore, using $\chi^2(\hat{\theta})$ instead of $\chi^2_{\min}$, a model will be rejected at a somewhat larger significance level as that required. Another justification for this substitution is the fact that the maximum likelihood and the minimum chi-square estimators are *asymptotically equal* in case the same class boundaries $\xi_0, \xi_1, \ldots, \xi_m$ are used (e.g. Cramér (1946)).

2.3. The minimum distance statistic

With grouped data, another important measure of the quality of fit is the *weighted Cramér-von Mises statistic*

$$K = \sum_{i=1}^{m} \frac{m}{F_0(\xi_i) \cdot S_0(\xi_i)} \cdot \left[F_i - F_0(\xi_i)\right]^2, \tag{2.9}$$

with $F_i = \frac{1}{n}\sum_{j=1}^{i} \lambda_j$, $i = 1, \ldots, m$, the empirical distribution function (e.g. Hogg and Klugman (1984), p. 135). For the "true" parameter vector θ, each term has a chi-square distribution with one degree of freedom, which justifies this statistic for empirical testing. Though substitution of an estimate $\hat{\theta}$ for θ will destroy the chi-square property, the K-statistic is an appealing measure. Each term makes an equal contribution to the total, and the weights $w(\xi_i) = m \cdot \left[F_0(\xi_i) \cdot S_0(\xi_i)\right]^{-1}$ are largest at the ends of the distribution. In particular, the K-statistic is useful for the analysis of long-tailed data.

2.4. The mean excess distance statistic

Consider the *mean excess function* of $X_0 = (X | X > \xi_0)$, that is

$$e(x) = E\left[X_0 - x | X_0 > x\right] = \frac{\pi_X(x)}{S_X(x)}, \quad x > \xi_0 \text{ ,where} \tag{2.10}$$

$$\pi_X(x) = \int_x^{\infty} S_X(t)\,dt \tag{2.11}$$

is the stop-loss transform of X, and its empirical counterpart

$$\hat{e}(\xi_i) = \frac{\sum_{j=i+1}^{m} (m\xi_j - m\xi_i) \cdot \lambda_j}{n - \sum_{j=1}^{m} \lambda_j}, \quad i = 1, \ldots, m-1. \tag{2.12}$$

In general, the mean excess plot of (2.12) exhibits an increasing slope for long-tailed data, a constant plot for exponential distributions, and a decreasing slope for short-tailed data. As our financial market data sets have shown, a convex plot may also occur quite frequently. Due to scarce observations in the tails, there may be a large uncertainty about the true behavior of $e(x)$, especially in the tails of the distribution. The best fit in this respect might not always lead to the best actuarial decision (Hogg and Klugman (1984), chap. 4). Despite of these and other shortcomings (lack in sampling distribution theory, see however Carriere (1992)), it seems useful to consider the *mean excess distance statistic*

$$ME = \sum_{i=1}^{m-1} \left[\frac{e(\xi_i) - \hat{e}(\xi_i)}{e(\xi_i)} \right]^2, \tag{2.13}$$

which should be as small as possible for a good fit.

2.5. The limited expected value distance statistic

In some situations it is impossible to calculate the mean excess function, for example when the mean of X does not exist, or it is impossible to compute the empirical mean excess function, for example when the data are censored. It is then useful to consider the *limited expected value function* of X_0 at x defined as the mean of X_0 censored at x through the expression

$$LE(x) = E\left[\min(X_0, x\right] = \xi_0 + \frac{\int_{\xi_0}^{x} S(t)dt}{S(\xi_0)} \tag{2.14}$$

If the mean excess function exists, one has the relationship

$$LE(x) = \xi_0 + e(\xi_0) - S_0(x) \cdot e(x). \tag{2.15}$$

The empirical counterpart of (2.15) is (e.g. Hogg and Klugman (1984), p. 151)

$$L\hat{E}(\xi_i) = \frac{1}{n} \sum_{j=1}^{i} \lambda_j \cdot m\xi_j + \xi_i \cdot [1 - F_1], \quad i = 1,...,m. \tag{2.16}$$

As a goodness-of-fit measure one uses the *limited expected value distance statistic*

$$LE = \sum_{i=1}^{m} \left[\frac{LE(\xi_i) - L\hat{E}(\xi_i)}{LE(\xi_i)} \right]^2, \tag{2.17}$$

which should be as small as possible.

3. The normal inverted gamma mixture or generalized Student t distribution.

The following distribution has been proposed to model financial returns by Praetz (1972) (see also Blattberg and Gonedes (1974), Kon (1984), Taylor (1992), Section 2.8). Its potential usefulness in actuarial science has been pointed out in Hürlimann (1995a).

If (X|θ) is conditional on θ normally distributed with mean μ and variance $1/\theta$, and θ follows a conjugate gamma prior $\Gamma(\frac{1}{2}c^2, a)$, $a > 0$, then X has the unconditional density (e.g. Hogg and Klugman (1984), p. 52-53, Heilmann (1989), example 3.7):

$$f_X(x) = \frac{\Gamma(a+\frac{1}{2})}{\sqrt{\pi}\cdot\Gamma(a)}\cdot\frac{1}{c}\left[\frac{c^2}{c^2+(x-\mu)^2}\right]^{a+\frac{1}{2}}, \tag{3.1}$$

with $\Gamma(x)$ the gamma function. We say that X has a normal inverted gamma mixture with parameters μ, c, a, abbreviated $NIG(\mu,c,a)$. The location-scale transform $Z = \frac{X-\mu}{c}$ has a Pearson type VII density (e.g. Johnson et al. (1995), Section 28.6)

$$f_z(z) = \frac{1}{B(a,\frac{1}{2})\cdot(1+z^2)^{a+\frac{1}{2}}}\ , \tag{3.2}$$

where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(\alpha+\mathrm{b})}$ is a beta coefficient. This can be viewed as a generalized Student t distribution because if $a = \frac{\upsilon}{2}$, $\upsilon = 1,2,3,\ldots$ is an integer, the random variable $\sqrt{\upsilon}\cdot Z$ has a Student t with υ degrees of freedom. In particular, $a = \frac{1}{2}$ is a Cauchy and $a = 1$ is a Bowers distribution (for the latter see Hürlimann (1993/95a/97/98b) among others). The substitution $t = \frac{z^2}{1+z^2}$ shows the integral identity

$$\int_0^x \frac{dz}{(1+z^2)^{a+\frac{1}{2}}} = \frac{1}{2}\int_0^{\frac{x^2}{1+x^2}} t^{-\frac{1}{2}}(1-t)^{a-1}\,dt, \tag{3.3}$$

from which it follows that the survival distribution satisfies the equivalent expressions

$$S_z(z) = \begin{cases} \frac{1}{2}\left[1-\beta\left(\frac{1}{2}, a; \frac{z^2}{1+z^2}\right)\right], & z \geq 0, \\ \frac{1}{2}\left[1+\beta\left(\frac{1}{2}, a; \frac{z^2}{1+z^2}\right)\right], & z \leq 0, \end{cases} \tag{3.4}$$

and

$$S_z(z) = \begin{cases} \frac{1}{2}\beta\left(\frac{1}{2}, a; \frac{1}{1+z^2}\right), & z \geq 0, \\ 1-\frac{1}{2}\beta\left(\frac{1}{2}, a; \frac{1}{1+z^2}\right), & z \leq 0, \end{cases} \tag{3.5}$$

where $\beta(a,b;x) = 1-\beta(b,a;1-x) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1}dt$ is a beta density. While the mean and skewness of Z are zero, the variance (if $a>1$) and kurtosis (if $a>2$) are given by

$$\sigma_z^2 = Var[Z] = \frac{1}{2(a-1)}, \quad \gamma_{2,z} = \frac{E[Z^4]}{Var[Z]^2} = 3 \cdot \left[\frac{a-1}{a-2}\right]. \tag{3.6}$$

The kurtos is takes values in $[3,\infty)$, and is therefore capable to model leptokurtic data. The scoring method for maximum likelihood estimation requires the knowledge of the partial derivatives of the survival distribution found in the Appendix.

The stop-loss transform of this statistical model reads (trivial exercise)

$$\pi_X(d) = \frac{c^2+(d-\mu)^2}{2a-1} \cdot f_X(d) - (d-\mu) \cdot S_X(d). \tag{3.7}$$

Setting $t_a = \sqrt{2(a-1)} \cdot Z$ if $a>1$, a standardized $NIG(0,1,a)$, one gets with $c = \sigma \cdot \sqrt{2(a-1)}$, $\sigma = \sigma_X$, that

$$\pi_X(d) = \frac{2(a-1)\sigma^2+(d-\mu)^2}{2a-1} \cdot \frac{1}{\sigma} f_{t_a}\left[\frac{d-\mu}{\sigma}\right] - (d-\mu) \cdot S_{t_a}\left[\frac{d-\mu}{\sigma}\right]. \tag{3.8}$$

As $a \to \infty$ one knows that t_a is a standardized normal random variable, hence (see also Hürlimann (1995a) for a special case)

$$\lim_{a\to\infty} \pi_X(d) = \sigma \cdot \varphi\left[\frac{d-\mu}{\sigma}\right] - (d-\mu) \cdot \overline{\Phi}\left[\frac{d-\mu}{\sigma}\right], \tag{3.9}$$

with $\Phi(x)$ the standard normal distribution, $\overline{\Phi}(x) = 1-\Phi(x)$ and $\varphi(x) = \Phi'(x)$. This is the stop-loss transform of a normal $N(\mu,\sigma)$ random variable. Using that the stop-loss transform uniquely determines the distribution function (e.g. Gerber (1979), Müller (1996), Hürlimann (2000)), one sees that a $NIG(\mu, c, a)$ is asymptotically normally distributed as $a \to \infty$.

Of interest is also the logarithmic version of the above. The random variable X such that $\ln(X) = \mu + c \cdot Z$, with Z a $NIG(0, 1, a)$ random variable, defines the *logarithmic normal inverted gamma mixture*, abbreviated $\ln NIG(\mu, c, a)$. Its density and survival distribution are given by

$$f_X(x) = \frac{1}{cx} f_Z\left[\frac{\ln(x)-\mu}{c}\right], \quad S_X(x) = S_Z\left[\frac{\ln(x)-\mu}{c}\right]. \tag{3.10}$$

Since $E\left[e^{cz}\right] = \infty$ the mean and stop-loss transform do not exist. However, since $NIG(0, 1, a)$ converges asymptotically as $a \to \infty$ to a normal distribution, the $\ln NIG(\mu, c, a)$ is a valuable alternative to the log-normal. This situation occurs in case the mixture is used as a Bayesian prediction model as in

Hürlimann (1995a). Indeed, if $D_n = (x_1, ..., x_n)$ is a sample of n observations from X, then the up-dated parameters of the predictive distribution $\ln NIG(\mu, c_n, \alpha_n)$ are (see Hürlimann (1995a), (2.2))

$$c_n = \sqrt{c^2 + \sum_{i=1}^{n} [\ln(x_i) - \mu]^2}, \ \alpha_n = \alpha + \frac{n}{2}. \tag{3.11}$$

For n sufficiently large, the predictive distribution will be very close to a log-normal.

4. The double Weibull distribution

Recently the Weibull distribution has received much attention in the modeling of financial returns (e.g. Mittnik and Rachev (1993)). This is due both to its theoretical capability to model the complexity of financial market data as well as its competitiveness in empirical fitting.

Applying the general location-scale transform $T(X) = \mu + \sigma \cdot Z$ with symmetric Z about zero, we are interested in the standardized double Weibull distribution with parameter $\alpha > 0$, abbreviated $SDW(\alpha)$, whose density and survival distribution are given by

$$f_Z(z) = \tfrac{1}{2}\alpha\lambda_\alpha |z|^{\alpha-1} \exp(-\lambda_\alpha |z|^\alpha), \ \lambda_\alpha = \Gamma(1+\tfrac{2}{\alpha})^{\frac{\alpha}{2}}, \tag{4.1}$$

$$S_Z(z) = \begin{cases} 1 - \tfrac{1}{2}\exp(-\lambda_\alpha |z|^\alpha), & z \le 0, \\ \tfrac{1}{2}\exp(-\lambda_\alpha |z|^\alpha), & z \ge 0, \end{cases} \tag{4.2}$$

Observe that the value of the parameter λ_α is chosen such that the variance is one. The special case $\alpha = 1, \lambda_1 = \sqrt{2}$ defines the Laplace distribution, which plays a central role in the geometric-multiplication stable scheme in Mittnik and Rachev (1993). It appears also as limiting case of the simple logarithmic modified double exponential model of financial returns in Hürlimann (1995b). The skewness of Z is clearly zero, and the kurtosis is

$$\gamma_{2,Z} = E\left[Z^4\right] = \frac{\Gamma(1+\frac{4}{\alpha})}{\Gamma(1+\frac{2}{\alpha})^2}. \tag{4.3}$$

Since $\gamma_{2,Z} \in [1,6]$ for $\alpha \ge 1$ and $\gamma_{2,Z} \in [6,\infty)$ for $\alpha \le 1$, this distribution covers the whole range of practical kurtosis values.

The logarithmic version of the above distribution is also considered. The random variable X such that $\ln(X) = \mu + \sigma \cdot Z$, with Z a $SDW(\alpha)$, defines the *logarithmic double Weibull distribution* with parameters μ, σ, α, abbreviated $\ln DW(\mu, \sigma, \alpha)$. It is a simple alternative to the log-normal model with density and survival distribution

$$f_X(x) = \frac{1}{\sigma x} f_Z\left[\frac{\ln(x) - \mu}{\sigma}\right], \ S_X(x) = S_Z\left[\frac{\ln(x) - \mu}{\sigma}\right]. \tag{4.4}$$

Though the stop-loss transform cannot be expressed in closed form, it can be evaluated using series representations.

Proposition 4.1. The stop-loss transform exists if $\alpha > 1$, or if $\alpha = 1$ and $\sigma < \lambda_1 = \sqrt{2}$, and is given by

$$\pi_X(x) = \begin{cases} \frac{1}{2} e^{\mu} \cdot \left[I\left(0; \sigma \lambda_\alpha^{-\frac{1}{\alpha}}\right) + J\left(\lambda_\alpha \left(\frac{\mu - \ln x}{\sigma}\right)^\alpha; \sigma \lambda_\alpha^{-\frac{1}{\alpha}}\right)\right] - x S_X(x), \ x \le e^{\mu}, \\ \frac{1}{2} e^{\mu} \cdot I\left(\lambda_\alpha \left(\frac{\ln x - \mu}{\sigma}\right)^\alpha; \sigma \lambda_\alpha^{-\frac{1}{\alpha}}\right) - x S_X(x), \ x \ge e^{\mu}, \end{cases} \tag{4.5}$$

where $I(x; y)$ and $J(x; y)$ are the infinite series

$$I(x;y) = \sum_{k=0}^{\infty} \Gamma\left(1 + \tfrac{k}{\alpha}\right) \cdot \left[1 - \Gamma\left(1 + \tfrac{k}{\alpha}; x\right)\right] \cdot \frac{y^k}{k!}, \tag{4.6}$$

$$J(x;y) = \sum_{k=0}^{\infty} \Gamma\left(1 + \tfrac{k}{\alpha}\right) \cdot \Gamma\left(1 + \tfrac{k}{\alpha}; x\right) \cdot (-1)^k \frac{y^k}{k!}, \tag{4.7}$$

and $\Gamma(\beta; x)$ is a gamma distribution with shape parameter β.

Proof. Consider first the case $x \ge e^{\mu}$. With the substitution $z = \left(\frac{\ln x - \mu}{\sigma}\right)^\alpha$ one obtains

$$\pi_X(x) = \tfrac{1}{2} \cdot \int_x^{\infty} \exp\left[-\lambda_\alpha \left(\tfrac{\ln x - \mu}{\sigma}\right)^\alpha\right] dx = \tfrac{1}{2} e^{\mu} \cdot \int_{\left(\frac{\ln x - \mu}{\sigma}\right)^\alpha}^{\infty} \tfrac{d}{dz}\left[e^{\sigma z^{\frac{1}{\alpha}}}\right] \cdot e^{-\lambda_\alpha z} dz.$$

With a partial integration one gets

$$\pi_X(x) = \tfrac{1}{2} e^{\mu} \int_{\lambda_\alpha \left(\frac{\ln x - \mu}{\sigma}\right)^\alpha}^{\infty} \exp\left[\sigma \lambda_\alpha^{-\frac{1}{\alpha}} u^{\frac{1}{\alpha}}\right] \cdot e^{-u} du - x S_X(x).$$

The expressions in (4.5) and (4.6) follow by noting that

$$I(x;y) = \int_x^{\infty} \exp\left[y \cdot u^{\frac{1}{\alpha}}\right] \cdot e^{-u} du = \sum_{k=0}^{\infty} \frac{y^k}{k!} \cdot \int_x^{\infty} u^{\frac{k}{\alpha}} \cdot e^{-u} du.$$

If $x \le e^{\mu}$ decompose $\pi_X(x) = \pi_X(e^{\mu}) + A(x)$ with

$$A(x) = \int_x^{e^{\mu}} S(x) dx = e^{\mu} - x - \tfrac{1}{2} \cdot \int_x^{e^{\mu}} \exp\left[-\lambda_\alpha \left(\tfrac{\mu - \ln x}{\sigma}\right)^\alpha\right] dx.$$

Proceed as above to get

$$A(x) = e^{\mu} - x - \tfrac{1}{2} e^{\mu} \cdot \int_0^{\left(\frac{\mu - \ln x}{\sigma}\right)^\alpha} \tfrac{d}{dz}\left[e^{-\sigma z^{\frac{1}{\alpha}}}\right] \cdot e^{-\lambda_\alpha z} dz$$

$$= e^{\mu} - x + x[1 - S(x)] - \tfrac{1}{2} e^{\mu} + \tfrac{1}{2} e^{\mu} \cdot \int_0^{\lambda_\alpha \left(\frac{\mu - \ln x}{\sigma}\right)} \exp\left[-\sigma \lambda_\alpha^{-\frac{1}{\alpha}} u^{\frac{1}{\alpha}}\right] \cdot e^{-u} du.$$

A series expansion of the last integral uses that

$$J(x;y)=\int_0^x \exp\left[-y\cdot u^{\frac{1}{\alpha}}\right]\cdot e^{-u}\,du=\sum_{k=0}^{\infty}(-1)^k\frac{y^k}{k!}\cdot\int_0^x u^{\frac{k}{\alpha}}\cdot e^{-u}\,du,$$

which is (4.7). Inserting further the expression $\pi_X(e^{\mu})=\frac{1}{2}e^{\mu}\cdot\left[I\left(0;\sigma\lambda_\alpha^{-\frac{1}{\alpha}}\right)-1\right]$, one obtains (4.5). ◊

