A STIN
B U L L E T I N

A Journal of the International Actuarial Association

CONTENTS

EDITORIAL AND ANNOUNCEMENTS

Editorial 133
The XXVIIth ASTIN Colloquium in Copenhagen, Denmark 135

ARTICLES

S. HAASTRUP, E. ARJAS
Claims Reserving in Continuous Time; A Nonparametric Bayesian Approach 139

M. JACQUES
On the Hedging Portfolio of Asian Options 165

W. HÜRLIMANN
Improved Analytical Bounds for Some Risk Quantities 185

J. DHAENE, M. GOOVAERTS
Dependency of Risks and Stop-Loss Order 201

K.-H. WALDMANN
Modified Recursions for a Class of Compound Distributions 213

B. SUNDT, J. DHAENE
On Bounds for the Difference between the Stop-Loss Transforms of Two Compound Distributions 225

WORKSHOPS

W. NEUHAUS
Optimal Estimation under Linear Constraints 233

K.D. SCHMIDT, A. SCHNAUS
An Extension of Mack’s Model for the Chain Ladder Method 247

Book Review 263
Announcements 265
Donation of Actuarial Books 271
Actuarial Vacancy 273

Ceuterick
EDITORIAL POLICY

_Austin 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.

_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

_Astin Bulletin_ is published and printed for ASTIN by Ceuterick s.a., Brusselsestraat 153, B-3000 Leuven, Belgium.

All queries and communications concerning subscriptions, including claims and address changes, and concerning back issues should be sent to Ceuterick.

The current subscription or back issue price per volume of 2 issues including postage is BEF 2 500.

Back issues up to issue 10 (= up to publication year 1979) are available for half of the current subscription price.

INDEX TO VOLUMES 1-20

The Cumulative Index to Volumes 1-20 is also published for ASTIN by Ceuterick at the above address and is available for the price of BEF 400.

Copyright © 1996 Ceuterick
I have been a Co-editor of *ASTIN Bulletin* for seven years, with special responsibility for AFIR-type papers. I have now handed over this responsibility to Dr Andrew Cairns, of Heriot-Watt University, Edinburgh. My former Editor colleagues have asked me to write a valedictory editorial, and I have great pleasure in doing so.

The last seven years have seen the formation of AFIR, the subsection of the IAA for the *Actuarial Approach to Financial Risks*, for what Hans Bühlmann has called “actuaries of the third kind”. Five AFIR Colloquia have taken place, and by the time this is published, the sixth in Nürnberg, will also have happened. Each has been exiting and interesting.

The number of papers of an AFIR type submitted to the *ASTIN Bulletin* has not been large, and the number published has been smaller. There may be several reasons for this: *ASTIN Bulletin* may still be seen as a journal for non-life-insurance mathematics, rather than for investment mathematics; there are many other journals devoted to investment mathematics or financial economics to which authors may prefer to submit papers; and it takes time for there to be a sufficient body of actuarial experts in this field who can make their own specifically actuarial contribution.

The AFIR Colloquia are open to all those who are interested, including non-actuaries. This has been of enormous benefit to the actuaries attending, because it has opened their eyes to the great amount of academic and practical literature in the field. A substantial proportion of papers has been contributed by the non-actuaries, and they have improved the quality of our discussions. But, since they are not members of the IAA, nor of AFIR, they do not normally receive or read *ASTIN Bulletin*, and may not think to contribute papers to it.

I may sound pessimistic, but this would be wrong: a new journal takes time to grow. The experience of the AFIR Colloquia, and other meetings, such as those of individual AFIR groups or investment sections in individual countries, shows that a great deal is happening. Many actuaries are now familiar with the basics of financial economics: mean-variance optimisation and all its developments; stochastic models for describing assets, both continuous and discrete; the pricing of options and other derivative securities; and the mathematics of stochastic calculus, time series, stochastic optimisation, stochastic programming, $\alpha$-stable (Lévy stable or stable Paretian or fractal) distributions, and other concepts which seven years ago were unknown, or seemed abstruse, difficult and irrelevant.

Actuarial education in several countries has also developed. Some of these topics have found their way into the formal education syllabuses of those actuarial associations that control their own educational programme (in North America, Australia and Britain, for example), and they may well have entered university actuarial syllabuses too. These developments have had the interesting effect of
pulling actuaries back from their comparative isolation into the mainstream of academic thought. A similar development has happened in the analysis of mortality statistics, and in the development of multi-state demographic models, such as models for AIDS, where actuaries have found themselves learning to talk the same language as other experts.

There are still die-hards. There are those who feel that actuaries should look after the liabilities, and others should concern themselves with the assets; there are those who are sceptical about the use of mathematical methods that they themselves do not understand; there are those who question some of the fundamental assumptions of, for example, utility theory or valuation using risk-neutral probabilities, and then attack the whole edifice when in my view they should see how the development would be changed if the basic assumptions were altered. But the continuing popularity of the AFIR Colloquia, and the high standard of papers presented, shows that the die-hards will not prevent the development of “actuaries of the third kind”.

On the other side, actuaries have begun to show that their traditional approach can also be of value in the world of financial economics. The concept that a risky enterprise needs risk capital (or solvency reserves) applies just as much to those who trade in derivatives as to insurance companies. Although reinsurance is not traded in the same sense as stock-exchange securities, many of the principles of a risk exchange to create an efficient portfolio of liabilities are similar, except that one wants a low mean cost of the liabilities (along with low variance or other risk measure), instead of a high mean return. Many insurance contracts contain implied options and can be valued using option-pricing principles and methodology. Not all securities are traded in a complete market (insurance contracts, reinsurance, pensions,...) is risk-neutral valuation principles may not apply, and one may need to consider utility-maximising principles. The list goes on.

Many problems are unsolved. Optimisation of investment objectives over multiple time periods can possibly be assisted by stochastic programming techniques, but these are very computer-intensive, and are not yet widespread; how to define the objective function to be optimised is not yet wholly clear. A reconciliation of continuous and discrete time-series models for investments remains incomplete; indeed there remains the argument between the theoretically nice concepts of efficient markets and rational expectations, and the empirically observed autoregressive models that may contradict the theories. There is the problem of how to optimise the investment strategy for a with-profits life insurance company, where the bonus distribution policy forms part of the strategy. And there may be the problem of how to construct an optimal investment strategy for the individual, taking into account all his/her assets and liabilities over his/her lifetime, most of which are not tradeable, or, like houses, are not infinitely divisible.

I look forward to many more numbers of ASTIN Bulletin containing AFIR-type papers, many more AFIR Colloquia, and continuing discussion with those who are interested in the application of financial economics within fields of interest to actuaries.

DAVID WILKIE
The 27th ASTIN Colloquium was held at the Moltkes Palæ in the very center of Copenhagen. More than 200 actuaries from 21 countries attended. The colloquium began on Sunday, 1st of September, with a “get-together party” at Assurandørenes Hus. The invited lectures and the working sessions took place on Monday, Wednesday and Thursday. Tuesday was reserved for an all-day excursion to Northern Zealand. The general assembly of ASTIN took place Wednesday afternoon. There CAS announced that Greg Taylor is to receive the Hachemeister price for his article *Modelling mortgage insurance claims distribution in the individual life model*, ASTIN Bull. Vol. 24, no. 1, 1994. Wednesday evening the Colloquium dinner was served at Restaurant Nimb in Tivoli. After the dinner the participants were invited to watch the traditional Tivoli fireworks. On that particular evening the letters “ASTIN” appeared as a part of the firework display.

The Colloquium was arranged by the Danish Actuarial Association. The members of the organizing committee were Lars Halling, Ulla Plesner, Jan Buschardt and Nikolaj Boysen. The scientific committee was formed by Ole Hesselager, Chresten Dengsøe, Nils Jespersen, Ulla Mønsted and Mette Rygaard.

The scientific program consisted of three invited lectures and thirty contributed papers. The proceedings, which include all contributed papers, were mailed to the participants prior to the Colloquium. Five speakers had signed up for the Speakers Corner. During the breaks computer packages were presented by C.R. Larsen, S. Bernegger, D. Pfeifer, O. Hesselager, and H. Panjer. In the following I shall give a brief review of the invited and contributed papers.

**S.P. Lowe & J.N. Stanard: An integrated dynamic financial analysis and decision support system for a property catastrophe reinsurer**

The authors presented a model for financial analysis. Both the assets risk and the insurance risk are modelled. In classical portfolio theory the standard deviation of the return is used as risk measure and thereby the Efficient Frontier is found. This approach is generalized by introducing a risk measure that combines assets and liability risk. Different risk measures are discussed, among these the “expected policy holder deficit”. The model is currently being used by a property catastrophe insurance company.

**N. Keidig, C. Andersen & P. Fledelius: The Cox regression model for survival data in non-life insurance: Description of claim occurrence and possibilities for experience based individual rating**

The Cox regression model is a standard tool in survival analysis. The model is used for studying the dependence of a hazard rate on both time and covariates.
authors apply the model to occurrence of claims in non-life insurance. The hazard rate is the occurrence intensity of an individual policy holder. It is taken to depend on calendar time, and on several covariates such as age, urbanization, and time elapsed since last claim. By an empirical study the authors show that there are dependencies between different covers (policies) on the same individual. If an individual has a high number of claims on one policy then the occurrence rate on his/her other policies will tend to be high too.

A. Kristiansen: On a system of minimum requirements for technical provisions in non-life insurance

In Norway the supervisory authorities have since 1992 stipulated minimum requirements for technical provisions in non-life insurance. Kristiansen presented relatively detailed models used to stipulate both minimum requirements for the overall technical provision, and minimum requirements for the loss provisions. The models allow for fluctuating claim frequencies, and variations in reinsurance structures. In Norway each non-life insurance company must have an actuary appointed by the supervisory authorities. From the audience it was asked whether it is reasonable to postulate minimum provisions, now that each company has its own expert.

Working session: Major hazards

A.M. Barfod & D. Lando price derivative contracts (e.g. options) on catastrophe losses. M. Rytgaard discusses in her first article the use of the empirical mean residual lifetime function. In her second article she presents techniques for calculating risk premiums for a certain type of reinsurance cover. Both T. Pentikäinen and Y. Romppainen address problems in credit insurance. Pentikäinen considers specific models, whereas Romppainen studies empirically the impact of booms and recessions in the national economy on credit insurance. J.D. Breen & E. Kromme consider calculation of stop-loss tariffs by use of simulation.

Working session: Profitability I

Working session: Solvency

B. Ajne discusses how to set aside solvency margin relatively to technical reserves and relatively to premium income. D.C.M. Dickson & A.D. Egídio dos Reis allow, in the classical surplus process, the insurer to borrow money in case of negative surplus. S. Haastrup and E. Arjas present a Bayesian model for continuous time claims reserving. E. Kremer applies threshold autoregressive models to claims reserving. C.R. Larsen approximates the distribution of a predictor (claim reserve) using Bootstrap methods. The approach does not depend on the reserving method used. F. Larsen, S. Monty & C. Clemmensen construct a dynamic loss reserving model. The model is a day by day model, and it allows for dynamic changes in both occurrences and reporting of claims. J. Paulsen and H.K. Gjessing find the ruin probability for a risk process with stochastic interest rate and stochastic inflation. B.L. Sandqvist presents an empirical non-parametric approach to prediction of reported but not settled claims. The work of Sandqvist is related to that of C.R. Larsen since they both work under a minimal set of assumptions.

Working session: Profitability II

P. De Angelis & N. Ettore D'Ortona discuss optimal reinsurance structures. A. Deis presents results from an analysis of permanent disability for a Danish population of impaired lives. J. Dhaene, G. Willmot & B. Sundt derive recursions for distribution functions and stop-loss transforms. C. Hipp illustrates how to calculate total claims distributions by use of a spreadsheet. W. Hürlimann contributes with eight papers. He treats tail cutting methods, recursive methods, mean scaled compound distributions, mean scaled individual risk models, bounds for expected financial payoffs, loss variance bounds, and stop-loss bounds for diatomic bivariate sums. J.P. Nielsen & P. Voldsgaard apply marker dependent hazard estimation to health dependent mortality. The methods used are related to the Cox regression model used by Keiding, Andersen & Fledelius. G. Taylor constructs a loss reserving model that treats the individual layers of an excess of loss reinsurance.

Working session: Speaker's corner

S. Wang implements proportional hazard transforms in rate making. A. Gisler considers the interaction between tariff-structure and bonus-malus. H. Schmitter asked for help to estimate change points in a tariff. Paul Johansen gave a talk with the title "mathematics v.s. statistics". He urged young actuaries to develop models for individual non-life policies. Gunnar Benktander gave an overview of the history of ASTIN. Gunnar ended his talk by citing Paul Johansens characterization of the difference between life- and non-life insurance: "you can only die once".

The 28th ASTIN Colloquium will take place 10-13 August 1997 in Cairns, Australia.

SVEND HAASTRUP, Copenhagen
CLAIMS RESERVING IN CONTINUOUS TIME; A NONPARAMETRIC BAYESIAN APPROACH

SVEND HAASTRUP AND ELJA ARJAS

Laboratory of Actuarial Mathematics, Universitetsparken 5, 2100 København Ø, Denmark
Department of Applied Mathematics and Statistics, University of Oulu, Linnanmaa 90570, Oulu, Finland

ABSTRACT

Occurrences and developments of claims are modelled as a marked point process. The individual claim consists of an occurrence time, two covariates, a reporting delay, and a process describing partial payments and settlement of the claim. Under certain likelihood assumptions the distribution of the process is described by 14 one-dimensional components. The modelling is nonparametric Bayesian. The posterior distribution of the components and the posterior distribution of the outstanding IBNR and RBNS liabilities are found simultaneously. The method is applied to a portfolio of accident insurances.

KEYWORDS

IBNR and RBNS liabilities, marked point process, missing data, Markov chain Monte Carlo.

1. INTRODUCTION

A major issue in non-life insurance is prediction of outstanding liabilities. Outstanding liabilities are traditionally divided into occurred but not reported (IBNR) claims and reported but not settled (RBNS) claims. At each time the insurance company has to predict the outstanding liabilities and provide a reserve correspondingly.

A vast number of articles have been written on the subject. In most models the data are assumed to be discretized. Arjas (1989), Jewell (1989), Norberg (1993a,b,c) and Hesselager (1994) model in continuous time. Norberg (1993a) describes occurrence and development of the claims by a marked Poisson process. In Norberg (1993c) and Kirkegaard (1994) different parametric specifications of the model are considered, and real insurance data are analyzed. Furthermore, Norberg (1993a) considers an extended model where the occurrence intensity is assumed to be a stochastic process, and he finds the best linear predictor of the outstanding liabilities.

The present paper deals, by way of a case-study, with a portfolio of accident insurances. The model used is close to that of Norberg (1993a). The claims process generating occurrences, covariates and developments of the claims is modelled as a marked Poisson process. Our approach to estimation and prediction is nonpara-
metric Bayesian. Adopting the methods of Arjas and Gasbarra (1994), the distribution governing the process of occurrences, covariates and developments is modelled by piecewise constant conditional intensities. The intervals on which the intensities are constant, and the values (levels) of the intensities on the different intervals, are then viewed as model parameters. In principle, such a parameter space is of infinite dimension. A prior distribution (a distribution on the parameter space) is attached to the intensities of the claims process.

Both the model parameters and the outstanding liabilities (RBNS and IBNR claims) are unknown. The problem is to find the conditional distribution of such unknowns given the observations. This distribution will be called the posterior; it covers the conditional distribution of the unknown parameters which by standard usage is called posterior, and the conditional distribution of future observables which by standard usage is called predictive distribution. In complex models, it is often difficult to identify the posterior. The posterior can always be determined up to proportionality, but it can be difficult to normalize, which is necessary e.g. for the calculation of means. Recently, a technique called Markov chain Monte Carlo (MCMC) integration has been used to solve this problem numerically in connection with large statistical models. A general review of the topic can be found in Smith and Roberts (1993). The idea is to generate a Markov chain which has the posterior distribution as its equilibrium distribution. Using such a chain, all quantities of interest can be estimated/predicted. For example, at each step of the chain a new value of the RBNS claims is sampled; the empirical distribution of these sampled RBNS claims then converges towards the predictive (posterior) distribution of the RBNS claims.

Section 2 below describes the data. A claim is described by an occurrence time, two covariates, and a development. The development contains a reporting delay, a settlement delay, and a partial payment process containing the partial payments made from reporting to the settlement of the claim.

In Section 3 distributional assumptions are made. Claims are assumed to occur in accordance with a Poisson process, and covariates and developments are modelled as marks associated with the occurrences. The distribution of both occurrences and marks is specified by piecewise constant conditional intensities, and a prior distribution of these intensities is chosen.

Section 4 describes the MCMC algorithm (sampling algorithm). A Markov chain with the desired properties is generated. The algorithm is close to the one described by Arjas and Gasbarra (1994).

Section 5 describes the estimations and predictions. Using the Markov chain generated in the sampling algorithm we approximate both the distribution of the claims process and the distribution of the outstanding liabilities.

2. The data

The data are a portfolio of accident insurances. In the following we first describe the general structure of the data, and then go on with a detailed description of the present data set.
2.1. Structure of the data

We use the set-up of Norberg (1993a). By a claim we understand a combination of a time of occurrence, a set of covariates, and a development. Let time 0 be the initial time, and let $\tau$ be the time at which IBNR and RBNS liabilities are to be predicted. Each individual policy is described by the covariates $(s, a)$ denoting sex and age of the policy holder. For each combination of calendar time $t$ in $(0, \tau]$, sex $s$ in $\{\text{male, female}\}$, and age $a$ in $\{1, 2, \ldots\}$, the number of policies covered is denoted $w(t, s, a)$. The function $w$ is called the exposure rate.

Using the notation of Norberg (1993a) the development of a claim can be described by

$$(U, (Y(v))_{v \in [0, \tau]}),$$

where $U$ is the waiting time from occurrence until notification (the reporting delay), $V$ is the waiting time from notification until final settlement, and $Y(v)$ is the amount paid $v$ time units after notification. The final claim amount is $Y(V)$, which is called $Y$ for short. Moreover, when a claim is reported, the occurrence time $T$ and the covariates sex and age $(S, A)$ are known, age being the age at the time of occurrence. So the complete description of a claim is

$$(T, S, A, U, X)$$

where, for short, $X = (Y(v))_{v \in [0, \tau]}$.

The partial payments process $X = (Y(v))_{v \in [0, \tau]}$ consists of a series of lump payments. An illustration can be found in Norberg (1993a), Figure 2.

Not all claims which occurred before time $\tau$ are actually observed. At time $\tau$ we have only observed the reported claims, i.e. claims with $T + U \leq \tau$, and for each of these we only know the development up to time $\tau$. This means that for a reported claim we always know $(T, S, A, U)$. Furthermore we observe,

for a settled claim,

$$(V, (Y(v))_{v \in [0, \tau]}), \quad \text{and} \quad (2.1)$$

for a reported but not settled claim,

$$(1_{\{V \leq \tau - T - U\}} = 0, (Y(v))_{v \in [0, \tau - T - U]}), \quad (2.2)$$

where $1_{\{V \leq \tau - T - U\}} = 0$ indicates that the claim is not settled. Note that the above is just the partial payment process, $(Y(v))_{v \in [0, \tau]}$, censored at calendar time $\tau$.

2.2. The present data set

The data are a portfolio of accident insurances, supplied by a Danish insurance company. There are four different kinds of claims; dental claims, spectacles claims, disability claims, and death claims. We have chosen to model the dental claims, and look at leisure time cover only. Claims which occurred between January 1, 1982 and December 31, 1990 and which were reported before March 3, 1992 are observed and contained in the data set. Here we have chosen to consider $\tau = 6$ years. This means
that only data recorded by December 31, 1987 are considered. The rest of the data can be used to check the validity of our model.

Some modifications are needed to give the data the desired structure. The exposure is only known by years, not days. An estimate of the exposure rate \( w \) is obtained by interpolation. Information about claim developments is almost as we need them. Some of the claims records do not contain information about the covariates, and for simplicity such claims records (as well as exposure) were removed. Furthermore, times of occurrence, notification, partial payments and settlements are only known by days. However, we shall model occurrences and developments in continuous time, and view the reported times of occurrence, notification, etc. as approximations to the true values. Finally, we have chosen to modify the observed partial payment processes. Looking at a partial payment process, two payments made during the same day on the same claim are lumped together and viewed as a single payment. Furthermore, about 2% of the partial payments were negative. Some of these had a corresponding positive payment on the same day, and in that case both these payments were disregarded. Negative payments that did not have a corresponding positive payment never exceeded the accumulated claim amount paid, and they were set off against the previous positive payments. For example, a recorded partial payment process containing two payments, one of 1200 DKK at January 5, 1985 and one of -500 DKK at January 16, 1985, is transformed into a partial payment process containing only one partial payment of 700 DKK made at January 5, 1985. Lumpung payments that are made during the same day is easy to justify, but to set off negative payments against previous positive payments requires some comments. We have chosen December 31, 1987 as the time of prediction. However, in our modification of the observed partial payments processes we included all partial payments made before March 3, 1992. This means that we use negative payments in the future to modify payments in the past. An insurance company can not do that as it does not know about payments in the future. Therefore, insurance companies with recorded negative payments will tend to slightly overestimate the outstanding liabilities if they use our model.

After these modifications, our data contain 434,000 exposure years. There are 2806 reported claims; of these 2191 are settled and 617 unsettled. There are 3718 observed partial payments, and they add up to 10,040,000 DKK.

3. Distributional assumptions

We model the distribution of all claims which occur in our portfolio and their full development. Let

\[
(T_i, S_i, A_i, U_i, X_i)_{i \geq 1}
\]

(3.1)

denote these claims. It is important to note that many of the claims (3.1) are not observed completely, and some of them not at all. As mentioned in the previous section, we only observe the reported claims and their development is censored at calendar time \( \tau \).
A Bayesian model is used. The modelling is done in two steps. First, the
distribution of the claims process (3.1) is described by intensities, henceforth
referred to as components. Our model will have 14 such components. Some describe
the distribution of the occurrences and covariates, some describe the distribution of
the reporting delays, and some the distribution of the partial payments processes.
Then, a prior distribution is chosen. The intensities (components) are assumed to be
piecewise constant. The intervals on which the intensities are constant, and the
values (levels) of the intensities on the different intervals are the parameters. A prior
distribution (a distribution on the parameter space) is attached to the intensities.

The two steps are described below. The first step contains ‘likelihood assump-
tions’, and here the major restrictions are made. The second step contains ‘prior
assumptions’. Since we have only little knowledge at hand, we try to add only little
structure in our choice of prior.

3.1. Likelihood assumptions
At calendar time \( t \) there are \( w(t, s, a) \) policies with sex \( s \) and age \( a \) in the portfolio.
We assume that for an individual policy, claims occur according to a Poisson
process with intensity

\[
f(t, s, a), \quad t \in (0, \tau]
\]

(age changing once a year). As a consequence we get that, amalgamating all policies
in our portfolio, the occurrence times and covariates follow a marked Poisson
process with intensity

\[
w(t, s, a)f(t, s, a), \quad (t, s, a) \in (0, \tau] \times \{\text{male, female}\} \times \{1, 2, \ldots\}.
\]

It is assumed that the intensity \( f \) can be written as

\[
f(t, s, a) = f_1(t)f_2(s, a).
\]

(3.2)

In this way the distribution of occurrences and covariates is described by three
components; a calendar time effect \( f_1 \), an age effect for males \( f_2(\text{male}, \cdot) \), and an age
effect for females \( f_2(\text{female}, \cdot) \).

The development of a claim contains a reporting delay and a partial payment
process. Following Norberg (1993a), we assume that the distribution of the devel-
opment of a claim depends on the past history of the process only through the
associated occurrence time and covariates. This kind of development distribution is
called position dependent, see Karr (1991). In the following the distribution of the
development of a claim, which occurred at time \( t \) and has covariates \((s, a)\), is
described by the distribution of the reporting delay \( U \), \( P_{U|t, s, a} \), and the distribution
\( P_{X|U, t, s, a} \) of the partial payment process \( X \) given \( U = u \).

The distribution \( P_{U|t, s, a} \) is assumed to be absolutely continuous (with respect to
the Lebesgue measure). It is modelled by the corresponding hazard rate

\[
g_{U|t, s, a}(u) = g(u),
\]

(3.3)

which is assumed independent of both the time of occurrence and the covariates;
that a probability $P$ on $\mathcal{R}_+$ has hazard rate $g$ means that $g(u)du = P(du)/P((u,\infty))$. There are many possible extensions. For example, we could assume that $P_{U,t,s,a}$ has hazard rate

$$g_{U,t,s,a}(u) = g_1(s,u)g_2(s,a)g_3(t),$$

containing 5 components. Both sexes have then their individual reporting delay component and age component, and then there is a calendar time component. Such an extension would cause no mathematical problems, but the computational effort would increase.

The partial payment process $X = (Y_v)_{v \in [0,v]}$ arising from a single claim is a jump process. Let $H_v = \{T, S, A, U, (Y(v'), 1_{(v'<v)})_{v \in [0,v]}\}$ be the history of a reported but not settled claim; here $v$ is the time since the claim was reported. Let $dy$ denote a small interval of length $dy$ around $y$. The distribution of $X$ is described using intensities. Let $h_{se}(v | H_{v-})$ be the intensity of settling at time $v$ without a partial payment at time $v$, let $h_{sep}(v, dy | H_{v-})$ be the intensity of settling at time $v$ with partial payment of size $dy$, and finally let $h_{p}(v, dy | H_{v-})$ be the intensity at time $v$ of having a partial payment of size $dy$ without settlement. We have to decide how the intensities $h_{se}$, $h_{sep}$ and $h_{p}$ depend on the history $H_{v-}$ just before time $v$.

The following information can be derived from $H_{v-}$:

- $N_v$ number of partial payments in $[0,v)$,
- $T_v$ time since the latest partial payment if any, else $T_v = v$.

We assume that the intensity of settling only is

$$h_{se}(v | H_{v-}) = \begin{cases} h_{se}^0(v) & \text{if } N_v = 0 \\ h_{se}^1(T_v) & \text{if } N_v > 0, \end{cases} \quad (3.4)$$

the intensity of settling with a partial payment of size $dy$ is

$$h_{sep}(v, dy | H_{v-}) = \begin{cases} h_{sep}^0(v)p_{sep}^0(dy) & \text{if } N_v = 0 \\ h_{sep}^1(T_v)p_{sep}^1(dy) & \text{if } N_v > 0, \end{cases} \quad (3.5)$$

and the intensity of having a partial payment of size $dy$ without settlement is

$$h_{p}(v, dy | H_{v-}) = \begin{cases} h_{p}^0(v)p_{p}^0(dy) & \text{if } N_v = 0 \\ h_{p}^1(T_v)p_{p}^1(dy) & \text{if } N_v > 0. \end{cases} \quad (3.6)$$

As a consequence, partial payments are distributed according to the densities $p_{sep}^0$, $p_{sep}^1$, $p_{p}^0$ and $p_{p}^1$. These will be described by their corresponding hazard rates denoted $q_{sep}^0$, $q_{sep}^1$, $q_{p}^0$ and $q_{p}^1$.

The assumptions (3.4)-(3.6) need some comments. The distribution of the partial payment process $P_{X|U,t,s,a}$ is assumed to be independent of both the time of occur-
rence \( t \), the reporting delay \( u \) and the covariates \((s, a)\). The intensities of the partial payment process depend on the past history of the process only through the number of partial payments \( (N_v) \) and the time since latest partial payment if any, else the time since notification. These assumptions are not crucial, and at this stage the model has many possible extensions. For example, it could be reasonable to assume that the size of a partial payment depends on the sizes of the previous partial payments. And often it is reasonable to assume that the size of a payment also depends on the time since the latest partial payment (if any). Such extensions can be done without mathematical difficulties, but the computational effort would increase. For example, making the size of a partial payment dependent on the cumulated amount of the previous payments, \( Y(v-) \), can be done assuming that the intensities \( q^1_{\text{sep}} \) and \( q^1_p \) have the structure

\[
q^1_{\text{sep}}(y \mid \mathcal{H}_v) = q^1_{\text{sep}}(y)q^1_{\text{sep}}(Y(v-))
\]

and

\[
q^1_p(y \mid \mathcal{H}_v) = q^1_p(y)q^1_p(Y(v-)).
\]

Even the product assumptions made here are not that crucial.

### 3.2. Prior assumptions

In the previous subsection the distribution of the claims process is described by 14 components (intensities). These are

\[
f_1, f_2(\text{male}, \cdot), f_2(\text{female}, \cdot), g,
\]

\[
h_{\text{sep}}, q_{\text{sep}}, h_p, q_p, h_{\text{sep}}, q_{\text{sep}}, h_p, q_p, h_{\text{sep}}, q_{\text{sep}}, h_p, q_p, h_{\text{sep}}.
\]

Now a prior distribution is to be chosen. The intensities are assumed piecewise constant. The intervals on which the intensities are constant, and the values (levels) of the intensities on the different intervals are the parameters. A prior distribution (a distribution on the parameter space) is attached to the intensities. The main idea is taken from Arjas and Gasbarra (1994). Following their notation, we shall denote unknown parameters by Latin letters and parameters in the prior distributions, the so-called hyperparameters, by Greek letters.

To begin with we look at the calendar time effect \( f_1 \). It is assumed to have a piecewise constant structure

\[
f_1(t) = \sum_{j \geq 0} I(s_j < t \leq s_{j+1}) b_j,
\]  

(3.7)

where \( b_j \), given \((b_0, \ldots, b_{j-1})\), follows a lognormal distribution with parameters \((\log(b_{j-1}), \sigma^2_{f_1})\) denoting the mean and the variance in the associated normal distribution, and \( 0 = S_0 < S_1 < S_2 < \ldots \) follows a Poisson process with intensity \( \lambda_{f_1} \). In short, we write

\[
b_j \mid (b_0, \ldots, b_{j-1}) \overset{D}{=} \text{LogN}(\log(b_{j-1}), \sigma^2_{f_1}), \quad (S_i)_{i \geq 0} \overset{D}{=} \text{Poisson}(\lambda_{f_1}).
\]  

(3.8)
For uniqueness it is assumed that the initial level \( b_0 \) is 1.

The prior structure (3.7)-(3.8) is essentially used to model all 14 components. A prior should reflect the knowledge one has about the problem at hand, and it is therefore reasonable to discuss what kind of prior information the structure (3.7)-(3.8) will represent. To save notation we drop the subscript \( f_i \) on the hyperparameters.

The intensity \( f_i \) is by (3.7) assumed to be a positive simple function (a piecewise constant function). The prior distribution of \( f_i \) should therefore be a distribution on the space of positive simple functions. The prior (3.8) is a possible choice. The prior expected number of changes in level per year is \( \lambda_i \), and the levels have a log-martingale structure,

\[
\log(b_j) = \log(b_{j-1}) + \varepsilon_j,
\]

where the \( \varepsilon \)'s are iid normally distributed with zero mean and variance \( \sigma^2 \). A small (large) value of \( \sigma^2 \) corresponds to a high (low) correlation between the levels at different times. We consider this choice of prior very vague. It is our experience that, with a reasonable amount of data, the data will ‘speak for themselves’. By adjusting the values of the hyperparameters we can, however, control the smoothness of the estimate, the estimate being the posterior mean, say. Now, our prior also has some weaknesses. The prior distribution of \( f_i \) is not stationary. The median in the prior distribution of \( b_j \) is \( b_0 \), but the mean is \( b_0 \exp(j\sigma^2/2) \). This could be a problem if we wish to predict the occurrence intensity in the years to come. There we have no data, and if we choose to predict using the posterior mean, we will therefore get an increasing estimator. The reader might ask: why not control the mean by multiplying \( b_j \) with \( \exp(-j\sigma^2/2) \)? If that is done, then it can be shown that \( b_j \) converges almost surely to 0, and new difficulties arise. As an alternative prior one could keep the distributional assumptions about the jump times, but assume that the levels are distributed as

\[
b_j = \bar{b} \zeta_j, \quad \text{where} \quad \bar{b} \overset{D}{=} \text{Gamma}(\gamma, \delta), \quad \text{and} \quad \zeta_j \overset{iid}{=} \text{Gamma}(\kappa, 1/\kappa). \quad (3.9)
\]

From a computational point of view, this prior gives no difficulties. Furthermore, it is stationary. The assumption is that the intensity varies around a level \( \bar{b} \). It does not allow for permanent changes in the intensity. As a consequence, the prior (3.9) adds more structure to the intensity (smaller variance), and that can be useful when predicting the future occurrence intensity. We are going to use prior (3.7)-(3.8). The prior is chosen in an attempt to add only little prior information.