To obtain expressions for the moments, first note that the mean equals (if $\alpha>1$, or if $\alpha=1$ and $\sigma<\lambda_1=\sqrt{2}$)

$$E[X]=\pi_X(0)=\tfrac{1}{2}e^{\mu}\cdot\left[I\left(0;\sigma\lambda_\alpha^{-\frac{1}{\alpha}}\right)+J\left(\infty;\sigma\lambda_\alpha^{-\frac{1}{\alpha}}\right)\right]=e^{\mu}\cdot\sum_{k=0}^{\infty}\Gamma\left(1+\tfrac{2k}{\alpha}\right)\cdot\frac{\left(\sigma\lambda_\alpha^{-\frac{1}{\alpha}}\right)^{2k}}{2k!}. \quad (4.8)$$

Since $E[X]=e^{\mu}\cdot E\left[e^{\sigma Z}\right]$, the series in (4.8) is nothing else than the moment generating function $M_Z(t)=E\left[e^{tZ}\right]$ evaluated at $t=\sigma$. From this observation one gets the higher order moments of X as

$$E\left[X^n\right]=e^{n\mu}\cdot E\left[e^{n\sigma Z}\right]=e^{n\mu}\cdot M_Z(n\sigma)=e^{n\mu}\cdot\sum_{k=0}^{\infty}\Gamma\left(1+\tfrac{2k}{\alpha}\right)\cdot\frac{\left(n\sigma\lambda_\alpha^{-\frac{1}{\alpha}}\right)^{2k}}{2k!},\ n=1,2,\ldots \quad (4.9)$$

In the attractive special case $\alpha=1$ (and $\sigma<\lambda_1=\sqrt{2}$) of the log-Laplace distribution, the relevant expressions can be given in closed form. One obtains

$$S_X(x)=\begin{cases}1-\frac{1}{2}(e^{-\mu}x)^{\frac{\sqrt{2}}{\sigma}}, & x\le e^{\mu},\\ \frac{1}{2}(e^{-\mu}x)^{-\frac{\sqrt{2}}{\sigma}}, & x\ge e^{\mu},\end{cases} \quad (4.10)$$

$$\pi_X(x)=\begin{cases}e^{\mu}\cdot\frac{2}{2-\sigma^2}-x+\frac{1}{2}\frac{\sigma}{\sqrt{2}+\sigma}(e^{-\mu}x)^{\frac{\sqrt{2}}{\sigma}+1}, & x\le e^{\mu},\\ \frac{1}{2}\frac{\sigma}{\sqrt{2}-\sigma}(e^{-\mu}x)^{-\left(\frac{\sqrt{2}}{\sigma}-1\right)}, & x\ge e^{\mu},\end{cases} \quad (4.11)$$

$$E\left[X^n\right]=e^{n\mu}\cdot\frac{2}{2-(n\sigma)^2},\quad n<\frac{\sqrt{2}}{\sigma}. \quad (4.12)$$

It is interesting to observe that the log-Laplace has Pareto tails with index $\frac{\sqrt{2}}{\sigma}$, and thus this simple special model is consistent in the tail region with Mandelbrot's Paretian hypothesis for financial returns (see Mandelbrot (1963), Fama (1963/65)). In particular, the mean excess function is linear in the tails and equals

$$e_X(x)=\frac{\pi_X(x)}{S_X(x)}=\frac{\sigma}{\sqrt{2}-\sigma}\cdot x,\quad x\ge e^{\mu}. \quad (4.13)$$

Since this function is increasing, and in accordance with extreme value theory (e.g. Embrechts et al. (1997)), the log-Laplace is thus susceptible to model long-tailed data. Concerning further properties and motivation, the interested reader is invited to have a look at Hürlimann (1995b).

5. Fitting non-life insurance data

To start with, it appears attractive to test the goodness-of-fit of the proposed models at data sets already examined in the actuarial literature. Our analysis concentrates on the theft claim size data in Hogg and Klugman (1984), Table 4.4, and on the industrial fire insurance claim statistics in Beard et al. (1984), Table 3.5.1.

5.1. Theft loss insurance data

The n = 32451 observations are grouped into m = 18 classes with boundaries, average losses and frequencies given in Table 5.1.

TABLE 5.1

Theft loss data

i	ξ_i	$m\xi_i$	λ_i
0	100	0	0
1	125	115	583
2	150	140	1368
3	156	154	280
4	175	166	1165
5	200	192	2082
6	211	206	631
7	250	232	2074
8	300	277	2285
9	350	327	1990
10	400	377	1646
11	500	452	2792
12	600	567	3271
13	850	713	4339
14	1100	972	2379
15	5100	1997	5181
16	10100	6870	286
17	25100	14354	91
18	50100	30430	8

Apart the quite good log-gamma, two parameter distributions do not seem to fit very well the present data. For example, Hogg and Klugman (1984) do not consider the log-normal, Pareto, Weibull and gamma as reasonable choices. The Benktander type I and II (see Benktander and Segerdahl (1960), Benktander (1970), Beard et al. (1984) and Embrechts et al. (1997)) defined uniquely by the mean excess functions

$$e_I(x) = \frac{\beta \cdot (1+x)}{1+2\alpha\beta \cdot \ln(1+x)}, \quad e_{II}(x) = \beta \cdot (1+x)^{1-\alpha}, \tag{5.1}$$

are similar unreasonable choices with high values of the goodness-of-fit statistics (see Table 5.3). However, note that the Benktander distributions were introduced to describe the excess losses over some higher threshold (for use in reinsurance) rather than the entire range of losses (for use in direct insurance), which is our main concern in the present study.

Among the three parameter distributions they consider, Hogg and Klugman (1984) found that a Burr provides the best fit with respect to the limited expected value criterion. The results of our parameter estimation are given in Table 5.2 and our goodness-of-fit analysis is summarized in Table 5.3.

Our implementation of the scoring method did not yield in a straightforward way the maximum likelihood estimators $\hat{\mu}$, $\hat{c}$, $\hat{a}$ for the $\ln NIG(\mu, c, a)$. Instead, and for comparisons, several possible fits were made. Each of the four lnNIG minimizes by varying a approximately one of the goodness-of-fit statistics. More precisely, overall rank 2 corresponds to a minimum χ^2, rank 3 to a minimum K, rank 4 to a maximum $\ln L$, and rank 5 to a minimum LE. Since maximum likelihood estimation for the $\ln DW(\mu, c, a)$ using the scoring method has been successful, such a distinction appears superfluous. The overall ranks in Table 5.2 match those in Table 5.3.

TABLE 5.2

PARAMETER VALUES OF THEFT LOSS DISTRIBUTIONS

Overall rank	**Distribution**	**Parameter values**
1	$\ln DW(\mu, \sigma, a)$	$\hat{a} = 1.270795$, $\hat{\mu} = 6.013325$, $\hat{\sigma} = 1.020931$
2	$\ln NIG(\mu, c, a)$	$a = 5.3$, $\hat{\mu} = 6.044392$, $\hat{c} = 2.966822$
3	$\ln NIG(\mu, c, a)$	$a = 5.1$, $\hat{\mu} = 6.043941$, $\hat{c} = 2.903233$
4	$\ln NIG(\mu, c, a)$	$a = 5.6$, $\hat{\mu} = 6.045009$, $\hat{c} = 3.059974$
5	$\ln NIG(\mu, c, a)$	$a = 4.1$, $\hat{\mu} = 6.041057$, $\hat{c} = 2.564565$
6	Burr (a, β, τ)	$\hat{\tau} = 1.66932$, $\hat{a} = 1.09626$, $\hat{\beta} = 2.6691.02903$
7	BenktanderI(a, β)	$\hat{a} = 0.00339$, $\hat{\beta} = 146.813$
8	BenktanderII(a, β)	$\hat{a} = 0.8745$, $\hat{\beta} = 364.6117$

TABLE 5.3

GOODNESS-OF-FIT OF THEFT LOSS DISTRIBUTIONS

Overall rank	**Distribution**	**$-\ln L$**	**χ^2**	**K**	**$10^3 \cdot LE$**	***ME***
1	lnDW	83551	914	0.4702	8.94	1.85
2	lnNIG	83648.60	1142.79	0.4362	3.00	–
3	lnNIG	83649.20	1143.09	0.4359	2.57	–
4	lnNIG	83648.30	1143.33	0.4376	3.67	–
5	lnNIG	83660.33	1156.86	0.4480	1.29	–
6	Burr	83672	1186	0.4645	1.04	0.46
7	BenktanderI	92902	18762	67.70	1193	1006
8	BenktanderII	83153	$12.8 \cdot 10^6$	7304.57	16.5	2.78

Though the fitted Burr has the lowest LE- and ME-statistics, it takes the worst overall rank among the three parameter distributions. It is also beaten by a non-optimal lnNIG with a similar LE-value. The preferred distribution is a lnDW. As the χ^2-values are rather high, a formal chi-square test, which would validate one or several of the models, is not undertaken. However, as demonstrated in Section 6, such a validation is sometimes possible.

5.2. Industrial fire loss data

The $n = 8324$ observations are grouped into $m = 29$ classes with boundaries and frequencies given in Table 5.4. The obtained parameter values and goodness-of-fit of four distributions are summarized in Table 5.5 and Table 5.6. In this situation maximum likelihood estimation using the scoring method has been successful, and a further distinction as in Section 5.1 appears akward.

The fitted Burr, whose mean does not exist, is beaten by three distributions and has here the highest LE-value. A simple two parameter log-normal fits better than the Burr. The log-normal is beaten by both the lnDW and the lnNIG. While the lnNIG has the lowest LE-value, the lnDW is the preferred distribution. No formal test is undertaken.

TABLE 5.4

INDUSTRIAL FIRE LOSS DATA

i	ξ_i	λ_i	i	ξ_i	λ_i
0	0	0	15	6310	323
1	10	283	16	10000	179
2	16	280	17	15849	173
3	25	157	18	25119	112
4	40	464	19	39811	94
5	63	710	20	63096	57
6	100	781	21	100000	39
7	158	530	22	158489	22
8	251	446	23	251189	17
9	398	491	24	398107	12
10	631	673	25	630957	5
11	1000	779	26	1000000	5
12	1585	741	27	1584890	3
13	2512	520	28	2511890	1
14	3981	425	29	6309570	2

TABLE 5.5

PARAMETER VALUES OF FIRE LOSS DISTRIBUTIONS

Overall rank	Distribution	Parameter values		
1	$\ln DW(\mu, \sigma, a)$	$\hat{a} = 1.41561$,	$\hat{\mu} = 5.79645$,	$\hat{\sigma} = 2.16935$
2	$\ln NIG(\mu, c, a)$	$\hat{a} = 25.7746$,	$\hat{\mu} = 5.89543$,	$\hat{c} = 15.23085$
3	$\ln N(\mu, \sigma)$	$\hat{\mu} = 5.90396$,	$\hat{\sigma} = 2.15982$	
4	$\mathrm{Burr}(\alpha, \beta, \tau)$	$\hat{\tau} = 0.80607$,	$\hat{\alpha} = 0.98114$,	$\hat{\beta} = 110.35718$

TABLE 5.6

GOODNESS-OF-FIT OF FIRE LOSS DISTRIBUTIONS

Overall rank	Distribution	$-\ln L$	χ^2	K	*LE*	*ME*
1	lnDW	24128	442	2.46	3.83	3.83
2	lnNIG	24213	637	1.86	2.53	–
3	lnN	24216	663	2.33	3.60	7.63
4	Burr	24281	758	1.71	9.09	–

6. FITTING FINANCIAL MARKET DATA

The distribution of the daily cumulative returns on a stock market index has been the subject of many past and current investigations. It is thus of great importance to look at the overall goodness-of-fit of the proposed models when compared with "good" competitors like the log-normal (justified by the model of Black and Scholes (1973)) and the symmetric α-stable distribution (justified by the work of Mandelbrot (1963), Fama (1965), and Peters (1994)). Our analysis is based on the SMI (Swiss Market Index) daily cumulative returns between September 29, 1998 and September 24, 1999.

The $n = 250$ observations are grouped into $m = 26$ classes with boundaries and frequencies given in Table 6.1. The parameter estimation is provided in Table 6.2 and the goodness-of-fit in Table 6.3. For the sake of comparisons, we distinguish between two NIG fits. The NIG minimizes approximately χ^2 by varying a and maximum likelihood estimation of μ, c while the NIG_2 uses maximum likelihood estimation of μ, c, a.

The fitted log-normal, with high χ^2- and K-values, seems unreasonable at first sight. The relative low LE- and ME-values in this example, which are quite smaller than the corresponding values of the two overall best fitted distributions, illustrate the apparent irrelevance of the LE- and ME-criteria (see however the comments to Figure 6.1). Also, the NIG_2, whose three parameters have been estimated with the maximum likelihood method, has high χ^2- and K-values and takes only overall rank 5. By the way, it has the lowest LE-value and a quite small ME-value. The other four distributions seem to fit quite well.

It is remarkable that the two-parameter log-Laplace is not significantly beaten by the lnDW. The preferred SMI distributions are a NIG followed by a symmetric α-stable distribution, abbreviated *SaS*, whose characterisitcs are summarized in the Appendix. The ranked 3 lnDW and the log-Laplace have smaller LE- and ME-values.

A quick look at the graphs of the empirical and fitted mean excess functions in Figure 6.1 is very instructive (similar observations hold for the non-life insurance data sets, an analysis which can be left to the reader). The behavior of the empirical graph is quite erratic in the right tail. The simplest fit for this is anticipated by a parabola or more generally a convex curve. In contrast to this, in non-life insurance, an increasing and concave curve fitting has been proposed, at least in the right tail (see Benktander and Segerdahl (1960), and Benktander (1970) on this point). The two best fitting distributions distinguish themselves from the others by a considerable slope in the right tail (in accordance with extreme value theory). This is the reason for the high ME- and LE-values of these fitted distributions. A suggestion for future work might be the definition of more adequate *weighted* LE- and ME-statistics, which take this phenomenon into account. This could perhaps also allow these statistics to enter into an extended goodness-of-fit test.

TABLE 6.1

SMI DAILY CUMULATIVE RETURNS

i	ξ_i	λ_i	i	ξ_i	λ_i
0	0.950	0	14	1.020	22
1	0.955	1	15	1.025	8
2	0.960	4	16	1.030	3
3	0.965	0	17	1.035	2
4	0.970	1	18	1.040	1
5	0.975	1	19	1.045	0
6	0.980	11	20	1.050	0
7	0.985	14	21	1.055	0
8	0.990	15	22	1.060	1
9	0.995	43	23	1.065	0
10	1.000	31	24	1.070	0
11	1.005	36	25	1.075	0
12	1.010	25	26	1.080	1
13	1.015	30			

TABLE 6.2

PARAMETER VALUES OF SMI DISTRIBUTIONS

Overall rank	Distribution	Parameter values
1	$NIG_1(\mu, c, a)$	$a = 1.875, \quad \hat{\mu} = 1.00063587, \quad \hat{c} = 0.022579$
2	$SaS(\mu, c, a)$	$a = 1.8, \quad \hat{\mu} = 1.0006, \quad \hat{c} = 0.01012$
3	$\ln DW(\mu, \sigma, a)$	$a = 1.005, \quad \hat{\mu} = 0.00038779, \quad \hat{\sigma} = 0.016489$
4	$\ln Laplace(\mu, \sigma)$	$\hat{\mu} = 0.00038277, \quad \hat{\sigma} = 0.016547$
5	$NIG_2(\mu, c, a)$	$\hat{a} = 6.71936, \quad \hat{\mu} = 1.00063727, \quad \hat{c} = 0.049471$
6	$\ln N(\mu, \sigma)$	$\hat{\mu} = 0.00058706, \quad \hat{\sigma} = 0.0151181$

TABLE 6.3

GOODNESS-OF-FIT OF SMI DISTRIBUTIONS

Overall rank	Distribution	$-\ln L$	χ^2	K	$10^7 \cdot LE$	ME
1	NIG_1	623.47	46.51	2.30	18.00	24.60
2	*Sas*	623.14	49.53	2.73	12.77	303
3	lnDW	629.40	58.25	3.35	5.85	3.90
4	lnLaplace	629.64	58.26	3.36	6.11	3.97
5	NIG_2	625.64	241.47	36.51	2.86	4.42
6	lnN	633.07	4490	485	5.96	5.87

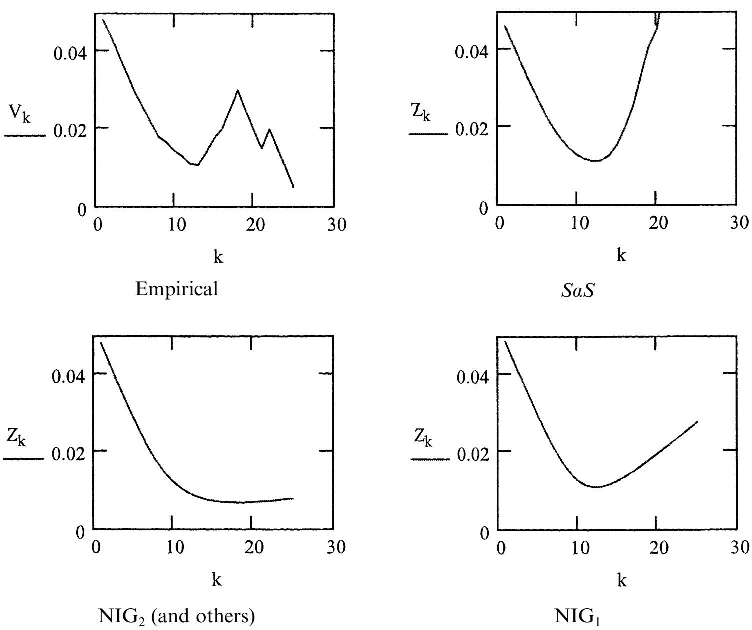

FIGURE 6.1: graphs of mean excess functions

An ultimate validation and selection among the above models must necessarily be based on a formal statistical test and consider other alternative decision rules for selection. To perform a correct formal chi-square test, the raw data in Table 6.1 must be grouped in a different way. According to Moore (1978/86) a number of rules are to be fulfilled (e.g. Klugman et al. (1998), p. 121). Recommended is an expected frequency of at least 1% in each class and a 5% expected frequency in 80% of the classes. In view of this, the raw data is regrouped as in Table 6.4.

TABLE 6.4

SMI DAILY CUMULATIVE RETURNS FOR CHI-SQUARE TEST

i	ξ_i	λ_i
0	0.95	0
1	0.97	6
2	0.985	26
3	1.00	89
4	1.015	91
5	1.03	33
6	1.08	5

Based on the parameter values in Table 6.2, the up-dated goodness-of-fit statistics, together with the p-value of the test, are found in Table 6.5, which order the distributions according to the new overall rank.

TABLE 6.5

GOODNESS-OF-FIT UNDER CHI-SQUARE TEST

Overall rank	Distribution	$-\ln L$	χ^2	p-value	K	$10^7 \cdot LE$	ME
1	NIG_2	351.87	0.75	0.69	0.014	7.76075	0.9382
2	*SaS*	352.85	1.34	0.51	0.056	5.82119	0.5710
3	NIG_1	353.28	2.63	0.27	0.055	11.14234	0.6414
4	lnN	353.11	3.08	0.38	0.061	2.67507	0.9077
5	lnDW	354.33	4.85	0.09	0.082	5.44844	0.6345
6	lnLaplace	354.42	5.00	0.17	0.084	5.45984	0.6275

With critical values of 5.99 (by 2 degrees of freedom for 3 parameters) and 7.82 (by 3 degrees of freedom for 2 parameters) for a 5% significance level, it is remarkable that all models are validated through this test. In view of the high χ^2- and K-values in Table 6.3, this could not be expected a priori for the NIG_2 and the lnN. This shows once more that informal decision rules are to be applied very carefully. The dramatic change in the proposed overall ranking coincides exactly with the brute χ^2-ranking and almost with the negative likelihood and K-ranking. The LE- and ME-values behave still quite erratically.