The age effects \( f_2(\text{male, } \cdot) \) and \( f_2(\text{female, } \cdot) \) are modelled as discrete versions of the construction (3.7)-(3.8). We assume that

\[
f_2(\text{male, } a) = \sum_{j \geq 0} 1(a \in \{K_j + 1, \ldots, K_{j+1}\}) \zeta_j, \quad a = 1, 2, 3, \ldots
\]

where \( (\zeta_j)_{j \geq 1} \) are modelled as \( (b_j)_{j \geq 1} \) above, but with new hyperparameters \( \sigma^2_{f_{2m}} \) and \( \lambda_{f_{2m}} \), and where \( (K_j + 1 - K_j)_{j \geq 1} \) are assumed to be iid geometrically distributed with parameter \( \kappa_{f_{2m}} \). Some readers might again find it reasonable to add the prior information that young males have a high intensity of making claims. Such
information is not added here, but our estimates will show that young males are a high risk group. The age effect for females $f_2(\text{female, } \cdot)$ is assumed to have the same structure as the age effect for males, but with new hyperparameters. Furthermore, we need to specify the prior of the initial levels $f_2(\text{male, 0})$ and $f_2(\text{female, 0})$. These are assumed to be lognormally distributed with parameters $(\mu_{0, f_{2m}}, \sigma_{0, f_{2m}}^2)$ for males and $(\mu_{0, f_{2f}}, \sigma_{0, f_{2f}}^2)$ for females; the $\mu$’s controlling the levels and the $\sigma$’s controlling the variability.

The remaining 11 components ($g$, $h_{\text{sep}}^0$, $q_{\text{sep}}^0$, etc.) are all assumed to have the same structure (3.7)-(3.8) as the calendar time effect $f_1$; each component having its individual hyperparameters. The initial levels are assumed to be lognormally distributed, again with individual hyperparameters for each component.

Finally, unknown parameters associated with different components are assumed to be independent with respect to the prior. This independence assumption is not that crucial, and in some branches it is not reasonable. It might be that a large number of occurrences induces relatively small claims. In that case $f_1$ should be positively correlated with the $q$’s. Also, it might not be reasonable to assume independence between the competing risks, $h_{\text{sep}}^0$, $h_{\text{sep}}^0$, and $h_0^0$, say. It could be that, if the intensity of settling only, $h_{\text{sep}}^0$, is high, then the intensity of settling with a partial payment, $h_{\text{sep}}^0$, tends to be high too. There are ways of modelling such dependencies, but for simplicity we have chosen not to consider them here.

### 3.3. Additional remarks

The distribution of the claims process (3.1) is described by 14 components. These components are modelled nonparametrically. Some readers might want to add more structure by using parametric models. For example, the distribution of the partial payments could be described by mixtures of lognormal distributions. Such a choice can be motivated easily. We work with dental claims; there are examinations and there are operations. Operations are more expensive than examinations, and therefore it can be argued that the distribution of a partial payment should have two peaks. The estimates will show that the distributions of the partial payments actually have two peaks. Also, as discussed previously, it could be useful to add more structure into the calendar time effect $f_1$ when predicting the occurrence intensity in the years to come. As time goes ($\tau$ increases) more and more claims are observed, and the estimates of all components that are independent of calendar time will be consistent (the posterior distributions will converge towards unit mass distributions). However, the uncertainty about the occurrence intensity in the years to come will remain. Assuming that the occurrence intensity does not change with calendar time ($f_1$ constant) is a possibility. This is, however, typically not realistic, and structures like (3.9) could be chosen.

Along the way we have pointed out possible extensions. Different branches of insurance (motor insurance, fire insurance, etc.) call for different specifications of the model. The insurance companies will have an idea about which dependencies the model should allow for, e.g. that a long settlement period typically induces large partial payments.
4. Markov Chain Monte Carlo Integration

We wish to approximate the posterior distribution of the unknowns: i.e. parameters and outstanding liabilities. To do that, Markov chain Monte Carlo (MCMC) integration is used. A Markov chain, which has the posterior distribution as its equilibrium distribution, is generated. The algorithm contains an arbitrarily chosen number of steps where, at each step, new values of the unknown parameters, the remaining developments of the reported but not settled claims (RBNS) and the occurrence times, covariates and full developments of the occurred but not reported claims (IBNR), are sampled. As the number of steps increases, the empirical distribution of these sampled quantities converges towards the posterior distribution. Thereby the posterior can be approximated, and based on this approximation the estimations and predictions are made. By increasing the number of steps we can make the approximation as exact as we wish.

Figure 1 shows a model graph. Quantities surrounded by squares are known; these are either hyperparameters chosen by us, or they are data. Quantities surrounded by ellipses are unknowns; these are either unknown parameters, IBNR claims or RBNS claims. Arrows indicate dependencies. In each step of the algorithm new values of all quantities surrounded by circles are sampled.

Before we go on with a description of the algorithm we shall again refer to Smith and Roberts (1993) for an introduction to MCMC integration.

4.1. Likelihood

In order to solve the Bayesian inferential problem we need the likelihood of the data, i.e. the distribution of the (observed) data given the unknown parameters.

The observations are the claims reported, including their development up to calendar time \( \tau \). We denote these observations

\[
(T^o_i, S^o_i, A^o_i, U^o_i, X^o_i)_{i \geq 1}.
\]

(4.1)

Note that \( X^o_i \) is the partial payment process censored at calendar time \( \tau \), i.e. \( \tau - T^o_i - U^o_i \) time units after notification; cf. (2.1)-(2.2). In the previous section the distribution of all claims which had occurred, including their full development were modelled. The results of Norberg (1993a) imply that, given the unknown parameters, the process generating the reported claims including their full development, even if these go beyond calendar time \( \tau \), has occurrence intensity

\[
w(t, s, a)f(t, s, a)e^{u(r - t)}.
\]

(4.2)

The conditional distribution of the reporting delay of a claim which occurred at time \( t \) is

\[
P_U(du)1(u \leq \tau - t)
\]

\[
P_U(\tau - t),
\]

(4.3)

and the distribution of the corresponding partial payment process is

\[
P_x(dx).
\]
The claims (4.1) occur with intensity (4.2). The distribution of a reporting delay of a claim which occurred at time \( t \) is given by (4.3). The distribution of the observed part of a partial payment process of a claim, which occurred at time \( t \) and was reported after \( u \) time units, is easily found using the 'intensity construction' of the distribution \( P_X \) of the partial payment process; see Subsection 3.1. We denote the distribution

\[
Pr_{X}^{\tau - t - u},
\]

where \( \tau - t - u \) refers to the censoring of the process \( \tau - t - u \) time units after notification. A likelihood is a density. In the following we use a somewhat sloppy notation which hopefully does not cause misunderstandings. The likelihood of the observations (4.1) is
Here the last part (4.5) dealing with the partial payment processes, has to be written in more detail. Recall that \( X_i^\circ \) is the observed part of the partial payment process of the \( i \)th claim. The superscript \( \tau - T_i - U_i^\circ \) refers to the censoring of the process at calendar time \( \tau \). From the observed partial payment processes \( (X_i^\circ)_{i \geq 1} \) some useful quantities are derived. First we consider exposure:

- \( W^0(v) \): The number of claims in which the waiting time from notification to first partial payment or settlement is at least \( v \). Note that \( W^0 \) is majorised by the number of reported claims and decreasing in \( v \). We have \( W^0(0) = 2806 \), the number of reported claims.

- \( W^1(v) \): The number of times any partial payment process is observed, with waiting time at least \( v \) since the latest partial payment. Also \( W^1 \) is decreasing in \( v \). We have \( W^1(0) = 1780 \). Here \( W^1 \) is less than the number of reported claims, but in principle \( W^1 \) can exceed the number of reported claims since a claim can have several partial payments.

Then ‘events’:

- \( (V_{se}^{0})_{j \geq 1} \): Waiting times from notification to settlement without payment. There are only 9 of these in the data.

- \( (V_{sep}^{0})_{j \geq 1} \) and \( (Y_{sep}^{0})_{j \geq 1} \): Waiting times from notification to settlement with one payment, and the corresponding amounts paid (in this case the payment equals the total claim amount). There are 1295 of these.
\((V_p^0)_{j \geq 1}, (V_p^1)_{j \geq 1}\): Waiting times from notification to the first partial payment without settlement, and the corresponding amounts paid. There are 1376 of these.

\((V_{se}^1)_{j \geq 1}\): Waiting times from a partial payment to settlement only (no partial payment at the time of settlement). There are 242 of these.

\((V_{sep}^1)_{j \geq 1}, (V_{sep}^0)_{j \geq 1}\): Waiting times from a partial payment to settlement with payment, and corresponding amount paid at settlement. There are 643 of these.

\((V_p^1)_{j \geq 1}, (V_p^1)_{j \geq 1}\): Waiting times from a partial payment to the next, when the latter does not settle the claim, and corresponding amounts. There are 404 of these.

As an illustration we have in Figure 2 shown two examples of partial payment processes. The first process settles within time \(\tau\). There are two payments made, and no payment is made at the time of settlement. The second process arises from an RBNS claim. There is one payment made and the partial payment process is censored (calendar time \(\tau\) is reached). Events and amounts are shown at the figure; waiting time between jumps are events, and sizes of jumps are amounts. Note that \(V\) is the waiting time from latest payment until censoring, and it only affects the exposure. The exposure functions are given by

\[
\text{Accumulated payments}
\]

\[
V_{p,1}^0, V_{p,1}^1, V_{se,1}^1
\]

\[
\text{Time since notification}
\]

\[
\text{Accumulated payments}
\]

\[
V_p^0, V_p^1, V_p^2, V\]

\[
\text{Time since notification}
\]

**Figure 2** Examples of observed partial payment processes; \(\times\) denotes a settlement.
\[
W^0(v) = \begin{cases} 
2 & v \in (0, V_p^{1,2}] \\
1 & v \in (V_p^{0,2}, V_p^{0,1}] \\
0 & v \in (V_p^{0,1}, \infty)
\end{cases}
\quad \text{and} \quad
W^1(v) = \begin{cases} 
3 & v \in (0, V_p^{1,1}] \\
2 & v \in (V_p^{1,1}, V_p^{1,2}] \\
1 & v \in (V_p^{1,2}, V_p^{2,1}] \\
0 & v \in (V_p^{2,1}, \infty)
\end{cases}
\]

As a function of exposure and events the last part (4.5) of the likelihood can be written as

\[
\prod_{i \geq 1} P_{i}^{\tau_i - T_i + U_i} (dX_i) \propto \prod_{j \geq 1} h_{\text{sc}}^0(V_{\text{sc}}^{0,j}) \prod_{j \geq 1} h_{\text{sep}}^0(V_{\text{sep}}^{0,j}) \prod_{j \geq 1} h_p^0(V_p^{0,j}) \times \prod_{j \geq 1} h_{\text{sc}}^1(V_{\text{sc}}^{1,j}) \prod_{j \geq 1} h_{\text{sep}}^1(V_{\text{sep}}^{1,j}) \prod_{j \geq 1} h_p^1(V_p^{1,j}) \times \exp \left( - \int_0^\infty W^0(v)(h_{\text{sc}}^0(v) + h_{\text{sep}}^0(v) + h_p^0(v)) \, dv \right) \times \prod_{j \geq 1} p_{\text{sc}}^0(Y_{\text{sc}}^{0,j}) \prod_{j \geq 1} p_{\text{sep}}^0(Y_{\text{sep}}^{0,j}) \times \prod_{j \geq 1} p_{\text{sc}}^1(Y_{\text{sc}}^{1,j}) \prod_{j \geq 1} p_{\text{sep}}^1(Y_{\text{sep}}^{1,j}) \times \exp \left( - \sum_{j \geq 1} \int_0^{Y_{\text{sep}}^{0,j}} q_{\text{sep}}^0(Y) \, dy \right). \tag{4.12}
\]

where the \( V, W \) and \( Y \) depend on the observed partial payment processes in an obvious way. The density \( p_{\text{sep}}^0 \), say, is given by its corresponding hazard rate \( q_{\text{sep}}^0 \). We have

\[
\prod_{j \geq 1} p_{\text{sep}}^0(Y_{\text{sep}}^{0,j}) \propto \prod_{j \geq 1} q_{\text{sep}}^0(Y_{\text{sep}}^{0,j}) \times \exp \left( - \sum_{j \geq 1} \int_0^{Y_{\text{sep}}^{0,j}} q_{\text{sep}}^0(Y) \, dy \right). \tag{4.12}
\]

The last part (4.5) of the likelihood function can now be substituted with (4.6)-(4.12).

We have chosen priors along the lines of Arjas and Gasbarra (1994). To copy their sampling algorithm it is necessary that, in each of the 14 components, the likelihood is proportional to an expression of the form

\[
\prod h(\cdot) \times \exp \left( - \int Z(\cdot)h(\cdot) \right),
\]

where \( h \) is a component, and \( Z \) does not depend (functionally) on \( h \). This is the case with all components but \( g \), which unfortunately occurs in the part (4.4) of the likelihood. A way to deal with this problem is to use data augmentation.

**4.2. Missing data**

The reason that the likelihood of the (observed) data does not have a tractable shape is that the occurred but not reported claims (IBNR) are not observed. Here data augmentation can be used. By sampling and adding to our data the occurrence times
and covariates of the IBNR claims, the likelihood obtains the desired shape. The occurrence times and covariates of the IBNR claims are called missing data. The missing data are denoted

$$(T_j^m, S_j^m, A_j^m)_{j \geq 1}. \quad (4.13)$$

Note that reporting delays and partial payment processes of the IBNR claims, and the remaining development of the RBNS claims, are not included as missing data. However, we are going to sample them at each step of the algorithm, but they are not used to sample unknown parameters. They are only used to predict outstanding liabilities, cf. Figure 1. The results of Norberg (1993a) imply that, given the unknown parameters, the missing data (4.13) are independent of the observations (4.1) and that the missing data are distributed as a marked Poisson process with intensity

$$w(t, s, a)f(t, s, a)(1 - P_U(\tau - t)), \quad (t, s, a) \in (0, \tau) \times \{\text{male, female}\} \times \{1, 2, \ldots\}. \quad (4.14)$$

The likelihood of the missing data only is

$$\Lambda(\text{mis}) \propto \prod_{j \geq 1} f(T_j^m, S_j^m, A_j^m)(1 - P_U(\tau - T_j^m))$$
where the last part is written in detail in (4.6)-(4.12). Now the likelihood has the desired shape.

4.3. Sampling algorithm

The sampling algorithm goes as follows: To begin with, initial values of the unknown parameters are determined. These can either be sampled from the prior distribution or chosen arbitrarily; asymptotically it does not matter. A general step of the algorithm contains three substeps:

Substep 1: Sample occurrence times and covariates of the IBNR claims (the missing data) given the unknown parameters sampled in the previous step of the algorithm;

Substep 2: Sample the remaining development of the reported but not settled claims (RBNS) and the full developments of the IBNR claims, given the unknown parameters sampled in the previous step of the algorithm, and given both observed and missing data where the latter were sampled in Substep 1;

Substep 3: Sample the unknown parameters given both the data (observations) and the missing data sampled in Substep 1.

These three substeps are discussed in the following.

4.3.1. Sampling the missing data

The missing data are the occurrence times and covariates of the IBNR claims, see (4.13). Given the present value of the unknown parameters, sampled in the previous step of the algorithm, the missing data are distributed as a marked Poisson process with intensity (4.14). It follows from Norberg (1993a), Theorem 1, that the total number of IBNR claims is Poisson distributed with mean

\[
W_{\text{IBNR}} = \int_0^\tau w(t, s, a)f_1(t)f_2(s, a)(1 - P_U(\tau - t))\,dt,
\]

and given this total number, the occurrence times and covariates of the IBNR claims are iid with density

\[
w(t, s, a)f_1(t)f_2(s, a)(1 - P_U(\tau - t))/W_{\text{IBNR}}
\]
on \((0, \tau] \times \{\text{male, female}\} \times \{1, 2, \ldots\}\). The missing data are sampled from this distribution.

4.3.2. Sample IBNR and RBNS claims

In Substep 2 the unknown parameters sampled earlier by the algorithm, the data, and the occurrence times and covariates of the IBNR claims sampled in Substep 1 (the missing data), are all held fixed.

First we sample the reporting delays of the IBNR claims. We already know their corresponding occurrence times \((T^n_j)_{j \geq 1}\). So, the \(j\)th reporting delay must exceed \(\tau - T^n_j\), and it is distributed according to the hazard rate \(g\). Now the \(j\)th reporting
delay is sampled from a distribution on \((\tau - T_j^n, \infty)\) with piecewise constant hazard rate \(g\).

The partial payment process of an IBNR claim can be sampled as follows. First the waiting time to the first ‘event’ is sampled. This waiting time is distributed according to the piecewise constant intensity \(h_{se}^0 + h_{sep}^0 + h_p^0\). Then the type of the event is sampled. With probability \(h_{sep}^0/(h_{se}^0 + h_{sep}^0 + h_p^0)\) it is a settlement only, with probability \(h_p^0/(h_{se}^0 + h_{sep}^0 + h_p^0)\) it is a settlement with partial payment, and with probability \(h_p^0/(h_{se}^0 + h_{sep}^0 + h_p^0)\) it is a partial payment only. If the first event includes a partial payment then the size of this payment is sampled. If the first event also includes settlement, then the partial payment has hazard rate \(q_{sep}^0\), else it has hazard rate \(q_p^1\). If the first event was a partial payment only then the waiting time to the next event and the type of this event are sampled. Here the components needed are \(h_{se}^1, h_{sep}^1, h_p^1, q_{sep}^1\) and \(q_p^1\). Waiting times and events are sampled until settlement is reached.

The remaining partial payment process of an RBNS claim is sampled using the same method as above. For each individual RBNS claim we need to know whether any previous partial payments have been made, and the time since the latest partial payment if any, else the time from reporting.

4.3.3. Sampling the unknown parameters

In this substep we condition on the observed data and the missing data sampled in Substep 1, and sample the unknown parameters (jump times and levels of the 14 components). The conditional distribution of these unknowns has a density that is proportional to the likelihood (4.15) of the combined observed and missing data, multiplied with the prior density of the unknown parameters. Direct sampling from this distribution is not possible. However, we can identify the conditional distributions of each of the unknown parameters given the remaining part of the unknown parameters and given the observed and missing data. From these one-dimensional distributions the parameters can be sampled one by one. A complete description of the sampling of the unknown parameters would take quite a few pages. We shall here give a short overview.

The unknown parameters are the jump times and levels of all the 14 components. To save notation we shall denote the unknown parameters by \(X_1, X_2, \ldots\) (how they are ordered will be discussed later). Assume that the present step is the \(k\)th, and denote by \(X_{1}^{k-1}, X_{2}^{k-1}, \ldots\) the values of the unknown parameters sampled in the previous step of the algorithm. They are now sampled one by one. As mentioned, the conditional distribution of each individual unknown parameter, given the remaining parts of the unknowns and given observed and missing data, is known (at least up to proportionality). First a new value \(X_1^k\) of \(X_1\) is sampled from the distribution of

\[
X_1 \text{ given } X_{2}^{k-1}, X_{3}^{k-1}, \ldots, \tag{4.18}
\]

then a new value \(X_2^k\) of \(X_2\) is sampled from the distribution of

\[
X_2 \text{ given } X_{1}^{k}, X_{3}^{k-1}, \ldots, \tag{4.19}
\]
and so the algorithm continues. The algorithm is called Gibbs sampling, cf. Smith and Roberts (1993). Recall that the unknown parameters are the levels and jump times of the 14 components. When the unknown parameter (an $X$) is a level, it is not tractable to sample from the above one-dimensional distributions and then Metropolis-Hastings is used. With this modification the algorithm is called a ‘variable-at-a-time Metropolis-Hastings’, cf. Chan and Geyer’s discussion of the paper by Tierney (1994).

The method used for doing the sampling (4.18)-(4.19) is taken from Arjas and Gasbarra (1994). The unknown parameters can be grouped into 14 groups according to the component they determine ($f_1, f_2(\text{male, } \cdot), f_2(\text{female, } \cdot), g, \ldots$). In Arjas and Gasbarra (1994) it is shown how the unknown parameters associated with one component can be sampled. Denote by $(b_j)$ and $(S_j)$ the levels and jump times of the component $f_j$, say. Arjas and Gasbarra order them as

$$b_0, S_1, b_1, S_2, \ldots, \tag{4.20}$$

and sample new values as indicated in (4.18)-(4.19) above. Their algorithm carries over with minor modifications; their levels are correlated gamma distributed, while our levels are correlated lognormally distributed, and when sampling the levels they use rejection sampling which corresponds to repeating Metropolis-Hastings until acceptance is reached. When sampling the unknown parameters associated with an individual component, we adopt the methods of Arjas and Gasbarra, and thereby the components are sampled one by one, following the order (4.20). But in which order should the different components be sampled? According to the prior the 14 components are independent, and by inspection of the likelihood (4.15) of the combined observed and missing data, we get that according to the conditional distribution of the unknown parameters given observed and missing data, all components but $f_1, f_2(\text{male, } \cdot)$ and $f_2(\text{female, } \cdot)$ are independent. Therefore, except for the mutual order of $f_1, f_2(\text{male, } \cdot)$ and $f_2(\text{female, } \cdot)$, the order in which the different components are sampled does not play a role. We have chosen to sample first $f_1$, then $f_2(\text{male, } \cdot)$, then $f_2(\text{female, } \cdot)$, and then the remaining 11 components. Asymptotically, the order does not play any role.

5. Results of the analysis

Suppose we stop the sampling algorithm after $n$ steps. What we then have is the first $n$ steps of a Markov chain. That the chain is Markov follows from the construction. Each step of the chain contains sampled values of the unknown parameters, occurrence times, covariates and full development of the IBNR claims, and remaining development of the RBNS claims.

By construction, the posterior distribution, i.e. the conditional distribution of the above unknowns given the observations, is invariant for that chain. It is obvious that
the chain is both irreducible and aperiodic. From Tierney (1994) it now follows that the chain is (positive) recurrent. We need the chain to be ergodic, and for that a stronger type of recurrence, called Harris recurrence, is needed. In their discussion of the paper by Tierney (1994), Chan and Geyer show that a certain type of algorithm, which they call a ‘variable-at-a-time Metropolis-Hastings’, is often Harris recurrent. Their proof carries over to our situation with only minor modifications. It then follows that our chain is ergodic. A survey of the theory can be found in Tierney (1994).

Let \((M_j)_{j=1,2,...}\) denote the chain, let \(h\) be a function on the space of possible values of the unknowns above, and assume that \(h\) has finite mean with respect to the posterior distribution. Then

\[
\frac{1}{n} \sum_{j=1}^{n} h(M_j) \overset{\text{a.s.}}{\rightarrow} E_{\text{posterior}} h. \tag{5.1}
\]

As an example, suppose we wish to find (approximate) the posterior distribution of the size of the outstanding liabilities. The outstanding liabilities are a function of the claims process (3.1). The liabilities are divided into liabilities arising from the IBNR claims,

\[
X_{IBNR} = \sum_{i \geq 1} 1(T_i \leq \tau < T_i + U_i) Y_i,
\]

and liabilities arising from the RBNS claims,

\[
X_{RBNS} = \sum_{i \geq 1} 1(T_i + U_i \leq \tau < T_i + U_i + \nu_i)(Y_i - Y_i(\tau - T_i - U_i)).
\]

Let \(X_{IBNR}^j\) and \(X_{RBNS}^j\) denote the values sampled in the \(j\)th step of the algorithm (both are functions of \(M_j\)). By (5.1) we have

\[
\frac{1}{n} \sum_{j=1}^{n} 1(X_{IBNR}^j + X_{RBNS}^j \leq x) \overset{\text{a.s.}}{\rightarrow} p_{\text{posterior}}(X_{IBNR} + X_{RBNS} \leq x), \tag{5.2}
\]

and thereby the posterior (predictive) distribution of the outstanding liabilities can be approximated by the left hand side of (5.2). The theory does not give an exact answer as to how large \(n\) should be chosen. However, what one usually does is to run several independent simulations and use these to evaluate the stability of the algorithm.

To begin with we have to choose values of the hyperparameters, and the initial values of the unknown parameters. Recall that each of the 14 components (intensities) can be written as

\[
\sum_{j \geq 0} 1(s_j < i \leq s_{j+1}) b_j. \tag{5.3}
\]

The hyperparameters \(\sigma_0^2\) and \(\mu_0\) determine the prior variance and mean of the initial level \((b_0)\) while \(\lambda\) (or \(\kappa\)) and \(\sigma^2\) determine the prior variability of the intensity; cf. the discussion in Subsection 3.2. Table 1 shows the chosen values of the hyper-
parameters. We shall not try to motivate the choice of all the values. However, a few comments are needed. The individual values of $\lambda$ and $\sigma^2$ were chosen as follows. We started out with some arbitrary values. If the estimate (posterior mean of the intensity) turned out either too ragged or too smooth, then new values of $\lambda$ and $\sigma^2$ were chosen. This way of choosing hyperparameters corresponds to how window size (bandwidth) is chosen in kernel density estimation, and it is not orthodox Bayesian in spirit. On the other hand, our choice of the hyperparameters $\tau$ and $\sigma^2$ represents prior knowledge. For example, we believe that small payments (from 0 to 25 DKK) are very unlikely, and therefore the values of the $q$'s should start out small. By choosing $\mu_0$ small we express the prior knowledge that the intensity (hazard rate) starts out small, and by choosing $\sigma_0^2$ small we express a high degree of belief in this knowledge. Other values of the hyperparameters were tried out. But, within what we believe were reasonable values, the hyperparameters had only minor influence on our estimations/predictions.

The unknown parameters are the jump times and levels (the $S$'s and $b$'s, cf. (5.3)) of each of the 14 components. Their initial values can, for example, be sampled from the prior distribution or be chosen arbitrarily. The theory says that asymptotically it does not matter what you do, which was also our experience. However, the algorithm turned out to be time consuming, and to save iterations we, therefore, wanted to start the algorithm at a place with high posterior probability. For each component, the initial values of the jump times were chosen such that areas with many observations (occurrences, delays or payments) had many jumps. Given the jump times, the maximum likelihood estimators of the levels were used as their initial values.
The algorithm was written in S-Plus and run on a DEC workstation. As mentioned, the algorithm turned out to be rather time consuming; about 1 hour for each 100 steps, but much could have been gained by writing parts of the algorithm in C, say. We did two runs; one of length 1000 and one of length 500. After a few iterations (40-60) the algorithm seemed stable (we plotted the approximations to the posterior mean of the individual 14 components, and found that these did not change much). As approximation to the posterior distribution of all unknowns we use the empirical distribution of all 1500 realizations.

Each of the 14 components was estimated by its (approximated) posterior mean (an average of the 1500 iterations). In Figure 3 the estimates of the components $f_1$, $f_2$ and $g$ are shown. From a mathematical point of view, it is easy to calculate (posterior) pointwise 95% credible intervals for each component, but the computational effort would increase. Even though we have no credible intervals, we allow ourselves to comment on the estimates. The occurrence intensity is the product of the calendar time effect $f_1$ and the age effect $f_2$. Looking at $f_1$, there seems to be no obvious seasonal effect. In observation year no. 3 (1984) the occurrence intensity is high (we have no explanation why), otherwise there are only minor variations. Looking at $f_2$ it is obvious that there is an age effect. Young males are a high risk group, while young females are a low risk group. From age 30 and on there are only minor differences between the sexes. The last graph shows the reporting delay hazard rate $g$. Based on the estimate from Figure 3, the 95% quantile of the

![Calendar time effect](calendar_time_effect.png)

![Age effect](age_effect.png)

![Reporting delay hazard rate](reporting_delay_hazard_rate.png)

**Figure 3** Posterior mean of $f_1$, $f_2$ and $g$
distribution of the reporting delays is 4 months, and almost 75% of the claims are reported within a month.

Figure 4 shows the estimates (posterior means) of the $h$'s and $q$'s. The posterior means of the $h$'s and $q$'s are smooth functions. However, for computational simplicity the estimates are calculated at a relatively small number of points, and the estimates may therefore appear a bit ragged. Before the first payment, the different types of events, settlement with payment, payment only, and settlement only, occur with intensities $h_{sep}^0$, $h_p^0$ and $h_{sep}^0$, respectively. The intensity of settling only is almost 0. The intensity of having a partial payment only has essentially the same structure. There is 5% probability that no events have occurred within 6 months from notification. After the first payment, the different types of events occur with intensities $h_{sep}^1$, $h_p^1$ and $h_{sep}^1$, respectively. Here the probability of settling only ($h_{sep}^1$) seems to be considerable, especially when the time since the latest payment is long. There is almost 50% probability that no events have occurred within 6 months from notification, and there is 5% probability that no events have occurred within 5 years. This means that for a claim with observed previous payments the expected settling time is long. Payments that are not made at the time of settlement are distributed with hazard rates $q_p^0$ and $q_p^1$. The corresponding densities are called $p_p^0$ and $p_p^1$, respectively. Now, 19% of all observed first-time
payments not made at the time of settlement were of the size of 178, 185, 186 or 189 DKK. These amounts are probably fees for some standard dental examination. Our estimate of $p^0_p$ therefore has a high peak in this area (the peak has been truncated on the graph), and it is questionable whether it is reasonable to model the distribution of these payments as a continuous distribution. None of the other types of payment (first-time payment at time of settlement, or subsequent payments) had such very frequently occurring sizes. The average first-time payment not made at time of settlement is approximately 1000 DKK. The distribution of the subsequent payments not made at time of settlement, $p^1_p$, has the same shape as $p^0_p$, except for the high peak. Based on the estimate from Figure 4, the mean of these payments is approximately 2100 DKK. The distributions of the payments that are made at the time of settlement, $p^0_{sep}$ and $p^1_{sep}$, seem to be independent of whether any previous payments have been made. The distributions have 2 peaks; one at approximately 200 DKK and one at approximately 3500 DKK. The average amounts are approximately 4200 and 4600 DKK.

Figure 5 contains our main results. Here the (approximated) predictive distribution of the outstanding liabilities are shown. Recall that the original data set contained information about occurrence times, covariates and reporting delays, on all claims which occurred between January 1, 1982 and December 31, 1990 and which were reported before March 3, 1992. Furthermore, all payments made upon

![Number of IBNR claims](image)

![IBNR liabilities](image)

![RBNS liabilities](image)

![IBNR and RBNS liabilities](image)

FIGURE 5 Predictive distribution of outstanding liabilities; observed accumulated run-off March 3, 1992 (4.2 years after time $\tau$) is shown with dotted lines
these claims before March 3, 1992, were observed. In our estimation/prediction we only use data that were available at December 31, 1987 ($\tau = 6$ years after January 1, 1982). The predictive distributions of the outstanding liabilities could therefore have been calculated at that time. The distributions are calculated using the method indicated in (5.2). We have the observed run-off for a period of 4.2 years. On the graphs the accumulated run-offs are shown with dotted vertical lines. There are 68 observed IBNR claims. This corresponds to the 39% quantile of the predictive distribution of the IBNR claims. The posterior distribution of the reporting delays has almost no mass for delays exceeding 3 years, and we therefore expect that all IBNR claims are reported within the 4.2 years. Table 2 shows mean, coefficient of variation (c.v.) and quantiles, from the predictive distributions of the outstanding liabilities, cf. Figure 5. The coefficient of variation is 0.22 in the predictive distribution of the outstanding IBNR liabilities, while it is only 0.09 for the RBNS liabilities. Intuitively, this is obvious; only the sizes of the RBNS claims are unknown, while for the IBNR claims also the number of claims is unknown. After 4.2 years the observed run-off for both the IBNR and the RBNS liabilities are below the 0.2 quantile of the predictive distribution of the outstanding liabilities. That was to be expected: 10 of the 68 observed IBNR claims were not yet settled after 4.2 years, and therefore there still remained some payments on the IBNR claims. Recall that RBNS claims with previous payments tend to have a long settlement delay. Of the observed 617 RBNS claims, 182 were not settled after 4.2 years. Some of these will settle without further payments, but still, the outstanding liabilities will exceed the observed 1,910,000 DKK. By combining the sampled values of the IBNR and the RBNS liabilities and using (5.2), we get an approximation to the predictive distribution of the outstanding IBNR and RBNS liabilities, shown at the last graph in Figure 5. The mean and quantiles are found in Table 2.

With the predictive distribution of the outstanding liabilities in hand, the insurance company can decide the size of the reserve to be set aside. Now, there is a possible extension at this stage. For deciding on an investment policy, it is useful for the company to know when the payments are due. This calls for the posterior (predictive) distribution of the run-off over calendar time. As previously, it is no mathematical problem to approximate this distribution; as a function of the un-

---

**TABLE 2**

<table>
<thead>
<tr>
<th></th>
<th>IBNR</th>
<th>RBNS</th>
<th>IBNR and RBNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>337,000 DKK</td>
<td>2,090,000 DKK</td>
<td>2,430,000 DKK</td>
</tr>
<tr>
<td>5% quantile</td>
<td>229,000 DKK</td>
<td>1,780,000 DKK</td>
<td>2,080,000 DKK</td>
</tr>
<tr>
<td>95% quantile</td>
<td>468,000 DKK</td>
<td>2,410,000 DKK</td>
<td>2,800,000 DKK</td>
</tr>
<tr>
<td>c.v.</td>
<td>0.22</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>4.2 years run-off</td>
<td>246,000 DKK</td>
<td>1,910,000 DKK</td>
<td>2,160,000 DKK</td>
</tr>
<tr>
<td>quantile for run-off</td>
<td>9%</td>
<td>18%</td>
<td>11%</td>
</tr>
</tbody>
</table>
knowns, sampled at each step of the algorithm, we get a value of the run-off over calendar time, and using these the desired distribution can be approximated.

CLOSING REMARKS

The distribution of the claims process was described by 14 one-dimensional components which were modelled in a nonparametric Bayesian way. We find the Bayesian approach very apt. A standard parametric approach to the prediction problem could be as follows. First, the distribution of the outstanding liabilities is found as a function of the parameters. Then, estimators (maximum likelihood, say) of the parameters are found. And finally, a reserve is calculated as a function of the estimators. Furthermore, the uncertainty about the parameter estimates can be incorporated into the reserve estimate. With the Bayesian approach the procedure takes place in a single step. The posterior distribution of unknown parameters and the predictive distribution of the outstanding liabilities are found simultaneously. Only the latter distribution is needed to predict the outstanding liabilities, since the uncertainty about the parameters is a part of the variation (variance) in that distribution.

We were a little less enthusiastic about the nonparametric modelling; the computations turned out very time consuming, and sometimes additional structure is needed (cf. the discussion in Subsection 3.3). In the future, we might want to model some components nonparametrically and some parametrically.

If we were to expand the model, then we would look at the partial payment processes. Claim handlers sometimes have additional information about the claims reported. Often, a claim handler forms an idea of the size of the total claim amount, and such information can be useful when predicting the outstanding payments on the RBNS claims.

ACKNOWLEDGMENT

Helpful comments and suggestions from Dario Gasbarra, Martin Jacobsen and Ragnar Norberg are acknowledged. Data were supplied by the Danish insurance company Topdanmark. This research was supported by The Academy of Finland.

REFERENCES


ON THE HEDGING PORTFOLIO OF ASIAN OPTIONS

MICHEL JACQUES
École d'actuariat, Université Laval

ABSTRACT

We give 2 explicit formulae for the hedging portfolio of Asian options. One is based on the usual Lognormal approximation, and the other on an Inverse Gaussian approximation. Both give excellent results as replicating strategies when the parameters of the model are in a reasonable range.

KEYWORDS

Option pricing theory; Asian option; replicating strategy; hedging portfolio.

1. INTRODUCTION

Asian options are path-dependent contingent claims since they settle against some kind of average of the quoted stock prices over a prescribed period. In the case where the average is the continuous (integral) average, Geman and Yor (1993) have obtained an expression for the Laplace transform of the price. The inverse Laplace transform however has to be taken numerically. In the case where the average is the discrete geometric average, exact expressions of the Black-Scholes type can be obtained (Kemna and Vorst (1990)).

In this paper we consider the case where the average is the (discrete) arithmetic average of the last $n$ prices. The challenge with these options is that the distribution of the average is not known when we make the usual assumption that the stock price follows a geometric Brownian motion. Up to now there is no way of computing exactly the price of these Asian options. Also, since the payoff considered here is different from the payoff of continuous average options, we should be careful if we were to approximate the price of the arithmetic average option by the price of the continuous average option. This is not what we are going to do here, since we prefer a more direct approach (of the Black-Scholes type) to a numerical inversion of a Laplace transform.

Different approaches for the evaluation of the price of (arithmetic average) Asian options have been taken. First Kemna and Vorst (1990) used Monte-Carlo simulation. This is however time-consuming and does not lead to the hedging portfolio. A faster approach based on an Edgeworth expansion around the Lognormal distribution has
been given by Turnbull and Wakeman (1991). Levy (1992) remarked that the higher terms in the Edgeworth expansion have a negligible numerical value for reasonable values of the parameters of the model and proposed to simply use a Lognormal approximation for pricing, giving closed-form formulae for the approximate price.

The aim of the present paper is to extend this approximate analysis to the construction of the hedging portfolio. If we had an exact formula for the price, standard delta hedging would provide a dynamical replicating strategy (Bergman (1985)). This is a rule indicating how to invest the price of the option so that the investment reproduces the value of the option at maturity. Since we have here an approximate formula for the price, we propose to study the accuracy of the approximation by computing the replicating strategy and by comparing its maturity value with the maturity value of the option. By the same procedure, we also remark that an approximation of the distribution of the average by an Inverse Gaussian distribution gives results comparable to those of the Lognormal approximation.

The paper is organized as follows. We first recall the recursive computation of the first two moments of the arithmetic average and then give evidence that an Inverse Gaussian approximation is just as good as a Lognormal approximation. Using these two approximations we derive explicit formulae for the hedging portfolio and show through numerical examples that these formulae are efficient in the sense that the replicating strategy is close to the intrinsic value of the option at maturity time.

2. COMPUTATION OF THE MOMENTS

We assume that there is on the market a stock with price described by a stochastic process \( \{S(t), t \geq 0\} \) following a geometric Brownian motion:

\[
dS(t) = aS(t) \, dt + bS(t) \, dB(t),
\]

where \( \{B(t), t \geq 0\} \) is a standard Brownian motion. The parameter \( a \) is the mean rate of return of the stock and the parameter \( b \) is called its volatility. There is also on the market a risk-free instrument (the bond) with a constant rate of return \( r \). We also assume that the bond and stock may be purchased in any (fractional) amount and that there is no restriction on short sales. By arbitrage arguments (Harrison and Pliska (1981), Harrison and Kreps (1979)), we have to change the probability measure so that \( a = r \) in all pricing computations.

The Asian options considered in this paper are derivatives which settle at a future date \( T \) against the arithmetic average of the last \( n \) quoted prices of the stock. The payoff of the option is thus

\[
\max\left\{0, \frac{S(T-n+1) + \ldots + S(T)}{n} - K\right\},
\]

where \( K \) is the option strike price. In the sequel, we assume that time is measured in units so that \( T \) is integral, hence the quoted prices are taken at integral times.

Because of (1), the random variable \( S(t) \) has a Lognormal distribution. The distribution of the sum of (dependent) Lognormal distributions is at present not known.
However, the moments of such a sum can be computed in a recursive way. Computing the moments directly by summing up the terms becomes complicated due to the dependence of the terms. An easy way of avoiding this is to work backwards, starting with the last summand and to isolate the (independent) increments of Brownian motion. This approach has been taken in Turnbull and Wakeman (1991) and is given here in order to fix notations.

Let

$$Y(r,b,n,T) = \frac{S(T-n+1) + \cdots + S(T)}{n}$$

be the arithmetic average of the last \( n \) stock prices. In the present section, we compute the moments of \( Y \), given that \( S(t) = x \) with \( t < T-n+1 \). (The case \( t \geq T-n+1 \) can be treated along the same lines, see section 4.) We have

$$Y(r,b,n,T) = \frac{1}{n} S(T-n+1) \left[ 1 + \exp \left( \frac{r-b^2}{2} + bN_{T-n+2} \right) + \cdots \right]$$

where \( \{N_t, t = T-n+2, \ldots, T\} \) are standardized Normal variables which are independent due to the independence of the increments of Brownian motion. Let then

$$Y_0 = 1 + \exp \left( \frac{r-b^2}{2} + bN_T \right)$$

for \( k = 1 \ldots n-2 \). Note that \( N_{T-k} \) is independent of \( Y_{k-1} \). We are interested in

$$Y(r,b,n,T) = \frac{1}{n} S(T-n+1) Y_{n-2}.$$
Note again that $S(T - n + 1)$ is independent of $Y_{n-2}$. Taking the first two moments of (2–4), we get

$$E[Y_0] = 1 + e^r,$$

$$E[Y_0^2] = 1 + 2e^r + e^{2r+b^2},$$

$$E[Y_k] = 1 + e^r E[Y_{k-1}],$$

$$E[Y_k^2] = 1 + 2e^r E[Y_{k-1}] + e^{2r+b^2} E[Y_{k-1}^2].$$

$$E[Y(r, b, n, T)] = \frac{1}{n} E[S(T - n + 1)] E[Y_{n-2}],$$

$$E[Y(r, b, n, T)^2] = \frac{1}{n^2} E[S(T - n + 1)^2] E[Y_{n-2}^2].$$

Proceeding recursively, or using symbolic software to solve for the recursive equations (5–8), we get

$$E[Y_{n-2}] = \frac{e^m - 1}{e^r - 1},$$

$$E[Y_{n-2}^2] = \frac{A(r, b, n)}{(e^r - 1)(e^{2r+b^2} - 1)(e^r - e^{2r+b^2})},$$

where the numerator $A(r, b, n)$ is given by the expression:

$$A(r, b, n) = -2e^{r(n+1)} + 2e^{r(n+3)+b^2} + e^{2r} - e^{2r(n+3)+b^2 + b^2} + e^{r(2n+1)+b^2}.$$

Let now $m_1(x, t; r, b, n)$ and $m_2(x, t; r, b, n)$ be the first and second moment of the arithmetic average of $n$ quoted prices of the stock, computed $t$ periods before the first price to enter in the arithmetic mean, given that the current stock price is $x$. Using (9–12), we immediately get

$$m_1(x, t; r, b, n) = \frac{xe^t(e^m - 1)}{n(e^r - 1)},$$

$$m_2(x, t; r, b, n) = \frac{x^2}{n} \frac{e^{(2r+b^2)t} A(r, b, n)}{(e^r - 1)(e^{2r+b^2} - 1)(e^r - e^{2r+b^2})}.$$

We will use the two explicit formulae (13) and (14) in order to obtain theoretical results in the sequel, but we will not use them for numerical evaluation, since (14) gives a ratio of extremely small numbers for reasonable values of the parameters $r$ and $b$ and is thus numerically unusable. The expression (14) could be rewritten in an algebraically equivalent form which is numerically usable. This form is given in the appendix of Levy (1992). However, we found it as easy to use the recursive formulae (5–10) to numerically evaluate the moments.
3. TWO APPROXIMATIONS TO THE DISTRIBUTION OF THE ARITHMETIC MEAN

Since the distribution of $Y(r, b, n, T)$ is not known, we have to use some kind of approximation based on our knowledge of the moments. Levy (1992) gave statistical tests on the third and fourth moments in order to substantiate the Lognormal approximation. One interesting point is that the recursive equation (3) is similar to the recursive equation used in Risk Theory to compute the distribution of the total portfolio claim in the individual model (Bowers et al. (1986)). The difference is that equation (3) is a relation between the independent variables $N_{T+k}$ and $Y_{k-1}$ which is more complicated than the usual sum of independent variables of Risk Theory. However it is natural to consider other approximations as is usually done in Risk Theory.

Different choices of distributions have been tried, most of them giving obviously bad results for the option price. However the approximation of the distribution of the arithmetic average by an Inverse Gaussian distribution (IG) gives prices comparable to those given by the Lognormal approximation (LN) when the parameters are chosen in the same range as in Levy (1992), Kemna and Vorst (1990) or Turnbull and Wakeman (1991).

Table 1 gives the price of an option computed with the formulae given later in the paper (see (29) and (46)), based on the two approximations. The parameters in Table 1 are chosen as an initial stock price $S(0) = 100$, an annual (nominal, daily compounded) interest rate of 9% (i.e. $r = \ln(1 + \frac{0.09}{365})$ daily), a maturity $T$ of 120 days and an averaging period $n$ of 30 days. The values of the volatility $b$ are also on an annual basis. Some of these numerical values may be considered as being high nowadays, but they are chosen to be in the same range as in the papers cited above.

<table>
<thead>
<tr>
<th>$b$</th>
<th>K</th>
<th>LN</th>
<th>IG</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>90</td>
<td>12.68</td>
<td>12.68</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>5.46</td>
<td>5.46</td>
</tr>
<tr>
<td>110</td>
<td></td>
<td>1.63</td>
<td>1.63</td>
</tr>
<tr>
<td>0.3</td>
<td>90</td>
<td>13.85</td>
<td>13.85</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>7.48</td>
<td>7.49</td>
</tr>
<tr>
<td>110</td>
<td></td>
<td>3.48</td>
<td>3.49</td>
</tr>
<tr>
<td>0.4</td>
<td>90</td>
<td>15.36</td>
<td>15.37</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>9.51</td>
<td>9.53</td>
</tr>
<tr>
<td>110</td>
<td></td>
<td>5.48</td>
<td>5.49</td>
</tr>
</tbody>
</table>

As we see from Table 1, the agreement between the prices computed according to the two approximations is excellent. We want to give further evidence for the use of these approximations by testing directly the distribution and not only its moments. This can be done by measuring the distance between a simulated sample of the true (unknown) distribution of the arithmetic mean and the distribution of the approximating variable.
First, let us recall that $X_1$ has a Lognormal distribution with parameters $\mu$ and $\sigma$ (LN($\mu$, $\sigma$) in short) if its density is

$$f_{X_1}(x) = \frac{\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)}{x\sigma\sqrt{2\pi}},$$

for $x > 0$. A straightforward calculation gives

$$E[X_1] = e^{\mu + \frac{\sigma^2}{2}},$$

$$E[X_1^2] = e^{2\mu + 2\sigma^2}.$$  

On the other hand, $X_2$ has an Inverse Gaussian distribution with parameters $\rho$ and $\beta$ (IG($\rho$, $\beta$) in short) if its density is

$$f_{X_2}(x) = \frac{\rho \exp\left(-\frac{(x-\rho)^2}{2\beta x}\right)}{\sqrt{2\pi}\beta},$$

for $x > 0$. This distribution (with an appropriate choice of parameters) appears as the distribution of the first time when Brownian motion hits a barrier. The reader may consult Johnson and Kotz (1970), chapter 15 for more information. Up to a transformation of parameters, the above density is equation (4.1), page 138 of that reference. A straightforward calculation gives

$$E[X_2] = \rho,$$

$$E[X_2^2] = \beta\rho + \rho^2.$$  

The approximations are obtained by identifying the first two moments of both distributions with the first two moments of $Y(r, b, n, T)$. We thus choose the parameters as

$$\mu = 2\ln m_1 - \frac{1}{2} \ln m_2,$$

$$\sigma^2 = \ln m_2 - 2\ln m_1,$$

for the Lognormal approximation and

$$\rho = m_1,$$

$$\beta = \frac{m_2 - m_1^2}{m_1},$$

for the Inverse Gaussian approximation. Notice that $m_1$ and $m_2$ are shorthand for $m_1(S(0), T - n + 1; r, b, n)$ and $m_2(S(0), T - n + 1; r, b, n)$ and, similarly, the 4 parameters $\mu$, $\sigma$, $\rho$ and $\beta$ are functions of $(S(0), T - n + 1; r, b, n)$.

We then proceed as follows. We generate $N$ trajectories of the geometric Brownian motion (stock price) and record for each trajectory the value of the arithmetic average of the last $n$ prices. This gives a sample of size $N$ of $Y(r, b, n, T)$, say $\{y_1, \ldots, y_N\}$. We can measure the distance between the cumulative distribution $F_N$ of this sample of the
exact distribution and the cumulative distribution $F$ of either a LN or IG variable by a kind of Cramér-von Mises distance:

$$d_1 = \sum_{i=1}^{N} (F_N(y_i) - F(y_i))^2.$$  

(See Hogg and Klugman (1984), page 83). In the computation of the option price, the small values of $Y$ (i.e. values less than $K$) do not appear. We might thus want to give more weight to the right tail of the distribution and use a weighted Cramér-von Mises distance:

$$d_2 = \sum_{i=1}^{N} \frac{(F_N(y_i) - F(y_i))^2}{1 - F(y_i)}.$$ 

For $N = 10,000$ and the same numerical values as in Table 1, the values of the two distances for the two approximations are given in Table 2.

<table>
<thead>
<tr>
<th>$b$</th>
<th>$d_1$ LN</th>
<th>$d_1$ IG</th>
<th>$d_2$ LN</th>
<th>$d_2$ IG</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.42</td>
<td>0.43</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>0.3</td>
<td>0.43</td>
<td>0.44</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>0.4</td>
<td>0.28</td>
<td>0.27</td>
<td>0.78</td>
<td>0.85</td>
</tr>
</tbody>
</table>

It is clear from Table 2 that the two approximating distributions are at the same distance from the distribution of the sample and this partially explains that the prices computed according to the two approximations are close to each other. Other candidate distributions have been tried and the corresponding distances are substantially larger. As a point of reference, the distances computed with a Normal distribution are 10 to 30 times larger than the distances reported in Table 2. The conclusion is that the IG approximation should thus be considered on the same footing as the usual LN approximation.

The above values of the distances have no absolute meaning and thus do not provide a measure of the quality of the approximations. However, if we make the assumption that $Y$ has one of the two approximate distributions, we can compute the hedging portfolio explicitly. Then we can check how well this portfolio replicates the value of the option along a random trajectory of the stock price. Pricing of options is performed in an arbitrage-free setting and arbitrage-free evaluation is tantamount to the existence of a replicating strategy. We thus propose to assess the quality of the approximations by using the very key ingredient of option pricing theory.
4. THE REPLICATING STRATEGY IN THE LOGNORMAL APPROXIMATION

Let \( u(x, t; r, b, n, T, K) \) be the price of the option sold at time \( t \), given that the price of the stock is \( S(t) = x \). If we build the replicating strategy determined by delta hedging, we need

\[
\xi(x; t; r, b, n, T, K) \equiv \frac{\partial u}{\partial x}(x, t; r, b, n, T, K).
\]  

(25)

Defining now \( \eta \) by

\[
\eta \equiv \frac{u(x, t; r, b, n, T, K) - \xi x}{S(0)e^{rT}},
\]  

(26)

we can interpret \( \xi \) as the number of shares of stock and \( \eta \) as the number of shares of the bond (measured in units of \( S(0) \)), since we have:

\[
u(x, t; r, b, n, T, K) = \xi x + \eta S(0)e^{rT}.
\]  

(27)

By arbitrage arguments (Harrison and Pliska (1981), Harrison and Kreps (1979)), it is well-known that the price \( u(x, t; r, b, n, T, K) \) is computed as

\[
u(x, t; r, b, n, T, K) = e^{-r(T-t)}\mathbb{E}\left[\max\{0, Y(r, b, n, T) - K\} | S(t) = x\right]
\]  

(28)

where \( f_Y \) is the probability density of \( Y \), approximated here by the density of a variable of type LN \((\mu, \sigma)\). In order to compute this integral, we need to relate the parameters \( \mu \) and \( \sigma \) to the option parameters. For this, we distinguish 3 cases.

4.1. The case \( t < T - n + 1 \)

If we are before the averaging period, \( \mu = \mu(x, T - t - n + 1; r, b, n) \) and \( \sigma = \sigma(x, T - t - n + 1; r, b, n) \) (see (21), (22)). A straightforward integration gives an analogue of the Black-Scholes formula:

\[
u(x, t; r, b, n, T, K) = e^{-r(T-t)}\Phi\left(\frac{\sigma^2 + \mu - \ln K}{\sigma}\right) - K\Phi\left(\frac{\mu - \ln K}{\sigma}\right),
\]  

(29)

where \( \Phi \) is the cumulative probability function of a standardized Normal variable and the parameters \( \mu \) and \( \sigma \) are evaluated at \((x, T - t - n + 1; r, b, n)\). Notice that the integration is well-known from the computation of Stop-Loss premiums. In order to get the replicating strategy, we need to differentiate \( u(x, t; r, b, n, T, K) \) with respect to \( x \). The only \( x \) dependence in \( u \) is through the parameters \( \mu \) and \( \sigma \). By (13) and (14), we have

\[
\frac{\partial m_1(x, t; r, b, n)}{\partial x} = \frac{m_1(x, t; r, b, n)}{x},
\]  

(30)

\[
\frac{\partial m_2(x, t; r, b, n)}{\partial x} = \frac{2m_2(x, t; r, b, n)}{x},
\]  

(31)
and we easily get, using (21) and (22), that
\[
\frac{\partial \mu(x,t; r, b, n)}{\partial x} = \frac{1}{x}, \quad (32)
\]
\[
\frac{\partial \sigma(x,t; r, b, n)^2}{\partial x} = 0. \quad (33)
\]
Differentiating (29) with (32) and (33) and simplifying, we have
\[
\xi(x,t; r, b, n, T, K) = e^{-r(T-t)} \frac{1}{x} e^{\mu + \sigma^2/2} \Phi \left( \frac{\sigma^2 + \mu - \ln K}{\sigma} \right), \quad (34)
\]
where the parameters \(\mu\) and \(\sigma\) are evaluated at \((x, T-t-n+1; r, b, n)\).

4.2. The case \(t \geq T - n + 1, t\) non-integral

Now some values of the stock price entering into the arithmetic mean are known. Let \(\tilde{t}\) be the integral part of \(t\). Then
\[
u(x,t; r, b, n, T, K) = e^{-r(T-t)} E\left[ \max \left\{ 0, \frac{1}{n} (S(T) + S(T-1) + \ldots + S(\tilde{t} + 1) + S(T-n+1) - K) | S(t) = x \right\} \right], \quad (35)
\]
where
\[
\gamma = \frac{S(T) + S(T-1) + \ldots + S(\tilde{t} + 1)}{T - \tilde{t}}, \quad (36)
\]
\[
K' = \frac{nK - (S(T-n+1) + S(T-n+2) + \ldots + S(\tilde{t}))}{T - \tilde{t}}. \quad (37)
\]
The price is similar to the price in (28), except that the arithmetic average is now taken over \(T - \tilde{t}\) future prices instead of \(n\) prices, the strike price is rescaled to \(K'\) and we compute the price \(1 - t + \tilde{t}\) periods before the first price to enter into the mean. If \(K' > 0\), the situation is similar to that in the previous subsection and, up to a global factor, \(u(x, t; r, b, n, T, K)\) takes the same form as in (29) with \(K\) replaced by \(K'\) and the parameters evaluated at \((x, 1 - t + \tilde{t}; r, b, T - \tilde{t})\). If \(K' \leq 0\), \(\max \{0, \gamma - K'\} = \gamma - K'\) and the expectation gives \(E[\gamma'] - K'\). Hence, if \(K' \leq 0\),
\[
u(x,t; r, b, n, T, K) = e^{-r(T-t)} \frac{T-\tilde{t}}{n} (e^{\mu + \sigma^2/2} - K'),
\]
with the parameters evaluated at \((x, 1 - t + \tilde{t}; r, b, T - \tilde{t})\). Differentiating with respect to \(x\) thus gives
\[
\xi(x,t; r, b, n, T, K) = e^{-r(T-t)} \frac{T-\tilde{t}}{n} \frac{1}{x} e^{\mu + \sigma^2/2} \Phi \left( \frac{\sigma^2 + \mu - \ln K'}{\sigma} \right), \quad (38)
\]
if \( K' > 0 \), and

\[
\xi(x, t; r, b, n, T, K) = e^{-r(T-t) \frac{T-t}{n}} \frac{1}{x} e^{\mu + \sigma^2 / 2},
\]

(39)

if \( K' \leq 0 \). In both equations the parameters \( \mu \) and \( \sigma \) are evaluated at \((x, 1-t+\tilde{t}; r, b, T-\tilde{t})\).

### 4.3. The case \( t \geq T - n + 1 \), \( t \) integral

The situation is essentially the same as in the previous subsection with \( \tilde{t} = t \). The price \( u(x, r, b, n, T, K) \) is as before with the parameters evaluated at \((x, 1; r, b, T- t)\). But now, by (37), \( K' \) contains \( S(t) = x \) and is thus a function of \( x \). We have immediately that

\[
\frac{\partial K'}{\partial x} = \frac{1}{T-t}.
\]

We thus pick up an extra term when differentiating with respect to \( x \):

\[
\xi(x, t; r, b, n, T, K) = e^{-r(T-t) \frac{T-t}{n}} \frac{1}{x} e^{\mu + \sigma^2 / 2} \Phi \left( \frac{\sigma^2 + \mu - \ln K'}{\sigma} \right)
\]

\[
+ \frac{1}{T-t} \Phi \left( \frac{\mu - \ln K'}{\sigma} \right),
\]

(41)

if \( K' > 0 \), and

\[
\xi(x, t; r, b, n, T, K) = e^{-r(T-t) \frac{T-t}{n}} \left( \frac{1}{x} e^{\mu + \sigma^2 / 2} + \frac{1}{T-t} \right),
\]

(42)

if \( K' \leq 0 \). In both equations the parameters \( \mu \) and \( \sigma \) are evaluated at \((x, 1; r, b, T- t)\).

### 4.4. Examples

We give in the present subsection examples of the hedging portfolio along randomly generated trajectories of the stock price process. The portfolio described in the preceding subsections is assumed to be updated continuously. If we update it only at discrete times, we expect it to only provide an approximation to the replicating (autofinancing) strategy for the intrinsic value of the option. We present here numerical examples when the portfolio is updated 100 times between two successive quotations entering into the average. For comparison sake, we also plot on the figure the intrinsic value of the option, if it were to attain maturity at the current time. This intrinsic value is \( \max\{0, (n\text{-periods running average}) - K\} \). For \( t < n \), we need values of the stock price at negative times to compute the arithmetic average and we assume that \( S(t) = S(0), t < 0 \). In the illustrations, we take the interest rate to be 9% per annum, the volatility to be 0.2 per annum, the maturity date to be 120 days and the averaging period to be 30 days. The initial price of the stock is 100 and the strike price of the option is 90.
For pricing purposes, we need to assume \( a = r \), but here we consider true "physical" trajectories. We have thus to generate the trajectories taking into account the real rate of return \( a \) of the stock. In the illustrations, \( a \) is taken as 15\%, significantly higher than the risk-free rate.

We assume that the price of the stock is monitored at discrete times \( \{ t_1, \ldots, t_N \} \) with \( N = kT \). Thus the portfolio will be updated \( k \) times between two successive quotations entering into the average. The hedging portfolio at time \( t_k \) is denoted \( \pi(t_k) \) and is initialized at the selling price of the option:

- \( \pi(0) = u(S(0), 0; r, b, n, T, K) \)
- compute \( \xi(0) \) according to (34) and buy \( \xi(0) \) shares of stock
- invest (or borrow) the rest in the bond: \( \eta(0) = \frac{\pi(0) - \xi(0)S(0)}{S(0)} \)

Next, at each monitoring time \( t_k \), just after the stock price is revealed,

- compute the value of the portfolio: \( \pi(t_k) = \xi(t_{k-1}) S(t_k) + \eta(t_{k-1}) S(0)e^{rt_k} \)
- compute \( \xi(t_k) \) according to (34), (38), (39), (41) or (42) and buy \( \xi(t_k) \) shares of stock
- invest (or borrow) the rest in the bond: \( \eta(t_k) = \frac{\pi(t_k) - \xi(t_k)S(t_k)}{S(0)e^{rt_k}} \)

Figure 1 presents a trajectory where the option ends up in the money (i.e. is exercised) while in Figure 3 the option ends up out of the money (i.e. is not exercised). The smooth curve is the current intrinsic value of the option (smoothness being caused by averaging) and the peaky curve is the value of the hedging portfolio \( \pi \). The value of the option \( u(x,t) \) at any time is also plotted on the figures, but it is undistinguishable from the hedging portfolio \( \pi \). This means that the hedging portfolio reproduces the price of the option along the path and the intrinsic value of the option at maturity with a very good precision for those particular trajectories. It is not easy to see the improvement gained from increasing the updating frequency \( k \) on a figure, so we leave this for the numerical analysis at the end of this subsection.

Figures 2 and 4 give the components of the hedging portfolio for the trajectories of Figures 1 and 3, respectively. The positive part is the stock part \( \xi \), whereas the negative part is the bond part \( \eta \). A negative bond means it is short-sold or borrowed. The behaviour of the two components in Figure 2 after the averaging has begun is to be noted. It is more or less linear; this seems to be characteristic of Asian options.

Hedge ratios for Asian options have also been considered by Vorst (1992), but his approximation is based on the geometric average option and is thus different from (and more complicated than) the approximation given here. Moreover he only gave the hedge ratio for the case \( t < T - n + 1 \) and then compared the result with Monte-Carlo simulation at a particular time. This information is clearly not sufficient to construct
the portfolio along the whole trajectory. Since we have only approximate formulae for the hedge ratio, it is of crucial importance to check that the approximate hedging portfolio is close to the intrinsic value of the option at maturity date. Indeed the error contained in the hedge ratio at a particular time could propagate in time and the dynamically constructed portfolio could substantially deviate from the intrinsic value of the option at maturity time. It is apparent from the trajectories in Figures 1 and 3 that it is not the case for those particular trajectories.

A more detailed analysis of the accuracy of the hedging portfolio can be made. Since the intrinsic value of the option at maturity can take arbitrarily small or arbitrarily large values, a natural candidate would be the relative error of the ultimate value of the hedging portfolio with respect to the intrinsic value of the option. However this excludes the cases where the intrinsic value vanishes. We are thus going to study the absolute error of the ultimate value of the hedging portfolio with respect to the intrinsic value of the option at maturity:

$$\varepsilon = \text{hedging portfolio} - \text{intrinsic value}.$$  

We randomly generate 1,000 trajectories of the stock price process with the same numerical values of the parameters as for the figures above. We record the corresponding errors, thus providing a sample of size 1,000 for \(\varepsilon\). The sample is generated with the hedging portfolio updated \(k\) times between two successive quotations entering into the average, with \(k = 1, 10, 100, 1000\) and 10000. The large values of \(k\) may be consi-
dered as excessive from a practical point of view, but the idea here is to study the quality of the approximation. In order to use the hedging portfolio for that purpose, we need in principle to update the portfolio continuously. The large values of $k$ should be seen as this continuous limit and not as a practical benchmark.

The proportion of trajectories in the sample with $|\epsilon| < 0.1$ is given in the LN column of Table 3. Recall from Table 1 that the price of the option for the parameter values considered is 12.68, meaning that the average payoff is about 10. An error of 0.1 may then be considered as small. Smaller values of $\epsilon$ would require larger values of the updating frequency $k$ in order to achieve good behaviour. This is also shown in Table 3 (LN column) for the proportion of trajectories with $|\epsilon| < 0.01$.

**TABLE 3**

<table>
<thead>
<tr>
<th>Error on the Hedging Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>$k$</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>1000</td>
</tr>
<tr>
<td>10000</td>
</tr>
</tbody>
</table>
We are thus entitled to conclude that, as long as the hedging portfolio is updated sufficiently often, the approximation is of extremely good quality. The results presented in Table 3 provide a far stronger assessment of the quality of the approximation than mere Monte-Carlo simulation at a particular time. It is to be noted that this assessment is indeed strong since the trajectories have been generated with a value of the stock mean rate of return \( a \) different from the risk-free interest rate \( r \) (the independence of the price from \( a \), guaranteed by the theory, could have been broken by the approximation).