Since there are different number of parameters, the p-value ranking differs. This is a reason for considering further alternative selection rules.

If parsimony is a concern, the best three parameter NIG_2 needs to be compared with the log-normal and log-Laplace. For this it is usual to apply a likelihood ratio test (though this is here only an informal decision rule). The test statistics are $2 \cdot (353.11 - 351.87) = 1.24$ (lnN versus NIG_2) and $2 \cdot (354.42 - 351.87) = 2.55$ (lnLaplace versus NIG_2). With one degree of freedom, the critical value is 3.84, and the log-normal and log-Laplace are selected first. A perhaps more appropriate alternative to this informal hypothesis test is a penalized likelihood scoring method, called Schwartz Bayesian Criterion (SBC) and introduced by Schwartz (1978) (see Klugman et al. (1998)). To the negative likelihood one adds the penalty $p \cdot [\ln(n) - \ln(2\pi)]$, where p is the number of estimated parameters and n is the sample size, to obtain the SBC-score, which decides upon ranking. The result of this SBC selection is reported in Table 6.6.

TABLE 6.6

SBC RANKING OF SMI DISTRIBUTIONS

SBC rank	Distribution	-lnL	penalty	SBC score
1	lnN	353.11	7.37	360.48
2	lnLaplace	354.42	7.37	361.79
3	NIG_2	351.87	11.05	362.92
4	*SaS*	352.85	11.05	363.90
5	NIG_1	353.28	11.05	364.73
6	lnDW	354.33	11.05	365.38

To the knowledge of the author, the above study should be a unique first one, which justifies statistically several different and important models of financial returns motivated through financial economic and other principles. It places Black-Scholes (1973) model at the top rank, and justifies also the simple limiting log-Laplace model in Hürlimann (1995b). Furthermore, it does not reject other good alternative choices like the normal inverted gamma mixture by Praetz (1972) (used by J.P. Morgan Stanley), the prominent symmetric α-stable distribution by Mandelbrot (1963) and Fama (1965), and the more recent double Weibull as geometric-multiplication stable scheme in Mittnik and Rachev (1993).

7. ON THE PREDICTION OF ONE-YEAR INDEX RETURNS FROM DAILY INDEX RETURNS

To conclude the present study with a practical illustration of the results in Section 6, it is interesting to compare the actual SMI index of 6966 at the end of the observation period with the SMI index resulting from a fitted distribution under a *strict white noise assumption* (independent and identically distributed daily returns). The obtained 250 days return for the fitted lnN, lnLaplace, NIG_2, *SaS*, NIG_1 and lnDW are respectively 19.19%, 13.94%, 17.26%, 16.18%, 17.22% and 14.05%. The average between the two extremes is 16.565%, which is quite close to the observed $100 \cdot \left(\frac{6966}{6020} - 1\right) = 15.71\%$. The range of variation for the SMI index at the end of the period is [7017 ± 158], where the midpoint is quite close to the actual index of 6966. The closest value to the actual index is 6994 for the *SaS*. Whether these extremely good fits are mere coincidence or of a deeper nature requires further investigations. Used as naive prediction value, if the index performs similarly in the next period, then the SMI index at the end of September 2000 should stay between 7945 and 8166. This corresponds approximately to the expected SMI index of 8200 at the end of year 2000 as predicted by a model of the Credit Suisse First Boston (see Tages-Anzeiger (1999)).

The above naive calculation is done under the strict white noise assumption, which is easy to test and refute from a pure statistical point of view (e.g. Taylor (1992), p. 19). However, it yields an acceptable value of the one-year return from a pure investment point of view. Must the independence hypothesis be rejected or can it be used for the present purpose? This well-known dilemma has been noted and studied in detail by Fama (1965), which states:

> "Dependence that is important from the trader's point of view need not be important from a statistical point of view, and conversely dependence which is important for statistical purposes need not be important for investment purposes."

Recall that Fama's tests did not reveal any evidence of important dependence from either an investment or a statistical point of view.

There exist some more formal mathematical calculations, which can justify the prediction of a one-year index return based on the distribution of the daily index returns. Under the made strict white noise assumption and for the distributions of Table 6.6 (except the *SaS*, for which more complex calculations are required), we have computed the four main characteristics of a distribution, namely the mean, standard deviation, skewness and kurtosis. For the $NIG(\mu, c, a)$ distributions, we have additionally calculated these characteristics for the Bayesian prediction models $NIG(\mu, c_n, a_n)$ with up-dated parameters $c_n = \sqrt{c^2 + \sum_{i=1}^{n} [x_1 - \mu]^2}$, $a_n = a + \frac{n}{2}$, as well as for the normal approximation to this prediction model as $a \to \infty$, where these models have been discussed in Section 3. The obtained results are summarized in Table 7.2. The required formulas for the first four moments $m_k, k = 1,2,3,4$, are straightforward and listed below in Table 7.1 for the convenience of the reader. The one-year return corresponds here to $T = 250$ days. The mean of the one-year return is then $\mu_T = m_1$, the standard deviation is $\sigma_T = \sqrt{m_2 - m_1^2}$, while the skewness γ_T and kurtosis $\gamma_{2,T}$ are calculated using the formulas

$$\gamma_T = \frac{m_3 - 3m_1 m_2 + 2m_1^3}{\sigma_T^3} \tag{7.1}$$

$$\gamma_{2,T} = \frac{m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4}{\sigma_T^4} \tag{7.2}$$

TABLE 7.1

MOMENT FORMULAS FOR ONE-YEAR INDEX RETURN PREDICTION MODELS

$\underline{\ln N(\mu,\sigma)}$

$$m_k = \exp\left[k\mu T + \tfrac{1}{2}(k\sigma)^2 T\right],\quad k=1,2,3,4$$

$\underline{\ln Laplace(\mu,\sigma)}$

$$m_k = \frac{\exp(k\mu T)}{1-\frac{1}{2}(k\sigma)^2 T},\quad k=1,2,3,4$$

$\underline{\ln DW(\mu,\sigma,a)}$

$$m_k = \exp(k\mu T)\cdot \sum_{j=0}^{\infty} \frac{\Gamma\left(1+\frac{2j}{a}\right)}{(2j)!}\cdot \left[(k\sigma)^2 \frac{T}{\Gamma\left(1+\frac{2}{a}\right)}\right]^2,\quad k=1,2,3,4 \text{ (this is formula (4.11))}$$

$\underline{NIG(\mu,c,a)}$

$$m_1=\mu^T, m_2=\left(\mu^2+\tfrac{1}{2}\tfrac{c^2}{a-1}\right)^T,\ m_3=\left(\mu^3+\tfrac{3}{2}\tfrac{\mu c^2}{a-1}\right)^T,\ m_4=\left(\mu^4+3\tfrac{\mu^2 c^2}{a-1}+\tfrac{3}{4}\tfrac{c^4}{(a-1)(a-2)}\right)^T,\quad a>2$$

$\underline{BNIG(\mu,c_n,a_n)}$ (Bayesian NIG prediction model)

The same formulas as for the NIG hold with c, a replaced by the up-dated parameters c_n, a_n calculated with the $n = 250$ daily observations.

$\underline{NNIG(\mu,\sigma_a)}$ (normal approximation to Bayesian NIG prediction model)

$$\sigma_a = \frac{c_n}{\sqrt{2(a_n-1)}}$$

$$m_1=\mu^T,\ m_2=(\mu^2+\sigma_a^2)^T,\ m_3=(\mu^3+3\mu\sigma_a^2)^T,\ m_4=(\mu^4+6\mu^2\sigma_a^2+3\sigma_a^4)^T$$

TABLE 7.2

ONE-YEAR RETURN PREDICTION FROM DAILY RETURN DISTRIBUTIONS

Distribution	μ_T	σ_T	γ_T	$\gamma_{2,T}$
LnN	19.19%	29.03%	0.75	4.003
LnLaplace	13.94%	32.36%	2.61	30.717
LnDW	14.05%	32.21%	2.54	28.686
NIG_1	17.22%	32.20%	0.84	4.144
$BNIG_1$	17.22%	28.70%	0.75	4.004
$NNIG_1$	17.22%	28.70%	0.75	4.004
NIG_2	17.27%	27.47%	0.71	3.918
$BNIG_2$	17.27%	28.63%	0.74	3.998
$NNIG_2$	17.27%	28.63%	0.74	3.998

The comparison of the figures in Table 7.2 are quite instructive. The Bayesian NIG models and their normal approximations have skewness and kurtosis parameters very close to the best SBC ranked log-normal model in Section 6, and these values are also closest to empirical values obtained from long-term one-year returns. The skewness and kurtosis parameters are overestimated by the log-Laplace and log-double Weibull models. From this perspective, only the log-normal and normal inverted gamma mixtures are selected for practical purposes. For a cautious prediction, the Bayesian normal inverted gamma mixture and its normal approximation should be the preferred models for prediction.

Appendix: comparative distributions

The formulas for the scoring method and the stop-loss transforms for the evaluation of the LE- and ME-statistics are listed.

Normal inverted gamma mixture

$S_X(x) = S_Z\left[\frac{x-\mu}{c}\right]$, with defined in (3.4)

$$\frac{\partial}{\partial\mu} S_X(x) = f_X(x), \quad \frac{\partial}{\partial c} S_X(x) = \left[\frac{x-\mu}{c}\right] \cdot f_X(x)$$

$$\frac{\partial}{\partial a} S_X(x) = \begin{cases} \left[\psi\left(a+\frac{1}{2}\right) - \psi(a)\right] \cdot S_X(x) + \frac{1}{4}(2a-1) \cdot \beta\left(a-1, \frac{1}{2}; \frac{c^2}{c^2+(x-\mu)^2}\right), & x \ge \mu, \\ -\left[\psi\left(a+\frac{1}{2}\right) - \psi(a)\right] \cdot F_X(x) - \frac{1}{4}(2a-1) \cdot \beta\left(a-1, \frac{1}{2}; \frac{c^2}{c^2+(x-\mu)^2}\right), & x \le \mu, \end{cases}$$

where $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$ is the digamma function or psi function.

$$\pi_X(x) = \frac{c^2+(x-\mu)^2}{2a-1} \cdot f_X(x) - (x-\mu) \cdot S_X(x)$$

Double Weibull distribution

$S_X(x) = S_Z\left[\frac{x-\mu}{c}\right]$, with $S_Z(z)$ defined in (4.2)

$$\frac{\partial}{\partial\mu} S_X(x) = f_X(x), \quad \frac{\partial}{\partial\sigma} S_X(x) = \left[\frac{x-\mu}{\sigma}\right] \cdot f_X(x)$$

$$\frac{\partial}{\partial a} S_X(x) = -\operatorname{sgn}\left[\frac{x-\mu}{\sigma}\right] \cdot \frac{1}{2}\left[\lambda_a' + \lambda_a \ln\left|\frac{x-\mu}{\sigma}\right|\right] \cdot \left|\frac{x-\mu}{\sigma}\right|^a \cdot \exp\left[-\lambda_a\left|\frac{x-\mu}{\sigma}\right|^a\right],$$

$$\lambda_a' = -\frac{1}{a}\Gamma(1+\tfrac{2}{a})^{\frac{a}{2}} \cdot \psi(1+\tfrac{2}{a})$$

$$\pi_X(x) = \sigma \cdot \pi_Z(\tfrac{x-\mu}{\sigma}), \text{ with } \pi_Z(d) = \tfrac{1}{2}\lambda_a^{-\frac{1}{a}}\Gamma(1+\tfrac{1}{a}) \cdot \left[1 - \Gamma\left(1+\tfrac{1}{a}; \lambda_a|d|^a\right)\right] - d \cdot S_Z(d),$$

where $\Gamma(\beta; x)$ is a gamma distribution with shape parameter β.

BenktanderI

$$S(x) = [1 + 2\alpha\beta \ln(1+x)] \cdot \exp\left\{-\left(\tfrac{\beta+1}{\beta}\right) \ln(1+x) - \alpha \cdot [\ln(1+x)]^2\right\}$$

$$\frac{\partial S(x)}{\partial \alpha} = \frac{2\beta \cdot \ln(1+x)}{1+2\alpha\beta \ln(1+x)} \cdot S(x) - [\ln(1+x)]^2 \cdot S(x)$$

$$\frac{\partial S(x)}{\partial \beta} = \frac{2\alpha \cdot \ln(1+x)}{1+2\alpha\beta \ln(1+x)} \cdot S(x) + \frac{1}{\beta^2} \cdot \ln(1+x) \cdot S(x)$$

$$\pi(x) = S(x) \cdot e(x),\ e(x) = \frac{\beta \cdot (1+x)}{1+2\alpha\beta \cdot \ln(1+x)}$$

BenktanderII

$$S(x) = (1+x)^{\alpha-1} \cdot \exp\left\{-\frac{1}{\alpha\beta}\left[(1+x)^{\alpha}-1\right]\right\}$$

$$\frac{\partial S(x)}{\partial \alpha} = \left[\frac{1}{\alpha^2 \beta}\left[(1+x)^{\alpha}-1\right] - \frac{1}{\alpha\beta} \ln(1+x)(1+x)^{\alpha} + \ln(1+x)\right] \cdot S(x)$$

$$\frac{\partial S(x)}{\partial \beta} = \frac{1}{\alpha\beta^2}\left[(1+x)^{\alpha}-1\right] \cdot S(x)$$

$$\pi(x) = S(x) \cdot e(x),\ e(x) = \beta \cdot (1+x)^{1-\alpha}$$

Burr

$$S(x) = \left[\frac{\beta}{\beta + x^{\tau}}\right]^{\alpha}$$

$$\frac{\partial S(x)}{\partial \alpha} = -\ln\left[\frac{\beta}{\beta + x^{\tau}}\right] \cdot S(x),\ \frac{\partial S(x)}{\partial \beta} = -\frac{\alpha x^{\tau}}{\beta^2} \cdot S(x)^{\frac{\alpha+1}{\alpha}},\ \frac{\partial S(x)}{\partial \tau} = -\frac{\alpha x^{\tau}}{\beta} \cdot \ln(x) \cdot S(x)^{\frac{\alpha+1}{\alpha}}$$

$$\pi(x) = \mu \cdot \left|1 - \beta\left(1+c, \alpha-c; \tfrac{x^{\tau}}{\beta+x^{\tau}}\right)\right| - x \cdot S(x), \text{ with the mean}$$

$$\mu = \frac{\beta^c \, \Gamma(\alpha-c) \cdot \Gamma(1+c)}{\Gamma(\alpha)},\ \alpha > c = \frac{1}{\tau}$$

$$LE(x) = \xi_0 + \frac{\beta^c}{S(\xi_0) \cdot \tau} \cdot \int_{(1+\beta^{-1} x^{\tau})^{-1}}^{(1+\beta^{-1}\xi_0^{\tau})^{-1}} y^{\alpha-c-1}(1-y)^{c-1} dy,\ \alpha \le c \text{ (case of infinite mean)}$$

Log-normal

$$S(x) = 1 - N\left(\frac{\ln x - \mu}{\sigma}\right)$$

$$\frac{\partial S(x)}{\partial \mu} = \frac{1}{\sigma} \cdot \Phi\left[\frac{\ln x - \mu}{\sigma}\right], \quad \frac{\partial S(x)}{\partial \sigma} = \frac{\ln x - \mu}{\sigma^2} \cdot \Phi\left[\frac{\ln x - \mu}{\sigma}\right]$$

$$\pi(x) = \exp(\mu + \tfrac{1}{2}\sigma^2) \cdot \left[1 - N\left(\frac{\ln x - \mu}{\sigma}\right) - \sigma\right] - x \cdot S(x)$$

Symmetric α-stable distribution

Explicit expressions exist only in the special cases $\alpha = 1$ (Cauchy) and $\alpha = 2$ (normal). Bergström (1952) developed series expansions that Fama and Roll (1968/71) and other authors applied in case $\alpha > 1$. The density and distribution of the normalized case $\mu = 0$, $c = 1$ is first stated.

$$f_Z(z) = \begin{cases} \dfrac{1}{\pi\alpha} \cdot \displaystyle\sum_{k=0}^{\infty} (-1)^k \frac{\Gamma\left(\frac{2k+1}{\alpha}\right)}{(2k)!} z^{2k}, & |z| \le 5\alpha - 4, \\ -\dfrac{1}{\pi} \cdot \displaystyle\sum_{k=0}^{n} (-1)^k \frac{\Gamma(\alpha k + 1)}{k! z^{\alpha k + 1}} \sin\left[\frac{\alpha k \pi}{2}\right] + R(z), & |z| > 5\alpha - 4. \end{cases}$$

The remainder satisfies $|R(z)| < C \cdot z^{-\alpha(n+1)-1}$, C a constant, and becomes smaller than the previous term in the summation as z gets larger.

$$F_Z(z) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{\pi\alpha} \cdot \displaystyle\sum_{k=0}^{\infty} (-1)^{k-1} \frac{\Gamma\left(\frac{2k-1}{\alpha}\right)}{(2k-1)!} z^{2k-1}, & |z| \le 5\alpha - 4, \\ 1 + \dfrac{1}{\pi} \cdot \displaystyle\sum_{k=0}^{n} (-1)^k \frac{\Gamma(\alpha k + 1)}{k! z^{\alpha \kappa}} \sin\left[\frac{\alpha k \pi}{2}\right] - \int_z^{\infty} R(t)\,dt, & |z| > 5\alpha - 4. \end{cases}$$

References

Beard, R.E., Pentikäinen, T. and Pesonen E. (1984) *Risk Theory*. Chapman and Hall.

Benktander, G. (1970) Schadenverteilung nach Grösse in der Nicht-Lebensversicherung. *Bulletin of the Swiss Association of Actuaries*, 263-84.

Benktander, G. and Segerdahl C.-O. (1960) On the analytical representation of claim distribution with special reference to excess of loss reinsurance. *Transactions of the International Congress of Actuaries*, 626-46.

Bergström, H. (1952) On some expansions of stable distribution functions. *Arkiv för Matematik* **2**, 375-78.

Black, F. and Scholes M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637-59.

Blattberg, R.C. and Gonedes N.J. (1974) A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business* **47**, 244-80.

Carriere, J. (1992) Limited expected value comparison tests. *Statistics and Probability Letters* **15**, 321-27.

Cramér, H. (1946) *Mathematical Methods of Statistics*. Princeton University Press.

Efron, B. (1982) Transformation theory: how normal is a family of distributions? (The 1981 Wald Memorial Lecture). *The Annals of Statistics* **10**, 323-339.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH TH. (1997) Modelling Extremal Events for Insurance and Finance. *Applications of Mathematics – Stochastic Modelling and Applied Probability*, vol. 33. Springer.

FAMA, E. (1963) Mandelbrot and the stable Paretian hypothesis. *J. of Business* **36**, 420-29.

FAMA, E. (1965) The behavior of stock market prices. *Journal of Business* **38**, 34-105.

FAMA, E. and ROLL R. (1968) Some properties of symmetric stable distributions. *Journal of the American Statistical Association* **63**, 817-36.

FAMA, E. and ROLL R. (1971) Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association* **66**, 331-38.

FISZ, M. (1973) *Wahrscheinlichkeitsrechnung und Mathematische Statistik*. VEB Deutscher Verlag der Wissenschaften, Berlin.

GERBER, H.-U. (1979) *An Introduction to Mathematical Risk Theory*. Hübner Foundation. University of Pennsylvania.

HEILMANN, W.-R. (1989) Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics* **8**, 77-95.

HOGG, R. and KLUGMAN S. (1984) *Loss Distributions*. John Wiley. New York.

HÜRLIMANN, W. (1993) Solvabilité et réassurance. *Bulletin of the Swiss Association of Actuaries*, 229-49.

HÜRLIMANN, W. (1995a) Predictive stop-loss premiums and Student's t-distribution. *Insurance: Mathematics and Economics* **16**, 151-59.

HÜRLIMANN, W. (1995b) *Is there a rational evidence for an infinite variance asset pricing model?* Proceedings of the 5th International AFIR Colloquium.

HÜRLIMANN, W. (1997) Fonctions extrémales et gain financier. *Elemente der Mathematik* **52**, 152-68.

HÜRLIMANN, W. (1998a) On the characterization of maximum likelihood estimators for location-scale families. *Communications in Statistics – Theory and Methods* **27(2)**, 495-508.

HÜRLIMANN, W. (1998b) *Extremal Moment Methods and Stochastic Orders*. Application in Actuarial Science. Monograph manuscript (available from the author).

HÜRLIMANN, W. (2000) Higher-degree stop-loss transforms and stochastic orders (I) Theory. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **XXIV(3)**, 449-63.

JOHNSON, N.L., KOTZ, A. and BALAKRISHNAN N. (1995) *Continuous Univariate Distributions*. (2nd ed.). John Wiley, New York.

KLUGMAN, S., PANJER, H. and WILLMOT G. (1998) *Loss Models. From Data to Decisions*. John Wiley, New York.

KON, S.J. (1984) Models of stock returns – a comparison. *Journal of Finance* **39**, 147-65.

MANDELBROT, B. (1963) The variation of certain speculative prices. *Journal of Business* **36**, 394-419.

MITTNIK, S. and RACHEV S.T. (1993) Modeling asset returns with alternative stable distributions. *Econometric Reviews* **12(3)**, 261-330.

MOORE, D. (1978) Chi-square tests. In Hogg, R. (Ed.). Studies in Statistics, vol. 19. *Mathematical Association of America*, 453-63.

MOORE, D. (1986) *Tests of chi-squared type*. In D'Agostino, R. and M. Stephens (Eds.). Goodness-of-Fit Techniques. Marcel Dekker, New York, 63-95.

MÜLLER, A. (1996) Ordering of risks: a comparative study via stop-loss transforms. *Insurance: Mathematics and Economics* **17**, 215-22.

PETERS, E. (1994) *Fractal Market Analysis*. Wiley Finance Editions.

PRAETZ, P.D. (1972) The distribution of share price changes. *Journal of Business* **45**, 49-65.

SCHWARTZ, G. (1978) Estimating the dimension of a model. *Annals of Statistics* **6**, 461-64.

TAGES-ANZEIGER (1999) *CSFB: Günstiger Ausblick auf 2000*. October 6, 1999, p. 43.

TAYLOR, S.J. (1992) *Modeling Financial Time Series* (3rd reprint). John Wiley.

WERNER HÜRLIMANN
Value and Risk Management
Winterthur Life and Pensions
Postfach 300
CH-8401 Winterthur

INTRODUCTION TO DYNAMIC FINANCIAL ANALYSIS

BY

ROGER KAUFMANN, ANDREAS GADMER AND RALF KLETT

ABSTRACT

In the last few years we have witnessed growing interest in Dynamic Financial Analysis (DFA) in the nonlife insurance industry. DFA combines many economic and mathematical concepts and methods. It is almost impossible to identify and describe a unique DFA methodology. There are some DFA software products for nonlife companies available in the market, each of them relying on its own approach to DFA. Our goal is to give an introduction into this field by presenting a model framework comprising those components many DFA models have in common. By explicit reference to mathematical language we introduce an up-and-running model that can easily be implemented and adjusted to individual needs. An application of this model is presented as well.

KEYWORDS AND PHRASES

Nonlife insurance, Dynamic Financial Analysis, Asset/Liability Management, stochastic simulation, business strategy, efficient frontier, solvency testing, interest rate models, claims, reinsurance, underwriting cycles, payment patterns.

1. WHAT IS DFA

1.1. Background

In the last few years, nonlife insurance corporations in the US, Canada and also in Europe have experienced, among other things, pricing cycles accompanied by volatile insurance profits and increasing catastrophe losses contrasted by well performing capital markets, which gave rise to higher realized capital gains. These developments impacted shareholder value as well as the solvency position of many nonlife companies. One of the key strategic objectives of a

[1] The article is partially based on a diploma thesis written in cooperation with Zurich Financial Services. Further research of the first author was supported by Credit Suisse Group, Swiss Re and UBS AG through RiskLab, Switzerland.

joint stock company is to satisfy its owners by increasing shareholder value over time. In order to achieve this goal it is necessary to get an understanding of the economic factors driving shareholder value and the cost of capital. This does not only include identifying the factors but investigating their random nature and interrelations to be able to quantify earnings volatility. Once this has been done various business strategies can be tested in respect of meeting company objectives.

There are two primary techniques in use today to analyze financial effects of different entrepreneurial strategies for nonlife insurance companies over a specific time horizon. The first one – *scenario testing* – projects business results under selected deterministic scenarios into the future. Results based on such a scenario are valid only for this specific scenario. Therefore, results obtained by scenario testing are useful only insofar as the scenario was correct. Risks associated with a specific scenario can only roughly be quantified. A technique overcoming this flaw is *stochastic simulation*, which is known as *Dynamic Financial Analysis* (DFA) when applied to financial cash flow modelling of a (nonlife) insurance company. Thousands of different scenarios are generated stochastically allowing for the full probability distribution of important output variables, like surplus, written premiums or loss ratios.

1.2. Fixing the Time Period

The first step to compare different strategies is to fix a time horizon they should apply to. On the one hand we would like to model over as long a time period as possible in order to see the long-term effects of a chosen strategy. In particular, effects concerning long-tail business only appear after some years and can hardly be recognized in the first few years. On the other hand, simulated values become more unreliable the longer the projection period, due to accumulation of process and parameter risk over time. A projection period of five to ten years seems to be a reasonable choice. Usually the time period is split into yearly, quarterly or monthly sub periods.

1.3. Comparison to ALM in Life Insurance

A DFA model is a stochastic model of the main financial factors of an insurance company. A good model should simulate stochastically the asset elements, the liability elements and also the relationships between both types of random factors. Many traditional ALM-approaches (ALM = Asset/Liability Management) in life insurance considered the liabilities as more or less deterministic due to their low variability (see for example Wise [43] or Klett [25]). This approach would be dangerous in nonlife where we are faced with much more volatile liability cash flows. Nonlife companies are highly sensitive to inflation, macroeconomic conditions, underwriting movements and court rulings, which complicate the modelling process while simultaneously making results less certain than for life insurance companies. In nonlife both the date of occurrence and the size of claims are uncertain. Claim costs in nonlife are inflation sensitive, whereas they are expressed in nominal terms for many traditional life insurance products. In order to cope with the stochastic nature of nonlife liabilities and assets, their number and their complex interactions, we have to rely on stochastic simulations.

1.4. Objectives of DFA

DFA is not an academic discipline per se. It borrows many well-known concepts and methods from economics and statistics. It is part of the financial management of the firm. As such it is committed to management of profitability and financial stability (risk control function of DFA). While the first task aims at maximizing shareholder value, the second one serves maintaining customer value. Within these two seemingly conflicting coordinates DFA tries to

- strategic asset allocation,
- capital allocation,
- performance measurement,
- market strategies,
- business mix,
- pricing decisions,
- product design,
- and others.

This listing suggests that DFA goes beyond designing an asset allocation strategy. In fact, portfolio managers will be affected by DFA decisions as well as underwriters. Concrete implementation and application of a DFA model depends on two fundamental and closely related questions to be answered beforehand:

1. Who is the primary beneficiary of a DFA analysis (shareholder, management, policyholders)?
2. What are the company individual objectives?

The answer to the first question determines specific accounting rules to be taken into account as well as scope and detail of the model. For example, those companies only interested in getting a tool for enhancing their asset allocation on very high aggregation level will not necessarily target a model that emphasizes every detail of simulating liability cash flows. Smith [39] has pointed out that making money for shareholders has not been the primary motivation behind developments in ALM (or DFA). Furthermore, relying on the Modigliani-Miller theorem (see Modigliani and Miller [34]) he put forward the hypothesis that a cost benefit analysis of asset/liability studies might reveal that costs fall on shareholders but benefits on management or customers. Our general conclusion is that company individual objectives – in particular with respect to the target group – have to be identified and formulated before starting the DFA analysis.

1.5. Analyzing DFA Results Through Efficient Frontiers

Before using a DFA model, management has to choose a financial or economic measure in order to assess particular strategies. The most common framework is the *efficient frontier* concept widely used in modern portfolio theory going back to Markowitz [32]. First, a company has to choose a return measure (e.g. expected surplus) and a risk measure (e.g. expected policyholder deficit, see Lowe and Stanard [30], or worst conditional mean as a coherent risk measure, see Artzner, Delbaen, Eber and Heath [2] and [3]). Then the measured risk and return of each strategy can be plotted as shown in Figure 1.1. Each strategy represents one spot in the risk-return diagram. A strategy is called *efficient* if there is no other one with lower risk at the same level of return, or higher return at the same level of risk.

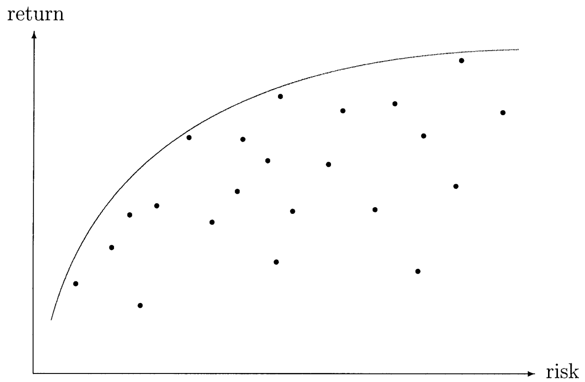

FIGURE 1.1: Efficient frontier.

For each level of risk there is a maximal return that cannot be exceeded, giving rise to an efficient frontier. But the exact position of the efficient frontier is unknown. There is no absolute certainty whether a strategy is really efficient or not. DFA is not necessarily a method to come up with an optimal strategy. DFA is predominantly a tool to *compare* different strategies in terms of risk and return. Unfortunately, comparison of strategies may lead to completely different results as we change the return or risk measure. A different measure may lead to a different preferred strategy. This will be illustrated in Section 4.

Though efficient frontiers are a good means of communicating the results of DFA because they are well-known, some words of criticism are in place. Cumberworth, Hitchcox, McConnell and Smith [10] have pointed out that there are pitfalls related to efficient frontiers one has to be aware of. They criticize that typical efficient frontier uses risk measures that mix together systematic risk (non-diversifiable by shareholders) and non-systematic risk, which blurs the shareholder value perspective. In addition to that, efficient frontiers might give misleading advice if they are used to address investment decisions once the concept of systematic risk has been factored into the equation.

1.6. Solvency Testing

A concept closely related to DFA is *solvency testing* where the financial position of the company is evaluated from the perspective of the customers. The central idea is to quantify in probabilistic terms whether the company will be able to meet its commitments in the future. This translates into determining the necessary amount of capital given the level of risk the company is exposed to. For example, does the company have enough capital to keep the probability of loosing $\alpha \cdot 100\%$ of its capital below a certain level for the risks taken? DFA provides a whole probability distribution of surplus. For each level α the probability of loosing $\alpha \cdot 100\%$ can be derived from this distribution. Thus DFA serves as a solvency testing tool as well. More information about solvency testing can be found in Schnieper [37] and [38].

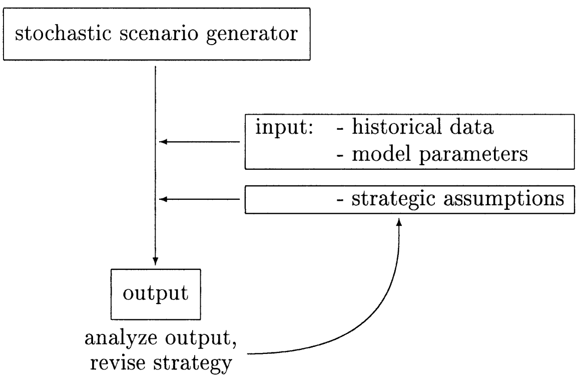

FIGURE 1.2: Main structure of a DFA model.

1.7. Structure of a DFA Model

Most DFA models consist of three major parts, as shown in Figure 1.2. The *stochastic scenario generator* produces realizations of random variables representing the most important drivers of business results. A realization of a random variable in the course of simulation corresponds to fixing a scenario. The second data source consists of company specific *input* (e.g. mean severity of losses per line of business and per accident year), assumptions regarding model parameters (e.g. long-term mean rate in a mean reverting interest rate model), and strategic assumptions (e.g. investment strategy). The last part, the *output* provided by the DFA model, can then be analyzed by management in order to improve the strategy, i.e. make new strategic assumptions. This can be repeated until management is convinced by the superiority of a certain strategy. As pointed out in Cumberworth, Hitchcox, McConnell and Smith [10] interpretation of the output is an often neglected and non-appreciated part in DFA modelling. For example, an efficient frontier leaves us still with a variety of equally desirable strategies. At the end of the day management has to decide for only one of them and selection of a strategy based on preference or utility functions does not seem to provide a practical solution in every case.

2. Stochastically Modelled Variables

A very important step in the process of building an appropriate model is to identify the key random variables affecting asset and liability cash flows. Afterwards it has to be decided whether and how to model each or only some of these factors and the relationships between them. This decision is influenced by considerations of a trade-off between improvement of accuracy versus increase in complexity which is often felt being equivalent to a reduction of transparency.

The risks affecting the financial position of a nonlife insurer can be categorized in various ways. For example, pure asset, pure liability and asset/liability risks. We believe that a DFA model should at least address the following risks:

- pricing or underwriting risk (risk of inadequate premiums),
- reserving risk (risk of insufficient reserves),
- investment risk (volatile investment returns and capital gains),
- catastrophes.

We could have also mentioned credit risk related to reinsurer default, currency risk and some more. For a recent, detailed DFA discussion of the possible impact of exchange rates on reinsurance contracts see Blum, Dacorogna, Embrechts, Neghaiwi and Niggli [5]. A critical part of a DFA model are the interdependencies between different risk categories, in particular between risks associated with the asset side and those belonging to liabilities. The risk of company losses triggered by changes in interest rates is called *interest rate risk*. We will come back to the question of modelling dependencies in Section 5.1. Our choice of company relevant random variables is based on the categorization of risks shown before.

A key module of a DFA model is an *interest rate generator*. Many models assume that interest rates will drive the whole model as displayed for example in Figure 4.1. An interest rate generator – or *economic scenario generator* as it is often called to emphasize the far reaching economic impact of interest rates – is necessary in order to be able to tackle the problem of evaluating interest rate risk. Moreover, nonlife insurance companies are strongly exposed to interest rate behaviour due to generally large investments in fixed income assets. In our model implementation we assumed that interest rates were strongly correlated with inflation, which itself influenced future changes in claim size and claim frequency. On the other hand, both of these factors affected (future) premium rates. Furthermore, we assumed correlation between interest rates and *stock returns*, which are generally an important component of investment returns.

On the liability side, we explicitly considered four sources of randomness: *non-catastrophe losses*, *catastrophe losses*, *underwriting cycles*, and *payment* *patterns*. We simulated catastrophes separately due to quite different statistical behaviour of catastrophe and non-catastrophe losses. In general the volume of empirical data for non-catastrophe losses is much bigger than for catastrophe losses. Separating the two led to more homogeneous data for non-catastrophe losses, which made fitting the data by well-known (right skewed) distributions easier. Also, our model implementation allowed for evaluating reinsurance programs. Testing different deductibles or limits is only possible if the model is able to generate sufficiently large individual losses. In addition, we currently experience a rapid development of a theory of distributions for extremal events (see Embrechts, Klüppelberg and Mikosch [16], and McNeil [33]). Therefore, we considered the separate modelling of catastrophe and non-catastrophe losses as most appropriate. For each of these two groups the number and the severity of claims were modelled separately. Another approach would have been to integrate the two kinds of losses by using heavy-tailed claim size distributions.

Underwriting cycles are an important characteristic of nonlife companies. They reflect market and macroeconomic conditions and they are one of the most important factors affecting business results. Therefore, it is useful to have them included in a DFA model set-up.

Losses are not only characterized by their (ultimate) size but also by their piecewise payment over time. This property increases the uncertainties of the claims process by introducing the time value of money and future inflation considerations. As a consequence, it is necessary not only to model claim frequency and severity but the uncertainties involved in the settlement process as well. In order to allow for reserving risk we used stochastic payment patterns as a means of estimating loss reserves on a gross and on a net basis.

In the abstract we pointed out that our intention was to present a DFA model framework. In concrete terms, this means that we present a model implementation that we found useful to achieve part of the goals outlined in Section 1.4. We do not claim that the components introduced in the remaining part of the paper represent a high class standard of DFA modelling. For each of the DFA components considered there are numerous alternatives, which might turn out to be more appropriate in particular situations. Providing a model framework means to present our model as a kind of suggested reference point that can be adjusted or improved individually.

2.1. Interest Rates

Following Daykin, Pentikäinen and Pesonen [15, p. 231] we assume strong correlation between general inflation and interest rates. Our primary stochastic driver is the (instantaneous) short-term interest rate. This variable determines bond return across all maturities as well as general inflation and superimposed inflation by line of business.

An alternative to the modelling of interest and inflation rates as outlined in this section and probably well-known to actuaries is the *Wilkie model*, see Wilkie [42], or Daykin, Pentikäinen and Pesonen [15, pp. 242-250].

2.1.1. Short-Term Interest Rate

There are many different interest rate models used by financial economists. Even the literature offering surveys of interest rate models has grown considerably. The following references represent an arbitrary selection: Ahlgrim, D'Arcy and Gorvett [1], Musiela and Rutkowski [35, pp. 281-302] and Björk [4]. The final choice of a specific interest rate model is not straightforward, given the variety of existing models. It might be helpful to post some general features of interest rate movements, which we took from Ahlgrim, D'Arcy and Gorvett [1]:

1. Volatility of yields at different maturities varies.
2. Interest rates are mean-reverting.
3. Rates at different maturities are positively correlated.
4. Interest rates should not be allowed to become negative.
5. The volatility of interest rates should be proportional to the level of the rate.