4.5. The limiting portfolio

One interesting question is to know the value that the stock component of the hedging portfolio approaches when time approaches maturity. For a standard Black-Scholes option with payoff at maturity \( \max( S(T) - K, 0) \), it is intuitive that the portfolio should ultimately be made of one share of stock if the option ends up in the money and nothing if the option ends up out of the money. Indeed, if the buyer of the option is to exercise the right to buy at a strike price lower than the current price of the stock, the seller of the option will have the share of stock at hand and will use the exercise price to cover the short position in the bond.

For Asian options, the intuition is not as clear since the comparison takes place with the arithmetic average of the last \( n \) quoted prices. A priori the ultimate stock component of the portfolio might be path dependent. An analysis of many graphs of the type presented in the previous subsection shows that it is not the case. The intrinsic value of
the option is \((S(T - n + 1) + \ldots + S(T - 1) + x) / n - K\) immediately before maturity if
the option is to end up in the money and 0 otherwise. If we are allowed to interchange
the differentiation process with the limit process, we get that \(\xi = 1 / n\) if the option
ends up in the money and \(\xi = 0\) otherwise. We proceed now to take this limit directly
on our expression for \(\xi\).

What we need is the limit of (38) or (39) when \(t\) approaches \(T\). Thus \(\tilde{t} = T - 1\) and,
after some algebraic manipulations, we see from (13) and (14) that
\[
\lim_{t \to T} m_1(x, 1 - t + \tilde{t}; r, b, T - \tilde{t}) = x,
\]
\[
\lim_{t \to T} m_2(x, 1 - t + \tilde{t}; r, b, T - \tilde{t}) = x^2.
\]

Hence, from (21) and (22),
\[
\lim_{t \to T} \mu(x, 1 - t + \tilde{t}; r, b, T - \tilde{t}) = \ln x,
\]
\[
\lim_{t \to T} \sigma(x, 1 - t + \tilde{t}; r, b, T - \tilde{t})^2 = 0.
\]

In the limit when \(t \to T\), the condition \(\mu - \ln K' > 0\) is thus equivalent to \(S(T) > nK -
(S(T - 1) + \ldots + S(T - n + 1))\), i.e. the option ends up in the money. Similarly, \(\mu - \ln K' < 0\)
is equivalent to the option ending out of the money. Using \(\Phi (+\infty) = 1, \Phi (-\infty) = 0\), we obtain
\[
\lim_{t \to T} \xi(x, t; r, b, n, T, K) = \begin{cases} 1 / n : \text{in the money}, \\ 0 : \text{out of the money}. \end{cases}
\]
(Note that the limit in (39) is immediate.) These ultimate values of $\xi$ are indeed the values appearing in Figures 2 and 4.

5. THE REPLICATING STRATEGY IN THE INVERSE GAUSSIAN APPROXIMATION

Since the conceptual steps for the Inverse Gaussian approximation are identical to those for the Lognormal approximation, we just cite the results and indicate the computational tricks.

5.1. The case $t < T - n + 1$

A straightforward integration (again well-known from the computation of Stop-Loss premiums) gives

$$u(x, t; r, b, n, T, K) = e^{-r(T-t)} \left[ (\rho - K) \Phi \left( \frac{\rho - K}{\sqrt{\beta K}} \right) + (\rho + K) e^{2\rho \beta} \Phi \left( -\frac{\rho + K}{\sqrt{\beta K}} \right) \right],$$

with the parameters $\rho$ and $\beta$ evaluated at $(x, T-t - n + 1; r, b, n)$ (see (23), (24)). We easily see that

$$\frac{\partial \rho}{\partial x} = \frac{\rho}{x},$$
$$\frac{\partial \beta}{\partial x} = \frac{\beta}{x},$$

and, after some simplifications, we get

$$\xi(x, t; r, b, n, T, K) = e^{-r(T-t)} \frac{\rho}{x} \left[ \Phi \left( \frac{\rho - K}{\sqrt{\beta K}} \right) + e^{2\rho \beta} \Phi \left( -\frac{\rho + K}{\sqrt{\beta K}} \right) \right],$$

with the parameters $\rho$ and $\beta$ evaluated at $(x, T-t - n + 1; r, b, n)$. 

5.2. The case $t \geq T - n + 1, t$ non-integral

The situation is the same as for the Lognormal approximation and we have

$$\xi(x, t; r, b, n, T, K) = e^{-r(T-t)} \frac{T-t}{n} \rho \left[ \Phi \left( \frac{\rho - K'}{\sqrt{\beta K'}} \right) + e^{2\rho \beta} \Phi \left( -\frac{\rho + K'}{\sqrt{\beta K'}} \right) \right],$$

if $K' > 0$, and

$$\xi(x, t; r, b, n, T, K) = e^{-r(T-t)} \frac{T-t}{n} \rho \left[ \Phi \left( \frac{\rho - K'}{\sqrt{\beta K'}} \right) + e^{2\rho \beta} \Phi \left( -\frac{\rho + K'}{\sqrt{\beta K'}} \right) \right],$$

if $K' \leq 0$. In both equations the parameters $\rho$ and $\beta$ are evaluated at $(x, 1 - t + T; r, b, T - t)$. 

5.3. The case \( t \geq T - n + 1, t \) integral

Again proceeding as in the Lognormal case, we get

\[
\xi(x, t; r, b, n, T, K) = e^{-r(T-t)} \frac{T-t}{n} \left[ \left( \frac{\rho + 1}{T-t} \right) \Phi \left( \frac{\rho - K}{\sqrt{\beta K'}} \right) + \left( \frac{\rho - 1}{T-t} \right) e^{2\rho/\beta} \Phi \left( -\frac{\rho + K}{\sqrt{\beta K'}} \right) \right],
\]

(50)

if \( K' > 0 \), and

\[
\xi(x, t; r, b, n, T, K) = e^{-r(T-t)} \frac{T-t}{n} \left( \frac{\rho}{x} + \frac{1}{T-t} \right),
\]

(51)

if \( K' \leq 0 \). In both equations the parameters \( \rho \) and \( \beta \) are evaluated at \((x, 1; r, b, T-t)\).

5.4. Examples

The same trajectories have been considered as for the Lognormal approximation. The formulae for the hedging portfolio using the Inverse Gaussian approximation present a numerical difficulty due to the factor \( e^{2\rho/\beta} \). For the numerical values used in section 4.4, the exponential factor blows up on most computer systems. One way to get rid of this problem is to use the Mill ratio for the tail of the Normal cumulative probability function (see for instance Resnick (1992), p. 487):

\[
1 - \Phi(x) \sim \frac{\varphi(x)}{x} \text{ as } x \to +\infty,
\]

\[
\Phi(x) \sim -\frac{\varphi(x)}{x} \text{ as } x \to -\infty,
\]

where \( \varphi \) is the probability density of a standardized Normal variable. This means that

\[
e^{2\rho/\beta} \Phi \left( -\frac{\rho + K}{\sqrt{\beta K'}} \right) \sim \frac{\sqrt{\beta K'}}{\rho + K} \Phi \left( \frac{\rho - K}{\sqrt{\beta K'}} \right) \text{ as } \beta \to 0.
\]

(52)

This approximation is used as soon as the ratio \( 2\rho/\beta \) attains a value that makes the exponential function blow up on the computer system under use.

With this additional feature, the resulting figures are completely undistinguishable from those presented in section 4.4: we could have given them on the same figures, but the reader would not have noticed them.

As far as the numerical analysis of the error of the hedging portfolio is concerned, the corresponding values are also given in the IG columns of Table 3. It is clear from those values that the Inverse Gaussian approximation should be considered as efficient as the Lognormal approximation. All the comparisons used in this paper do not allow for a preference between the two approximations.
5.5. The limiting portfolio

Using (52) together with
\[
\lim_{t \to T} \rho(x, 1 - t + i; r, b, T - i) = x, \\
\lim_{t \to T} \beta(x, 1 - t + i; r, b, T - i) = 0,
\]
we arrive at the same conclusion as in the Lognormal approximation:
\[
\lim_{t \to T} \xi(x, t; r, b, n, T, K) = \begin{cases} 
1/n &: \text{in the money,} \\
0 &: \text{out of the money.}
\end{cases}
\]
(53)

6. CONCLUSION

This paper gives explicit formulae for building the hedging portfolio for Asian (arithmetic average) options. The formulae are based on the approximation of the arithmetic average of correlated Lognormal variables by either a Lognormal or an Inverse Gaussian variable. Replicating strategies (a key ingredient of option pricing theory) have been built with the help of these formulae and have been used as a measure of the quality of the approximations. The numerical examples given in section 4.4 and 5.4 provide a strong assessment of the good quality of both approximations. Since the distribution of the average is not known, it is important to consider all relevant approximations. The comparisons between the two approximations show that they are equally efficient.

ACKNOWLEDGEMENT

The author wishes to thank Jacques Rioux for useful discussions and the anonymous referees for constructive remarks that helped to produce a better paper.

REFERENCES


Journal of Banking and Finance 14, 113-129.

LEVY, EDMOND (1992) "Pricing European average rate currency options". Journal of International Money 
and Finance 11, 474-491.


TURNBULL, S. M. AND WAKEMAN, L. M. (1991) "A Quick Algorithm for Pricing European Average Op-

VORST, TON (1992) "Prices and Hedge Ratios of Average Exchange Rate Options". International Review of 
Financial Analysis 1, 179-193.

MICHEL JACQUES
École d'actuarial, Université Laval
Québec, Canada G1K 7P4
mjacques@act.ulaval.ca
IMPROVED ANALYTICAL BOUNDS FOR SOME RISK QUANTITIES

WERNER HÜRLIMANN
Allgemeine Mathematik, Paulstr. 9, Winterthur-Leben

ABSTRACT
Simple analytical lower and upper bounds are obtained for stop-loss premiums and ruin probabilities of compound Poisson risks in case the mean, variance and range of the claim size distribution are known. They are based on stop-loss extremal distributions and improve the bounds derived earlier from dangerous extremal distributions. The special bounds obtained in case the relative variance of the claim size is unknown, but its maximal value is known, are related to other actuarial results.

KEYWORDS
analytical bounds, stop-loss, ruin probability, stochastic orderings, relative variance, atomic distributions

1. INTRODUCTION
A main topic of risk theory under incomplete information is the construction of divers bounds for risk quantities using stochastic orderings. This well-established technic has an important impact on insurance practice. The present paper is devoted to the problem of finding simple analytical lower and upper bounds for stop-loss premiums and ruin probabilities of classical compound Poisson risks in case only the mean, variance and range of the claim size distribution are known.

In this situation the distributions with minimal and maximal stop-loss premiums of the claim size depend on the deductible and cannot be used to bound the stop-loss premiums or ruin probabilities of the induced compound Poisson risks uniformly for each deductible or initial reserve. However larger classes of claim size distributions may be allowed for ordering comparisons, which lead to uniform lower and upper bounds. For example Kaas and Goovaerts (1986) have derived distributions, which are the best lower and upper bound with respect to the dangerousness ordering criterion for any distribution with fixed mean, variance and range. In the present paper we derive similar distributions, which are best with respect to the stop-loss ordering criterion. They lead to uniformly better and even simpler analytical bounds than the previous ones.

In Section 2 the stop-loss ordered extremal distributions are introduced and compared with the dangerous extremal distributions. A main result states that the stop-loss order maximum precedes in dangerousness the dangerous order maximum, and that the dangerous order minimum precedes in dangerousness the stop-loss order minimum. In Section 3 improved analytical bounds for stop-loss premiums

and ruin probabilities of compound Poisson risks are constructed through discretization of the stop-loss ordered maximal distribution applying the technic of mass dispersion. Then we comment on two special situations. In Section 4.1 the upper bounds in case of small deductibles and initial reserves are shown to coincide with the best upper bounds derived by Kaas (1991). A discussion of the obtained bounds provided the relative variance of the claim size is unknown, but its maximal value is known, is given in Section 4.2. One recovers the safest diatomic risk with fixed mean and known range from Bühlmann et al. (1977), which can be regarded as a positive answer to the following modified Schmitter problem. Given that the individual claims have given mean and maximal variance, which claims distribution maximizes the ruin probability for a given initial reserve? In case the deductible equals the mean of the compound Poisson risk, the obtained stop-loss upper bound is shown to be closely related to earlier investigations by Benktander (1977). Finally, in Section 5 the substantial improvement in the new stop-loss bounds is demonstrated numerically.

2. STOP-LOSS ORDERED EXTREMAL VERSUS DANGEROUS EXTREMAL DISTRIBUTIONS

Consider a risk \( X \), representing claim sizes, from the set \( D = D(\mu; \sigma) \) of all random variables with fixed mean \( \mu \), variance \( \sigma^2 \), and support contained in the interval \( I_x = [0, b] \). By relative variance we mean the square of the coefficient of variation. The following notations are used for relative variances and a ratio thereof:

\[
\nu = \frac{(\sigma^2)}{\mu^2} : \text{relative variance of } X \\
\nu_o = \frac{(b - \mu)}{\mu} : \text{maximal relative variance of risks with known mean } \mu \text{ and range } [0, b] \\
\nu_r = \frac{\nu}{\nu_o} : \text{ratio of relative variances}
\]

To simplify the presentation and calculations, let us work in the standardized risk scale defined by the transformation \( Z = \frac{(X - \mu)}{\sigma} \), which is interpreted as relative signed mean deviation. Then the support of \( Z \) is \( I_z = [-1, \nu_o] \).

Extremal random variables \( X^\nu, X^\ell \), having distributions \( F^\nu(x), F^\ell(x) \), with respect to the usual stochastic dominance partial order relation \( \preceq_{st} \) have been given in Goovaerts and Kaas (1986):

\[
X^\nu \preceq_{st} X \preceq_{st} X^\ell, \text{ for all } X \in D,
\]

\[
\iff F^\ell(x) \leq F(x) \leq F^\nu(x), \text{ for all } x \in I_x.
\]

In the transformed risk scale \( Z^\nu = \frac{(X^\nu - \mu)}{\sigma}, Z^\ell = \frac{(X^\ell - \mu)}{\sigma} \), have distributions \( F^\nu(z), F^\ell(z) \) as described in Table 1.

Extremal distributions \( X^-, X^+ \) with respect to the dangerousness order relation \( \preceq_D \) are constructed in Kaas and Goovaerts (1986), and have the property \( X^- \preceq_D X \preceq_D X^+ \), for all \( X \in D \). Their standardized distributions are displayed in Table 2.
Improvement Analytical Bounds for Some Risk Quantities

Table 1
Extremal Distributions with Respect to Stochastic Dominance

<table>
<thead>
<tr>
<th>z</th>
<th>$F^t(z)$</th>
<th>$F^u(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1 \leq z \leq -\nu_r$</td>
<td>$0$</td>
<td>$v/(v + z^2)$</td>
</tr>
<tr>
<td>$-\nu_r \leq z \leq -\nu$</td>
<td>$[v_r/(1 + v_o)] \cdot (v_r + z)/(1 + z)$</td>
<td>$[v_o/(1 + v_o)] \cdot (v_o + 1 - v_r - z)/(v_o - z)$</td>
</tr>
<tr>
<td>$v \leq z &lt; v_o$</td>
<td>$\frac{z^2}{(v + z^2)}$</td>
<td>$1$</td>
</tr>
<tr>
<td>$z = v_o$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

Stop-loss ordered extremal distributions $X_*, X^*$, with the property $X_* \leq_{sd} X \leq_{sd} X^*$ can be constructed following an idea expressed in H"{u}rlimann (1993/95). If $\pi_*(x)$, $\pi^*(x)$, $0 \leq x \leq b$, are the minimal and maximal stop-loss premiums when $X$ varies over $D$, then these extremal distributions are defined using the one-to-one correspondence between a distribution and its stop-loss transform, namely as

$$F_*(x) = 1 + (d/dx)\pi_*(x), \text{ and}$$

$$F^*(x) = 1 + (d/dx)\pi^*(x), 0 \leq x \leq b.$$  

The extreme value functions $\pi_*(x)$, $\pi^*(x)$ have been obtained in De Vylder and Goovaerts (1982) (see also Goovaerts et al. (1984), p.316, Kaas et al. (1994), chap. X.2). After a straightforward calculation, one gets the standardized distributions of Table 3.

It may also be useful to know the expected values and variances of the dangerous and stop-loss ordered extremal distributions.

Table 2
Dangerous Extremal Distributions

<table>
<thead>
<tr>
<th>z</th>
<th>$F^-(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1 \leq z \leq -\nu_r$</td>
<td>$0$</td>
</tr>
<tr>
<td>$-\nu_r \leq z \leq 0$</td>
<td>$F^t(z) = [v_r/(1 + v_o)] \cdot (v_r + z)/(1 + z)$</td>
</tr>
<tr>
<td>$0 &lt; z \leq v$</td>
<td>$F^u(z) = [v_o/(1 + v_o)] \cdot (v_o + 1 - v_r - z)/(v_o - z)$</td>
</tr>
<tr>
<td>$v \leq z \leq v_o$</td>
<td>$F^w(z) = 1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>z</th>
<th>$F^+(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1 \leq z \leq z(\alpha^+)$</td>
<td>$F^u(z) = v/(v + z^2)$</td>
</tr>
<tr>
<td>$z(\alpha^+) \leq z \leq z(\beta^+)$</td>
<td>$F^w(z(\alpha^+)) = v/(v + z(\alpha^+)^2)$</td>
</tr>
<tr>
<td>$z(\beta^+) \leq z &lt; v_o$</td>
<td>$F^w(z) = \frac{z^2}{(v + z^2)}$</td>
</tr>
<tr>
<td>$z = v_o$</td>
<td>$F^w(z) = 1$</td>
</tr>
</tbody>
</table>

$z(\alpha^+) = (\alpha^+ - \mu)/\mu = (v - v_r - [v_r(1 + v)(v_r + v_o)]^{1/2})/(1 + v_r)$

$z(\beta^+) = (\beta^+ - \mu)/\mu = (v - v_r + [v_r(1 + v)(v_r + v_o)]^{1/2})/(1 + v_r)$
TABLE 3
STOP-LOSS ORDERED EXTREMAL DISTRIBUTIONS

<table>
<thead>
<tr>
<th>$z$</th>
<th>$F_*(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1 \leq z &lt; -\nu_r$</td>
<td>0</td>
</tr>
<tr>
<td>$-\nu_r \leq z &lt; \nu$</td>
<td>$\nu_0/(1 + \nu_0)$</td>
</tr>
<tr>
<td>$\nu \leq z \leq \nu_0$</td>
<td>1</td>
</tr>
<tr>
<td>$-1 \leq z \leq z(\alpha^*)$</td>
<td>$\nu/(1 + \nu)$</td>
</tr>
<tr>
<td>$z(\alpha^<em>) \leq z \leq z(\beta^</em>)$</td>
<td>$\frac{1}{2}(1 + z/(\nu + z^2))$</td>
</tr>
<tr>
<td>$z(\beta^*) \leq z &lt; \nu_0$</td>
<td>$\nu_0/(\nu_r + \nu_0)$</td>
</tr>
<tr>
<td>$z = \nu_0$</td>
<td>1</td>
</tr>
<tr>
<td>$z(\alpha^<em>) = (\alpha^</em> - \mu)/\mu = \frac{1}{2}(\nu - 1)$</td>
<td></td>
</tr>
<tr>
<td>$z(\beta^<em>) = (\beta^</em> - \mu)/\mu = \frac{1}{2}(\nu_0 - \nu_r)$</td>
<td></td>
</tr>
</tbody>
</table>

Lemma 2.1. The following expressions hold:

\[ E[Z^-] = E[Z_+] = E[Z^+] = 0 \]  \hspace{1cm} (2.1)

\[ \text{Var}[Z^-] = \nu \cdot (2 - \nu_r) + 2\nu_0 \cdot (1 - \nu_r) \cdot \ln\{1 - \nu_r\} \]  \hspace{1cm} (2.2)

\[ \text{Var}[Z_] = \nu \cdot \nu_r \]  \hspace{1cm} (2.3)

\[ \text{Var}[Z^*] = \nu \cdot (1 - \frac{1}{2} \cdot \ln\{\nu_r\}) \]  \hspace{1cm} (2.4)

\[ \text{Var}[Z^+] = \nu \cdot (1 + \ln\{\nu \cdot (\nu_r + \nu_0) \cdot (1 + \nu)\} - \ln\{\nu_r \cdot (\nu + z(\alpha^*)^2) \cdot (\nu + z(\beta^*)^2)\}) \]  \hspace{1cm} (2.5)

Proof. The elementary calculations are left to the reader. The variances are best obtained replacing Stieltjes integrals by Riemann integrals (e.g. Kaas et al. (1994), Theorem 1.3.1.1):

\[ E[Z^2] = 1 - F(-1) + 2 \cdot \int_{-1}^{\nu_0} z(1 - F(z))dz. \]

Our aim is to show the following main result.

Theorem 2.1. The dangerous and stop-loss ordered extremal random variables satisfy the following stochastic order relations between random variables with equal mean:

\[ X^- \leq_D X_* \leq_{st} X \leq_{sl} X^* \leq_D X^+, \text{ for all } X \in D. \]  \hspace{1cm} (2.6)

Proof. These ordering relations are a consequence of the defining ordering inequalities between $X_*$, $X$ and $X^*$, and the auxiliary results below, whose overall content is...
intuitively immediate from the visualizations in Figures 1 and 2, and for which a more rigorous proof is reported to the Appendix.

Lemma 2.2. The stop-loss ordered maximum satisfies the following properties:

\[ Z'' \leq_{st} Z^* \leq_{st} Z^f, \]  

or equivalently

\[ Z'' \leq_{st} Z^* \leq_{st} Z^f, \]  

or equivalently

\[ (2.7) \]
\[ F'(z) < F'(z) < F'(z), \text{ for all } z \in I = [-1, v_o]. \] (2.8)

\[ F'(1) = F'(1), F'(v_0) = F'(v_0). \] (2.9)

**Lemma 2.3.** The stop-loss ordered maximum is less dangerous than the dangerous maximum, that is \( Z^* \leq_D Z^+. \) More precisely one has

\[ F'(z) \leq F'(z), -1 \leq z \leq z_o, \]
\[ F'(z) \geq F'(z), z_o \leq z \leq v_o, \] (2.10)

where the crossing point is determined by

\[ z_o = \frac{1}{2} \sqrt{v} \cdot \sqrt{v / z(\alpha^+) - z(\alpha^+)} / \sqrt{v} \cdot \text{sgn}\{1 - z(\alpha^+)^2 / v\}, \] (2.11)

or in the original \( x \)-scale

\[ x_o = (1 + z_o) \mu = \mu + \frac{1}{2} \sigma \cdot \sigma / (\alpha^+ - \mu) - (\alpha^+ - \mu) / \sigma \cdot \text{sgn}\{1 - (\alpha^+ - \mu)^2 / \sigma^2\}. \] (2.12)

**Lemma 2.4.** The stop-loss ordered minimum is more dangerous than the dangerous minimum, that is \( Z^- \leq_D Z^- \). More precisely one has

\[ F^- (z) \leq F_*(z), -1 \leq z \leq 0, \]
\[ F^- (z) \geq F_*(z), 0 \leq z \leq v_o. \] (2.13)

3. **Bounds for stop-loss premiums and ruin probabilities**

An important issue in Practical Risk Theory is the construction of more or less accurate bounds on stop-loss premiums and ruin probabilities for compound random sums \( S = X_1 + \ldots + X_N \), where the claim number \( N \) is fixed, say Poisson(\( \lambda \)), and the claim sizes \( X_i = d X \in D \) are independent and identically distributed, \( X_i \) independent from \( N \).

3.1. **Stop-loss and ruin probability inequalities**

Since the crossing condition between dangerously ordered random variables is a sufficient condition for stop-loss order, all random variables in (2.6) are stop-loss ordered with equal means. In particular the variances of Lemma 2.1 satisfy the following inequalities

\[ \text{Var}[X^-] \leq \text{Var}[X_*] \leq \text{Var}[X] = \sigma^2 \leq \text{Var}[X^*] \leq \text{Var}[X^+]. \] (3.1)

Let \( S^-, S_*, S^* \), \( S^+ \) the compound random sums obtained when replacing \( X \) by \( X^- \), \( X_*, X^* \) in \( S \). Following Kaas (1991) (see also Kaas et al. (1994), chap. XI), the stop-loss ordering relations imply the following inequalities between stop-loss
premiums and ultimate ruin probabilities:
\[ \pi(S^-; d) \leq \pi(S_*; d) \leq \pi(S; d) \leq \pi(S^+; d), \]
uniformly for all deductibles \( d \geq 0 \), all \( X \in D \),
\[ (3.2) \]
\[ \psi(S^-; u) \leq \psi(S_*; u) \leq \psi(S; u) \leq \psi(S^+; u), \]
uniformly for all initial reserves \( u \geq 0 \), all \( X \in D \),
\[ (3.3) \]
where one uses the notations \( \pi(S; d) = E[(S-d)+], \psi(S; u) = 1 - Pr(L \leq u), \)
\[ L = \max\{S(t) - ct : t \geq 0\} \]
the maximal aggregate loss associated to the aggregate claims up to time \( t : S(t) = X_1 + \ldots + X_{N(t)} \), \( N(t)_{t \geq 0} \) the Poisson process with intensity \( \lambda, c = \lambda \mu (1 + \theta) \) the constant premium rate with security loading \( \theta \geq 0 \).

In particular (3.2), (3.3) imply that the previous theoretical, numerical and analytical bounds based on the dangerous extremal risks \( X^-, X^+ \) can be improved by using instead the stop-loss extremal risks \( X_-, X^* \). For practical reasons let us restrict our attention to the evaluation of analytical bounds, which improve the previous bounds by Steenackers and Goovaerts (1991).

### 3.2. Discrete approximations

To obtain analytical bounds, discrete approximations to the claim size distribution are constructed. For \( F^-(x), F^+(x) \) one finds discrete approximations \( X^-_d, X^+_d \) in Steenackers and Goovaerts (1991). Since \( F_+(x) \) is already a diatomic distribution, it remains to discuss the discretization of \( F^+(x) \).

By means of mass dispersion over the interval \([z(\alpha^*), z(\beta^*)]\), let us construct the following 4-atomic random variable \( Z^*_d = \{-1, z(\alpha^*), z(\beta^*), v_o\} \), which is necessarily more dangerous than \( Z^* \), and thus also stop-loss larger (see e.g. Gerber (1979), chap.7, Example 3.2). The probabilities of \( Z^*_d \) are defined by the following conditions:

\[ p(-1) = F^*(-1) = v/(1 + v) \]
\[ p(\alpha^*) + p(\beta^*) = F^*(z(\beta^*)) - F^*(z(\alpha^*)) \]
\[ z(\alpha^*) \cdot p(\alpha^*) + z(\beta^*) \cdot p(\beta^*) = \int_{z(\alpha^*)}^{z(\beta^*)} zdF^*(z) \]
\[ p(v_o) = 1 - F^*(z(\beta^*)) = v_o/(v_r + v_o) \]

The condition (3.5) preserves the probability mass over \([z(\alpha^*), z(\beta^*)]\) while (3.6) preserves the mean. The right-hand side in (3.6) equals

\[ -(1 - F^*(Z))]_{z(\alpha^*)}^{z(\beta^*)} + \int_{z(\alpha^*)}^{z(\beta^*)} (1 - F^*(z))dz, \]
where the first term is
\[ \frac{1}{2} \frac{(v - 1)}{(v + 1)} - \frac{1}{2} \frac{(v - v_r)}{(v_o + v_r)} \]  
(3.9)

and the second one is
\[ \frac{1}{2} \left[ z - (v + z^2) \right] \right] \frac{1}{2} = \frac{1}{2} (1 - v_r), \]  
(3.10)

where use has been made of the relations
\[ v + z(c^*)^2 = \left( v + 1 \right)^2, \quad v + z(\beta^*)^2 = \frac{1}{4} (v_o + v_r)^2. \]  
(3.11)

It remains to solve the linear system of equations
\[ p(c^*) + p(\beta^*) = A, \quad z(\alpha^*) \cdot p(c^*) + z(\beta^*) \cdot p(\beta^*) = B, \]  
with
\[ A = v_o/(v_o + v_r) - v/(v + 1), \quad B = v/(1 + v) - v/(v_r + v_o). \]  
(3.12)

A calculation shows that
\[ p(\alpha^*) = \frac{A z(c^*) - B}{z(\beta^*) - z(c^*)} = \frac{(v_o - v)}{(1 + v_o) \cdot (1 + v)}; \]  
(3.14)

\[ p(\beta^*) = \frac{B - A z(c^*)}{z(\beta^*) - z(c^*)} = \frac{(v_o - v)}{(1 + v_o) \cdot (v_r + v_o)}. \]  
(3.15)

Since \( v \leq v_o \) the probabilities are always non-negative, as should be.

3.3. The analytical bounds
A compound Poisson(\( \lambda \)) risk \( S \) with discrete claim size support \( X = \{0, x_1, x_2, x_3\} \), and probabilities \( p_0, p_1, p_2, p_3 \) can be expressed as (e.g. Gerber (1979), chap.1, Section 7, or Bowers et al. (1986), Theorem 11.2):
\[ S = x_i N_i + x_2 N_2 + x_3 N_3, \]  
(3.16)

where the numbers of occurrences \( N_i \) of claim size \( x_i \) are independent Poisson with parameter \( \lambda p_i, i = 1, 2, 3 \). From this representation one gets after some well-known calculations the needed analytical formulas:
\[ \pi(S; d) = E[(S - d)^+] = \lambda \mu - d \]
\[ + \exp\{-\lambda(1 - p_0)\} \sum_{n=0}^{\infty} \left[ (\lambda p_1)^n / n_1 ! \right] \cdot (d - \sum_{i=1}^{3} n_i x_i)^+, \]  
(3.17)
\[
\psi(S; u) = 1 - Pr(L \leq u) = \\
1 - \theta/(1 + \theta) \sum_{n = 0}^{\infty} \left[ p_1^n/\ln p_2^n/\ln p_3^n/\ln p_3^n \right] \cdot \exp\{z(1 - p_0)\} \cdot (-z)^n + n_2 + n_3, \\
\text{with } z = (u - \sum_{i = 1}^{3} n_i x_i)^+ / \mu(1 + \theta). 
\]

It is important to observe that these infinite series representations are always finite sums because summation occurs only for \( \sum_{i = 1}^{3} n_i x_i < d, u \).

Table 4 provides a unified overview of the discrete approximations used to get analytical bounds for compound Poisson risks. A numerical illustration, which demonstrates the obtained substantial improvement is found in Section 5.

4. SPECIAL CASES

It is interesting and useful to relate our results to various other considerations made so far in the actuarial literature. We illustrate with two examples.

4.1. The upper bounds for small deductibles and initial reserves

Suppose the deductible of a stop-loss contract is small such that \( d < \alpha^* = \frac{1}{2}(1 + v)\mu \). Then the infinite series representation (3.17) shrinks to the only term \( n_1 = n_2 = n_3 = 0 \), and therefore one has

\[
\pi(S_d; d) = \lambda\mu - d + \exp\{-\lambda(1 - p_0)\} \cdot d. 
\]

In terms of the mean \( \mu_s = \lambda\mu \) and the relative variance \( v_s = (1 + v)\lambda \) of a compound Poisson(\( \lambda \)) risk \( S \) with claim size \( X \in D \), this can be rewritten as

\[
\pi(S_d; d) = \lambda\mu - d + \exp\{-1/v_s\} \cdot d, \ d \leq \frac{1}{2}(1 + v)\mu. 
\]

From Kaas (1991), p.141, one knows that the maximizing claim size distribution over \( D \) is the diatomic risk with support \{\( x_1, x_2 \)\} = \{0, (1 + v)\mu\} and probabilities \( p_1 = v/(1 + v), \ p_2 = 1/(1 + v) \). The corresponding (compound) Poisson risk \( S = (1 + v)\mu N, \ N \) Poisson(\( \lambda \)), has the same stop-loss premiums \( \pi(S_d; d) = \pi(S_d^*; d) \) provided \( d \leq \frac{1}{2}(1 + v)\mu \). A similar result holds for the ultimate ruin probabilities. Therefore our analytical upper bounds obtained from the stop-loss ordered maximal distribution coincides in the special case of small deductibles and initial reserves with the optimal (=best upper) bounds.

4.2. The bounds by unknown relative variance

Suppose only the mean \( \mu \) and the maximal relative variance \( v_o \) are known, but the true relative variance \( v \) is unknown. Equivalently \( \mu \) and the upper end point \( b = (1 + v_o)\mu \) of the interval \( I_x \) are known. The true \( v \) satisfies the inequality \( 0 \leq v \leq v_o \). Choosing \( v = v_o \) in the formulas determining \( X_o^* \), one gets \( \alpha^* = \beta^* = \frac{1}{2}(1 + v_o)\mu \), \( p(\alpha^*) = p(\beta^*) = 0 \), and thus \( X_d^* \) is a diatomic risk with
<table>
<thead>
<tr>
<th>claim size $X_0$</th>
<th>$X_*$</th>
<th>$X_0^*$</th>
<th>$X_0^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>atoms</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$-\mu \ln {1 - v_r}(1 - v_r)/v_r$</td>
<td>$\mu(1 - v_r)$</td>
<td>$\alpha^*$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$\mu(1 + \nu)$</td>
<td>$\beta^*$</td>
<td>$\beta^+$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$\mu(1 + v_o) + \mu v_o \ln {1 - v_r}(1 - v_r)/v_r$</td>
<td>$d + 1$, or $\mu(1 + v_o)$</td>
<td>$\mu(1 + v_o)$</td>
</tr>
<tr>
<td></td>
<td>$u + 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| $p_0$            | 0       | $\nu/(1 + \nu)$ | $\sqrt{\nu}(\mu/x_1)\text{Arctan}\{\sqrt{\nu}(x_1/\mu)/(\nu + 1 - (x_1/\mu))\}$ |
| $p_1$            | $v_r v_o/(1 + v_o)$ | $v_o/(1 + v_o)p(\alpha^*)$ | $v/(\nu + (x_1/\mu - 1)^2) - p_0$ |
| $p_2$            | $1 - v_r$ | $1/(1 + v_o)$ | $p(\beta^*)$ | $1 - p_0 - p_1 - p_3$ |
| $p_3$            | $v_r/(1 + v_o)$ | $1$ | $v_r/(v_r + v_o)$ | $\sqrt{\nu}(v_o + 1 - x_2/\mu)$ |

$\alpha^* = \frac{1}{2} \mu(1 + \nu)$, $\beta^* = \mu(1 + \frac{1}{\nu}(v_o - v_r))$
$p(\alpha^*) = (v_o - v)/(1 + v_o)(1 + \nu)$, $p(\beta^*) = (v_o - v)/(1 + v_o)(v_r + v_o)$

\[
\alpha^+ = \mu(1 + v_r - v_r(1 + v_r)(v_r + v_o))/(1 + v_r) \\
\beta^+ = \mu(1 + [v - v_r - (v_r(1 + v_r)(v_r + v_o)]/(1 + v_r))
\]

$\pi(S_{\mu}^*: d) = \mu_z - d + \exp \{-1/v_o, s\} \sum_{n=0}^{no(d)}(d - n v_o, s \mu_z)/[(v_o, s)^{2n} \cdot n!]$, with \[no(d) = [d/v_o, s \mu_z][[x] : \text{greatest integer less than } x), \quad (4.3)\]

where $v_o, s = (1 + v_o)/\mu$ denotes the maximal relative variance of a compound Poisson($\lambda$) risk with mean claim size $\mu$ and maximal relative variance $v_o$ of the claim size. Similarly the ultimate ruin probability should be calculated as

\[
\psi(S_{\mu}^*; u) = 1 - \theta/(1 + \theta) \sum_{m=0}^{mo(u)} \exp \{z/v_o, s\}(-z/v_o, s)^m/m!, \text{ with} \quad z = (u - n v_o, s \mu_z)/(1 + \theta) \mu_z, mo(u) = [u/v_o, s \mu_z]. \quad (4.4)
\]

In particular the latter formula can be viewed as a positive answer to the following modified Schmitter problem (discussion papers on this topic are Brockett et al. (1991) and Kaas (1991)). Given that the individual claims have mean $\mu$ and
maximal variance $v_0 \mu^2$, which claims distribution maximizes the ruin probability for a given initial reserve $u$?

On the other side it is interesting to look at the special deductible $d = \mu_s$, the mean of a compound Poisson($\lambda$) risk. In this situation the safest stop-loss upper bound (4.3) can be rewritten as

$$\pi(S_d^\#; \mu_s) = \mu_s \cdot \exp(-\lambda_0 + [\lambda_0]1n\{\lambda_0\})/\Gamma(\lambda_0 + 1), \ \lambda_0 = 1/v_0.s.$$  \hspace{1cm} (4.5)

This is the special stop-loss premium of a compound Poisson($\lambda_0$) risk with individual claims of equal size $\mu_s v_0.s$. It is approximately equal to the special stop-loss premium of a Gamma($\lambda_0$, $\lambda_0/\mu_s$) distributed risk with mean $\mu_s$ and variance $v_0.s \mu_s^2$.

Applying Stirling's approximation formula for the Gamma function

$$\Gamma(\lambda_0 + 1) \approx \exp(-\lambda_0 + \lambda_01n\{\lambda_0\})(2\pi\lambda_0)^{1/2} \cdot (1 + 1/12\lambda_0 + ...),$$  \hspace{1cm} (4.6)

one gets approximately

$$\pi(S_d^\#; \mu_s) \approx \mu_s (2\pi\lambda_0)^{-1/2}/(1 + 1/12\lambda_0 + ...) < \mu_s (v_0.s/2\pi)^{1/2}. \hspace{1cm} (4.7)$$

The upper bound on the right-hand side is the special stop-loss premium of a Normal ($\mu_s$, $v_0.s \mu_s^2$) distributed risk, to which $\pi(S_d^\#; \mu_s)$ converges when the relative variance $v_0.s$ goes to zero. The latter property is known to be true asymptotically (that is in large portfolios) for arbitrary compound Poisson risks (e.g. Daykin et al. (1994), p.64). Note that the above investigation corresponds to the findings of Benktander (1977), where as an important additional complement, the parameter

$$1/\lambda_0 = v_0.s = (1 + v_0)/\lambda \hspace{1cm} (4.8)$$

should be equal to the maximal relative variance of the considered compound Poisson risk, given that the true relative variance is unknown.

5. Numerical illustration

Though it would be possible to compute close numerical lower and upper bounds to the considered risk quantities for all of the compound Poisson risks with claim sizes $X^-, X^+, X^*$, $X^+$, we restrict our numerical investigation to a comparison of the corresponding analytical bounds as summarized in Table 4.

The same examples as in Steenackers and Goovaerts (1991) have been calculated. The claim size distribution is assumed to be uniform (1,3) with range [0,3], hence $\mu = 2$, $v = 1/12$, $v_o = 1/2$, $v_r = 1/6$, and the Poisson parameter $\lambda$ is 1, or 10. The safety loading is $\theta = 0.2$. The results are found in Tables 5, 6, 7. The improvement is very substantial for the new stop-loss bounds, but less spectacular for the ruin probabilities. Concerning the previous bounds, small numerical inaccuracies for the deductible $d = 60$ in Table 6, as well as in Table 7 have been located, probably due to rounding errors.
TABLE 5
BOUNDS FOR STOP-LOSS PREMIUMS WITH CLAIM-RANGE [0,3], MEAN 2, VARIANCE 1/3 AND CLAIM NUMBER POISSON (1)

<table>
<thead>
<tr>
<th>d</th>
<th>exact value</th>
<th>bound $X_d^-$</th>
<th>bound $X_*$</th>
<th>bound $X_d^+$</th>
<th>bound $X_d^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8.277·10^{-1}</td>
<td>89.3%</td>
<td>93.8%</td>
<td>107.3%</td>
<td>113.9%</td>
</tr>
<tr>
<td>4</td>
<td>2.689·10^{-1}</td>
<td>78.4%</td>
<td>87.2%</td>
<td>115.7%</td>
<td>127.2%</td>
</tr>
<tr>
<td>6</td>
<td>7.184·10^{-2}</td>
<td>67.2%</td>
<td>77.6%</td>
<td>122.7%</td>
<td>138.1%</td>
</tr>
<tr>
<td>8</td>
<td>1.627·10^{-2}</td>
<td>56.6%</td>
<td>72.1%</td>
<td>141.9%</td>
<td>188.6%</td>
</tr>
<tr>
<td>10</td>
<td>3.253·10^{-3}</td>
<td>46.0%</td>
<td>63.0%</td>
<td>164.3%</td>
<td>237.8%</td>
</tr>
<tr>
<td>12</td>
<td>5.815·10^{-4}</td>
<td>36.5%</td>
<td>51.9%</td>
<td>190.9%</td>
<td>279.2%</td>
</tr>
<tr>
<td>14</td>
<td>9.346·10^{-5}</td>
<td>28.5%</td>
<td>44.6%</td>
<td>220.2%</td>
<td>388.5%</td>
</tr>
<tr>
<td>16</td>
<td>1.366·10^{-5}</td>
<td>21.9%</td>
<td>36.5%</td>
<td>260.1%</td>
<td>533.2%</td>
</tr>
<tr>
<td>18</td>
<td>1.840·10^{-6}</td>
<td>16.5%</td>
<td>28.4%</td>
<td>322.0%</td>
<td>670.0%</td>
</tr>
<tr>
<td>20</td>
<td>2.302·10^{-7}</td>
<td>12.2%</td>
<td>22.9%</td>
<td>379.2%</td>
<td>934.5%</td>
</tr>
</tbody>
</table>

APPENDIX

Proof of Lemma 2.2. (2.9) is immediate. Let us show (2.8) in two steps.

Step 1: \( F^\ell(z) \leq F^c(z) \)
(a) \(-1 \leq z \leq v:\)
\[
F^\ell(z) \leq F^\ell(v) = v/(1 + v) = F^c(-1) \leq F^c(z)
\]
(b) \(z \geq v:\)
(b1) \(v_o \geq (1 + \sqrt{2})\sqrt{v} \Rightarrow v \leq \frac{1}{2}(v_o - v_r)\)

TABLE 6
BOUNDS FOR STOP-LOSS PREMIUMS WITH CLAIM-RANGE [0,3], MEAN 2, VARIANCE 1/3 AND CLAIM NUMBER POISSON (10)

<table>
<thead>
<tr>
<th>d</th>
<th>exact value</th>
<th>bound $X_d^-$</th>
<th>bound $X_*$</th>
<th>bound $X_d^+$</th>
<th>bound $X_d^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>5.757</td>
<td>99.0%</td>
<td>99.1%</td>
<td>101.7%</td>
<td>103.3%</td>
</tr>
<tr>
<td>20</td>
<td>2.626</td>
<td>95.6%</td>
<td>96.7%</td>
<td>105.4%</td>
<td>110.1%</td>
</tr>
<tr>
<td>25</td>
<td>9.321·10^{-1}</td>
<td>91.6%</td>
<td>92.3%</td>
<td>112.5%</td>
<td>123.3%</td>
</tr>
<tr>
<td>30</td>
<td>2.563·10^{-1}</td>
<td>81.8%</td>
<td>85.7%</td>
<td>123.6%</td>
<td>145.3%</td>
</tr>
<tr>
<td>35</td>
<td>5.507·10^{-2}</td>
<td>74.7%</td>
<td>77.4%</td>
<td>139.9%</td>
<td>179.7%</td>
</tr>
<tr>
<td>40</td>
<td>9.383·10^{-3}</td>
<td>61.0%</td>
<td>68.1%</td>
<td>162.6%</td>
<td>232.3%</td>
</tr>
<tr>
<td>45</td>
<td>1.289·10^{-3}</td>
<td>53.1%</td>
<td>58.2%</td>
<td>193.7%</td>
<td>313.4%</td>
</tr>
<tr>
<td>50</td>
<td>1.449·10^{-4}</td>
<td>40.0%</td>
<td>48.4%</td>
<td>236.1%</td>
<td>439.0%</td>
</tr>
<tr>
<td>55</td>
<td>1.355·10^{-5}</td>
<td>33.3%</td>
<td>39.2%</td>
<td>293.6%</td>
<td>634.2%</td>
</tr>
<tr>
<td>60</td>
<td>1.067·10^{-6}</td>
<td>23.3%</td>
<td>31.2%</td>
<td>371.8%</td>
<td>945.9%</td>
</tr>
</tbody>
</table>
TABLE 7
BOUNDS FOR RUIN PROBABILITIES WITH CLAIM-RANGE [0,3], MEAN 2, VARIANCE 1/3 AND
SAFETY MARGIN 20%

<table>
<thead>
<tr>
<th>( u )</th>
<th>( \text{bound } X_d^- )</th>
<th>( \text{bound } X_d^+ )</th>
<th>( \text{bound } X_d^* )</th>
<th>( \text{bound } X_d^{**} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.747184</td>
<td>0.747184</td>
<td>0.755158</td>
<td>0.764982</td>
</tr>
<tr>
<td>2</td>
<td>0.617238</td>
<td>0.625370</td>
<td>0.663538</td>
<td>0.682110</td>
</tr>
<tr>
<td>3</td>
<td>0.523757</td>
<td>0.526666</td>
<td>0.566954</td>
<td>0.587361</td>
</tr>
<tr>
<td>4</td>
<td>0.437362</td>
<td>0.441446</td>
<td>0.492510</td>
<td>0.517595</td>
</tr>
<tr>
<td>5</td>
<td>0.366885</td>
<td>0.371088</td>
<td>0.425256</td>
<td>0.452560</td>
</tr>
<tr>
<td>6</td>
<td>0.307327</td>
<td>0.311606</td>
<td>0.367586</td>
<td>0.395577</td>
</tr>
<tr>
<td>7</td>
<td>0.257467</td>
<td>0.261752</td>
<td>0.317711</td>
<td>0.346172</td>
</tr>
<tr>
<td>8</td>
<td>0.215718</td>
<td>0.219854</td>
<td>0.274574</td>
<td>0.302710</td>
</tr>
<tr>
<td>9</td>
<td>0.180725</td>
<td>0.184666</td>
<td>0.237313</td>
<td>0.264791</td>
</tr>
<tr>
<td>10</td>
<td>0.151413</td>
<td>0.155110</td>
<td>0.205100</td>
<td>0.231598</td>
</tr>
<tr>
<td>20</td>
<td>0.025798</td>
<td>0.027111</td>
<td>0.047693</td>
<td>0.060689</td>
</tr>
<tr>
<td>30</td>
<td>0.004396</td>
<td>0.004739</td>
<td>0.011090</td>
<td>0.015903</td>
</tr>
<tr>
<td>40</td>
<td>0.000749</td>
<td>0.000828</td>
<td>0.002579</td>
<td>0.004167</td>
</tr>
<tr>
<td>50</td>
<td>0.000130</td>
<td>0.000149</td>
<td>0.000600</td>
<td>0.001092</td>
</tr>
</tbody>
</table>

If \( \frac{1}{2} (v - 1) \leq z < \frac{1}{2} (v_o - v_r) \) one checks that

\[
F^t(z) = \frac{z^2}{v + z^2} \leq F^*(z) = \frac{1}{2} \left( 1 + \frac{z}{v + z^2} \right) ^{\frac{1}{2}}
\]

If \( v \leq \frac{1}{2} (v_o - v_r) \leq z < v_o \) one has

\[
F^t(z) \leq F^t(v_o -) = v_o/(v_r + v_o) = F^*(z)
\]

(b2) \( v_o \leq (1 + \sqrt{2}) \sqrt{v} = v \geq \frac{1}{2} (v_o - v_r) \)

One concludes as in the second if-part of (b1)

Step 2: \( F^*(z) \leq F^u(z) \)

(a) \(-1 \leq z \leq -v_r:\)

(a1) \(1/v_o \leq \frac{1}{2} (1 - v)/v \Rightarrow \frac{1}{2} (v - 1) \leq -v_r \)

If \( -1 \leq z \leq \frac{1}{2} (v - 1) \) then one has

\[
F^*(z) = v/(1 + v) = F^u(-1) \leq F^u(z)
\]

If \( \frac{1}{2} (v - 1) \leq z \leq -v_r \leq \frac{1}{2} (v_o - v_r) \) one checks that

\[
F^*(z) = \frac{1}{2} \left( 1 + \frac{z}{v + z^2} \right) ^{\frac{1}{2}} \leq v/(v + z^2) = F^u(z)
\]

(a2) \(1/v_o \geq \frac{1}{2} (1 - v)/v \Rightarrow \frac{1}{2} (v - 1) \geq -v_r \)

One concludes as in the first if-part of (a1)

(b) \(-v_r \leq z \leq v_o:\)

\[
F^*(z) \leq F^*(v_o -) = v_o/(v_r + v_o) = F^u(-v_r) \leq F^u(z)
\]
Proof of Lemma 2.3. From Lemma 2.2 and the expression for $F^+(z)$ one has

$$F(z) \leq F^+(z) \leq F^+(z), \text{ for all } z \in [-1, v_0)$$

$$F^+(-1) = F^+(-1), \quad F^+(v_0-) = F^+(v_0-).$$

Since $F^+(z)$ is continuous and non-decreasing on $[-1, v_0)$, it follows from undergraduate calculus that $F^+(z)$ takes any value between $F^+(-1)$ and $F^+(v_0-)$. In particular there exists $z_0$ such that $F^+(z_0) = F^+(z(\alpha^+))$. By construction of $F^+$, one obtains

$$F^+(z_0) = F^+(z), \quad -1 \leq z \leq z_0,$$

$$F^+(z) \geq F^+(z), \quad z_0 \leq z \leq v_0,$$

which shows that $Z^* \leq Z^+$. The crossing point $z_0$ is obtained by solving its defining equation.

Proof of Lemma 2.4. This is immediately seen from the expressions for $F^-(z), F^*(z)$ given in Tables 2 and 3.

Note added in proof. After this paper has been submitted, the author has found the stop-loss ordered extremal distributions of Table 3 in Stoyan (1973).

REFERENCES


IMPROVED ANALYTICAL BOUNDS FOR SOME RISK QUANTITIES


DEPENDENCY OF RISKS AND STOP-LOSS ORDER

JAN DHAENE AND MARC J. GOOVAERTS

ABSTRACT

The correlation order, which is defined as a partial order between bivariate distributions with equal marginals, is shown to be a helpfull tool for deriving results concerning the riskiness of portfolios with pairwise dependencies. Given the distribution functions of the individual risks, it is investigated how changing the dependency assumption influences the stop-loss premiums of such portfolios.

KEYWORDS

Dependent risks; Bivariate distributions; Correlation order; Stop-loss order.

1. INTRODUCTION

Consider the individual risk theory model with the total claims of the portfolio during some reference period (e.g. one year) given by

\[ S = \sum_{i=1}^{n} X_i \]

where \( X_i \) is the claim amount caused by policy \( i \) (\( i = 1, 2, \ldots, n \)). In the sequel we will always assume that the individual claim amounts \( X_i \) are nonnegative random variables and that the distribution functions \( F_i \) of \( X_i \) are given.

Usually, it is assumed that the risks \( X_i \) are mutually independent because models without this restriction turn out to be less manageable. In this paper we will derive results concerning the aggregate claims \( S \) if the assumption of mutually independence is relaxed. More precisely, we will assume that the portfolio contains a number of couples (e.g. wife and husband) with non-independent risks. Therefore, we will rearrange and rewrite (1) as

\[ S = \sum_{i=1}^{m} (X_{2i-1} + X_{2i}) + \sum_{i=2m+1}^{n} X_i \]

1 Work performed under grant OT/93/5 of Onderzoeksfonds K.U.Leuven
2 Katholieke Universiteit Leuven
3 Katholieke Universiteit Leuven and Universiteit Amsterdam

ASTIN BULLETIN, Vol. 26, No. 2, 1996, pp. 201-212
with \( m \) the number of coupled risks. For any \( i \) and \( j \) \((i, j = 1, 2, \ldots n; i \neq j)\) we assume that \( X_i \) and \( X_j \) are independent risks, except if they are members of the same couple \((X_{2k-1}, X_{2k}), (k = 1, 2, \ldots m)\). The class of all multivariate random variables \((X_1, \ldots, X_n)\) with given marginals \( F_i \) of \( X_i \) and with the pairwise dependency structure as explained above, will be denoted by \( R(F_1, \ldots, F_n) \).

It is clear that for any \((X_1, \ldots, X_n)\) belonging to \( R(F_1, \ldots, F_n) \), the riskiness of the aggregate claims \( S = X_1 + \ldots + X_n \) will be strongly dependent on the way of dependency between the members of couples.

In order compare the riskiness of the aggregate claims of different elements of \( R(F_1, \ldots, F_n) \), we will use the stop-loss order.

**Definition 1** A risk \( S_1 \) is said to precede a risk \( S_2 \) in stop-loss order, written \( S_1 \preceq_{st} S_2 \), if their stop-loss premiums are ordered uniformly:

\[
E(S_1 - d)_+ \leq E(S_2 - d)_+
\]

for all retentions \( d \geq 0 \).

Let \((X_1, \ldots, X_n)\) and \((Y_1, \ldots, Y_n)\) be two elements of \( R(F_1, \ldots, F_n) \). and denote their respective sums by

\[
S_1 = \sum_{i=1}^{m} (X_{2i-1} + X_{2i}) + \sum_{i=2m+1}^{n} X_i
\]

and

\[
S_2 = \sum_{i=1}^{m} (Y_{2i-1} + Y_{2i}) + \sum_{i=2m+1}^{n} Y_i
\]

We want to find ordering relations between the corresponding couples of \( S_1 \) and \( S_2 \) which imply a stop-loss order for \( S_1 \) and \( S_2 \). More precisely, we are looking for a partial order \( \preceq_{ord} \) between bivariate distributed random variables which has the following property:

\[
(X_{2k-1}, X_{2k}) \preceq_{ord} (Y_{2k-1}, Y_{2k}) \quad (k = 1, 2, \ldots, m)
\]

implies

\[
S_1 \preceq_{st} S_2
\]

A well-known property of stop-loss ordering is that it is preserved under convolution of independent risks, see e.g. Goovaerts et al. (1990). Hence, a sufficient condition for \((4)\) to be true is

\[
X_{2k-1} + X_{2k} \preceq_{st} Y_{2k-1} + Y_{2k}
\]

\((k = 1, 2, \ldots, m)\)

So it follows immediately that we can restrict ourselves to the following problem: Find a partial order \( \preceq_{ord} \) between bivariate distributed random variables \((X_1, X_2)\) and \((Y_1, Y_2)\) with the same marginal distributions, for which the following property holds:

\[
(X_1, X_2) \preceq_{ord} (Y_1, Y_2)
\]

implies
It is clear that an ordering \( \preceq_{ord} \) for which (6) implies (7) will immediately lead to a solution of the problem described by (3) and (4).

Part of the results in this paper are generalisations of results in Dhaene et al. (1995) where the individual life model is considered, i.e. the case where each individual risk has a two-point distribution in zero and some positive value.

2. A PARTIAL ORDER FOR BIVARIATE DISTRIBUTIONS

2.1. Correlation order

Let \( R(F_1, F_2) \) be the class of all bivariate distributed random variables with given marginals \( F_1 \) and \( F_2 \). For any \( (X_1, X_2) \in R(F_1, F_2) \) we have
\[
F_1(x) = \text{Prob}(X_1 \leq x) \quad F_2(x) = \text{Prob}(X_2 \leq x)
\]

We also introduce the following notation for the bivariate distribution function:
\[
F_{X_1, X_2}(x_1, x_2) = \text{Prob}(X_1 < x_1, X_2 < x_2)
\]

In the sequel we will always restrict ourselves to the case of non-negative risks. Futher, if we use stop-loss premiums or covariances, we will always silently assume that they are well-defined.

Now let \( (X_1, X_2) \) and \( (Y_1, Y_2) \) be two elements of \( R(F_1, F_2) \). In order to investigate an order between these bivariate distributed random variables which implies stop-loss order for \( X_1 + X_2 \) and \( Y_1 + Y_2 \), we could start by comparing \( \text{Cov}(X_1, X_2) \) and \( \text{Cov}(Y_1, Y_2) \).

At first sight, one could consider the following inequality
\[
\text{Cov}(X_1, X_2) < \text{Cov}(Y_1, Y_2)
\]

and investigate whether this implies
\[
X_1 + X_2 \preceq_{st} Y_1 + Y_2
\]

Although it is customary to compute covariances in relation with dependency considerations, one number alone cannot reveal the nature of dependency adequately, and hence (8) will not imply (9) in general, a counterexample is given in Dhaene et al. (1995). However, in the special case that \( F_1 \) and \( F_2 \) are two-point distributions with zero and some positive value as mass points, (8) and (9) are equivalent, see also Dhaene et al. (1995).

Instead of comparing \( \text{Cov}(X_1, X_2) \) and \( \text{Cov}(Y_1, Y_2) \) one could compare \( \text{Cov}(f(X_1), g(X_2)) \) with \( \text{Cov}(f(Y_1), g(Y_2)) \) for all non-decreasing functions \( f \) and \( g \), see e.g. Barlow et al. (1975).

**Definition 2** Let \( (X_1, X_2) \) and \( (Y_1, Y_2) \) be elements of \( R(F_1, F_2) \). Then we say that \( (X_1, X_2) \) is less correlated than \( (Y_1, Y_2) \), written \( (X_1, X_2) \preceq_{c} (Y_1, Y_2) \), if
\[
\text{Cov}(f(X_1), g(X_2)) \leq \text{Cov}(f(Y_1), g(Y_2))
\]
for all non-decreasing functions \( f \) and \( g \) for which the covariances exist. The correlation-order is a partial order over joint distributions in \( \mathcal{R}(F_1, F_2) \) and expresses the idea that two random variables with given marginals are more 'positively dependent' or 'positively correlated' when they have some joint distribution than some other one.

2.2. An alternative definition

In this subsection we will derive an alternative definition for the correlation order introduced above. First, we will recall and prove a lemma contained in Hoeffding (1940), which we will need for the derivation of the alternative definition, see also Jodgeo (1982), p. 326. The proof will be repeated here because it is instructive for what follows.

Lemma 1. For any \((X_1, X_2) \in \mathcal{R}(F_1, F_2)\) we have

\[
\text{Cov}(X_1, X_2) = \int_0^\infty \int_0^\infty (F_{X_1, X_2}(u, v) - F_1(u)F_2(v)) \, du \, dv
\]

Proof: Let \( I \) denote the indicator function, then the following well-known identity holds

\[
x - z = \int_0^\infty [I(z \leq u) - I(x \leq u)] \, du \quad (x, z \geq 0)
\]

Hence, for \( x_1, x_2, z_1, z_2 \geq 0 \) we find

\[
(x_1 - z_1)(x_2 - z_2) = \int_0^\infty \int_0^\infty \{I(z_1 \leq u)I(z_2 \leq v) + I(x_1 \leq u)I(x_2 \leq v) - I(z_1 \leq u)I(x_2 \leq v) - I(x_1 \leq u)I(z_2 \leq v)\} \, du \, dv
\]

Now let \((X_1, X_2)\) and \((Z_1, Z_2)\) be independent identically distributed pairs, then we have

\[
2 \text{Cov}(X_1, X_2) = E((X_1 - Z_1)(X_2 - Z_2))
\]

so that we find (11) from (13). Q.E.D

Now we are able to state an equivalent definition for the correlation order considered in definition 2.

Theorem 1. Let \((X_1, X_2)\) and \((Y_1, Y_2)\) be elements of \( \mathcal{R}(F_1, F_2) \). Then the following statements are equivalent:

(a) \((X_1, X_2) \leq_c (Y_1, Y_2)\)

(b) \(F_{X_1, X_2}(x_1, x_2) \leq F_{Y_1, Y_2}(x_1, x_2)\) for all \(x_1, x_2 \geq 0\)

Proof: Assume that (a) holds and choose \( f(u) = I(u > x_1) \) and \( g(u) = I(u > x_2) \). Then we find from (10) that

\[
E(I(X_1 > x_1, X_2 > x_2)) \leq E(I(Y_1 > x_1, Y_2 > x_2))
\]
or equivalently

\[ \text{Prob}(X_1 > x_1, X_2 > x_2) \leq \text{Prob}(Y_1 > x_1, Y_2 > x_2) \]

from which (b) can easily be derived.

Now, suppose that (b) holds. It follows immediately that, for non-decreasing functions \( f \) and \( g \),

\[ \text{Prob}(f(X_1) \leq x_1, g(X_2) \leq x_2) \leq \text{Prob}(f(Y_1) \leq x_1, g(Y_2) \leq x_2) \]

for all \( x_1, x_2 \geq 0 \), so that (a) follows as an immediate consequence of Lemma 1 and Definition 2. Q.E.D

Statement (b) in Theorem 1 asserts roughly that the probability that \( X_1 \) and \( X_2 \) both realize 'small' values is not greater than the probability that \( Y_1 \) and \( Y_2 \) both realize 'equally small' values, suggesting that \( Y_1 \) and \( Y_2 \) are more positively interdependent than \( X_1 \) and \( X_2 \). The statement (b) is equivalent with each of the following statements, each understood to be valid for all \( x_1 \) and \( x_2 \):

(c) \( \text{Prob}(X_1 < x_1, X_2 > x_2) \leq \text{Prob}(Y_1 < x_1, Y_2 > x_2) \)

(d) \( \text{Prob}(X_1 > x_1, X_2 \leq x_2) \leq \text{Prob}(Y_1 > x_1, Y_2 \leq x_2) \)

(e) \( \text{Prob}(X_1 > x_1, X_2 > x_2) \leq \text{Prob}(Y_1 > x_1, Y_2 > x_2) \)

Each of these statements can be interpreted similarly in terms of 'more positively interdependence' of \( Y_1 \) and \( Y_2 \). Hence, the equivalence of (a) and (b) in Theorem 1 has some intuitive interpretation.

References related to the correlation order defined above are Barlow et al. (1975), Cambanis et al. (1976) and Tchen (1980). For economic applications, see also Epstein et al. (1980) and Aboudi et al. (1993, 1995).

2.3. Correlation order and stop-loss order

In this subsection we will prove that the correlation order between bivariate distributions implies stop-loss order between the distributions of their sums.

Lemma 2 For any \((X_1, X_2) \in R(F_1, F_2)\) we have

\[ E(X_1 + X_2 - d)_+ = E(X_1) + E(X_2) - d + \int_0^d F_{X_1, X_2}(x, d-x)dx \]

Proof: We have that

\[ E(X_1 + X_2 - d)_+ = E(X_1) + E(X_2) - d + E(d - X_1 - X_2)_+ \]

For non-negative real numbers \( x_1 \) and \( x_2 \) the following equality holds

\[ (d - x_1 - x_2)_+ = \int_0^d 1(x_1 \leq x, x_2 \leq d - x)dx \]

so that
\[ E(d - X_1 - X_2)_+ = \int_0^d E(I(X_1 \leq x, X_2 \leq d - x)) \, dx \]

which proves the lemma. Q.E.D

Now we are able to state the following result.

**Theorem 2** Let \( (X_1, X_2) \) and \( (Y_1, Y_2) \) be two elements of \( R(F_1, F_2) \). Then

\[ (X_1, X_2) \preceq (Y_1, Y_2) \]

implies

\[ X_1 + X_2 \preceq Y_1 + Y_2 \]

**Proof:** The proof follows immediately from Theorem 1 and Lemma 2. Q.E.D

From Theorem 2 we conclude that the correlation order is a useful tool for comparing the stop-loss premiums of sums of two non-independent risks with equal marginals.

3. **Riskiest and Safest Dependency Between Two Risks**

Consider again the class \( R(F_1, F_2) \) of all bivariate distributed random variables with given marginals \( F_1 \) and \( F_2 \) respectively. For every \( (X_1, X_2) \) and \( (Y_1, Y_2) \in R(F_1, F_2) \) we will compare their respective riskiness by comparing the stop-loss premiums of \( X_1 + X_2 \) and \( Y_1 + Y_2 \). More precisely, we will say that \( (X_1, X_2) \) is less risky than \( (Y_1, Y_2) \) if

\[ X_1 + X_2 \preceq Y_1 + Y_2 \]

In this section we will look for the riskiest and the safest elements of \( R(F_1, F_2) \). Use will be made of the following well-known result which is usually attributed to both Hoeffding and Fréchet, see e.g. Fréchet (1951).