In addition to these characteristics there are some practical issues raised by Rogers [36]. According to Rogers an interest rate model should be

- *flexible* enough to cover most situations arising in practice,
- *simple* enough that one can compute answers in reasonable time,
- *well-specified*, in that required inputs can be observed or estimated,
- *realistic*, in that the model will not do silly things.

It is well-known that an interest rate model meeting all the criteria mentioned does not exist. We decided to rely on the one-factor Cox-Ingersoll-Ross (CIR) model. CIR belongs to the class of equilibrium based models where the instantaneous rate is modelled as a special case of an Ornstein-Uhlenbeck process:

$$dr = \kappa(\theta - r)dt + \sigma r^{\gamma} dZ, \tag{2.1}$$

By setting $\gamma = 0.5$ we arrive at CIR also known as the *square root process*

$$dr_t = a(b - r_t)dt + s\sqrt{r_t}\, dZ_t, \tag{2.2}$$

where

r_t = instantaneous short-term interest rate,
b = long-term mean,
a = constant that determines the speed of reversion of the interest rate toward its long-run mean b,
s = volatility of the interest rate process,
(Z_t) = standard Brownian motion.

CIR is a mean-reverting process where the short rate stays almost surely positive. Moreover, CIR allows for an affine model of the term structure making the model analytically more tractable. Nevertheless, some studies have shown (see Rogers [36]) that one-factor models in general do not satisfactorily fit empirical data and restrict term structure dynamics. Multifactor models like Brennan and Schwartz [6] or Longstaff and Schwartz [29] or *whole yield* approaches like Heath-Jarrow-Morton [20] have proven to be more appropriate in this respect. But this comes at the price of being much more involved from a theoretical and a practical implementation point of view. Our decision for CIR was motivated by practical considerations. It is an easy to implement model that gave us reasonable results when applied to US market data. Moreover, it is a standard model and in widespread use, in particular in the US.

Actually, we are interested in simulating the short rate dynamics over the projection period. Hence, we discretized the mean reverting model (2.2) leading to

$$r_t = r_{t-1} + a(b - r_{t-1}) + s\sqrt{r_{t-1}}\, Z_t, \tag{2.3}$$

where
r_t = the instantaneous short-term interest rate at the beginning of year t,
$Z_t \sim \mathcal{N}(0,1)$, Z_1, Z_2, ... i.i.d.,
a, b, s as in (2.2).

Cox, Ingersoll and Ross [9] have shown that rates modelled by (2.2) are positive almost surely. Although it is hard for the short rate process to go negative in the discrete version of the last equation the probability is not zero. To be sure we changed equation (2.3) to

$$r_t = r_{t-1} + a(b - r_{t-1}) + s\sqrt{r_{t-1}^{\;+}}\, Z_t. \tag{2.4}$$

A generalization of CIR is given by the following equation, where setting g = 0.5 yields again CIR:

$$r_t = r_{t-1} + a(b - r_{t-1}) + s(r_{t-1}^{\;+})^g Z_t. \tag{2.5}$$

This general version provides more flexibility in determining the degree of dependence between conditional volatility of interest rate changes and the level of interest rates.

The question of what an appropriate level for g might be leads to the field of model calibration which we will encounter at several places within DFA modelling. In fact, the problem plays a dominant role in DFA tempting many practitioners to state that DFA is all about calibration. Calibrating an interest rate model of the short rate refers to determining parameters – a, b, s and g in equation (2.5) – so as to ensure that modelled spot rates (based on the instantaneous rate) correspond to empirical term structures derived from traded financial instruments. Björk [4] calls the procedure to achieve this *inversion of the yield curve*. However, the parameters can not be uniquely determined from an empirical term structure and term structure of volatilities resulting in a non-perfect fit. This is a general feature of equilibrium interest rate models. Whereas this is a critical point for valuing interest rate derivatives, the impact on long-term DFA results may be limited.

With regard to calibrating the inflation model it should be mentioned that building models of inflation based on historical data may be a feasible approach. But it is unclear whether the future evolution of inflation will follow historical patterns: DFA output will probably reflect the assumptions with regard to inflation dynamics. Consequently, some attention needs to be paid to these assumptions. Neglecting this is a common pitfall of DFA modelling. In order to allow for stress testing of parameter assumptions, the model should not only rely on historical data but on economic reasoning and actuarial judgment of future development as well.

2.1.2. Term Structure

Based on equation (2.2) we calculated the prices $F(t, T, (r_t))$ being in place at time t of zero-coupon bonds paying 1 monetary unit at time of maturity $t + T$, as

(2.6) $$F(t, T, (r_t)) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_{t+s}\,ds} \middle| r_t \right] = e^{\log A_T - r_t B_T},$$

where

$$A_T = \left(\frac{2Ge^{(a+G)\,T/2}}{(a+G)(e^{GT}-1)+2G} \right)^{2ab/s^2},$$

$$B_T = \frac{2\,(e^{GT}-1)}{(a+G)(e^{GT}-1)+2G},$$

$$G = \sqrt{a^2 + 2s^2}\,.$$

A proof of this result can be found in Lamberton and Lapeyre [27, pp. 129-133]. Note, that the expectation operator is taken with respect to the martingale measure $\mathbb{Q}$ assuming that equation (2.2) is set up under the martingale measure $\mathbb{Q}$ as well. The continuously compounded spot rates $R_{t,T}$ at time t derived from equation (2.6) determine the modelled term structure of zero-coupon yields at time t:

(2.7) $$R_{t,T} = -\frac{\log F(t, T, (r_t))}{T} = \frac{r_t B_T - \log A_T}{T},$$

where T is the time to maturity.

2.1.3. General Inflation

Modelling loss payments requires having regard to inflation. Following our introductory remark to Section 2.1 we simulated general inflation i_t by using the (annualized) short-term interest rate r_t. We did this by using a linear regression model on the short-term interest rate:

(2.8) $$i_t = a^I + b^I r_t + \sigma^I \varepsilon_t^I,$$

where

$\varepsilon_t^I \sim \mathcal{N}(0,1)$, $\varepsilon_1^I, \varepsilon_2^I, \ldots$ i.i.d.,

a^I, b^I, σ^I: parameters that can be estimated by regression, based on historical data.

The index I stands for *general inflation*.

2.1.4. Change by Line of Business

Lines of business are affected differently by general inflation. For example, car repair costs develop differently over time than business interruption costs. Claims costs for specific lines of business are strongly affected by legislative and court decisions, e.g. product liability. This gives rise to so-called *superimposed inflation*, adding to general inflation. More on this can be found in Daykin, Pentikäinen and Pesonen [15, p. 215] and Walling, Hettinger, Emma and Ackerman [41].

To model the change in loss frequency δ_t^F (i.e. the ratio of number of losses divided by number of written exposure units), the change in loss severity δ_t^X, and the combination of both of them, δ_t^P, we used the following formulas:

$$\delta_t^F = \max(a^F + b^F i_t + \sigma^F \varepsilon_t^F, -1), \tag{2.9}$$

$$\delta_t^X = \max(a^X + b^X i_t + \sigma^X \varepsilon_t^X, -1), \tag{2.10}$$

$$\delta_t^P = \left(1 + \delta_t^F\right)\left(1 + \delta_t^X\right) - 1, \tag{2.11}$$

where

$\varepsilon_t^F \sim \mathcal{N}(0,1)$, $\varepsilon_1^F, \varepsilon_2^F, \ldots$ i.i.d.,

$\varepsilon_t^X \sim \mathcal{N}(0,1)$, $\varepsilon_1^X, \varepsilon_2^X, \ldots$ i.i.d., $\varepsilon_{t_1}^F, \varepsilon_{t_2}^X$ independent $\forall t_1, t_2$,

$a^F, b^F, \sigma^F, a^X, b^X, \sigma^X$: parameters that can be estimated by regression, based on historical data.

The variable δ_t^P represents changes in loss trends triggered by changes in inflation rates. δ_t^P is applied to premium rates as will be explained in Section 3, see (3.2). Its construction through (2.11) ensures correlation of aggregate loss amounts and premium levels that can be traced back to inflation dynamics.

The technical restriction of setting δ_t^F and δ_t^X to at least –1 was necessary to avoid negative values for numbers of losses and loss severities.

We modelled changes in loss frequency dependent on general inflation because empirical observations revealed that under specific economic conditions (e.g. when inflation is high) policyholders tend to report more claims in certain lines of business.

The corresponding cumulative changes $\delta_t^{F,c}$ and $\delta_t^{X,c}$ can be calculated by

$$\delta_t^{F,c} = \prod_{s=t_0+1}^{t} (1 + \delta_s^F), \tag{2.12}$$

$$\delta_t^{X,c} = \prod_{s=t_0+1}^{t} (1 + \delta_s^X), \tag{2.13}$$

where

$t_0 + 1$ = first year to be modelled.

2.2. Stock Returns

The major asset classes of a nonlife insurance company comprise fixed income type assets, stocks and real estate. Here, we confine ourselves to a description of the model employed for stocks. Modelling stocks can start either with concentrating on stock prices or stock returns (although both methods should turn out to be equivalent in the end). We followed the last approach since we could rely on a well established theory relating stock returns and the risk-free interest rate: the *Capital Asset Pricing Model* (CAPM) going back to Sharpe-Lintner, see for example Ingersoll [22].

In order to apply CAPM we needed to model the return of a portfolio that is supposed to represent the stock market as a whole, the *market portfolio*. Assuming a significant correlation between stock and bond prices and taking into account multi-periodicity of a DFA model we came up with the following linear model for the stock market return in projection year t conditional on the one-year spot rate $R_{t,1}$ at time t.

$$\mathbb{E}\left[r_t^M \middle| R_{t,1}\right] = a^M + b^M (e^{R_{t,1}} - 1), \tag{2.14}$$

where

$e^{R_{t,1}} - 1$ = risk-free return, see (2.7),

a^M, b^M = parameters that can be estimated by regression, based on historical data and economic reasoning.

Since we modelled sub periods of length one year, we conditioned on the *one-year* spot rate. Note that r_t^M must not be confused with the instantaneous short-term interest rate r_t in CIR. Note also that a negative value of b^M means that increasing interest rates entail expected stock prices falling.

Now we can apply the CAPM formula to get the conditional expected return on an arbitrary stock S:

$$\mathbb{E}\left[r_t^S \middle| R_{t,1}\right] = \left(e^{R_{t,1}} - 1\right) + \beta_t^S \left(\mathbb{E}\left[r_t^M \middle| R_{t,1}\right] - \left(e^{R_{t,1}} - 1\right)\right), \tag{2.15}$$

where

$$\begin{aligned} e^{R_{t,1}} - 1 &= \text{risk-free return,} \\ r_t^M &= \text{return on the market portfolio,} \\ \beta_t^S &= \beta\text{-coefficient of stock S} \\ &= \frac{\mathrm{Cov}(r_t^S, r_t^M)}{\mathrm{Var}(r_t^M)}. \end{aligned}$$

If we assume a geometric Brownian Motion for the stock price dynamics we get a lognormal distribution for $1 + r_t^S$:

$$1 + r_t^S \sim \text{lognormal}\,(\mu_t, \sigma^2),\ r_1^S,\ r_2^S,\ \ldots \text{ independent,} \tag{2.16}$$

with μ_t chosen to yield

$$m_t = e^{\mu_t + \sigma^2/2},$$

where

$m_t = 1 + \mathbb{E}\left[r_t^S \middle| R_{t,1}\right]$ see (2.15),

σ^2 = estimated variance of logarithmic historical stock returns.

Again, we would like to emphasize that our method of modelling stock returns represents only one out of many possible approaches.

2.3. Non-Catastrophe Losses

Usually, non-catastrophe losses of various lines of business develop quite differently compared to catastrophe losses, see also the introductory remarks of Section 2. Therefore, we modelled non-catastrophe and catastrophe losses separately and per line of business. For simplicity's sake, we will drop the index denoting line of business in this section.

Experience shows that loss amounts depend also on the age of insurance contracts. The *aging phenomenon* describes the fact that the loss ratio – i.e. the ratio of (estimated) total loss divided by earned premiums – decreases when the age of policy increases. For this reason we divided insurance business into three classes, as proposed by D'Arcy, Gorvett, Herbers, Hettinger, Lehmann and Miller [13]:

- new business (superscript 0),
- renewal business – first renewal (superscript 1), and
- renewal business – second and subsequent renewals (superscript 2).

More information about the aging phenomenon can be found in D'Arcy and Doherty [11] and [12], Feldblum [19], and in Woll [44].

Disregarding the time of incremental loss payment for the moment, the two main stochastic factors affecting total claim amount are: number of losses and severity of losses, see for instance Daykin, Pentikäinen and Pesonen [15]. The choice of a specific claim number and claim size distribution depends on the line of business and is the result of fitting distributions to empirical data requiring foregoing adjustments of historical loss data. In this section we shall demonstrate our model of non-catastrophe losses by referring to a negative binomial (claim number) and a gamma (claim size) distribution.

To simulate loss numbers N_t^j and mean loss severities $X_t^j = \frac{1}{N_t^j} \sum_{i=1}^{N_t^j} X_t^j(i)$ for period t and renewal category j we utilized mean values $\mu^{F,j}$, $\mu^{X,j}$ and standard deviations $\sigma^{F,j}$, $\sigma^{X,j}$ of historical data for loss frequencies and mean loss severities. We took also into account inflation and written exposure units. Because loss frequencies behave more stable than loss numbers, we used estimations of loss frequencies instead of relying on estimates of loss numbers.

As an example of a distribution for claim numbers N_t^j we consider the negative binomial distribution with mean $m_t^{N,j}$ and variance $v_t^{N,j}$. Generally, we reserved the variables m and v for mean and variance of different factors. These factors were referred by attaching a superscript (N, X, Y, …) to m or v:

$$\begin{aligned} &N_t^j \sim \mathrm{NB}\,(a, p), \;\; j = 0, 1, 2, \\ &N_1^j, N_2^j, \ldots \text{ independent}, \end{aligned} \tag{2.17}$$

with a and p chosen to yield

$$\begin{aligned} m_t^{N,j} &= \mathbb{E}\left[N_t^j\right] = \frac{a(1-p)}{p}, \\ v_t^{N,j} &= \mathrm{Var}(N_t^j) = \frac{a(1-p)}{p^2}, \end{aligned} \tag{2.18}$$

where

$m_t^{N,j} \;=\; w_t^j \mu^{F,j} \delta_t^{F,c}$,

$v_t^{N,j} \;=\; \left(w_t^j \sigma^{F,j} \delta_t^{F,c}\right)^2$,

$w_t^j \;=$ written exposure units; introduced in more detail and modelled in (3.3),

$\mu^{F,j} \;=$ estimated frequency, based on historical data,

$\sigma^{F,j}$ = estimated standard deviation of frequency, based on historical data,
$\delta_t^{F,c}$ = cumulative change in loss frequency, see (2.12).

Negative binomial distributed variables N exhibit over-dispersion: $\mathrm{Var}(N) \geq \mathbb{E}[N]$. Consequently, this distribution yields a reasonable model only if $v_t^{N,j} \geq m_t^{N,j}$.

Historical data are a good basis to calibrate this model as long as there had been no significant structural changes within a line of business in prior years. Otherwise, explicit consideration of exposure data may be a better basis for calibrating the claims process.

In the following we will present an example of a claim size distribution for high frequency, low severity losses. Due to the fact that the density function of the gamma distribution decreases exponentially under appropriate choice of parameters it is a distribution serving our purposes well:

$$X_t^j \sim \text{Gamma}\ (\alpha, \theta),\ \ j = 0, 1, 2, \qquad (2.19)$$
$$X_1^j,\ X_2^j,\ \ldots \text{ independent,}$$

with α and θ chosen to yield

$$m_t^{X,j} = \mathbb{E}\left[X_t^j\right] = \alpha\theta,$$
$$v_t^{X,j} = \mathrm{Var}(X_t^j) = \alpha\theta^2,$$

where

$m_t^{X,j}$ = $\mu^{X,j}\,\delta_t^{X,c}$,
$v_t^{X,j}$ = $\left(\sigma^{X,j}\,\delta_t^{X,c}\right)^2 / \delta_t^{F,c}$,
$\mu^{X,j}$ = estimated mean severity, based on historical data,
$\sigma^{X,j}$ = estimated standard deviation, based on historical data,
$\delta_t^{X,c}$ = cumulative change in loss severity, see (2.13),
$\delta_t^{F,c}$ = cumulative change in loss frequency, see (2.12).

By multiplying the number of losses with the mean severity, we got the total (non-catastrophic) loss amount in respect of a certain line of business: $\sum_{j=0}^{2} N_t^j X_t^j$.

2.4. Catastrophes

We are turning now to losses triggered by catastrophic events like windstrom, flood, hurricane, earthquake, etc. In Section 2 we mentioned that we could have integrated non-catastrophic and catastrophic losses by using heavy-tailed distributions, see Embrechts, Klüppelberg and Mikosch [16]. Nevertheless, we decided for separate modelling, see our reasons given in Section 2.

There are different ways of modelling the number of catastrophes, e.g. negative binomial, poisson, or binomial distribution with mean m^M and variance v^M. We assumed that there were no trends in the number of catastrophes:

$$\begin{aligned} & M_t \sim \text{NB, Pois, Bin, } \dots \text{ (mean } m^M\text{, variance } v^M\text{)}, \\ & M_1, M_2, \dots \text{ i.i.d.,} \end{aligned} \tag{2.20}$$

where

m^M = estimated number of catastrophes, based on historical data,
v^M = estimated variance, based on historical data.

Contrary to the modelling of non-catastrophe losses, we simulated the *total* (economic) loss (i.e. not only the part the insurance company in consideration has to pay) for each *catastrophic event* $i \in \{1,\dots, M_t\}$ separately. Again, there are different probability distributions, which prove to be adequate for this purpose, in particular GPD (generalized Pareto distribution) $G_{\xi,\beta}$. GPD's play an important role in Extreme Value Theory, where $G_{\xi,\beta}$ appears as the limit distribution and Mikosch [16, p. 165]. In the following equation Y_t^i describes the total economic loss caused by catastrophic event $i \in \{1,\dots, M_t\}$ in projection period t.

$$\begin{aligned} & Y_{t,i} \sim \text{lognormal, Pareto, GPD, } \dots \text{ (mean } m_t^Y\text{, variance } v_t^Y\text{)}, \\ & Y_{t,1}, Y_{t,2}, \dots \text{ i.i.d.,} \\ & Y_{t_1,i_1}, Y_{t_2,i_2} \text{ independent } \forall (t_1,i_1) \neq (t_2,i_2), \end{aligned} \tag{2.21}$$

where

$m_t^Y = \mu^Y \delta_t^{X,c}$,
$v_t^Y = \left(\sigma^Y \delta_t^{X,c}\right)^2$,
μ^Y = estimated loss severity, based on historical data,
σ^Y = estimated standard deviation, based on historical data,
$\delta_t^{X,c}$ = cumulative change in loss severity, see (2.13).