**Lemma 3** For any \( (X_1, X_2) \in R(F_1, F_2) \) we have that

\[ \max[F_1(x_1) + F_2(x_2) - 1; 0] \leq F_{X_1, X_2}(x_1, x_2) \leq \min[F_1(x_1), F_2(x_2)] \]  

(14)

The upper and lower bounds are themselves bivariate distributions with marginals \( F_1 \) and \( F_2 \) respectively.

Now we can state the following result concerning the riskiest and the safest elements of \( R(F_1, F_2) \).

**Theorem 3** Let \( (Y_1, Y_2) \) and \( (Z_1, Z_2) \) be elements of \( R(F_1, F_2) \) with distribution functions given by

\[ F_{Y_1, Y_2}(x_1, x_2) = \max[F_1(x_1) + F_2(x_2) - 1; 0] \]

and

\[ F_{Z_1, Z_2}(x_1, x_2) = \min[F_1(x_1), F_2(x_2)] \]
respectively. Then for any \((X_1, X_2) \in R(F_1, F_2)\) we have that
\[
Y_1 + Y_2 \leq_{sl} X_1 + X_2 \leq_{sl} Z_1 + Z_2
\]

**Proof:** The inequalities follow immediately from Theorems 1 and 2 from Lemma 3. Q.E.D

From Theorem 3 we can conclude that the random variables \((Y_1, Y_2)\) and \((Z_1, Z_2)\) are safest and the riskiest elements of \(R(F_1, F_2)\) respectively.

Let us now look at the special case that the two marginal distributions are equal. From Theorem 3, we find that a most risky element in \(R(F, F)\) is \((Z_1, Z_2)\) with
\[
F_{Z_1, Z_2}(x_1, x_2) = \min[F(x_1), F(x_2)]
\]
which leads to
\[
F_{Z_1, Z_2}(x, d-x) = \begin{cases} F(x) & \text{if } x \leq d/2 \\ F(d-x) & \text{if } x > d/2 \end{cases}
\]
From Lemma 2 we find
\[
E(Z_1 + Z_2 - d)_+ = E(Z_1)_+ + E(Z_2)_+ - d + \int_{0}^{d/2} F(x)dx + \int_{d/2}^{d} F(d-x)dx
\]
\[
= E(Z_1)_+ + E(Z_2)_+ - 2 \int_{0}^{d/2} (1 - F(x))dx
\]
\[
= 2E(Z_1 - d/2)_+
\]
so that we find the following corollary to Theorem 3.

**Corollary 1**  For any \((X_1, X_2) \in R(F, F)\) we have that
\[
E(X_1 + X_2 - d)_+ \leq 2E(X_1 - d/2)_+
\]
Furthermore, the upperbound is the stop-loss premium with retention \(d\) of \(Z_1 + Z_2\) where \((Z_1, Z_2) \in R(F, F)\) with distribution function (15).

Now assume that \(F\) is an exponential distribution with parameter \(\alpha \geq 0\).
i.e.
\[
F(x) = 1 - e^{-\alpha x} \quad x > 0
\]
Then we obtain from Corollary 1 that for any \((X_1, X_2) \in R(F, F)\), we have
\[
E(X_1 + X_2 - d)_+ \leq 2 \int_{d/2}^{\infty} (1 - F(x))dx = \frac{2}{\alpha} e^{-\alpha d/2}
\]
(16)
This upperbound for the exponential case can be found in Heilmann (1986). He derived this result by using some techniques described in Meilijson et al. (1979). Heilmann also considers riskiest elements in \(R(F_1, F_2)\) where \(F_1\) and \(F_2\) are exponential distributions with different parameters. This result can also be found from our Lemma 2 and Theorem 3.
4. POSITIVE DEPENDENCY BETWEEN RISKS

In a great many situation, certain insured risks tend to act similarly. For instance, in group life insurance the remaining life-times of a husband and his wife can be shown to possess some ‘positive dependency’. Several concepts of bivariate positive dependency have appeared in the mathematical literature, see Tong (1980) for a review, for actuarial applications see Norberg (1989) and Kling (1993). We will restrict ourselves to positive quadrant dependency.

**Definition 3** The random variables \( X_1 \) and \( X_2 \) are said to be positively quadrant dependent, written \( \text{PQD}(X_1, X_2) \), if

\[
\Pr(X_1 < x_1, X_2 < x_2) \geq \Pr(X_1 < x_1)\Pr(X_2 < x_2)
\]

for all \( x_1 \geq 0, x_2 \geq 0 \).

It is clear that \( \text{PQD}(X_1, X_2) \) is equivalent with saying that \( X_1 \) and \( X_2 \) are more correlated (in the sense of Definition 2) than if they were independent.

Positive quadrant dependency can be defined in terms of covariances, as is shown in the following lemma, see also Epstein et al. (1980).

**Lemma 4** Let \( X_1 \) and \( X_2 \) be two random variables. Then the following statements are equivalent:

(a) \( \text{PQD}(X_1, X_2) \)

(b) \( \text{Cov}(f(X_1), g(X_2)) \geq 0 \) for all non-decreasing real functions \( f \) and \( g \) for which the covariance exists

**Proof:** The result follows immediately from Definitions 1 and 3, and Theorem 1. Q.E.D.

Remark that \( \text{PQD}(X_1, X_2) \) implies that \( \text{Cov}(X_1, X_2) \geq 0 \). Equality only holds if \( X_1 \) and \( X_2 \) are independent.

As is shown in the following theorem, the notion of positive quadrant dependency can be used for considering the effect of the independence assumption, when the risks are positively dependent actually.

**Theorem 4** Let \( (X_1, X_2) \) and \( (Y_1^{\text{ind}}, Y_2^{\text{ind}}) \) be two elements of \( R(F_1, F_2) \) with \( \text{PQD}(X_1, X_2) \) and where \( Y_1^{\text{ind}} \) and \( Y_2^{\text{ind}} \) are mutually independent. Then

\[
Y_1^{\text{ind}} + Y_2^{\text{ind}} \leq_{\text{st}} X_1 + X_2
\]

**Proof:** The result follows immediately from Theorems 1 and 2. Q.E.D.

Theorem 4 states that when the marginal distributions are given, and when \( \text{PQD}(X_1, X_2) \), then the independence assumption will always underestimate the actual stop-loss premiums.
Let us now consider the special case that $F_i$ is a two-point distribution in 0 and $\alpha_i > 0$ ($i = 1, 2$). For any $(X_1, X_2) \sim R(F_1, F_2)$ with $\text{Cov}(X_1, X_2) \geq 0$, we have that

$$P_r(X_1 = \alpha_1, X_2 = \alpha_2) \geq P_r(X_1 = \alpha_1) P_r(X_2 = \alpha_2)$$

This inequality can be transformed into

$$P_r(X_1 = 0, X_2 = 0) \geq P_r(X_1 = 0) P_r(X_2 = 0)$$

from which we find

$$P_r(X_1 \leq x_1, X_2 \leq x_2) \geq P_r(X_1 \leq x_1) P_r(X_2 \leq x_2) \quad x_1 \geq 0, x_2 \geq 0$$

We can conclude that in this special case $P(QD)(X_1, X_2)$ is equivalent with $\text{Cov}(X_1, X_2) \geq 0$.

From Theorem 4 we find that when the marginal distributions $F_i$ are given two-point distributions in 0 and $\alpha_i > 0$ ($i = 1, 2$) and when $\text{Cov}(X_1, X_2) \geq 0$, making the independence assumption will underestimate the actual stop-loss premiums. This result can also be found in Dhaene et al. (1995).

5. NUMERICAL EXAMPLE AND CONCLUDING REMARKS

As stipulated in Section 1 the results that we have derived for two risks can also be used for considering the riskiness of portfolios where the only non-independent risks can be classified into a given number of couples. Several theorems, together with the stop-loss preservation property for convolutions of independent risks, immediately lead to statements about the stop-loss premiums of such portfolios. Take Theorem 4 as an example. Consider a portfolio with given distribution functions of the individual risks where the only non-independent risks appear in couples and where the risks of each couple are positive quadrant dependent. Then we find from Theorem 4 that taking the independence assumption will always lead to underestimated values for the stop-loss premiums of the portfolio under consideration.

Let us now illustrate the effect of introducing dependencies between risks in an insurance portfolio by a numerical example. We will use Gerber's (1979) life insurance portfolio which is represented in the following table.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>GERBER'S PORTFOLIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>claim probability</td>
<td>amount at risk</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.03</td>
<td>2</td>
</tr>
<tr>
<td>0.04</td>
<td>-</td>
</tr>
<tr>
<td>0.05</td>
<td>-</td>
</tr>
<tr>
<td>0.06</td>
<td>-</td>
</tr>
</tbody>
</table>

The portfolio consists of 31 risks. Each risk can either produce no claim or a fixed positive claim amount (the amount at risk) during a certain reference period. The claim probability is the probability that the risk produces a claim during the reference
period. The expectation of the aggregate claims equals 4.49. We label the risks from 1 to 31, row by row. Hence, risks 1 and 2 have claim probability 0.03 and a conditional claim amount (given that a claim occurs) equal to 1: risks 3, 4 and 5 have claim probability 0.03 and conditional claim amount 2, ... .

In Table 2 several independency assumptions for this portfolio are considered.

### Table 2: Description of Several Independency Assumptions

<table>
<thead>
<tr>
<th>Situation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>All risks</td>
<td>(1,2)</td>
<td>(24,31)</td>
<td>(1,2)</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Mutually independent</td>
<td>(3,4)</td>
<td>14.23</td>
<td>(3,4)</td>
<td>dependency</td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td>(5,6)</td>
<td>(29,30)</td>
<td>(5,6)</td>
<td>assumptions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7,8)</td>
<td>(21,22)</td>
<td>(7,8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(9,10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11,12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(13,14)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In situation 1 it is assumed that all risks are mutually independent. Situation 2 corresponds to the case that the only couples that occur in the portfolio are (1, 2), (3, 4), (5, 6) and (7, 8). In situation 3 there are also 4 couples. Comparing situations 2 and 3, we see that in the latter case the couples have higher claim probabilities and higher conditional claim amounts. Situation 4 is an extension of situation 2 in the sense that it not only contains the couples of situation 2, but also some others. Finally, situation 5 corresponds to the case that no independency assumptions are made so that all risks can be dependent. The results that will be stated for this situation can be found in Dhaene et al. (1995).

In the following table the ratio (multiplied by 100) of the maximal stop-loss premium (according to Theorem 3) divided by the stop-loss premium in the independent case (assumption 1) is given for the situations considered in Table 2.

### Table 3: Relative Height of the Maximal Stop-Loss Premiums Under Several Independency Assumptions

<table>
<thead>
<tr>
<th>Retention</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>101.6</td>
<td>103.8</td>
<td>103.9</td>
<td>146.6</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>103.8</td>
<td>116.5</td>
<td>110.9</td>
<td>239.3</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>108.0</td>
<td>137.6</td>
<td>122.1</td>
<td>412.6</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>112.8</td>
<td>169.1</td>
<td>137.7</td>
<td>778.6</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>120.7</td>
<td>206.4</td>
<td>159.8</td>
<td>1549.8</td>
</tr>
<tr>
<td>12</td>
<td>100</td>
<td>130.1</td>
<td>226.4</td>
<td>191.2</td>
<td>3336.3</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
<td>143.8</td>
<td>354.2</td>
<td>233.3</td>
<td>7604.2</td>
</tr>
</tbody>
</table>
From this table we can conclude that in any situation the relative increase of the stop-loss premium is an increasing function of the retention. For the higher rententions the effect will be most dramatically. Comparing the assumptions 2 and 3, we see that increasing the claim probabilities and the claim amounts of the couples leads to an increased effect. Of cours, increasing the number of coupled risks will increase the relative effect on the maximal stop loss premiums, as can be seen from comparing the assumptions 2 and 4. Finally, from the last column we can conclude that assuming no independency at all, and hence allowing all possible kinds of dependencies, the extremal stop-loss premiums increase astronomically. The specific dependency relations that give rise to this extremal stop-loss premiums for a life insurance portfolio are derived in Dhaene et al. (1995).

Finally, we remark that in this paper we have only derived results for bivariate dependencies. The special, but important bivariate case will often be sufficient to describe dependencies in portfolios but is also provides a theoretical stepping stone towards the concept of dependence in the multivariate case. Some notions of dependence in the multivariate case can be found in Barlow et al. (1975). One of the notions of multivariate dependency which is often used in actuarial science is the exchangeability of risks, see e.g. Jewell (1984). It is a (remarkable) pity that the usefulness of other notions of multivariate dependency has hardly been considered in the actuarial literature.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous referees for their helpful comments.

REFERENCES


MODIFIED RECURSIONS FOR A CLASS OF COMPOUND DISTRIBUTIONS

KARL-HEINZ WALDMANN

Inst. f. Wirtschaftstheorie und Operations Research, Universität Karlsruhe, D-76128 Karlsruhe

ABSTRACT

Recursions are derived for a class of compound distributions having a claim frequency distribution of the well known (a,b)-type. The probability mass function on which the recursions are usually based is replaced by the distribution function in order to obtain increasing iterates. A monotone transformation is suggested to avoid an underflow in the initial stages of the iteration. The faster increase of the transformed iterates is diminished by use of a scaling function. Further, an adaptive weighting depending on the initial value and the increase of the iterates is derived. It enables us to manage an arbitrary large portfolio. Some numerical results are displayed demonstrating the efficiency of the different methods. The computation of the stop-loss premiums using these methods are indicated. Finally, related iteration schemes based on the cumulative distribution function are outlined.

KEYWORDS

Collective risk model, aggregate claims distribution, Panjer's algorithm.

1. INTRODUCTION

Compound distributions are used extensively in modeling the total amount of claims, X, in an insurance portfolio. Based on a claim frequency distribution satisfying the recursion

\[ p_n = \left( a + \frac{b}{n} \right) p_{n-1}, \quad n \in \mathbb{N} := \{1, 2, \ldots\} \]  

the probability mass function \( g(x) = P(X = x), x \in \mathbb{N}, \) is often evaluated recursively as

\[ g(x) = \sum_{i=1}^{x} \left( a + b \frac{i}{x} \right) f(i) g(x - i) \]  

starting with

\[ g(0) = p_0, \]

where \( f(i), i \in \mathbb{N}, \) denotes the probability mass function of the iid claim sizes \( Y_1, Y_2, \ldots \)

Applying this well known recursion (see, e.g., Panjer and Willmot (1992), Sundt (1991) for details) to a portfolio with a large number of contracts, the initial value \( g(0) \) is close to zero. This fact may cause an underflow (on a computer with standard software) followed by an abort or irregular running of the procedure. Panjer and Willmot (1986) (and Waldmann (1994, 1995) within the setting of an individual life model) suggest the use of a scaling function to stabilize the algorithm with respect to underflow/overflow. Moreover, Panjer and Wang (1993) study the stability of this type of recursion from a more theoretical point of view.

To overcome the problem of underflow in the initial and final stages of the iteration, we reformulate iteration scheme (2) with the probability mass function \( g(x) \) replaced by the distribution function \( G(x) = P(X \leq x) \). The resulting recursion has the nice property of producing increasing values lying within the unit interval. However, an underflow of the initial values is still possible. Therefore we transform \( G(x) \) to \( H(x) = \frac{G(x) - G(0)}{G(0)} \) avoiding an underflow in the initial stage of the algorithm. The stronger increase of the transformed values \( H(x) \) may lead to an overflow in the final stage of the algorithm. This difficulty, however, can be partially managed by retransforming \( H(x) \) to \( G(x) \) for some \( x_0 \in \mathbb{N} \) and continuing with the iteration scheme for \( G(x) \). Moreover, the increase of the transformed values \( H(x) \) can be diminished by use of a scaling function of type \( \exp(-\alpha - \beta x) \) for suitable constants \( \alpha \) and \( \beta \). Scaling functions of this type considerably extend the range of applicability of the recursion but cannot avoid a breakdown by letting the expected number of claims tend to infinity. Therefore, we also present an adaptive transformation of \( G(x), x \in \mathbb{N} \), which enables us to manage an arbitrary large portfolio. The flexibility of the transformation results from its recursive definition depending on the initial value and the increase of the iterates. It is realized by dividing the range of \( G(0), G(1), \ldots \) into \( L \) layers and iterating in these layers successively. To make each layer representable on the computer, a scaling function is used, which is constant within a layer and suitably adapted by switching from layer \( \ell \) to \( \ell + 1 \).

The paper is organized as follows. The iteration scheme is given in Section 2. Section 3 contains the transformed iteration schemes. Some numerical results are displayed in section 4 demonstrating the efficiency and applicability of the different methods. In Section 5 we extend our approach to a claim frequency distribution satisfying recursion (1) for \( n = m + 1, m + 2, \ldots \) and some \( m \in \mathbb{N} \) only. The calculation of the stop-loss premiums using the methods of Sections 2 and 3 are indicated in section 6. Finally, Section 7 is devoted to a set of iteration schemes based on the cumulative distribution function \( \hat{G}(x) := \sum_{i=0}^{x} G(i) \).

2. AN ITERATION SCHEME FOR \( G(x) \)

In the following let an empty sum \( \sum_{i=1}^{0} \ldots \) be defined to be zero. By slightly modifying a standard approach in deriving the iteration scheme for \( g(x), x \in \mathbb{N} \), we are in a position to obtain the following recursion for \( G(x) \).
Theorem 1: \( G(x), x \in \mathbb{N}, \) can be evaluated recursively as

\[
xG(x) = r_1(x) + r_2(x).
\]

where \( G(0) = p_0 \) and, for all \( x \in \mathbb{N}, \)

\[
\begin{align*}
r_1(x) &= r_1(x - 1) + G(x - 1) \\
r_2(x) &= a \sum_{i=1}^{x-1} f(i) r_2(x - i) + (a + b) \sum_{i=1}^{x} i f(i) G(x - i)
\end{align*}
\]

with \( r_1(0) = 0. \)

Proof. Introduce the generating functions \( \varphi(z) = \sum_{x=0}^{\infty} g(x)z^x, \Phi(z) = \sum_{x=0}^{\infty} G(x)z^x, \) and \( \hat{\Phi}(z) = \sum_{x=0}^{\infty} \hat{G}(x)z^x. \) Note that \( \Phi(z) = \varphi(z)/(1 - z), \hat{\Phi}(z) = \hat{\varphi}(z)/(1 - z). \) Further, let \( \Psi(z) = \sum_{x=0}^{\infty} f(x)z^x \) be the generating function of the claim size distribution.

To derive the recursion formula for \( G(x) \) we start with the well known identity

\[
\varphi(z) = \sum_{x=0}^{\infty} g(x)z^x = \Phi(z)/(1 - z)
\]

Differentiating both sides with respect to \( z \) we obtain

\[
(1 - z)\Phi'(z) - \Phi(z) = \sum_{n=1}^{\infty} np_n \Psi(z)^{n-1} \Psi'(z)
\]

\[
= a \sum_{n=1}^{\infty} (n - 1)p_{n-1} \Psi(z)^{n-1} \Psi'(z) + (a + b) \sum_{n=1}^{\infty} p_{n-1} \Psi(z)^{n-1} \Psi'(z)
\]

\[
= a\Psi(z)((1 - z)\Phi'(z) - \Phi(z)) + (a + b)\Psi'(z)(1 - z)\Phi(z)
\]

Now, multiplying both sides by \( z/(1 - z), \) the last equation can also be written as

\[
z\Phi'(z) = z\Phi(z) + a\Psi(z) [z\Phi'(z) - z\hat{\Phi}(z)] + (a + b)z\Psi'(z)\Phi(z)
\]

Finally, using \( z\Phi'(z) = \sum_{x=1}^{\infty} xG(x)z^x \) and an analogous representation of \( z\Psi'(z), \) a comparison of the coefficients of \( z^x, x \in \mathbb{N}, \) leads to the identity

\[
xG(x) = r_1(x) + r_2(x)
\]

where

\[
r_1(x) = \sum_{j=0}^{x-1} G(j) = r_1(x - 1) + G(x - 1)
\]
\begin{align*}
r_2(x) &= a \sum_{i=1}^{x-1} f(i) \left[ (x-i)G(x-i) - \sum_{j=0}^{x-i-1} G(j) \right] + (a+b) \sum_{i=1}^{x} i f(i)G(x-i) \\
&= a \sum_{i=1}^{x-1} f(i) r_2(x-i) + (a+b) \sum_{i=1}^{x} i f(i)G(x-i) \\
\end{align*}

(with $G(0) = p_0, r_1(0) = 0$).

It easily follows from (3) that $r_1(x)$ and $r_2(x)$ are nonnegative and increasing functions of $x$. $r_1(x)$ can be implemented as a single number to be adapted at each step of iteration. Additionally, if both $a$ and $a+b$ are nonnegative, which holds for the important cases of a Poisson counting distribution ($a = 0, b = \lambda$) and a negative binomial counting distribution ($a = p, b = p(\gamma - 1)$), $r_2(x)$ is the result of additions and multiplications of nonnegative real-valued numbers.

It is not necessary to recursively determine $r_2(x)$. By rearranging its defining terms in (5) we obtain

**Corollary 1:** $G(x), x \in \mathbb{N}$, can be evaluated as in Theorem 1 with (4) replaced by

\begin{equation}
r_2(x) = -a \sum_{i=1}^{x-1} f(i) r_1(x-i) + \sum_{i=1}^{x} (ax + bi) f(i) G(x-i) \tag{4'}
\end{equation}

Note, however, that the numbers to be added/multiplicated in (4') are no longer nonnegative.

Looking at the binomial counting distribution ($a = -p/(1-p), b = -(n+1)a$), (4) and (4') have both positive and negative terms. The numerical results, however, which will be displayed in section 4 below give no hint for an instability with respect to rounding errors.

For a geometric counting distribution the recursion for $G(x)$ can already be found in Sundt (1991), p. 114.

**Corollary 2:** In case of a geometric counting distribution ($a = p, b = 0$) it holds that

\begin{equation}
G(x) = 1 - p + p \sum_{i=1}^{x} f(i) G(x-i) \tag{6}
\end{equation}

with $G(0) = 1 - p$.

**Proof.** With (10.14) in Sundt (1991) and $r_1(x)$ as in Theorem 1 we infer by induction on $x$

\[ G(x) - p \sum_{i=1}^{x} f(i) G(x-i) = \frac{1}{x} \left( r_1(x) - p \sum_{i=1}^{x} f(i) r_1(x-i) \right) = 1 - p, \]

which is the desired result. \qed
3. Stabilization of the Algorithm With Respect to Underflow/Overflow

The recursion for \( G(x) \) has the nice property of being monotone, but the initial value \( p_0 \) may cause an underflow followed by an abort or irregular running of the procedure. Our first step in guaranteeing a regular running of the procedure is based on the following theorem.

**Theorem 2:** The transformed values \( H(x) = \frac{[G(x) - G(0)]}{G(0)}, x \in \mathbb{N} \), can be computed recursively via

\[
x H(x) = h_1(x) + h_2(x),
\]

where \( H(0) = 0 \), and, for \( x \in \mathbb{N} \),

\[
\begin{align*}
  h_0(x) &= h_0(x - 1) + x f(x) \\
  h_1(x) &= h_1(x - 1) + H(x - 1) \\
  h_2(x) &= (a + b) h_0(x) + a \sum_{i=1}^{x-1} f(i) h_2(x - i) + (a + b) \sum_{i=1}^{x-1} i f(i) H(x - i)
\end{align*}
\]

with \( h_0(0) = 0 \) and \( h_1(0) = 0 \).

**Proof.** Set \( h_0(0) = h_1(0) = h_2(0) = 0 \). Then, together with Theorem 1, for \( x \in \mathbb{N} \),

\[
\begin{align*}
  h_0(x) &= \sum_{i=1}^{x} i f(i) = h_0(x - 1) + x f(x) \\
  h_1(x) &= [r_1(x) - x G(0)] / G(0) = h_1(x - 1) + [G(x - 1) - G(0)] / G(0) \\
  h_2(x) &= x H(x) - h_1(x) \\
  &= [x G(x) - x G(0) - r_1(x) + x G(0)] / G(0) \\
  &= r_2(x) / G(0) \\
  &= a \sum_{i=1}^{x-1} f(i) r_2(x - i) / G(0) + (a + b) \sum_{i=1}^{x} i f(i) [(G(x - i) - G(0)) + G(0)] / G(0) \\
  &= a \sum_{i=1}^{x-1} f(i) h_2(x - i) + (a + b) \left[ \sum_{i=1}^{x-1} i f(i) H(x - i) + h_0(x) \right]
\end{align*}
\]

giving the desired recursion. \( \Box \)

The function \( h_0(x) \) avoids that the sequence \( H(x) \) degenerates to a sequence that has all its elements equal to zero. \( h_0(x) \) is further a measure for the increase of the iterates. Since \( H(x) \to (1 - p_0) / p_0 \) for \( x \to \infty \), it may be necessary to retransform \( H(x) \) to \( G(x) \) for some \( x_0 \in \mathbb{N} \) and to continue with the recursive computation of \( G(x) \), \( x \geq x_0 \).

Moreover, the increase of \( H(x) \) can be diminished by weighting \( H(x) \) by \( \exp(- (\alpha + \beta x)) \) for suitable parameters \( \alpha \in \mathbb{R} := (-\infty, \infty), \beta \geq 0 \). The resulting recursion is given in the following Theorem.
Theorem 3: For $\alpha \in \mathbb{R}$, $\beta \geq 0$, the transformed values $\tilde{H}(x) = H(x)e^{-(\alpha+\beta x)}$, $x \in \mathbb{N}$, can be evaluated recursively as
\[
x\tilde{H}(x) = \tilde{h}_1(x) + \tilde{h}_2(x)
\] (8)
where $\tilde{H}(0) = 0$, and, for $x \in \mathbb{N}$,
\[
\begin{align*}
0 & = h_0(x) = h_0(x - 1) + x f(x) \\
\tilde{h}_1(x) & = e^{-\beta}[\tilde{h}_1(x - 1) + \tilde{H}(x - 1)] \\
\tilde{h}_2(x) & = (a + b)e^{-(\alpha+\beta x)}h_0(x) + a \sum_{i=1}^{x-1} e^{-\beta i}\tilde{h}_2(x - i) + (a + b) \sum_{i=1}^{x-1} e^{-\beta i}\tilde{H}(x - i)
\end{align*}
\]
with $h_0(x) = 0$ and $\tilde{h}_1(0) = 0$.

The parameter $\alpha$ of the scaling function $\exp(-\alpha - \beta x)$ gives a constant weight to $h_0(x)$ and can be used to reduce the order of $h_0(x)$. In addition, the parameter $\beta$ can be utilized to diminish the increase of $h_0(x)$ and the resulting $\tilde{H}(x)$. The parameter $\beta$, however, is much more sensitive than $\alpha$. If $\beta$ is too large, isotonicity of $\tilde{H}(x)$ does no longer hold for all $x \in \mathbb{N}$. In such a case things may change and the transformation may lead to an earlier abort on account of an underflow.

The use of an exponential scaling function considerably extends the range of applicability of the recursion but cannot avoid a breakdown by letting the expected number of claims tend to infinity. We next present an adaptive transformation of $G(x)$, $x \in \mathbb{N}$, which enables us to manage an arbitrary large portfolio. The flexibility of the transformation results from its recursive definition depending on the initial value and the increase of the iterates. It is realized by dividing the range of $G(0)$, $G(1), \ldots$ into $L$ layers and iterating in these layers successively. To make each layer representable on the computer, a scaling function is used, which is constant within a layer and suitably adapted by switching from layer $\ell$ to $\ell + 1$.

Let $\omega$ and $\Omega$ denote the smallest and greatest positive numbers, respectively, that can be represented on the computer using standard software. We interpret the interval $[\omega, \Omega]$ as the size of a layer. Further we introduce a subinterval $[10^{-t}, 10^{T}]$ of $[\omega, \Omega]$ for suitable constants $t$, $T > 0$. The interval $[10^{-t}, 10^{T}]$ is the region in a layer, in which the iteration is started (resp. restarted) and continued (up to some value greater than $10^{T}$). Clearly, to avoid rounding errors, the set $[10^{-t}, 10^{T}]$ has to be chosen 'smaller' than $[\omega, \Omega]$.

In addition to $t$ and $T$, the number $L$ of layers depends on $p_0$. Set
\[
c := -\log_{10} p_0
\]
(i.e. $10^{-c} = p_0$). Then $L$ can be chosen such that
\[
t + (L - 2)(T + t) < c \leq t + (L - 1)(T + t)
\]
holds. If $L = 1$, i.e. $p_0 \geq 10^{-t}$, there is no need for a transformation. Therefore assume $L > 1$. Finally, let $\xi$ be the largest $x \in \mathbb{N}$ with $f(x) > 0$ (and $\infty$ if there is no such one).

The resulting transformed iterates $G^*(0)$, $G^*(1)$, \ldots can now be recursively defined as follows:
(a) **Layer 1.** Set

\[ G^*(0) = 10^{-t} \]

and compute \( G^*(1), G^*(2), \ldots \) up to some \( x_1 \), say, with \( G^*(x_1) > 10^T \) according to

\[ xG^*(x) = r_1^*(x) + r_2^*(x), \tag{9} \]

where

\[ r_1^*(x) = r_1^*(x-1) + G^*(x-1) \]

\[ r_2^*(x) = a \sum_{i=1}^{x-1} f(i)r_2^*(x-i) + (a + b) \sum_{i=1}^{x} i f(i)G^*(x-i) \]

with \( r_1^*(0) = 0 \).

(b) **Layer \ell.** (2 \( \leq \ell \leq L - 1 \)). Reset

\[ G^*(x) = 10^{-(T+t)}G^*(x) \]

\[ r_\nu^*(x) = 10^{-(T+t)} r_\nu^*(x), \quad \nu \in \{1, 2\} \]

for all \( x_{\ell-1} - \xi < x \leq x_{\ell-1} \) (with \( G^*(x) \), \( r_1^*(x) \), \( r_2^*(x) \) equal to zero if they are less than \( \omega \)) and compute \( G^*(x_{\ell-1} + 1), G^*(x_{\ell-1} + 2), \ldots \) up to some \( x_\ell \), say, with \( G^*(x_\ell) > 10^T \) according to (9).

(c) **Layer L.** Set \( \gamma = c - t - (L - 2)(T + t) \). Reset

\[ G^*(x) = 10^{-\gamma}G^*(x) \]

\[ r_\nu^*(x) = 10^{-\gamma} r_\nu^*(x), \quad \nu \in \{1, 2\} \]

for all \( x_{L-1} - \xi < x \leq x_{L-1} \) and compute \( G^*(x_{L-1} + 1), G^*(x_{L-1} + 2), \ldots \) according to (9).

Summarizing the steps of iteration we immediately obtain

**Theorem 4:** Let \( L > 1 \). Evaluate \( G^*(0), G^*(1), \ldots \) as above. Then, for all \( x > x_{L-1} - \xi \),

\[ G(x) = G^*(x) \]

\[ r_\nu(x) = r_\nu^*(x), \quad \nu \in \{1, 2\}. \]

Further, for all \( s = 1, \ldots, L - 1 \), and all \( x_{L-s-1} - \xi < x \leq x_{L-s} - \xi \)

\[ G(x) = 10^{-\gamma-(s-1)(T+t)}G^*(x) \]

\[ r_\nu(x) = 10^{-\gamma-(s-1)(T+t)} r_\nu^*(x), \quad \nu \in \{1, 2\} \]

(with \( x_0 = 0 \) and \( G(x') = 0 \) for \( x' < 0 \)).
Our numerical results are ascertained with a computer program written in Turbo Pascal 5.0. We used real-valued variables of type 'extended' having a range from $1.9 \times 10^{-4951}$ to $1.1 \times 10^{4932}$. Thus $\omega = 1.9 \times 10^{-4951}$ and $\Omega = 1.1 \times 10^{4932}$.

We consider as a starting point the portfolio of $\bar{n} = 31$ independent life insurance policies discussed in Gerber (1979), p. 53. Each policy is supposed to have an amount at risk $i \in I = 1, \ldots, 5$ and a mortality rate $\bar{q}_j$ with $j \in J = 1, \ldots, 4$. Further $\bar{n}_{ij}$ denotes the number of all policies with amount at risk $i$ and mortality rate $\bar{q}_j$. Note that the expected number of claims is 1.4.

This individual life model is approximated by a compound Poisson model with $a = 0$, $b = \bar{n}\lambda = \sum_{i \in I} \sum_{j \in J} \bar{n}_{ij}\bar{q}_j$ and $f(i) = \sum_{j \in J} \bar{n}_{ij}\bar{q}_j/\bar{n}\lambda$, $i \in I$, and by a compound binomial model with $a = -\lambda/(1 - \lambda)$, $b = (\bar{n} + 1)\lambda/(1 - \lambda)$, and $f(i)$ as in the compound Poisson model (cf., e.g., Kuon, Radke, and Reich (1993)).

Since the portfolio consists of 31 policies only, there is no need for a stabilization of (3) with respect to underflow. We therefore expand the portfolio by considering $k\bar{n}_{ij}$ policies in place of $\bar{n}_{ij}$ (for all $i \in I$ and $j \in J$).

The recursions (7) and (8) on which our transformed iterates $H(x)$ and $\tilde{H}(x)$ are based have been carried out up to some $x_0 \in \mathbb{N}$ guaranteeing $G(x_0) > 10^{-4000}$. For $x > x_0$, after retransforming the relevant data, recursion (3) has been applied to determine $G(x)$ directly. Moreover, we have distinguished between recursions (4) and (4') when calculating $r_2(x)$. Based on (4) and (4') also the transformed iterates $H(x)$ and $\tilde{H}(x)$ have been studied separately.

We say that a recursion is stable if the algorithm does not stop with an underflow or overflow and that both $|E'(X) - E''(X)|/E''(X) \leq 10^{-5}$ and $|\text{Var}'(X)^{1/2} - \text{Var}''(X)^{1/2}|/\text{Var}''(X)^{1/2} \leq 10^{-5}$ hold, where $E'(X)$ and $\text{Var}'(X)$ are determined with the help of the probability mass function of $X$ and $E''(X)$, $\text{Var}''(X)$ result from the moments of the counting distribution and claim size distribution together with the properties of expectation and variance. The maximal $k$ and the associated number of policies we have obtained in this way are displayed in Table 2. Although the recursions for $G(x)$ and $H(x)$ nearly work within the same range of $k$, there is an essential difference. By increasing $k$, the recursion for $G(x)$ aborts with an underflow resulting from the initial value $G(0)$, the one for $H(x)$ starts with stable initial values and aborts with an overflow. The reduction of the

<table>
<thead>
<tr>
<th>Mortality Rate</th>
<th>Amount at Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.03</td>
<td>2</td>
</tr>
<tr>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>
increase of \( H(x) \) as realized by use of \( \tilde{H}(x) \) then gives for both the compound Poisson model and the compound binomial model (independent of the use of (4) or (4')) stable solutions for a portfolio with nearly 2 million contracts.

We already mentioned that an \( L \) layer model can be applied to an arbitrary large portfolio. To give some insight into the increase of the number \( L \) of layers when the number of contracts is increased, we have used the interval \([10^{-r}, 10^r] = [10^{-4000}, 10^{+4000}]\) for carrying out the iterations.

For \( k = 10^4, 10^5, 10^6 \) (which corresponds to \( 3.1 \cdot 10^5, 3.1 \cdot 10^6, 3.1 \cdot 10^7 \) contracts) the number \( L \) of layers needed is displayed in Table 3.

### 5. MODIFICATION OF THE CLAIM NUMBER DISTRIBUTION

The class of counting distributions can be extended by supposing the recursion

\[
p_n = \left( a + \frac{b}{n} \right) p_{n-1}, \quad n = m + 1, m + 2, \ldots
\]

(10)

to hold for some \( m \in \mathbb{N}_0 := 0, 1, 2, \ldots \) only. In this more general situation the iteration scheme for the probability mass function of \( X, g(x) \), reads (cf., e.g., Panjer and Willmot (1992), Corollary 6.16.1)

### TABLE 3

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\textbf{\( k \)} & \textbf{Poisson} & \textbf{binomial} \\
\hline
\( 10^4 \) & \( L = 1 \) & \( L = 1 \) \\
\( 10^5 \) & \( L = 8 \) & \( L = 8 \) \\
\( 10^6 \) & \( L = 77 \) & \( L = 78 \) \\
\hline
\end{tabular}
\end{table}
\[ g(x) = \frac{1}{x} \left\{ \sum_{i=1}^{x} (a + bi)f(i)g(x - i) \right\} + \sum_{n=1}^{m} q_n f^{n*}(x), \]

where \( f^{n*} \) denotes the \( n \)-fold convolution of \( f \) with itself and where \( q_n \) is defined by

\[ q_n := p_n - \left( a + \frac{b}{n} \right) p_{n-1}, \quad n = 1, \ldots, m. \]

Essentially the same arguments as in the proof of Theorem 1 give

**Theorem 5:** \( G(x), x \in \mathbb{N}, \) can be evaluated recursively as

\[ xG(x) = r_1(x) + r_2(x), \]

where

\[ r_1(x) = r_1(x-1) + G(x-1) + \sum_{n=1}^{m} q_n x f^{n*}(x), \quad x \in \mathbb{N}, \]

(with \( G(0) = p_0, r_1(0) = 0 \)) and \( r_2(x) \) as in (4).

Being interested in extending Theorem 2, we only have to replace the recursion for \( h_1(x) \) by \( h_1(x) = h_1(x-1) + H(x-1) + \sum_{n=1}^{m} (q_n/p_0) x f^{n*}(x). \) Analogously, in Theorem 3, we have to redefine \( h_1(x) \) by

\[ h_1(x) = e^{-\beta} [h_1(x-1) + \tilde{H}(x-1)] + e^{-(\alpha + \beta x)} \sum_{n=1}^{m} (q_n/p_0) x f^{n*}(x). \]

The \( L \)-layer approach does not work in case of the claim frequency distribution (10).

### 6. Evaluation of the stop-loss premiums

Let us begin with the claim frequency distribution (1). It is well known that the stop-loss premium \( SL(\tau), \) \( SL(\tau) := \sum_{x=\tau+1}^{\infty} (x - \tau)g(x), \) with retention \( \tau \in \mathbb{N} \) can be written as

\[ SL(\tau) = E(X) - \tau + \tilde{G}(\tau - 1), \quad \tau \in \mathbb{N}. \] (11)

Using Theorem 1 to determine \( G(x), \) then \( r_1(x) = \tilde{G}(x - 1) \) is obtained as a byproduct. Thus the results of Sections 2 and 3 can also be utilized to compute the stop-loss premiums for specified retentions. In particular, using Theorem 1, \( SL(\tau) = E(X) - \tau + r_1(\tau). \) Using the transformed iterates \( H(x) \) and \( \tilde{H}(x), \) \( r_1(x) \) follows from \( r_1(x) = (h_1(x) + x)p_0 \) and \( r_1(x) = [e^{\alpha + \beta x} h_1(x) + x]p_0, \) respectively. Applying the \( L \)-layer method, \( r_1(x) \) results from \( r_1^L(x) \) and is given explicitly in Theorem 4.

In case of the more general claim frequency distribution (10), \( r_1(x) = \tilde{G}(x - 1) + \sum_{n=1}^{m} q_n \sum_{i=1}^{x} i f^{n*}(i). \) The transformed iterates \( H(x) \) and \( \tilde{H}(x) \) (with the recursions for \( h_1(x) \) and \( \tilde{h}_1(x) \) as defined in Section 5) give \( r_1(x) \) as in the case of the claim frequency distribution (1). The \( L \)-layer approach, however, does not work.
7. Iteration schemes based on $\hat{G}(x)$

The iteration schemes which will be presented in this section are based on the cumulative distribution function $\hat{G}(x)$. Forming the first and second differences of $\hat{G}(x)$, $\Delta \hat{G}(x) := \hat{G}(x + 1) - \hat{G}(x) = G(x + 1)$ and $\Delta^2 \hat{G}(x) := \Delta^2 \hat{G}(x + 1) - \Delta \hat{G}(x) = g(x + 2)$, we immediately obtain the distribution function $G(x)$ and the probability mass function $g(x)$, respectively.

Note that $\hat{G}(x)$ has the nice property of being an increasing and convex function. In case of the claim frequency distribution (1) the recursion reads

Theorem 6: Let $m = 0$. Then $\hat{G}(x), x \in \mathbb{N}$, can be evaluated recursively as

$$x \hat{G}(x) = \hat{r}_1(x) + \hat{r}_2(x),$$

where $\hat{G}(0) = p_0$ and, for all $x \in \mathbb{N},$

$$\hat{r}_1(x) = \hat{r}_1(x - 1) + 2\hat{G}(x - 1)$$

$$\hat{r}_2(x) = a \sum_{i=1}^{x-1} f(i) \hat{r}_2(x - i) + (a + b) \sum_{i=1}^{x} i f(i) \hat{G}(x - i)$$

with $\hat{r}_1(0) = 0$.

Proof. Starting with the identity (cf. proof of Theorem 1)

$$(1 - z)^2 \hat{\Phi}(z) = \sum_{n=0}^{\infty} p_n \hat{\Psi}(z)^n$$

similar arguments as in the proof of Theorem 1 give the desired recursion.

Since (3) and (12) (formally) differ in a factor 2 only, the methods of Section 3 can be adapted easily. Using (11) also the stop-low premiums follow immediately by retransforming the transformed iterates $H(x), H'(x)$ and $G^*(x)$, say, to $\hat{G}(x)$.

In case of the more general claim frequency distribution (10) the iterates $\hat{G}(x)$ and its transformed versions can be obtained in a straightforward manner following the approach given in section 5.

REFERENCES


ON BOUNDS FOR THE DIFFERENCE BETWEEN THE STOP-LOSS TRANSFORMS OF TWO COMPOUND DISTRIBUTIONS

BJØRN SUNDT¹ AND JAN DAENE²

ABSTRACT

In the present note we deduce a class of bounds for the difference between the stop-loss transforms of two compound distributions with the same severity distribution. The class contains bounds of any degree of accuracy in the sense that the bounds can be chosen as close to the exact value as desired; the time required to compute the bounds increases with the accuracy.

1. INTRODUCTION

During the last twenty years, there has grown up a large literature on approximations and inequalities for stop-loss premiums under various assumptions. One way of approximation is to approximate the original distribution with another distribution that makes the evaluation simpler. In such cases it is useful to have bounds for the difference between the exact stop-loss premium and the approximation, that is, it is of interest to have bounds for the difference between the stop-loss transforms of two distributions.

When approximating the stop-loss transform of a compound distribution, it is sometimes convenient to replace the counting distribution with another distribution, e.g. a Bernoulli distribution or a Poisson distribution, and keep the severity distribution unchanged. Such approximations are discussed by i.a. Dhaene & Sundt (1996).

In the present note we deduce classes of bounds for the difference between the stop-loss transforms of two compound distributions with the same severity distribution. The classes contain bounds of any degree of accuracy in the sense that the bounds can be chosen as close to the exact value as desired; the time required to compute the bounds increases with the accuracy.

2. NOTATION AND CONVENTIONS

Let $\mathcal{R}_+$ and $\mathbb{Z}_+$ denote the sets of respectively the non-negative real numbers and the non-negative integers, and $\mathcal{P}_{\mathbb{R}_+}$ and $\mathcal{P}_{\mathbb{Z}_+}$ the classes of probability distributions with

¹ University of Bergen, Norway
² Katholieke Universiteit Leuven, Belgium

finite mean on respectively $\mathcal{R}_+$ and $\mathbb{Z}$. For distributions in $\mathcal{P}_{\mathcal{R}_+}$, we shall denote the cumulative distribution function with a capital letter and, for distributions in $\mathcal{P}_{\mathbb{Z}+}$, the discrete density function with the corresponding lower case letter. The stop-loss transform of a distribution will be denoted by a horizontal bar on the top of the symbol of the distribution, that is, for a distribution $F \in \mathcal{P}_{\mathcal{R}_+}$, we have the stop-loss transform $\overline{F}$ given by

$$\overline{F}(x) = \int_x^\infty (y-x)dF(y) = \int_x^\infty (1-F(y))dy. \quad (x \geq 0)$$

The mean of $F$ is denoted by $\mu_F$, that is,

$$\mu_F = \overline{F}(0) = \int_0^\infty ydF(y) = \int_0^\infty (1-F(y))dy.$$

We shall denote a compound distribution with counting distribution $P \in \mathcal{P}_{\mathbb{Z}+}$ and severity distribution $H \in \mathcal{P}_{\mathcal{R}_+}$ by $P \vee H$, that is,

$$P \vee H = \sum_{n=0}^\infty p(n)H^n*,$$

where $H^n*$ denotes the $n$-fold convolution of $H$.

For $F \in \mathcal{P}_{\mathcal{R}_+}$ and $r \in \mathcal{R}_+$, we define the approximation $F_r$ by

$$F_r(x) = \begin{cases} F(x) & (0 \leq x < r) \\ 1 & (x \geq r) \end{cases}$$

This approximation can be interpreted as the distribution obtained by setting all observations greater than $r$ equal to $r$. The limiting cases $r = 0$ and $r = \infty$ correspond to respectively the distribution concentrated in zero and the original distribution $F$.

We shall interpret $\sum_{i=a}^b v_i = 0$ when $b < a$.

3. RESULTS

3.1. The following lemma is proved as formula (38) in De Pril & Dhaene (1992) for the special case $r = 1$; the proof is easily extended to the general case.

**Lemma 1** For $H \in \mathcal{P}_{\mathcal{R}_+}$, $r, m \in \mathbb{Z}$ such that $r \leq m$, and $x \in \mathcal{R}_+$, we have

$$(m-r)\overline{H}(x) \leq \overline{H^m*}(x) - \overline{H^r*}(x) \leq (m-r)\mu_H.$$
Proof. We have
\[ P \vee H(x) - Q \vee H(x) = \sum_{n=r}^{\infty} p(n) (H^{n*}(x) - H^{*}(x)). \]

Application of Lemma 1 gives
\[ \sum_{n=r}^{\infty} p(n) (n - r) H(x) \leq P \vee H(x) - Q \vee H(x) \leq \sum_{n=r}^{\infty} p(n) (n - r) \mu_H, \]
from which we obtain (1). Q.E.D.

The second inequality in (1) was proved under more general assumptions by Sundt (1991), who also showed that \( 0 \leq P \vee H(x) - Q \vee H(x), \) which is weaker than the first inequality in (1).

If \( P_r = P, \) that is, \( P(r) = 1, \) then the bounds in (1) become equal to zero.

Lemma 1 appears as a special case of Lemma 2 by letting \( P \) be the distribution concentrated in \( m. \)

3.2. For \( P, Q \in \mathcal{P}_\mathcal{R}, H \in \mathcal{P}_\mathcal{R}, r \in \mathcal{Z}, \) and \( x \in \mathcal{R}, \) we introduce
\[ B_r(x; P, Q, H) = P_r \vee H(x) - Q_r \vee H(x) + \mu_H P(r) - H(x) Q(r), \]
which can also be written as
\[ B_r(x; P, Q, H) = \sum_{n=1}^{r-1} (p(n) - q(n)) H^{n*}(x) - (P(r - 1) - Q(r - 1)) H^{*}(x) + \mu_H P(r) - H(x) Q(r). \]

Theorem 1 For \( P, Q \in \mathcal{P}_\mathcal{R}, H \in \mathcal{P}_\mathcal{R}, r \in \mathcal{Z}, \) and \( x \in \mathcal{R}, \) we have
\[ -B_r(x; Q, P, H) \leq P \vee H(x) - Q \vee H(x) \leq B_r(x; P, Q, H). \]

Proof. Application of Lemma 2 gives
\[ P \vee H(x) - Q \vee H(x) \leq P \vee H(x) + \mu_H P(r) - Q \vee H(x) - H(x) Q(r) = B_r(x; P, Q, H), \]
which proves the second inequality in (4). The first inequality follows by symmetry.

This completes the proof of Theorem 1. Q.E.D.

We shall look at some special cases of Theorem 1:

1. As
\[ B_r(x; P, P_r, H) = \mu_H P(r) \quad B_r(x; P_r, P, H) = -H(x) P(r), \]
we see that Lemma 2 (and thus also Lemma 1) is a special case of Theorem 1.
2. From (3) we obtain
\[ B_1(x; P, Q, H) = -(p(0) - q(0)) \bar{H}(x) + \mu_H \bar{P}(1) - \bar{H}(x) \bar{Q}(1) = \left( \mu_H - \bar{H}(x) \right) \bar{P}(1) + \bar{H}(x) \left( \mu_P - \mu_Q \right). \]

For \( H \in \mathcal{P}_z \), this case is discussed in Dhaene & Sundt (1996).

3. From (2) we obtain
\[ B_0(x; P, Q, H) = \mu_H \mu_P - \bar{H}(x) \mu_Q. \]  
(5)

4. If \( P(r) = Q(r) = 1 \), then \( P_r = P \) and \( Q_r = Q \), and from (2) we obtain
\[ B_r(x; P, Q, H) = -B_r(x; Q, P, H) = \bar{P}(x) + \bar{H}(x), \]
that is, in this case Theorem 1 becomes trivial.

5. From (2) we obtain
\[ B_r(x; P, P, H) = \left( \mu_H - \bar{H}(x) \right) \bar{P}(r), \]  
(6)

that is, unfortunately the bounds in Theorem 1 do not in general become equal to zero when comparing two identical compound distributions.

3.3. Let \( D_1(x; P, Q, H) \) denote the difference between the upper and lower bound in Theorem 1, that is,
\[ D_1(x; P, Q, H) = B_r(x; P, Q, H) - B_{r-1}(x; Q, P, H). \]  
(7)

Then
\[ D_r(x; P, Q, H) = \left( \mu_H - \bar{H}(x) \right) \left( \bar{P}(r) + \bar{Q}(r) \right). \]  
(8)

We see that \( D_r(x; P, Q, H) \) decreases to zero when \( r \) increases to infinity, that is, we can make the difference between the upper and lower bound in Theorem 1 as small as desired by making \( r \) sufficiently large.

We see that \( D_r(x; P, Q, H) \) increases from zero to \( \mu_H \left( \bar{P}(r) + \bar{Q}(r) \right) \) when \( x \) increases from zero to infinity. Thus our bounds are most accurate for low values of \( x \). Furthermore, if for some \( \varepsilon > 0 \) we choose \( r \) such that
\[ \bar{P}(r) + \bar{Q}(r) < \frac{\varepsilon}{\mu_H}, \]
then \( D_r(x; P, Q, H) < \varepsilon \) for all \( x \in \mathcal{R}_+ \).

3.4. Let
\[ b_r(x; P, Q, H) = B_r(x; P, Q, H) - B_{r+1}(x; P, Q, H). \]
From (3) and trivial calculus we obtain

\[ b_r(x; P, Q, H) = (P(r) - Q(r))(\overline{H^{(r+1)*}}(x) - \overline{H^{*}}(x)) + \mu_H(1 - P(r)) - \overline{H}(x)(1 - Q(r)). \]  

(9)

By rewriting (9) as

\[ b_r(x; P, Q, H) = (1 - P(r))\left[ \mu_h + \overline{H^{*}}(x) - \overline{H^{(r+1)*}}(x) \right] + 
\[ (1 - Q(r))\left[ \overline{H^{(r+1)*}}(x) - \overline{H^{*}}(x) - \overline{H}(x) \right] \]

and application of Lemma 1, we see that \( b_r(x; P, Q, H) \geq 0 \). Thus \( B_r(x; P, Q, H) \) is non-increasing in \( r \). This implies that in (4), the upper bound is non-increasing and the lower bound is non-decreasing in \( r \), and as \( D_r(x; P, Q, H) \) goes to zero when \( r \) goes to infinity, both bounds converge to \( P \vee \overline{H}(x) - Q \vee \overline{H}(x) \).

Formula (9) can be applied for recursive evaluation of \( B_r(x; P, Q, H) \). Furthermore, when we have found \( B_r(x; P, Q, H) \), we easily obtain \( B_r(x; Q, P, H) \) from (7) and (8).

3.5. The main purpose of the present subsection is to deduce an improvement of the bounds in Theorem 1. For doing that, we shall need the following lemma.

**Lemma 3** For \( P, Q \in \mathcal{P}_{\mathbb{Z}^+} \), \( H \in \mathcal{P}_{\mathbb{R}^+} \), \( r \in \mathbb{Z}^+ \), and \( x \in \mathbb{R}^+ \), we have

\[ b_r(x; P, Q, H) \geq \left( \mu_h - \overline{H}(x) \right)(1 - \max(P(r), Q(r))). \]  

(10)

**Proof.** We apply Lemma 1 in (9). If \( P(r) \geq Q(r) \), then

\[ b_r(x; P, Q, H) \geq (P(r) - Q(r))\overline{H}(x) + \mu_h(1 - P(r)) - \overline{H}(x)(1 - Q(r)), \]

that is,

\[ b_r(x; P, Q, H) \geq \left( \mu_h - \overline{H}(x) \right)(1 - P(r)). \]  

(11)

Analogously, if \( P(r) < Q(r) \), then

\[ b_r(x; P, Q, H) \geq (P(r) - Q(r))\mu_H + \mu_H(1 - P(r)) - \overline{H}(x)(1 - Q(r)), \]

that is,

\[ b_r(x; P, Q, H) \geq \left( \mu_h - \overline{H}(x) \right)(1 - Q(r)). \]  

(12)

From (11) and (12) we obtain (10).  

Q.E.D.
Theorem 2 For $P, Q \in \mathcal{P}_{\mathbb{Z}^+}, H \in \mathcal{P}_{\mathbb{R}_+}, r \in \mathbb{Z}_+, \text{ and } x \in \mathcal{R}_+, \text{ we have}$

$$B_r(x; Q, P, H) + \left(\mu_H - \overline{H}(x)\right) \sum_{k=r}^{\infty} \left(1 - \max\{P(k), Q(k)\}\right) \leq$$

$$\overline{P \lor H(x)} - \overline{Q \lor H(x)} \leq$$

$$B_r(x; P, Q, H) - \left(\mu_H - \overline{H}(x)\right) \sum_{k=r}^{\infty} \left(1 - \max\{P(k), Q(k)\}\right). \tag{13}$$

Proof. For $s \in \mathbb{Z}_+$ such that $s \geq r$, we obtain by applying successively Theorem 1 and Lemma 3

$$\overline{P \lor H(x)} - \overline{Q \lor H(x)} \leq B_s(x; P, Q, H) =$$

$$B_r(x; P, Q, H) - \sum_{k=r}^{s-1} b_k(x; P, Q, H) \leq$$

$$B_r(x; P, Q, H) - \left(\mu_H - \overline{H}(x)\right) \sum_{k=r}^{s-1} \left(1 - \max\{P(k), Q(k)\}\right). \tag{14}$$

By letting $s$ go to infinity, we obtain the second inequality in (13); the first inequality follows by symmetry.

This completes the proof of Theorem 2. Q.E.D.

We see that in (13), like in (4), the lower bound is non-decreasing in $r$ and the upper bound is non-increasing in $r$.

The infinite summation in (13) may seem complicated. However, as all the terms are non-negative, we obtain weaker bounds by including only a finite number of terms. Furthermore, if $P$ or $Q$ has a finite support, then only a finite number of terms are non-zero.

In the following corollary we consider another case where the summation obtains a particularly simple form.

Corollary 1 Let $P, Q \in \mathcal{P}_{\mathbb{Z}^+}, H \in \mathcal{P}_{\mathbb{R}_+}, \text{ and } x \in \mathcal{R}_+$. If there exists a non-negative integer $s$ (possibly equal to infinity) such that

$$Q(y) \leq P(y), \quad (y = 0, 1, \ldots, s-1)$$

$$Q(y) \geq P(y), \quad (y = s, s+1, \ldots) \tag{14}$$

then

$$- B_r(x; Q, P, H) + \left(\mu_H - \overline{H}(x)\right)\left(\overline{P}(r) - \overline{P}(s) + \overline{Q}(s)\right) \leq$$

$$\overline{P \lor H(x)} - \overline{Q \lor H(x)} \leq$$

$$B_r(x; P, Q, H) - \left(\mu_H - \overline{H}(x)\right)\left(\overline{P}(r) - \overline{P}(s) + \overline{Q}(s)\right) \tag{15}$$

(r = 0, 1, \ldots, s-1)
DIFFERENCE STOP-LOSS ORDER AND TWO COMPOUND DISTRIBUTIONS

\[ -B_r(x; Q, P, H) + \left( \mu_H - \overline{H}(x) \right) \overline{Q}(r) \leq \overline{P \vee H}(x) - \overline{Q \vee H}(x) \leq \\
Br(x; P, Q, H) - \left( \mu_H - \overline{H}(x) \right) \overline{Q}(r). \quad (r = s, s + 1, \ldots) \tag{16} \]

Proof. For \( r = 0, 1, \ldots, s-1 \), we have

\[
\sum_{k=r}^{\infty} \left( 1 - \max(P(k), Q(k)) \right) = \sum_{k=r}^{s-1} (1 - P(k)) + \sum_{k=s}^{\infty} (1 - Q(k)) = \\
\overline{P}(r) - \overline{P}(s) + \overline{Q}(s),
\]

and insertion in (13) gives (15). The inequalities (16) are proved analogously.

This completes the proof of Corollary 1. Q.E.D.

If we in addition to (14) assume that \( \mu_Q \leq \mu_P \), then we have the stop-loss orderings \( Q \leq P \) and \( Q \vee H \leq P \vee H \); for proofs cf. e.g. Goovaerts, Kaas, van Heerwaarden, & Bauwelincx (1990).

At the end of subsection 3.2. we pointed out that unfortunately the bounds in Theorem 1 do not become equal to zero when \( Q = P \). From (6) and (15) we see that the improved bounds of Theorem 2 do not have this deficiency.

ACKNOWLEDGEMENT

The authors are grateful to an anonymous referee for useful suggestions.

REFERENCES


BJÖRN SUNDT
Department of Mathematics
University of Bergen
Allégaten 55
N-5007 Bergen
Norway

JAN DHAENE
Departement Toegepaste Economische Wetenschappen
Katholieke Universiteit Leuven
Huis Eghen Heerd
Minderbroederstraat 5
B-3000 Leuven
Belgium
This paper shows how a multivariate Bayes estimator can be adjusted to satisfy a set of linear constraints. In the direct approach, the constraint is enforced by a restriction on the class of admissible estimators. In an alternative approach, the constraint is merely encouraged by a mixed risk function which penalises misbalance between the estimator and the constraint. The adjustment to the optimal unconstrained estimator is shown to depend on the risk function and the linear constraints only, not on the probability model underlying the Bayes estimator. Two practical examples are given, one of which involves reconciliation of independently assessed share values with current market values.

**Keywords**

Bayes Estimation, Linear Bayes Estimation, Credibility Theory, Share Valuation

**I. Introduction**

Actuaries often need to reconcile the estimates they have arrived at, with the data used to calculate the estimates. In the estimation of pure premiums, for instance, the actuary would always check that the total premiums calculated are sufficient to cover the total cost of claims.

The concept of balanced linear estimators was introduced by Neuhaus (1995). In that paper, a linear estimator is called balanced if it satisfies certain linear constraints involving the original data; furthermore, the optimal linear estimator is called the credibility estimator, and the optimal balanced linear estimator is called the balanced credibility estimator.

This paper generalises the balancing concept in two directions. The first generalisation, presented in Section 2, involves balancing arbitrary estimators, i.e. estimators which are not necessarily linear in the data. By this approach one arrives at an optimal balanced estimator. As a by-product one obtains a much shorter derivation of the balanced credibility estimator than in Neuhaus (1995). A simple example of the calculations needed is given in Section 3.

Section 4 provides a more practical example. Given independent valuations of the different shares in a market, the actuary could wish to reconcile these values to the market value of the portfolio currently held, as well as the overall value of the portfolio.
market. We apply the balancing formula to calculate the optimal adjustment of individually assessed share values.

In Section 5 we compare the balanced credibility estimator with the homogeneous unbiased credibility estimator in the Bühlmann-Straub model. The latter estimator has long been known to be balanced, see e.g. Gisler (1987).

The second generalisation, briefly presented in Section 6, involves the use of a mixed risk function, where balancing is not enforced but misbalance is penalised. The mixed risk function is similar to the optimisation criterion used in Whittaker-Henderson graduation (see e.g. Taylor, 1992). A mixed risk function has also been used by Sundt (1992) as a way of smoothing a sequence of credibility estimators.

In both generalisations, the necessary adjustment to the unconstrained Bayes estimator turns out to be independent of the probability model underlying the Bayes estimator. This allows the actuary to balance the estimates after the unconstrained Bayes estimator has been calculated, and without reference to the model used.

Finally, a few words on terminology: Actuaries’ frequent need to balance a set of estimators against a set of data, is the prime motivation for studying linearly constrained estimators. Since in the general context it is easy, however, to construct linear constraints that do not comply with any sensible notion of balancing, we will simply talk of constrained estimators in the balance of this paper.

2. OPTIMAL ESTIMATION UNDER BINDING LINEAR CONSTRAINTS

We assume the existence of a latent random vector (the estimand),

$$b^{\text{est}} := (b_1, \ldots, b_p)'$$

(2.1)

as well as the existence of an observed random vector (the statistic)

$$X^{\text{est}} := (X_1, \ldots, X_n)'$$

(2.2)

Assume that $b$ and $X$ are defined over the same probability space and square integrable, and assume also that the joint distribution of $(b, X)$ is known.

An estimator $\hat{b}$ is any measurable function

$$\hat{b} : \mathbb{R}^n \rightarrow \mathbb{R}^p : x \rightarrow \hat{b}(x),$$

(2.3)

such that $\hat{b}(X)$ is square integrable. The criterion (risk function) we use to measure the performance of a given estimator is generalised mean squared error,

$$r(\hat{b}) := E[(\hat{b}(X) - b)' W (\hat{b}(X) - b)],$$

(2.4)

with $W^{p \times p}$ an some fixed, positive definite risk weighting matrix.

It is well known that the optimal estimator in the absence of any constraints, which we denote by $\tilde{b}$, is the conditional mean of $b$ given $X$:

$$\tilde{b}(X) = E[b \mid X].$$

(2.5)
The risk of that estimator is
\[ r(\hat{b}) = \text{tr}[W \cdot \text{E} \text{Cov}[b | X]]. \] (2.6)

The optimality of \( \hat{b} \) and its risk (2.6) are a direct consequence of the decomposition
\[
\begin{align*}
    r(\hat{b}) &= \text{E}[S(\hat{b} - b) ' W(\hat{b} - b) | X] \\
    &= \text{tr}[W \cdot \text{E} \text{Cov}(b | X)] + \text{E}[(\hat{b} - b) ' W(\hat{b} - b)] \\
    &= r(b) + \text{E}[S(\hat{b} - b)' W(\hat{b} - b) | X].
\end{align*}
\] (2.7)

We refer to the last term in the above expression as the excess risk of the estimator \( \hat{b} \).

Assume now that a constraint of the following general form has been imposed on the class of admissible estimators:
\[ f(\hat{b}(X)) = g(X) \quad \text{a.s.}, \] (2.8)
where \( f : \mathbb{R}^p \to \mathbb{R}^q \) and \( g : \mathbb{R}^n \to \mathbb{R}^q \) are known, fixed functions. The constraint may equivalently be stated as
\[ \hat{b}(X) \in f^{-1}\{g(X)\}, \] (2.9)
provided that \( f^{-1}\{g(X)\} \neq \emptyset \) a.s. An estimator will be called constrained if it satisfies (2.8).