After having generated Y_t^i we split it into pieces reflecting the loss portions of different lines of business:

$$Y_{t,i}^k = a_{t,i}^k Y_{t,i}, \quad k = 1,\dots, l, \tag{2.22}$$

where

k = line of business,

l = total number of lines considered,

$\forall i \in \{1,\ldots, M_t\}$: $\left(a^1_{t,i},\ldots,a^l_{t,i}\right) \in \left\{x \in [0,1]^l, \|x\|_1 = 1\right\} \subset \mathbb{R}^l$ is a random convex combination, whose probability distribution within the $(l-1)$ dimensional tetraeder can be arbitrarily specified.

Simulating the percentages $a^k_{t,i}$ stochastically over time varies the impact of catastrophes on different lines favoring those companies, which are well diversified in terms of number of lines written.

Knowing the market share of the nonlife insurer and its reinsurance structure permits calculation of loss payments allowing as well for catastrophes. Although random variables were generated independently our model introduced differing degrees of dependence between aggregate losses of different lines by ensuring that they were affected by same catastrophic events (although to different degrees).

2.5. Underwriting Cycles

More or less irregular cycles of underwriting results several years in length are an intrinsic characteristic of the (deregulated) nonlife insurance industry. Cycles can vary significantly between countries, markets and lines of business. Sometimes their appearance is masked by smoothing of published results. There are probably many potential background factors, varying from period to period, causing cycles. Among others we mention

- time lag effect of the pricing procedure
- trends, cycles and short-term variations of claims,
- fluctuations in interest rate and market values of assets.

Besides having introduced cyclical variation driven by interest rate movements – remember that short-term interest rates are the main factor affecting all other variables in the model – we added a sub-model concerned with premium cycles induced by competitive strategies. In this section we shall describe this approach.

We used a homogeneous Markov chain model (in discrete time) similar to D'Arcy, Gorvett, Hettinger and Walling [14]: We assign one of the following states to each line of business for each projection year:

1 weak competition,
2 average competition,
3 strong competition.

In state 1 (weak competition) the insurance company demands high premiums being aware that it can most likely increase its market share. In state 3 (strong competition) the insurance company has to accept low premiums in order to at least keep its current market share. Assuming a stable claim environment, high premiums are equivalent to high profit margin over pure premium, and low premiums equal low profit margin. Changing from one state to another might cause significant changes in premiums.

The transition probabilities p_{ij}, $i, j \in \{1, 2, 3\}$, which denote the probability of changing from state i to state j from one year to the next are assumed to be equal for each projection year. This means that the Markov chain is homogeneous. The p_{ij}'s form a matrix T:

$$T = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}.$$

There are many different possibilities to set these transition probabilities p_{ij}, $i, j \in \{1, 2, 3\}$. It is possible to model the p_{ij}'s depending on current market conditions applicable to each line of business separately. If the company writes l lines of business this will imply 3^l states of the world. Because business cycles of different lines of business are strongly correlated, only few of the 3^l states are attainable. Consequently, we have to model $L \ll 3^l$ states, where the transition probabilities p_{ij}, $i, j \in \{1,\dots, L\}$ remain constant over time. It is possible that some of them are zero, because there may exist some states that cannot be attained directly from certain other states. When L states are attainable, the matrix T has dimension $L \times L$:

$$T = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1L} \\ p_{21} & p_{22} & \cdots & p_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ p_{L1} & p_{L2} & \cdots & p_{LL} \end{pmatrix}.$$

In order to fix the transition probabilities p_{ij} in any of the above mentioned cases each state i should be treated separately and probabilities assigned to the variables $p_i,\dots,p_{iL}$ such that $\sum_{j=1}^{L} p_{ij} = 1 \; \forall i$. Afterwards, the stationary probability distribution π has to be considered which the chosen probability distribution generally converges to, irrespective of the selected starting point, given that the Markov chain is irreducible and positive recurrent. We took advantage of the fact that $\pi = \pi T$ to check whether the estimated values for the transition probabilities are reasonable because it is easier to estimate the stationary probability distribution π than to find suitable values for the p_{ij}'s. Since it is extremely delicate to estimate the transition probabilities in an appropriate way, one should not only rely on historical data but use experience based knowledge as well.

It is crucial to set the initial market conditions correctly in order to produce realistic financial projections of the insurance entity.

2.6. Payment Patterns

So far we have been focusing on claim numbers and severities. This section is dedicated to explaining how we managed to model the uncertainties of the claim settlement process, i.e. the random time to payment, as indicated in Section 2. We considered a whole loss portfolio belonging to a specific line of business and its aggregate yearly loss payments in different calendar years (or *development periods*). The piecewise (or incremental) payment of aggegrate losses stemming from one and the same *accident year* forms a *payment pattern*. An (incremental) payment pattern is a vector with length equal to an assumed number of development periods. The i-th vector component describes the percentage of estimated ultimate loss amount (on aggregate portfolio level) to be paid out in the $(i-1)$-st development year. If we consider yearly loss payments pertaining to a specific accident year t then the i-th development year refers to calendar year $t + i$.

In the following we will denote accident years by t_1 and development years by t_2. For simplicity's sake, we will drop the index representing line of business for the most part of this section.

Very often one finds payment patterns treated as being deterministic in DFA models. This will be justified by pointing out that payment patterns do not change significantly from one year to the next. We believe that in order to account for reserving risk in a DFA model properly one has to have a stochastic model for the timing of loss payments as well.

Generally, for each prior accident year considered, the loss amounts which have been paid to date are known. Figure 2.1 displays this in graphical format. The triangle formed by the area on the left hand side of the bold line – the *loss triangle* – represents empirical, i.e. known, loss payments whereas the remaining parts represent outstanding and future loss payments, which are unknown. For example, if we assume to be at the end of calendar year 2000 ($t_0 = 2000$) considering accident year 1996 ($= t_0 - 4$), we know the loss amounts pertaining to accident year 1996, which have been paid out in calendar years 1996, 1997,…, 2000. But we do not know the amounts that will be paid in calendar year 2001 and later. Some very popular actuarial techniques for estimating outstanding loss payments – which are characterized by those cell

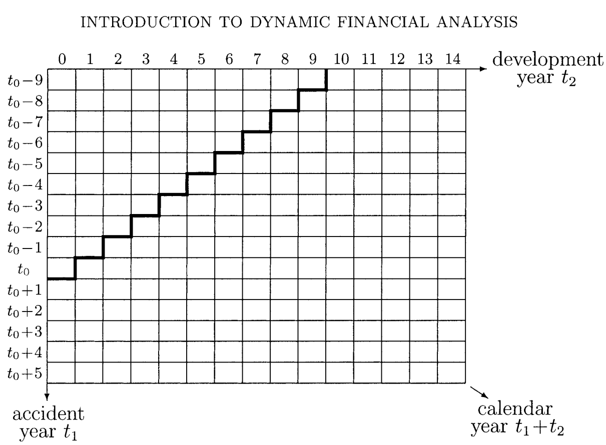

FIGURE 2.1: Paid losses (upper left triangle), outstanding loss payments and future loss payments.

entries (t_1, t_2), $t_1 \le t_0$, belonging to the right hand side of the bold line – are based on deriving an average payment pattern from loss payments represented by the loss triangle.

In the simplified model description of this section we will not take into account the empirical fact that payment patterns of single large losses differ from those of aggregate losses. We will also disregard changes in future claim inflation, although it might have a strong impact on certain lines of business.

For each line we assumed an ultimate development year τ when all claims arising from an accident year would be paid completely. Incremental claim payments denoted by Z_{t_1,t_2} are known for previous years $t_1 + t_2 \le t_0$. Ultimate loss amounts $Z_{t_1}^{\text{ult}} := \sum_{t=0}^{\tau} Z_{t_1,t}$ vary by accident year t_1. In order to determine loss reserves taking into account reserving risk we first had to simulate random loss payments Z_{t_1,t_2}. As a second step we needed to have a procedure for estimating ultimate loss amounts $Z_{t_1}^{\text{ult}}$ at each future time.

We distinguished two cases. First we will explain the modelling of outstanding loss payments pertaining to previous accident years followed by a description to model loss payments in respect of future accident years.

For previous accident years ($t_1 \le t_0$) payments Z_{t_1,t_2}, with $t_1 + t_2 \le t_0$ are known. We used them as a basis for predicting outstanding payments. We used a chain-ladder type procedure (for the chain-ladder method, see Mack [31]), i.e. we applied ratios to cumulative payments per accident year. The following type of loss development factor was defined

(2.23) $$d_{t_1,t_2} := \frac{Z_{t_1,t_2}}{\sum_{t=0}^{t_2-1} Z_{t_1,t}}, \quad t_2 \ge 1.$$

Note that this ratio is not a typical chain-ladder link ratio. When mentioning *loss development factors* in this section we are always referring to factors defined by (2.23).

Since a lognormal distribution usually provides a good fit to historical loss development factors, we used the following model for outstanding loss payments in calendar years $t_1 + t_2 \ge t_0 + 1$ for accident years $t_1 \le t_0$:

(2.24) $$Z_{t_1,t_2} = d_{t_1,t_2} \cdot \sum_{t=0}^{t_2-1} Z_{t_1,t},$$

where

$d_{t_1,t_2} \sim$ lognormal $(\mu_{t_2}, \sigma_{t_2}^2)$,

μ_{t_2} = estimated logarithmic loss development factor for development year t_2, based on historical data,

σ_{t_2} = estimated logarithmic standard deviation of loss development factors, based on historical data.

This loss payment model is able to provide realistic loss payments as long as there have been no significant structural changes in the loss history. However, if for an accident year $t_1 \leq t_0$ a high percentage of ultimate claim amount had been paid out in one of the first development years $t_2 \leq t_0 - t_1$, this approach would increase the reserve due to higher development factors leading to overestimation of outstanding payments. Consequently, single large losses should be treated separately. Sometimes changes in law affect insurance companies seriously. Such unpredictable structural changes are an important risk. A well-known example are health problems caused by buildings contaminated with asbestos. These were responsible for major losses in liability insurance. Such extreme cases should perhaps be modelled by separate scenarios.

Ultimate loss amounts for accident years $t_1 \leq t_0$ were calculated as

$$Z_{t_1}^{\text{ult}} = \sum_{t=0}^{\tau} Z_{t_1,t} \tag{2.25}$$

The second type of loss payments are due to future accident years $t_1 \geq t_0 + 1$. the components determining total loss amounts in respect of these accident years have already been explained in Sections 2.3 and 2.4:

$$Z_{t_1}^{\text{ult}}(k) = \sum_{j=0}^{2} N_{t_1}^{j}(k) X_{t_1}^{j}(k) + b_{t_1}(k) \sum_{i=1}^{M_{t_1}} Y_{t_1,i}^{k} - R_{t_1}(k), \tag{2.26}$$

where

$N_{t_1}^{j}(k)$ = number of non-catastrophe losses in accident year t_1 for line of business k and renewal class j, see (2.17),

$X_{t_1}^{j}(k)$ = severity of non-catastrophe losses in accident year t_1 for line of business k and renewal class j, see (2.19),

$b_{t_1}(k)$ = market share of the company in year t_1 for line of business k,

M_{t_1} = number of catastrophes in accident year t_1, see (2.20),

$Y_{t_1,i}^{k}$ = severity of catastrophe i in line of business k in accident year t_1, see (2.22),

$R_{t_1}(k)$ = reinsurance recoverables; a function of the $Y_{t_1,i}^{k}$'s, depending on the company's reinsurance program.

It remains to model the incremental payments of these ultimate loss amounts over the development periods. Therefore, we simulated incremental percentages A_{t_1,t_2} of ultimate loss amount by using a beta probability distribution with parameters based on payment patterns of previous calendar years:

$$A_{t_1,t_2} = \begin{cases} B_{t_1,0} & \text{for } t_2 = 0, \\ B_{t_1,t_2}\left(1 - \sum_{t=0}^{t_2-1} A_{t_1,t}\right) & \text{for } t_2 \geq 1, \end{cases} \tag{2.27}$$

where

B_{t_1,t_2} = incremental loss payment due to accident year t_1 in development year t_2 in relation to the sum of remaining incremental loss payments pertaining to the same accident year
~ beta(α, β), α, $\beta > -1$.

Here α and β are chosen to yield

$$m_{t_1,t_2} = \mathbb{E}\left[B_{t_1,t_2}\right] = \frac{\alpha+1}{\alpha+\beta+2},$$
(2.28)
$$v_{t_1,t_2} = \mathrm{Var}\left(B_{t_1,t_2}\right) = \frac{(\alpha+1)(\beta+1)}{(\alpha+\beta+2)^2(\alpha+\beta+3)},$$

where

m_{t_1,t_2} = estimated mean value of incremental loss payment due to accident year t_1 in development year t_2 in relation to the sum of remaining incremental loss payments pertaining to the same accident year, based on $\frac{A_{t_1-1,t_2}}{\sum_{t=t_2}^{\tau} A_{t_1-1,t}}, \frac{A_{t_1-2,t_2}}{\sum_{t=t_2}^{\tau} A_{t_1-2,t}}, \ldots,$

v_{t_1,t_2} = estimated variance, based on the same historical data.

It can happen that $\alpha > -1$, $\beta > -1$ satisfying (2.28) do not exist. This means that the estimated variance reaches or exceeds the maximum variance $m_{t_1,t_2}(1 - m_{t_1,t_2})$ possible for a beta distribution with mean m_{t_1,t_2}. In this case, we resorted to a Bernoulli distribution for B_{t_1,t_2} because the Bernoulli distribution marks a limiting case of the beta distribution:

$$B_{t_1,t_2} \sim Be(m_{t_1,t_2}).$$

This approach limited the maximum variance to $m_{t_1,t_2}(1 - m_{t_1,t_2})$.

For each future accident year ($t_1 \geq t_0$) we finally calculated loss payments in development year t_2 by:

$$Z_{t_1,t_2} = A_{t_1,t_2} Z_{t_1}^{\mathrm{ult}}. \tag{2.29}$$

So far we have been dealing with the simulation of incremental claim payments due to an accident year. We still have to explain how we arrived at reserve estimates at each time during the projection period. For each accident year t_1 we estimated the ultimate claim amount in each development year t_2 through:

$$\hat{Z}_{t_1,t_2}^{\mathrm{ult}} = \prod_{t=t_2+1}^{\tau} (1+e^{\mu_t}) \sum_{t=0}^{t_2} Z_{t_1,t}, \tag{2.30}$$

where

μ_t = estimated logarithmic loss development factor for development year t, based on historical data,

$Z_{t_1,t}$ = simulated losses for accident year t_1, to be paid in development year t, see (2.24) and (2.29).

Note that (2.30) is an *estimate* at the end of calendar year $t_1 + t_2$, whereas (2.26) represents the *real* future value. Reserves in respect of accident year t_1 at the end of calendar year $t_1 + t_2$ are determined by the difference between estimated ultimate claim amount $\hat{Z}^{\text{ult}}_{t_1,t_2}$ and paid to date losses in respect of accident year t_1. Reserving risk materializes through variations of the difference between the simulated (real) ultimate claim amounts and the estimated values.

Similarly, at the end of calendar year $t_1 + t_2$ we got an estimate for *discounted* ultimate losses for each accident year t_1. Note that only future loss payments are discounted whereas paid to date losses are taken at face value:

$$\hat{Z}^{\text{ult, disc}}_{t_1,t_2} = \left(1 + e^{-R_{t_1+t_2,1}} e^{\mu_{t_2+1}} + \sum_{s=t_2+2}^{\tau} e^{-R_{t_1+t_2,s-t_2}} e^{\mu_s} \prod_{t=t_2+1}^{s-1} (1 + e^{\mu_t})\right) \sum_{t=0}^{t_2} Z_{t_1,t}, \tag{2.31}$$

where

$R_{t,T}$ = T year spot rate at time t, see (2.7),

μ_t = estimated logarithmic loss development factor for development year t, based on historical data,

$Z_{t_1,t}$ = simulated losses for accident year t_1, paid in development year t, see (2.24) and (2.29).

Interesting references on stochastic models in loss reserving are Christofides [8] and Taylor [40].

3. The Corporate Model: From Simulations to Financial Statements

As pointed out in Section 1.4, DFA is an approach to facilitate and help justify management decisions. These are driven by a variety of considerations: maximizing shareholder value, constraints imposed by regulators, tax optimization and rankings by rating agencies and analysts. Parties outside the company rely on financial reports in making decisions regarding their relationship with the company. Therefore, a DFA model has to bridge the gap between stochastic simulation of cash flows and financial statements (pro forma balance sheets and income statements). The accounting process helps organize cash flow simulations into a readily understood and consistent financial structure. This requires a substantial number of accrual items to be generated in order to develop accounting entries for the model's financial statements.

A DFA model has to allow for a statutory accounting framework if it wants to address solvency requirements imposed by regulators thoroughly. If the focus is on shareholder value the model should predominantly be concerned with economic values, implying, for example, assets being marked-to-market and all policy liabilities being discounted. While statutory accounting focuses on solvency and balance sheet, generally accepted accounting principles (GAAP) emphasize income statements and comparability between entities of different nature. Consequently, a perfect DFA model should, among other things, include different accounting frameworks (i.e. statutory, GAAP and economic). This increases implementation costs substantially. A less burdensome approach would be to concentrate on GAAP accounting taking into account solvency requirements by introducing them as constraints to the model where appropriate. Our DFA implementation focused on an economic perspective.

In order to keep the exposition simple and within reasonable size we will mention only some key relationships of the corporate model. A much more comprehensive description is given in Kaufmann [24].

One of the fundamental variables is (economic) surplus U_t, defined as the difference between the market value of assets and the market value of liabilities (derived by discounting loss reserves and unearned premium reserves). The amount of available surplus reflects the financial strength of an insurance company and serves as a measure for shareholder value. We consider a company as being insolvent once $U_t < 0$.

Change in surplus is determined by the following cash flows:

$$\Delta U_t = P_t + (I_t - I_{t-1}) + (C_t - C_{t-1}) - Z_t - E_t - (R_t - R_{t-1}) - T_t \tag{3.1}$$

where

P_t = earned premiums,

I_t = market value of assets (including realized capital gains in year t),

C_t = equity capital,

Z_t = losses paid in calendar year t,

E_t = expenses,

R_t = (discounted) loss reserves,

T_t = taxes.

Note that $C_t - C_{t-1}$ describes the result of capital measures like issuance of new equity capital or capital reduction.

We derived *earned* from *written* premiums. For each line of business, written premiums P_t^j for renewal class j should depend on change in loss trends, the position in the underwriting cycle and on the number of written exposures. This leads to written premium $\tilde{P}_t^j$ of

$$\tilde{P}_t^j = (1+\delta_t^P)(1+c_{m_{t-1},m_t})\frac{w_t^j}{w_{t-1}^j}\tilde{P}_{t-1}^j, \quad j=0,1,2, \tag{3.2}$$

where

δ_t^P = change in loss trends, see remarks after (2.11),

m_t = market condition in year t, see Section 2.5,

$c_{A,B}$ = constant that describes how premiums develop when changing from market condition A to B; $c_{A,B}$ can be estimated based on historical data,

w_t^0 = written exposure units for new business,

w_t^1 = written exposure units for renewal business, first renewal,

w_t^2 = written expoure units for renewal business, second and subsequent renewals.

Description of the calculation of initial values $\tilde{P}_{t_0}^j$ in (3.2) will be deferred to the paragraph subsequent to equation (3.4). Variables $c_{A,B}$ have to be available as input parameters at the start of the DFA analysis. When estimating the percentage change of premiums implied by changing from market condition A to B it seems plausible to assume that the final impact is zero if market conditions change back from B to A. This translates into $(1+c_{A,B})\,(1+c_{B,A}) = 1$. Also, the impact on premium changes triggered by changing from market condition A to B and from B to C afterwards should be the same as changing from A to C directly: $(1+c_{A,B})\,(1+c_{B,C}) = (1+c_{A,C})$. We assumed an autoregressive process of order 1, AR(1), for the modelling of exposure unit development:

$$w_t^j = (a^j + b^j w_{t-1}^j + \varepsilon_t^j)^+, \quad j=0,1,2, \tag{3.3}$$

where

$\varepsilon_t^j \sim \mathcal{N}(0,(\sigma^j)^2),\ \varepsilon_1^j, \varepsilon_2^j, \ldots$ i.i.d.,

a^j, b^j, σ^j = parameters that can be estimated based on historical data.

The initial values $w_{t_0}^j$ are known since they represent the current number of exposure units. Choosing parameter $b^j < 1$ ensures stationarity of the AR(1) process (3.3). When deriving parameters a^j and b^j, prior adjustments to historical data might be necessary if jumps in number of exposure units had occurred caused by acquisition or transfer of loss portfolios. We found it helpful to admit deterministic modelling of exposure growth as well in order to allow for these effects, which are mostly anticipated before changes in the composition of the portfolio become effective.

Setting premium rates based on knowledge of past loss experience and exposure growth as expressed in (3.2) leaves us still with substantial uncertainties with regard to the adequacy of premiums. These uncertainties are conveyed in the term *underwriting risk*. Note that written premiums represented by equation (3.2) would come close to be adequate if the realizations of all random variables referring to projection year t ($\delta_t^P, c_{m_{t-1},m_t}, w_t^j$) were known in advance and assuming adequacy of current premiums $\tilde{P}_{t_0}^j$. Unfortunately, premiums to be charged in year t have to be determined prior to the beginning of year t. Therefore, random variables in (3.2) have to be replaced by estimations in order to model written premiums P_t^j, which would be charged in projection year t.

$$P_t^j = \left(1+\hat{\delta}_t^P\right)\left(1+\hat{c}_{m_{t-1},m_t}\right)\frac{\hat{w}_t^j}{w_{t-1}^j}\tilde{P}_{t-1}^j, \quad j=0,1,2, \tag{3.4}$$

where we got the estimates via their expected values:

$\hat{\delta}_t^P = [1+a^X+b^X(a^I+b^I(ab+(1-a)r_{t-1}))]\,[1+a^F+b^F(a^I+b^I(ab+(1-a)r_{t-1}))]-1,$
see (2.11), (2.10), (2.9), (2.8) and (2.4).

$$\hat{c}_{m_{t-1},m_t} = \sum_{m=1}^{l(k)} p_{m_{t-1},m}\, c_{m_{t-1},m},$$

$l(k)$ = number of states for line of business k, see Section 2.5,

$p_{m_{t-1},m}$ = transition probability, see Section 2.5.

$\hat{w}_t^j$ = $a^j + b^j w_{t-1}^j$, see (3.3).

While (3.2) represents a random variable that describes (almost) adequate premiums, (3.4) is the expected value of this random variable representing actually written premiums. Note that the time index $t = t_0$ refers to the year prior to the first projection year. By combining (3.2) and (3.4) we deduce that the initial values $\tilde{P}_{t_0}^j$ can be calculated via $P_{t_0}^j$:

$$\tilde{P}_{t_0}^j = \frac{1+\delta_{t_0}^P}{1+\hat{\delta}_{t_0}^P}\,\frac{1+c_{m_{t_0-1},m_{t_0}}}{1+\hat{c}_{m_{t_0-1},m_{t_0}}}\,\frac{w_{t_0}^j}{\hat{w}_{t_0}^j}\,P_{t_0}^j, \quad j=0,1,2. \tag{3.5}$$

$P_{t_0}^j$ represent written premiums charged for the last year and still valid just before the start of the first projection year. We assumed that premiums $P_{t_0}^j$ were adequate and based on established premium principles allowing for the cost of capital to be earned. An alternative of setting starting values according to (3.5) would be to use business plan data instead. This is an approach applicable at several places of the model.

By using written premiums $P_t^j(k)$ as given in (3.4) where the index k denotes line of business, we got the following expression for total earned premiums of all lines and renewal classes (see explanation in Section 2.3) combined:

$$P_t = \sum_{k=1}^{l} \sum_{j=0}^{2} a_t^j(k) P_t^j(k) + (1 - a_{t-1}^j(k)) P_{t-1}^j(k), \tag{3.6}$$

where

$a_t^j(k)$ = percentage of premiums earned in year written, estimated based on historical data.

We restricted ourselves to modelling only the most important asset classes, i.e. fixed income type investments (e.g. bonds, policy loans, cash), stocks, and real estate. Modelling of stock returns has already been mentioned in Section 2.2, future prices of fixed income investments can be derived from the generated term structure explained in Section 2.1. Our approach of modelling real estate was very similar to the stock return model of Section 2.2.

Future investment profits depend not only on the development of market values of assets currently on the balance sheet but also on decisions how new funds will be reinvested. In order to build a DFA model that really deserves to be called *dynamic* we should account for potential changes of asset allocation in future years compared to a pure static approach that keeps the asset allocation unchanged. This requires defining investment rules depending on specific economic conditions.

Capital measures $\Delta C_t = C_t - C_{t-1}$ were modelled as additions or deductions from surplus depending on a target reserves-to-surplus ratio. A purely deterministic approach that increased or decreased equity capital by a certain amount at specific times would have been an alternative.

Aggregate loss payments in projection year t were calculated based on variables defined in Section 2.6:

$$Z_t = \sum_{k=1}^{l} \sum_{t_2=0}^{\tau(k)} Z_{t-t_2,\, t_2}(k), \tag{3.7}$$

where

$Z_{t-t_2,\, t_2}(k)$ = losses for accident year $t - t_2$, paid in development year t_2; see (2.24) and (2.29),

$\tau(k)$ = ultimate development year for this line of business,

k = line of business.

We used a simple approach for modelling general expenses E_t. They were calculated as a constant plus a multiple of written exposure units $w_t^j(k)$. The appropriate intercept $a^E(k)$ and slope $b^E(k)$ were determined by linear regression:

$$E_t = \sum_{k=1}^{l} \left(a^E(k) + b^E(k) \sum_{j=0}^{2} w_t^j(k) \right). \tag{3.8}$$

For loss reserves R_t we got

$$R_t = \sum_{k=1}^{l} \sum_{t_2=0}^{\tau(k)} \left(\hat{Z}_{t-t_2,t_2}^{\text{ult,disc}}(k) - \sum_{s=0}^{t_2} Z_{t-t_2,s}(k) \right), \tag{3.9}$$

where

$\hat{Z}_{t-t_2,t_2}^{\text{ult,disc}}(k)$	=	estimation in calendar year t for discounted ultimate losses in accident year $t-t_2$; see (2.31),
$Z_{t-t_2,s}(k)$	=	losses for accident year $t-t_2$, paid in development year s; see (2.24) and (2.29),
$\tau(k)$	=	ultimate development year,
k	=	line of business.

An important variable to be considered are taxes, T_t, because many management decisions are tax driven. The proper treatment of taxes depends on the accounting framework. We used a rather simple tax model allowing for current income taxes only, i.e. neglecting the possibility of deferred income taxes for GAAP accounting.

4. DFA in Action

The aim of this section is to give an example of potential applications of DFA. Figure 4.1 displays the model logic of the approach introduced in this paper in graphical format. By providing a simple example we will show how to analyze surplus and ruin probabilities. It was not intended to describe a specific effect when using the parameters given below. The parameters were made up, i.e. they were not based on a real case.

Simplifying assumptions

- Only one line of business.
- New business and renewal business are not modelled separately.
- Payment patterns are assumed to be deterministic.
- No transaction costs.
- No taxes.
- No dividends paid.

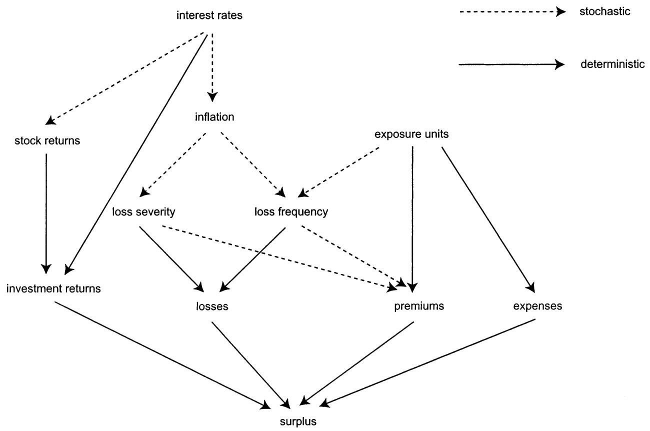

FIGURE 4.1: Schematic description of the modelling process: stochastic and deterministic influences on surplus.

Model choices

- Number of non-catastrophe losses ~ NB (154, 0.025).
- Mean severity of non-catastrophe losses ~ Gamma (9.091, 242), inflation-adjusted.
- Number of catastrophes ~ Pois (18).
- Severity of individual catastrophes ~ lognormal (13, 1.5^2), inflation-adjusted.
- Optional excess of loss reinsurance with deductible 500000 (inflation-adjusted), and cover ∞.
- Underwriting cycles: 1 = weak, 2 = average, 3 = strong. State in year 0: 1 (weak). Transition probabilities: p_{11} = 60%, p_{12} = 25%, p_{13} = 15%, p_{21} = 25%, p_{22} = 55%, p_{23} = 20%, p_{31} = 10%, p_{32} = 25%, p_{33} = 65%.
- All liquidity is reinvested. There are only two investment possibilities:
  1) buy a risk-free bond with maturity one year,
  2) buy an equity portfolio with a fixed beta.
- Market valuation: assets and liabilities are stated at market value, i.e. assets are stated at their current market values, liabilities are discounted at the appropriate term spot rate determined by the model.

Model parameters

- Interest rates, see (2.4): $a = 0.25$, $b = 5\%$, $s = 0.1$, $r_1 = 2\%$.
- General inflation, see (2.8): $a^I = 0\%$, $b^I = 0.75$, $\sigma^I = 0.025$.
- No inflation impacting the number of claims.
- Inflation impacting severity of claims, see (2.10): $a^X = 3.5\%$, $b^X = 0.5$, $\sigma^X = 0.02$.
- Stock returns, see (2.14), (2.15) and (2.16): $a^M = 4\%$, $b^M = 0.5$, $\beta_t^S \equiv 0.5$, $\sigma = 0.15$.
- Market share: 5%.

- Expenses: 28.5% of written premiums.
- Premiums for reinsurance: 175 000 p.a. (inflation-adjusted).

Historical data

- Written premiums in the last year: 20 million.
- Initial surplus: 12 million.

Strategies considered

- Should the company buy reinsurance coverage or not?
- How should the reinvestment of excess liquidity be split between fixed income instruments and stocks?

Projection period

- 10 years (yearly intervals).

Risk and return measures

- Return measure: expected surplus $\mathbb{E}[U_{10}]$.
- Risk measure: ruin probability, defined as $\mathbb{P}[U_{10} < 0]$.

We ran this model 10 000 times for the twelve strategies summarized in Figure 4.2. The first three rows represent a fixed asset allocation. The remaining ones are characterized by an upper limit for the amount of money allowed to be invested in bonds. The amount exceeding this limit is invested in stocks. For each strategy we evaluated the expected surplus and the probability of ruin. Figure 4.3 rules out only one strategy definitely, based on the selected risk and return measures: strategy 1b has lower return but higher risk than strategy 6a.

If we replace the return measure "expected surplus" by the median surplus, and evaluate the same twelve strategies, we get a completely different picture.

		a with reinsurance	b without reinsurance
1	100 % bonds 0 % stocks	23.17 mio. 0.49 %	23.29 mio. 1.15 %
2	50 % bonds 50 % stocks	25.28 mio. 2.14 %	25.51 mio. 2.48 %
3	0 % bonds 100 % stocks	27.17 mio. 9.69 %	27.70 mio. 10.13 %
4	$\leq$ 5 mio. bonds rest stocks	26.48 mio. 6.08 %	26.79 mio. 6.52 %
5	$\leq$10 mio. bonds rest stocks	25.74 mio. 3.64 %	26.06 mio. 4.49 %
6	$\leq$20 mio. bonds rest stocks	24.62 mio. 0.90 %	24.95 mio. 1.65 %

FIGURE 4.2: Simulated expected surplus and ruin probability for the evaluated strategies.

Figure 4.4 shows that by choosing the median surplus as return measure and ruin probability as risk measure all six strategies with a ruin probability above 3% (i.e. strategies 3a, 3b, 4a, 4b, 5a and 5b) are clearly outperformed by the strategies 2a and 2b, where half of the money is invested in bonds and the other half in stocks.

An advantage of median surplus is the fact that one can easily calculate confidence intervals for this return measure. In Figure 4.5 we plotted confidence intervals, based on the 10 000 simulations performed. These intervals should be interpreted as 95% confidence intervals for ruin probability given a specific strategy and 95% confidence intervals for median surplus given a specific strategy. Note that Figure 4.5 does not attempt to give joint confidence areas. Furthermore it is important to be aware of the fact that a 95% confidence interval for median surplus does not mean that 95% of the simulations at the end of the projection period result in an amount of surplus that lies in this

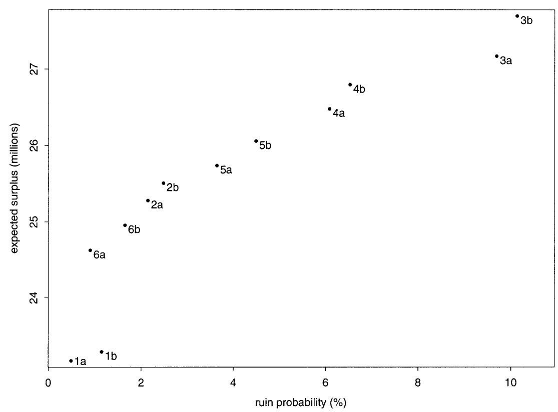

FIGURE 4.3: Graphical comparison of ruin probabilities and expected surplus for selected business strategies.

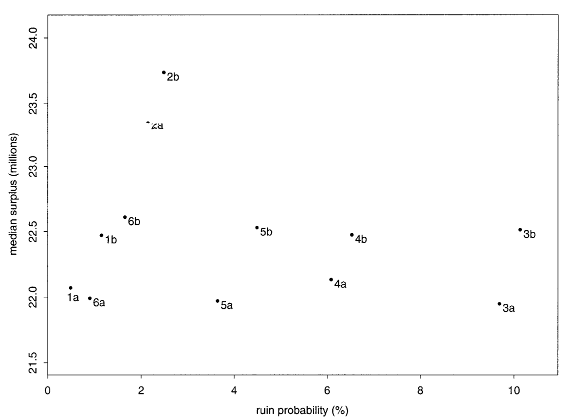

FIGURE 4.4: Graphical comparison of ruin probabilities and median surplus for selected business strategies.

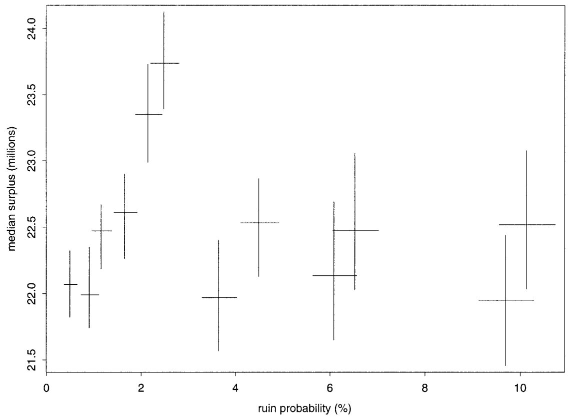

FIGURE 4.5: 95% confidence intervals for ruin probability and median surplus, based on 10 000 simulations for each strategy.

interval. The correct interpretation is that given our observed sample of 10 000 simulations, the probability for median surplus lying in this interval is 95%.

5. SOME REMARKS ON DFA

5.1. Discussion Points

This introductory paper discussed only the most relevant issues related to DFA modelling. Therefore, we would like to mention briefly some additional points without necessarily being exhaustive.

5.1.1. Deterministic Scenario Testing

In Section 1 we mentioned the superiority of DFA compared to deterministic scenario testing. This does not imply that the latter method is useless at all. On the contrary, deterministic scenario testing is a very useful thing, in particular when it comes to assess the impact of extreme events at pre-defined dates or when specific macroeconomic influences are to be evaluated. It is a very useful feature of a DFA tool being able to switch off stochasticity and return to deterministic scenarios.

5.1.2. Macroeconomic Environment

In life insurance financial modelling interest rates are often considered to be the only macroeconomic factor affecting the values of assets and liabilities. Hodes, Feldblum and Neghaiwi [21] have pointed out that in nonlife insurance, interest rates are only one of various other factors affecting liability values. In Worker's Compensation in the US, for instance, unemployment rates and industrial capacity utilization have greater effects on loss costs than interest rates have, while third-party motor claims are correlated with total volume of traffic and with sales of new cars. Although rarely done it might be worthwhile modelling specific macroeconomic drivers like industrial capacity utilization or traffic volume separately. This would require a foregoing econometric analysis of the dynamics of particular factors.

5.1.3. Correlations

DFA is able to allow for dependencies between different stochastic variables. Before starting to implement these dependencies one should have a sound understanding of existing dependencies within an insurance enterprise. Estimating correlations from historical (loss) data is often not feasible due to aggregate figures and structural changes in the past, e.g. varying deductibles, changing policy conditions, acquisitions, spin-offs, etc. Furthermore, recent research, see for example Embrechts, McNeil and Straumann [17] and [18], and Lindskog [28], suggests that linear correlation is not appropriate to model dependencies between heavy-tailed and skewed risks.

We suggest modelling dependencies implicitly, as a result of a number of contributory influences, for example, catastrophes that impact more than one line of business or interest rate changes affecting only specific lines. The majority of these relations should be implemented based on economic and actuarial wisdom, see for instance Kreps [26].

5.1.4. Separate Modelling of New and Renewal Business

In the model outlined in this paper we allowed for separate modelling of new and renewal business, see Section 2.3. Hodes, Feldblum and Neghaiwi [21] pointed out that this makes perfectly sense due to different stochastic behaviour of the respective loss portfolios. Furthermore, having this split allows a deeper analysis of value drivers within the portfolio and marks an important step towards determining an *appraised value* for a nonlife insurance company.

5.1.5. Model Validation

What is finally a good DFA model and what is not? Experience, knowledge and intuition of users from actuarial, economic and management side play a dominant role in evaluating a DFA model. A danger in this respect might be that non-intuitive results could be blamed on a bad model instead of wrong assumptions. A further possibility to evaluate a model is to test results coming out of the DFA model against empirical results. This will only be feasible in very few restricted cases because it would require keeping track of data for several years. However, model validation should deserve more attention. This needs to be recommended in particular to those practitioners dealing with software vendors of DFA tools who do not intend to justify their decision of buying an expensive DFA product by referring to the software design only.

5.1.6. Model Calibration

We have already touched on this at several places and pointed to its importance within a DFA analysis. However sophisticated a DFA tool or model might be, it has to be fed with data and parameter values. Studies have shown that the major part of a DFA analysis had been devoted to this exercise.

Usually, the calibration part is an ongoing process during the course of an analysis in order to fine-tune the model.

5.1.7. Interpretation of Output
We mentioned in Section 1.5 that the interpretation process of DFA output follows very often traditional patterns, e.g. efficient frontier analysis, which might lead to false or at least questionable conclusions, see Cumberworth, Hitchcox, McConnell and Smith [10]. Another example showing how critical interpretation of results can be is this: A net present value (NPV) analysis applied to model office cash flows can generate or destroy a huge amount of shareholder value by making slight changes to CAPM assumptions, which are often used for determining the discount rate. A way to keep feet on sound economic ground and simultaneously remove a great deal of arbitrariness is through resorting to *deflators*, see Jarvis, Southall and Varnell [23]. The use of this concept, originating in the work of Arrow and Debreu, has been promoted by Smith [39] and is further discussed in Bühlmann, Delbaen, Embrechts and Shiryaev [7]. The cited references might be evidence for growing awareness that our toolbox for interpreting and understanding DFA results needs to be renovated in order to enhance the use of DFA.

5.2. Strength and Weaknesses of DFA

DFA models provide generally deeper insight into risks and potential rewards of business strategies than scenario testing can do. DFA marks a milestone towards evaluating business strategies when compared to old-style analysis of considering only key ratios. DFA is virtually the only feasible way to model an entire nonlife operation on a cash flow basis. It allows for a high degree of detail including analysis of the reinsurance program, modelling of catastrophic events, dependencies between random elements, etc. DFA can meet different objectives and address different management units (underwriting, investments, planning, actuarial, etc.).

Nevertheless, it is worth mentioning that a DFA model will never be able to capture the complexity of the real-life business environment. Necessarily, one has to restrict attention during the model building process to certain features the model is supposed to reflect. However, the number of parameters which have to be estimated beforehand and the number of random variables to be modelled even within medium-sized DFA models contribute a big deal of process and parameter risk to a DFA model. Furthermore one has to be aware that results will strongly depend on the assumptions used in the model set-up. A critical question is: How big and sophisticated should a DFA model be? Everything comes at a price and a simple model that can produce reasonable results will probably be preferred by many users due to growing reluctance of using non-transparent “black boxes”. In addition, smaller models tend to be more in line with intuition, and make it easier to assess the impact of specific variables. A good understanding and control of uncertainties and approximations is vital to the usefulness of a DFA model.

5.3. Closing Remarks

We wanted to give an introduction into DFA by hinting to pitfalls and emphasizing important issues to be taken into account in the modelling process. Our intention was to provide the uninformed reader with a simple DFA approach enabling these readers to implement DFA using our approach as a kind of reference model. Many commercial DFA tools are roughly structured as the model outlined in this paper. Specific concepts and concrete implementation of the model components are often different. We are absolutely aware that there are numerous alternatives to each of the sub-models introduced in this paper. Some of them might be much more powerful or flexible than our approach. We wanted to provide a framework leaving it up to the reader to complete the DFA house by making adjustments or amendments at his/her discretion. Although we did not necessarily target the DFA experts our exposition might have also served to give an impression of the complexity of a fully fledged DFA model.

Acknowledgement

We would like to thank Paul Embrechts, Peter Blum and the anonymous referees for numerous comments on an earlier version of the paper. We also benefited substantially from discussions on DFA with Allan Kaufman and Stavros Christofides.

References

1. Ahlgrim K.C., D'Arcy S.P. and Gorvett R.W. (1999) Parametrizing Interest Rate Models, *Casualty Actuarial Society Forum*, 1-50
2. Artzner P., Delbaen F., Eber J. and Heath D. (1997) Thinking Coherently, *RISK* **10**, 68-71.
3. Artzner P., Delbaen F., Eber J. and Heath D. (1999) Coherent Measures of Risk, *Mathematical Finance* **9(3)**, 203-228.
4. Björk T. (1996) Interest Rate Theory. In *Financial Mathematics* (ed. W. Runggaldier), Lecture Notes in Mathematics **1656**, 53-122, Springer, Berlin.
5. Blum P., Dacorogna M., Embrechts P., Neghaiwi T. and Niggli H. (2001) *Using DFA for Modelling the Impact of Foreign Exchange Risks on Reinsurance Decisions*, Paper presented at the Casualty Actuarial Society 2001 Reinsurance Meeting on Using Dynamic Financial Analysis to Optimize Ceded Reinsurance Programs and Retained Portfolios, Washington D.C., July 2001. Available as ETH Zürich Preprint.
6. Brennan M.J. and Schwartz E.S. (1982) An Equilibrium Model of Bond Pricing and a Test of Market Efficiency, *Journal of Financial and Quantitative Analysis* **17**, 301-329.
7. Bühlmann H., Delbaen F., Embrechts P. and Shiryaev A.N. (1998) On Esscher Transforms in Discrete Finance Models, *ASTIN Bulletin* **28(2)**, 171-186.
8. Christofides S. (1990) Regression Models Based on Log-Incremental Payments, *Claims Reserving Manual* **2**, Institute of Actuaries, London.
9. Cox J.C., Ingersoll J.E. and Ross S.A. (1985) A Theory of the Term Structure of Interest Rates, *Econometrica* **53**, 385-407.
10. Cumberworth M.P., Hitchcox A.M., McConnell W.M. and Smith A.D. (1999) Corporate Decisions in General Insurance: Beyond the Frontier, available from http://www.actuaries.org.uk/library/sessional_meeting_papers.html.

11. D'Arcy S.P. and Doherty N. (1989) The Aging Phenomenon and Insurance Prices, *Proceedings of the Casualty Actuarial Society* **76**, 24-44.
12. D'Arcy S.P. and Doherty N. (1990) Adverse Selection, Private Information and Lowballing in Insurance Markets, *Journal of Business* **63**, 145-164.
13. D'Arcy S.P., Gorvett R.W., Herbers J.A., Hettinger T.E., Lehmann S.G. and Miller M.J. (1997) Building a Public Access PC-Based DFA Model, *Casualty Actuarial Society Forum*, 1-40.
14. D'Arcy S.P., Gorvett R.W., Hettinger T.E. and Walling R.J. (1998) Using the Public Access DFA Model: A Case Study, *Casualty Actuarial Society Forum*, 55-118.
15. Daykin C.D., Pentikäinen T. and Pesonen M. (1994) *Practical Risk Theory for Actuaries*, Chapman & Hall, London.
16. Embrechts P., Klüppelberg C. and Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.
17. Embrechts P., McNeil A.J. and Straumann D. (1999) Correlation: Pitfalls and Alternatives, *RISK* **12(5)**, 69-71.
18. Embrechts P., McNeil A.J. and Straumann D. (1999) Correlation and Dependence in Risk Management: Properties and Pitfalls, *Preprint ETH Zürich*, available from http://www.math.ethz.ch/~embrechts.
19. Feldblum S. (1996) Personal Automobile Premiums: An Asset Share Pricing Approach for Property/Casualty Insurance, *Proceedings of the Casualty Actuarial Society* **83**, 190-296.
20. Heath D., Jarrow R. and Morton A. (1992) Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation, *Econometrica* **60**, 77-105.
21. Hodes D.M., Feldblum S. and Neghaiwi A.A. (1999) The Financial Modeling of Property-Casualty Insurance Companies, *North American Actuarial Journal* **3(3)**, 41-69.
22. Ingersoll J.E. (1987) *Theory of Financial Decision Making*, Rowman & Littlefield Studies in Financial Economics, New Jersey.
23. Jarvis S., Southall F.E. and Varnell E. (2001) Modern Valuation Techniques, *Staple Inn Actuarial Society*, available from http://www.sias.org.uk/progold.htm.
24. Kaufmann R. (1999) *DFA: Stochastische Simulation zur Beurteilung von Unternehmensstrategien bei Nichtleben-Versicherungen*, Master Thesis, ETH Zürich.
25. Klett, R. (1994) *Asset-Liability-Management im Lebensversicherungsbereich*, Master Thesis, University of Freiburg.
26. Kreps, R.E. (2000) A Partially Comonotonic Algorithm for Loss Generation, *Proceedings of XXXIst International ASTIN Colloquium*, 165-176, Porto Cervo, Italy.
27. Lamberton D. and Lapeyre B. (1996) *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall, London.
28. Lindskog F. (2000) *Modelling Dependence with Copulas and Applications to Risk Management*, Master Thesis, ETH Zürich.
29. Longstaff F.A. and Schwartz E.S. (1992) Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model, *Journal of Finance* **47**, 1259-1282.
30. Lowe S.P. and Stanard J.N. (1997) An Integrated Dynamic Financial Analysis and Decision Support System for a Property Catastrophe Reinsurer, *ASTIN Bulletin* **27(2)**, 339-371.
31. Mack T. (1997) *Schadenversicherungsmathematik*, Verlag Versicherungswirtschaft E.V., Karlsruhe.
32. Markowitz H.M. (1959) *Portfolio Selection: Efficient Diversification of Investments*, John Wiley, New York.
33. McNeil A.J. (1997) Estimating the Tails of Loss Severity Distributions using Extreme Value Theory, *ASTIN Bulletin* **27(1)**, 117-137.
34. Modigliani M. and Miller M. (1958) The Cost of Capital, Corporation Finance, and the Theory of Investment, *American Economic Review* **48**, 261-297.
35. Musiela M. and Rutkowski M. (1998) *Martingale Methods in Financial Modelling*, 2nd edition, Springer, Berlin.
36. Rogers L.C.G. (1995) Which Model for Term-Structure of Interest Rates Should One Use? In *Mathematical Finance, IMA Volume* **65**, 93-116, Springer, New York.
37. Schnieper R. (1997) Capital Allocation and Solvency Testing, *SCOR Notes*, 55-104.

38. SCHNIEPER R. (1999) Solvency Testing, *Mitteilungen der Schweizerischen Aktuarvereinigung*, 11-45.
39. SMITH A.D. (1996) How Actuaries Can Use Financial Economics, *British Actuarial Journal* **2(V)**, 1057-1193.
40. TAYLOR G.C. (2000) *Loss Reserving: An Actuarial Perspective*, Kluwer Academic Publishers, Boston.
41. WALLING R.J., HETTINGER T.E., EMMA C.C. and ACKERMAN S. (1999) Customizing the Public Access Model Using Publicly Available Data, *Casualty Actuarial Society Forum*, 239-266.
42. WILKIE A.D. (1995) More on a Stochastic Asset Model for Actuarial Use, *British Actuarial Journal* **1(V)**, 777-964.
43. WISE A.J. (1984) The Matching of Assets to Liabilities, *Journal of the Institute of Actuaries* **111**, 445-485.
44. WOLL R.G. (1987) Insurance Profits: Keeping Score, *Financial Analysis of Insurance Companies, Casualty Actuarial Society Discussion Paper Program*, 446-533.

ROGER KAUFMANN
RiskLab
Department of Mathematics
ETH Zentrum
CH-8092 Zürich
Switzerland
kaufmann@math.ethz.ch

ANDREAS GADMER
Zürich Kosmos Versicherungen
Schwarzenbergplatz 15
A-1015 Wien
Austria
andreas.gadmer@zurich.com

RALF KLETT
Zurich Financial Services
Mythenquai 2
CH-8022 Zürich
Switzerland
ralf.klett@zurich.com

BOOK REVIEWS

Hartmut MILBRODT, Manfred HELBIG (1999): *Mathematische Methoden der Personenversicherung*. de Gruyter. IBSN 3-11-014226-0

The book “Mathematische Methoden der Personenversicherung” by Hartmut Milbrodt and Manfred Helbig is a major textbook about life insurance mathematics and has the ambitious aim to cover a large part of the classic and modern life insurance mathematics in German. It is aimed at actuarial students, life insurance professionals and at research fellows.

In order to reach this aim, the monograph has over 600 pages and 13 chapters:

- Versicherungsmathematik: Teil der Versicherungswissenschaft
- Elementare Finanzmathematik: Der Zins als Rechnungsgrundlage
- Ausscheideordnungen in der Lebensversicherung
- Stochastische Prozesse in der Personenversicherung
- Versicherungsleistungen in der Lebensversicherung
- Versicherungsleistungen in der allgemeinen Personenversicherung
- Berechnung erwarteter Barwerte spezieller Versicherungsleistungen mittels Kommuationszahlen
- Prämien
- Das Deckungskapital einer Versicherung eines unter einem einzigen Risiko stehenden Lebens
- Das Deckungskapital in der allgemeinen Personenversicherung
- Überschuss und Überschussanalyse in der Lebensversicherung
- Mathematischer Anhang.

From the above table of contents, it is seen that this book covers a large amount of things an actuary in a life insurance has to know such as commutation functions, smoothing of moralities, bonus schemes and multi-state model for life insurance. From this point of view, the book is necessary for each library. A particular highlight of this book is the treatment of markov models in life insurance in a very general way. The theory is as well illustrated by practical examples. On the other hand, the book is rather long and not as concise as for example “Life insurance mathematics” by Hans Gerber.

One reason for being so long stems from the aim of the authors to present all theorems in the most general framework. Therefore the definitions, propositions and theorems become rather involved and it is possible get lost. The exercises are either very theoretical (mathematical) or bound to earth and so there is something for every type of reader. The solutions are unfortunately missing.

ASTIN BULLETIN, Vol. 31, No. 1, 2001, pp. 251-254

On the other hand, this book is unique because it tries to present the traditional and the modern life insurance mathematics within one book and therefore I think that is in particular helpful for people who want to know both types of life insurance mathematics.

MICHAEL KOLLER

G.E. WILLMOT AND X. SHELDON LIN (2000): *Lundberg Approximations for Compound Distributions with Insurance Applications*. Springer Lecture Notes in Statistics, 156. ISBN 0 387 95135 0.

Contents:

In its broadest interpretation, one can say that Lundberg approximations yield exponential inequalities and first order asymptotic expansions for compound distributions. Typical applications include ruin estimation in risk theory and approximations for the total claim amount over a given period of time. Similar problems occur in dam theory, queueing theory and reliability. The present text mainly uses techniques from the latter field to augment the classical insurance results. The various chapters typically start with some general results on the relevant topic; these results are then exemplified under specific distributional assumptions. Though the original Lundberg approximations were established for short-tailed distributions (as claim size, say), also the long-tailed case (like the Pareto) is discussed.

The text is well written; proofs and examples are given very much in detail. Consequently, the text can be used to augment a course on risk theory for instance through the discussion of specific examples

PAUL EMBRECHTS

J. Grandell: *Mixed Poisson Processes.* Chapman & Hall, London, 1997, 260 pages, ISBN 0 412 78700 8.

Mixed Poisson distributions and processes can loosely be regarded as Poisson distributions or processes with random intensity parameters. The distributions of these parameters are called *structure distributions*. It is surprising that such a simple construction has a lot of applications and serves as a source for further generalizations. By author's words, "the present book can be looked upon as a detailed survey, and contains no essential new results". One can agree with these modest words only on the understanding that the author gave a deep insight in the topic and related fields, provided many examples and counter-examples, historical remarks, and a comprehensive bibliography resulting in an excellent book.

In order to feel a flavour of the book, let us briefly consider its contents. Chapter 1 informally introduces readers into the subject. It contains relevant references and comments about the history of the problem. The mixed Poisson distribution is accurately defined in Chapter 2. Its various properties (e.g., the infinite divisibility) and relationships with other distributions are examined. Chapter 3 contains a mathematical background: point and Markov processes, martingales. In Chapter 4, the mixed Poisson processs is introduced, its basic properties are established, and relevant examples are given. As the author indicates, this chapter "is, to a great extent, a slightly (this adjective seems not to be adequate – *V.K.*) modernized summary of Lundberg's work" [*On random processes and their application to sickness and accident statistics*, 1940]. Various random processes such as infinitely divisible, Hoffman, Yule, birth, Pólya, and others are considered in the light of their relations to mixed Poisson processes. Chapter 5 is of special theoretical and applied interest. It is devoted to Cox, Gauss-Poisson, and mixed renewal processes regarded as important generalizations of mixed Poisson processes that can be viewed as approximations of a wide class of point processes. The emphasis is placed on constructive definitions of these processes. In particular, the author considers the *thinning* allowing to characterize the Cox and Gauss-Poisson processes. Various characterizations of mixed Poisson processes are given in Chapter 6. They are stated within sets of birth, stationary point, and general point processes. Chapter 7 deals with certain aging properties of the structure distributions. These properties are used in Chapter 8 for bounds, asymptotic formulae, and recursive evaluation of mixed Poisson distributions. The last Chapter 9 is devoted to applications to risk business with the emphasis on ruin probabilities, where contribution of the author is outstanding. Readers can also find there other interesting topics, e.g., associated with subexponential distributions.

This compact book is well-balanced as it combines rigorous mathematical treatments with informal discussions. It brings together many facts published in journals and other issues and contains a comprehensive bibliography on the subject and related topics. Certainly, it will serve as a valuable source of facts and inspiration for actuaries, applied mathematicians, students, and researchers.

Vladimir Kalashnikov

UNIVERSITY OF TORONTO

DEPARTMENT OF STATISTICS

The Department of Statistics invites applications for a tenure-stream position in Actuarial Science, rank open. The position is to begin September 1, 2001 or as soon as possible thereafter. Duties will include teaching courses in the actuarial science program at both the undergraduate and graduate level, conducting research in actuarial science, and service to the professional actuarial associations. Qualifications required are a Ph.D. in actuarial science, statistics, mathematics or a related area, professional accreditation in the CIA, SOA or CAS, and an active research program. Salary and rank are commensurate with experience.

We will begin reviewing applications on June 30, 2001 and continue until the position is filled.

Letters of application with curriculum vitae should be sent to

Professor Nancy Reid
Department of Statistics
University of Toronto
100 St. George Street, 6th Floor
Toronto, Ontario CANADA M5S 3G3
reid@utstat.utoronto.ca

Applications at the rank of Assistant Professor should arrange to have three letters of reference sent under separate cover to Professor Reid. Other applicants should include the names of three referees with their letter of application.

Information on the Actuarial Science program and the Department of Statistics at the University of Toronto, is available on the department's web page at www.utstat.toronto.edu.

The University of Toronto is strongly committed to diversity within its community. The University especially welcomes applications from visible minority group members, women, Aboriginal persons, persons with disabilities, and others who may contribute to further diversification of ideas.