From the decomposition (2.7) and the observation that both \( \hat{b} \) and \( \hat{b} \) are functions of the statistic \( X \), it is evident that the optimal constrained estimator, which we denote by \( \tilde{b}(X) \), can be found by pointwise minimisation for each possible realisation \( X = x \):
\[ \tilde{b}(x) = \text{proj}_W(\hat{b}(x) | f^{-1}\{g(x)\}), \] (2.10)
where \( \text{proj}_W(a | B) \) denotes a projection of a vector \( a \) into a set \( B \), with respect to the metric derived from the inner product \( \langle a, b \rangle = a'Wb \). For general functions \( f, g \), this projection need not be unique and may not even exist; however, the projection does exist if \( f \) is a continuous function. In particular, if \( f \) and \( g \) are linear functions, the projection has the explicit formula given in the next theorem.

**Theorem**
Assume that the two constraining functions are linear,
\[ f(\hat{b}) = L\hat{b}, \quad g(x) = P x, \] (2.11)
where \( L^{q \times p} \) and \( P^{q \times n} \) are known, fixed matrices, and \( L \) is of full row rank \( q \leq p \); then the optimal constrained estimator is
\[ \hat{b}(X) = \hat{b}(X) + W^{-1}L'Q^{-1}(PX - L\hat{b}(X)), \quad (2.12) \]

with \( Q = LW^{-1}L' \).

**Proof**
Consider a fixed value of the statistic \((X = x)\) and note that both \( \hat{b} \) and \( \hat{b} \) are functions of \( x \). Define the Lagrange functional
\[ F = \frac{1}{2} (\hat{b} - \hat{b}) 'W(\hat{b} - \hat{b}) - \lambda' (L\hat{b} - Px), \quad (2.13) \]
with \( \lambda \in \mathbb{R}^n \) a vector of Lagrange multipliers. Solving the equation
\[ \frac{\partial F}{\partial \hat{b}} = (\hat{b} - \hat{b}) 'W - \lambda' L = 0 \quad (2.14) \]
yields
\[ \hat{b}(\lambda) = \hat{b} + W^{-1}L'\lambda. \quad (2.15) \]
Now use the constraint \((2.11)\) to determine \( \lambda = Q^{-1}(Px - L\hat{b}) \); substitute this expression in \((2.15)\) to find \((2.12)\).

**Remarks**
Two remarks on the hypotheses of the theorem are in order. Firstly, the full rank assumption on \( L \) is needed in order to ensure that the equation \((2.8)\) is consistent. Secondly, the assumption that \( g \) is linear has not been used at all; thus the estimator \((2.12)\) may easily be extended to more general functions \( g \); however, the assumption that \( g \) is linear will allow us to derive transparent formulas for the excess risk, which is our next stopping point.
Using \((2.12)\), the excess risk of the optimal constrained estimator over the unconstrained Bayes estimator is easily seen to be
\[ \text{E}[ (\hat{b} - \bar{b}) 'W(\hat{b} - \bar{b}) ] = \text{tr}[Q^{-1}E(PX - L\bar{b})(PX - L\bar{b})']. \quad (2.16) \]
One can write
\[ E(PX - L\bar{b})(PX - L\bar{b})' = E[E(PX - Lb | X) \cdot E' (PX - Lb | X)] \]
\[ = \text{CovE}[PX - Lb | X] + E(PX - Lb) \cdot E' (PX - Lb) \]
\[ = \text{Cov}(PX - Lb) - E\text{Cov}[PX - Lb | X] \]
\[ + E(PX - Lb) \cdot E' (PX - Lb) \]
\[ = E(PX - Lb)(PX - Lb)' - L \cdot E\text{Cov}[b | X] \cdot L' \]
\[ = P \cdot E\text{Cov}[X | b] \cdot P' - L \cdot E\text{Cov}[b | X] \cdot L' \]
\[ + E[(P \cdot E(X | b) - Lb)(P \cdot E(X | b) - Lb)']. \quad (2.17) \]
In the important special case where \( n = p \) and \( \mathbb{E}(X \mid b) = b \) and \( L = P \), the last term drops out and the expression is reduced to \( L(\text{E Cov}[X \mid b] - \text{E Cov}[b \mid X]) \) \( L' \).

Note that the adjustment matrix in (2.12), namely
\[
J = W^{-1}LQ^{-1},
\]  
(2.18)
does not depend on the probability distribution of \((b, X)\). As a consequence, one can calculate the constraining adjustment after the unconstrained Bayes estimator has been calculated, and without reference to the model used to derive that estimator. Unlike the unconstrained Bayes estimator, however, the optimal constrained estimator \( \mathbf{b} \) depends on the risk weighting matrix \( W \) and, of course, the constraints.

An important special case is where \( W = \text{diag}(w_1, \ldots, w_p) \) is a diagonal matrix and there is only one constraint \( (q = 1) \). In that case one can write \( L = (l_1, \ldots, l_p) \) and \( P = (P_1, \ldots, P_n) \). Inserted into (2.12), this gives the following formula for the \( i \)th component of the optimal constrained estimator:
\[
\tilde{b}_i = \bar{b}_i + \frac{l_i}{w_i} \left( \sum_{j=1}^{p} \frac{l_j^2}{w_j} \right)^{-1} \left( \sum_{j=1}^{p} p_jX_j - \sum_{j=1}^{p} l_j\bar{b}_j \right).
\]  
(2.19)
where \( \Delta \) is the 'amount of misbalance' exhibited by the unconstrained Bayes estimator. The excess risk in this case becomes
\[
r(\mathbf{b}) - r(\tilde{\mathbf{b}}) = \left( \sum_{j=1}^{p} \frac{l_j^2}{w_j} \right)^{-1} \cdot \mathbb{E}(\Delta^2). \]  
(2.20)
Neuhaus (1995) treated the case where, on top of linear constraints of the form (2.11), one imposes the additional constraint that the estimator \( \tilde{\mathbf{b}} \) be a linear function of the statistic \( X \) (i.e. a 'credibility' estimator). Using Lagrange minimisation, the resulting 'balanced credibility estimator' was shown to be of the same form as (2.12), with \( \tilde{\mathbf{b}} \) the credibility estimator rather than the Bayes estimator. Given the present result, a simple reasoning leading to that result goes as follows: since the risk \( r(\mathbf{b}) \) of a linear estimator \( \tilde{\mathbf{b}} \) depends on the distribution of \((b, X)\) only through its moments of first and second order, the balanced credibility estimator can only depend on those moments, too. But those moments could have been generated by a Gaussian distribution, in which case even the optimal constrained estimator is a linear function of \( X \). Being selected from a wider class of admissible estimators, the optimal constrained estimator in the Gaussian case must coincide with the balanced credibility estimator; which in turn coincides with the balanced credibility estimator in any other model that generates the same first and second order moments.

The excess risk (2.16) measures the average cost of constraining the estimator in the long run by repeated independent estimation situations. One could argue that the expectation should be dropped and that the pointwise increase of loss is what matters. If one takes this view, the pointwise increase at \( X = x \) can be easily
calculated by the formula
\[(b(x) - \tilde{b}(x))'W(b(x) - \tilde{b}(x)) = (Px - L\tilde{b}(x))'Q^{-1}(Px - L\tilde{b}(x)), \quad (2.21)\]
being simply the distance between \(b\) and its projection onto \(f^{-1}\{g(x)\}\).

3. Example: Optimal Constrained Estimation of Lognormal Means

As we have seen, the optimal constrained estimator (2.12) always has a component that is linear in the statistic \(X\). Let us now consider a model in which the optimal unconstrained estimator is non-linear in \(X\), and quantify the excess risk generated by the constraint.

Assume that the portfolio under consideration consists of stochastically independent policies labelled by \(i = 1, \ldots, p\). For policy no. \(i\), assume we have observed claim amounts \(X_{ij} : j = 1, \ldots, n_i\), with \(n_i\) fixed. Now assume that the \(X_{ij}\) are conditionally independent, given the value \(\theta_i\) of a hidden random parameter \(\Theta_i\), and that under the same conditional distribution,
\[Y_{ij} := \log(X_{ij}) \sim \text{Normal}(\theta_i, \phi), \quad (3.1)\]
with a fixed value of \(\phi\). Assume that the hidden risk parameters \(\Theta_i\) are independent with
\[\Theta_i \sim \text{Normal}(\mu, \lambda). \quad (3.2)\]
The properties of the lognormal distribution are summarised in e.g. Hogg & Klugman (1984). In particular, the conditional mean of \(X_{ij}\), given \(\Theta_i\), is
\[b_i = \mathbb{E}[X_{ij} | \Theta_i] = e^{\theta_i} + \frac{\phi}{\lambda}. \quad (3.3)\]
Now assume that it is our intention to estimate the vector of lognormal means,
\[b := (b_1, \ldots, b_p)' \quad (3.4)\]
under the constraint that the weighted sum of the estimates must equal the sum of claims:
\[\sum_{i=1}^{p} n_i b_i = \sum_{i=1}^{p} \sum_{j=1}^{n_i} X_{ij} = \sum_{i=1}^{p} n_i X_i, \quad (3.5)\]
where \(X_i := n_i^{-1} \sum_j X_{ij}\) is the average of claims against policy no \(i\).

It is well known that the conditional distribution of \(\Theta_i\), given \(X_{i1}, \ldots, X_{in_i}\), is again a normal with conditional mean
\[\mu_i = \frac{n_i \lambda}{n_i \lambda + \phi} \left\{ 1 + \frac{\phi}{n_i \lambda + \phi} \right\} \mathbb{E}[\log(X_{ij})] + \frac{\phi}{n_i \lambda + \phi} \mu =: z_i \bar{Y}_i + (1 - z_i) \mu \quad (3.6)\]
and conditional variance
\[ \lambda_i = \frac{\lambda \phi}{n_i \lambda + \phi} = (1 - z_i) \lambda. \]  
(3.7)

The Bayes estimator of \( b_i \) is its conditional mean
\[ \bar{b}_i = \mathbb{E}[b_i | X_{i1}, \ldots, X_{in_i}] = e^{\mu + \frac{\lambda \phi}{n_i \lambda + \phi}}, \]  
(3.8)

and its conditional variance is
\[ \text{Var}[b_i | X_{i1}, \ldots, X_{in_i}] = e^{2\mu + \phi + \lambda_i (e^{\lambda_i} - 1)}. \]  
(3.9)

Using (3.9) and the marginal distribution of \( \bar{Y}_i \), which is Normal(\( \mu, \lambda + \phi/n_i \)), we find after some tedious manipulation the Bayes risk for estimating \( b_i \):
\[ \mathbb{E}\text{Var}[b_i | X_{i1}, \ldots, X_{in_i}] = e^{2\mu + 2\lambda + \phi (1 - e^{-\lambda_i})}. \]  
(3.10)

Let us assume that the matrix \( W \) is diagonal. Inserting \( l_i = p_i = n_i (i = 1, \ldots, p) \) and \( \bar{b}_i \) given by (3.8) into (2.19), the optimal constrained estimator can be read off directly.

In order to calculate the excess risk using (2.16) and (2.17), we must also find
\[ \mathbb{E}\text{Var}[X_i | b_i] = \frac{1}{n_i} e^{2\mu + 2\lambda + \phi (e^\phi - 1)}. \]  
(3.11)

Now inserting the expressions (3.11) and (3.10) into (2.17), we find the excess risk generated by the constraint:
\[ r(\bar{b}) - r(b) = e^{2\mu + 2\lambda + \phi \sum_{i=1}^{p} n_i (e^\phi - 1) - n_i (1 - e^{-\lambda_i})]. \]  
(3.12)

Using Jensen's inequality one can check that each of the summands on the right hand side is non-negative.

The weighted Bayes risk is
\[ r(\bar{b}) = e^{2\mu + 2\lambda + \phi \sum_{i=1}^{p} w_i (1 - e^{-\lambda_i})}. \]  
(3.13)

In the case where \( n_1 = \ldots = n_p = n \) and \( W = I \), it is easy to calculate the relative excess risk,
\[ \frac{r(\bar{b}) - r(b)}{r(b)} = \frac{1}{p} \left[ \frac{1}{n} \cdot \frac{e^\phi - 1}{1 - e^{-\lambda_i} - 1} \right]. \]  
(3.14)
with \( \lambda_1 = \lambda \phi/(n \lambda + \phi) = (1 - z_1) \lambda \). The relative excess risk can become arbitrarily large. One can also show that
\[
\lim_{n \to \infty} \frac{r(\mathbf{b}) - r(\mathbf{b}^*)}{r(\mathbf{b})} = \frac{1}{p} \left[ \frac{e^\phi - 1}{\phi} - 1 \right] > 0,
\]
(3.15)
independent of \( \lambda \). Thus constraints should never be applied uncritically, and special care must be taken when the original observations have a heavy-tailed distribution.

4. Example: Combining Share Valuations with Market Values

Consider an actuary who has been charged with an analysis of a portfolio of \( p \) shares. Assume that in addition to the market values at any time \( t \), which we denote by
\[
\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))',
\]
(4.1)
the actuary has access to an individual valuation of the shares, based on an analysis of the economic fundamentals. Let us assume that the individually assessed share values reflect the analyst’s conditional expectation, given all the information available to him, and denote the analyst’s best bet by
\[
\tilde{\mathbf{b}}(t) = (\tilde{b}_1(t), \ldots, \tilde{b}_p(t))'.
\]
(4.2)

Of course it is possible that \( \tilde{b}_i(t) = X_i(t) \) for the major stocks and those stocks that have not had the attention of the analyst.

In order not to stray too far from the market value of the shares, the actuary now wishes to ensure that at least the overall value of the shares coincides with their overall market value. Thus assume that the number of shares listed is
\[
\mathbf{n}(t) = (n_1(t), \ldots, n_p(t))',
\]
(4.3)
while the number of shares held by the company (or pension fund) is
\[
\mathbf{m}(t) = (m_1(t), \ldots, m_p(t))',
\]
(4.4)
Assume that \( \mathbf{n}(t) \) and \( \mathbf{m}(t) \) are linearly independent. We must also assume that shares have a common nominal value.

There could conceivably be two different constraints the actuary wishes to obey:
\[
\begin{align*}
(1) & : \sum_{i=1}^p n_i(t) \tilde{b}_i(t) = \sum_{i=1}^p n_i(t) X_i(t), \\
(2) & : \sum_{i=1}^p m_i(t) \tilde{b}_i(t) = \sum_{i=1}^p m_i(t) X_i(t).
\end{align*}
\]
(4.5)
The first constraint prevents any deviation from a market weighted index, while the second constraint ensures that the values used in the actuary’s analysis add up to the
total value the company must show in its books. There could be additional constraints, for example a constraint to prevent deviation from a major sub-index like the All Industrials.

The two constraints in (4.5) are formalised by the matrices

$$L^{2x2}(t) = P^{2x2}(t) = \begin{pmatrix} n'(t) \\ m'(t) \end{pmatrix}.$$ (4.6)

We suppress the argument $(t)$ in the rest of this section.

The optimal adjustment to make to the analyst's set of estimates, follows directly from (2.12):

$$\delta - \bar{b} = W^{-1}L'Q^{-1}\Delta,$$ (4.7)

where the vector

$$\Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} \sum_i n_i(X_i - \bar{b}_i) \\ \sum_i m_i(X_i - \bar{b}_i) \end{pmatrix}$$ (4.8)

contains the misbalance of the analyst’s estimates against each of the two constraints.

If $W$ is diagonal, one can derive the following expression for the adjustment to vector of share values:

$$\delta - \bar{b} = \begin{pmatrix} \sum_i \frac{m_i}{w_i} \left[ \sum_j \frac{n_j^2}{w_{ij}} \Delta_1 + \sum_j \frac{n_j}{w_{ij}} \Delta_2 - \sum_j \frac{n_j m_j}{w_{ij}} \left[ \frac{m_i}{w_i} \Delta_1 + \frac{n_i}{w_i} \Delta_2 \right] \right] \\ \sum_j \frac{n_j^2}{w_{ij}} \cdot \sum_j \frac{m_j^2}{w_{ij}} - \left[ \sum_j \frac{m_j n_j}{w_{ij}} \right]^2 \end{pmatrix}^i = 1,...,p$$ (4.9)

In particular if mean squared error is weighted by the number of shares listed ($w_i = n_i$), we obtain the simple adjustment

$$\delta - \bar{b} = \left( \frac{[\bar{s} - s_i] \delta_1 + [s_i - \bar{S}] \delta_2}{\bar{s} - \bar{S}} \right)^i = 1,...,p,$$ (4.10)

with $N := \sum j n_j$ the total number of shares listed, $M := \sum j m_j$ the total number of shares held, $s_i := m_i/n_i$ the stake held in stock $i$, $\bar{s} := M^{-1} \sum j m_j s_j$ the average stake held and $S := M/N$ the overall stake in the stock market; $\delta_1 := \Delta_1/N$ and $\delta_2 := \Delta_2/M$ denote the relative (per share) deviations between the market values and the analyst’s values.

From (4.10) one sees how the optimal adjustment depends on the relative misbalances and the company’s relative exposure to the different stocks. In particular, if $\delta_2 > \delta_1$, then the adjustment to be made to share price no. $i$, is an increasing function of $s_i$, the stake held in stock no. $i$. Roughly speaking, $\delta_2 > \delta_1$ means that the market likes the portfolio held by the company better than the analyst; in that case it seems reasonable that the value of stocks of which the company has a major
holding, should be assessed more highly than the value of stocks of which the company only has a small holding. The opposite holds if $\delta_2 < \delta_1$; if $\delta_2 = \delta_1 = \delta$, then each share price is adjusted with the same amount $\delta$.

5. Balanced vs. Homogeneous Credibility Estimation in the Bühlmann-Straub Model

For the purpose of this section only, we revert to using the term ‘balanced estimators’, since the constraint applied here neatly fits in with the intuitive notion of balancing. Moreover, the estimators analysed in this section satisfy several different constraints, so that the simple term ‘constrained estimator’ without a number of qualifiers would be highly ambiguous.

Assume that the actuary is charged with estimating the pure premiums of $n$ independent insurance policies. For policy no. $i$, what has been observed is a measure of exposure, denoted by $p_i$, and the total claims cost, denoted by $S_i$. The empirical pure premium per unit of exposure of policy no. $i$ is then $X_i = S_i/p_i$.

Assume now that the probability distribution of $X_i$ is governed by an unobserved random parameter $\theta_i$ coming from a distribution $U$, in such a way that

$$E_{\theta_i}(X_i) = b(\theta_i),$$
$$\text{Var}_{\theta_i}(X_i) = \nu(\theta_i)/p_i. \tag{5.1}$$

Define the following structural parameters:

$$\beta = E(b(\Theta)), \quad \phi = E(\nu(\Theta)), \quad \lambda = \text{Var}(b(\Theta)). \tag{5.2}$$

If a diagonal risk weighting matrix $W = \text{diag}(w_1, \ldots, w_n)$ is used, the balanced credibility estimator under the constraint

$$\sum_{i=1}^{n} p_i \bar{b}_i = \sum_{i=1}^{n} p_i X_i \tag{5.3}$$

is given by

$$\bar{b}_i = z_i X_i + (1 - z_i)\beta + \frac{p_i}{w_i} \left( \sum_{j=1}^{n} \frac{p_j^2}{w_j} \right)^{-1} \sum_{j=1}^{n} p_j (1 - z_j) (X_j - \beta), \tag{5.4}$$

($i = 1, \ldots, n$), and its risk is

$$r(\bar{b}) = \lambda \sum_{j=1}^{n} w_j (1 - z_j) + \phi \left( \sum_{j=1}^{n} \frac{p_j^2}{w_j} \right)^{-1} \sum_{j=1}^{n} p_j (1 - z_j), \tag{5.5}$$

where $z_j = \lambda p_j / (\lambda p_j + \phi)$ are the credibility factors. The results (5.4) and (5.5) are proved in Neuhaus (1995).

Another linear estimator that is known to balance (i.e. satisfy the constraint (5.3)), is the optimal homogeneous and unbiased credibility estimator,
This fact has been noted by e.g. Gisler (1987).

Since the optimal homogeneous and unbiased estimator has two constraints to satisfy and just happens to be balanced as well, there is no prize for guessing that its risk will exceed that of the optimal balanced estimator. The question is just, by how much the risks differ.

Specialising equation (5.27) of Neuhaus (1995) or using the representation (5.6) directly, one easily shows that the risk of $\hat{b}$ is

$$r(b) = \lambda \sum_{j=1}^{n} w_j(1 - z_j) + \lambda \left( \sum_{j=1}^{n} z_j \right)^{-1} \sum_{j=1}^{n} w_j(1 - z_j)^2.$$  \hspace{1cm} (5.7)

After some manipulations one can write the difference in risk as

$$r(\hat{b}) - r(b) = \frac{\sigma^2}{\lambda} \left[ \left( \sum_{j=1}^{n} z_j \right)^{-1} \sum_{j=1}^{n} z_j \frac{z_j w_j}{p_j^2} - \left( \sum_{j=1}^{n} \frac{z_j^2}{z_j w_j} \right)^{-1} \sum_{j=1}^{n} z_j \right] \geq 0.$$  \hspace{1cm} (5.8)

The inequality is a consequence of Jensen’s inequality applied to the convex function $x \rightarrow x^{-1}$ ($x > 0$), and will be strict unless all $z_j w_j/p_j^2$ are identical.

6. **Optimal estimation under a penalty for misbalance**

In the example of Section 3 we noted that the relative excess risk introduced by the constraint, can become arbitrarily large. Thus one must consider whether the benefit of constraining the estimator is worth the added risk.

A compromise approach would be not to enforce the constraint, but merely to ‘encourage’ it by a suitable modification of the risk function, so that the modified risk function reflects our preference of constrained estimators. This approach is similar to that taken in Whittaker-Henderson graduation, see e.g. Taylor (1992). It is also very similar to the smoothing approach proposed by Sundt (1992).

Let us therefore introduce the following, mixed risk function:

$$r_\alpha(\hat{b}) = (1 - \alpha) \cdot E(\hat{b} - b)'W(\hat{b} - b) + \alpha \cdot E(L\hat{b} - PX)'V(L\hat{b} - PX),$$  \hspace{1cm} (6.1)

with $W$, $V$ fixed, positive definite matrices, and $\alpha \in [0, 1]$ a parameter which quantifies the trade-off between estimator precision and estimator constraints.

One can write

$$r_\alpha(\hat{b}) = EE[(1 - \alpha) \cdot (\hat{b} - b)'W(\hat{b} - b) + \alpha \cdot (L\hat{b} - PX)'V(L\hat{b} - PX) | X].$$  \hspace{1cm} (6.2)

Thus the optimal estimator may be determined pointwise for each possible realisation of $X = x$, and after having decomposed the first term in (6.2) as in (2.7), it is
easy to verify that the optimal estimator is

$$\tilde{b}_\alpha(x) = ( (1 - \alpha)W + \alpha L'VL )^{-1} [ (1 - \alpha)W\tilde{b}(x) + \alpha L'VPx ] .$$  \hspace{1cm} (6.3)

Using tedious but straightforward matrix transformations, we can derive an equivalent expression,

$$\tilde{b}_\alpha(x) = b(x) + J_\alpha (Px - L\tilde{b}(x)) ,$$  \hspace{1cm} (6.4)

with

$$J_\alpha = \alpha W^{-1}L'((1 - \alpha)V^{-1} + \alpha Q)^{-1}.$$  \hspace{1cm} (6.5)

It is plain to see that

$$\lim_{\alpha \to 1} \tilde{b}_\alpha(x) = \tilde{b}(x) ,$$  \hspace{1cm} (6.6)

i.e. the optimal estimator under the risk function $r_\alpha$ converges to the constrained estimator when the relative penalty for misbalance increases.

7. CONCLUSION

The results of the author's previous paper (Neuhaus, 1995) have been generalised. Constrained estimators solve a practical problem faced by most actuaries and, as it turns out, the necessary adjustment is often very simple to compute. However, the warning about constraints creating an excess risk cannot be put too strongly; in unfavourable cases, an elaborate search for a realistic model and the optimal estimator may well have been in vain, if subsequent use of constraints greatly increases the risk of the estimator.

An interesting aspect concerns the use of constraints in empirical Bayes estimation and empirical credibility estimation. The normal procedure followed by actuaries is to estimate the distribution of $(b, X)$ (or its first and second order moments if only a credibility estimator is sought), and then to act as if the estimated model was the true model. In that case the constraining adjustment is still appropriate because, as we have seen, it is independent of the model used to derive the Bayes (or credibility) estimator. However, the resulting constrained estimator will not be the optimal constrained estimator (or balanced credibility estimator), only an approximation of it.

The last argument may be extended to arbitrary estimators. The derivation in Section 2 rests fully on a pointwise minimum distance projection of the optimal estimator into the space of constrained estimators. Now if one redefines the optimisation problem from one of finding the optimal constrained estimator, to one of finding the constrained estimator with minimum distance from a given
estimator, one still arrives at the same constraining adjustment. A similar argument is valid for the weighted estimator of Section 6.

ACKNOWLEDGEMENTS

This paper was written during the author's stay at the Laboratory of Actuarial Mathematics. Thanks go to the Laboratory staff for their hospitality and in particular to Ragnar Norberg for making my visit possible. Thanks also to Ragnar Norberg for the probing questions that lead to this paper, and to Thomas Mack (Münchener Rück) for posing the question that is answered in Section 5. Finally, thanks to two anonymous referees for their thoughtful comments.

References


Address for correspondence
Walther Neuhaus,
UNI Storebrand ASA,
P.O. Box 1380
Vika, N-0114 Oslo,
Norway
AN EXTENSION OF MACK'S MODEL FOR THE CHAIN LADDER METHOD

KLAUS D. SCHMIDT AND ANJA SCHNAUS

ABSTRACT

The chain ladder method is a simple and suggestive tool in claims reserving, and various attempts have been made aiming at its justification in a stochastic model. Remarkable progress has been achieved by Schnieper and Mack who considered models involving assumptions on conditional distributions. The present paper extends the model of Mack and proposes a basic model in a decision theoretic setting. The model allows to characterize optimality of the chain ladder factors as predictors of non-observable development factors and hence optimality of the chain ladder predictors of aggregate claims at the end of the first non-observable calendar year. We also present a model in which the chain ladder predictor of ultimate aggregate claims turns out to be unbiased.

1. INTRODUCTION

The chain ladder method is a simple and suggestive tool in claims reserving, and various attempts have been made aiming at its justification in a stochastic model. Remarkable progress has been achieved by Schnieper [1991] and Mack [1993, 1994a, 1994b] who considered models involving assumptions on conditional distributions.

The present paper proposes a basic model in a decision theoretic setting (Section 2) which is analyzed on the background of a general result on conditional prediction (Section 3). The model allows to characterize optimality of the chain ladder factors as predictors of non-observable development factors and hence optimality of the chain ladder predictors of aggregate claims at the end of the first non-observable calendar year (Section 4).

The model considered here is exclusively based on assumptions on the conditional joint distribution (with respect to the past over all occurrence years) of the collection of all development factors from a given development year; by contrast, the model of Mack assumes unconditional independence of the occurrence years and certain properties of the conditional distributions of single development factors. Since our model properly extends the model of Mack (Section 5), we obtain a justification of the chain ladder method under strictly weaker assumptions.

We also present a partial solution to the prediction problem for ultimate aggregate claims: It is shown that in another model which again properly extends the model of
Mack the chain ladder predictor of ultimate aggregate claims is unbiased but shares this property with many other predictors (Section 6). Optimality of the chain ladder predictor of ultimate aggregate claims remains an open problem.

Throughout this paper, let \((\Omega, \mathcal{F}, P)\) be a probability space. We assume that all random variables under consideration have finite second moments.

2. THE PREDICTION PROBLEM AND THE BASIC MODEL

Consider a family of random variables \(\{S_{i,k}\}_{i,k \in [0,1,\ldots,n]}\). The random variable \(S_{i,k}\) is interpreted as the aggregate claim size of all claims which occur in occurrence year \(i\) and which are settled before the end of calendar year \(i + k\). We also interpret the subscript \(k\) as the development year.

We assume that the aggregate claims \(S_{i,k}\) are strictly positive and that they are observable for \(i + k \leq n\) but non-observable for \(i + k > n\). The observable aggregate claims can be represented by the run-off triangle:

<table>
<thead>
<tr>
<th>Occurrence year</th>
<th>0</th>
<th>1</th>
<th>\ldots</th>
<th>(n - i)</th>
<th>(n - i + 1)</th>
<th>\ldots</th>
<th>(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(S_{0,0})</td>
<td>(S_{0,1})</td>
<td>\ldots</td>
<td>(S_{0,n-i})</td>
<td>(S_{0,n-i+1})</td>
<td>\ldots</td>
<td>(S_{0,n})</td>
</tr>
<tr>
<td>1</td>
<td>(S_{1,0})</td>
<td>(S_{1,1})</td>
<td>\ldots</td>
<td>(S_{1,n-i})</td>
<td>(S_{1,n-i+1})</td>
<td>\ldots</td>
<td>(S_{1,n})</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
</tr>
<tr>
<td>(i - 1)</td>
<td>(S_{i-1,0})</td>
<td>(S_{i-1,1})</td>
<td>\ldots</td>
<td>(S_{i-1,n-i+1})</td>
<td>(S_{i-1,n-i})</td>
<td>\ldots</td>
<td>(S_{i-1,n})</td>
</tr>
<tr>
<td>(i)</td>
<td>(S_{i,0})</td>
<td>(S_{i,1})</td>
<td>\ldots</td>
<td>(S_{i,n-i})</td>
<td>(S_{i,n})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
</tr>
<tr>
<td>(n - 1)</td>
<td>(S_{n-1,0})</td>
<td>(S_{n-1,1})</td>
<td>\ldots</td>
<td>(S_{n-1,n})</td>
<td>(S_{n-1,n-1})</td>
<td>\ldots</td>
<td>(S_{n-1,n-1})</td>
</tr>
<tr>
<td>(n)</td>
<td>(S_{n,0})</td>
<td>(S_{n,1})</td>
<td>\ldots</td>
<td>(S_{n,n-1})</td>
<td>(S_{n,n-1})</td>
<td>\ldots</td>
<td>(S_{n,n-1})</td>
</tr>
</tbody>
</table>

The problem is to predict the non-observable aggregate claims from the observable ones.

The chain ladder method consists in using the chain ladder predictors

\[
\hat{S}_{i,m} := S_{i,n-i} \cdot \prod_{l=n-i+1}^{m} \hat{F}_l
\]

for all \(i \in \{1, \ldots, n\}\) and \(m \in \{n - i + 1, \ldots, n\}\), where the chain ladder factors \(\hat{F}_l\) are defined by

\[
\hat{F}_l := \frac{\sum_{i=0}^{n-l} S_{i,l}}{\sum_{i=0}^{n-l} S_{i,l-1}}
\]

for all \(l \in \{1, \ldots, n\}\).
In order to study the properties of the chain ladder factors and of the chain ladder predictors, we introduce the development factors

\[ F_{i,l} := \frac{S_{i,l}}{S_{i,l-1}} \]

for all \( i \in \{0, 1, \ldots, n\} \) and \( l \in \{1, \ldots, n\} \). Then the aggregate claims satisfy

\[ S_{i,m} = S_{i,n-i} \cdot \prod_{l=n-i+1}^{m} F_{i,l} \]

for all \( i \in \{0, 1, \ldots, n\} \) and \( m \in \{n-i+1, \ldots, n\} \), and the chain ladder factors can be written as

\[ \hat{F}_l = \sum_{i=0}^{n-l} \sum_{j=0}^{n-l} \frac{S_{i,l-j}}{S_{j,l-j}} F_{i,l} \]

for all \( l \in \{1, \ldots, n\} \).

Let us now change the point of view by turning from occurrence years to development years.

<table>
<thead>
<tr>
<th>Occurrence year</th>
<th>Development year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>( n-k )</td>
<td>( n-k )</td>
</tr>
<tr>
<td>( n-k+1 )</td>
<td>( n-k+1 )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>( n-l )</td>
<td>( n-l )</td>
</tr>
<tr>
<td>( n )</td>
<td>( n )</td>
</tr>
</tbody>
</table>

First of all, it is easy to see that for each \( k \in \{1, \ldots, n\} \) the chain ladder factor \( \hat{F}_k \) minimizes the expression

\[ \sum_{i=0}^{n-k} \sum_{j=0}^{n-k} \frac{S_{i,k-j}}{S_{j,k-j}} (F_{i,k} - \delta)^2 \]

over all random variables \( \delta \). Thus, for development year \( k \), the chain ladder factor \( \hat{F}_k \) is the best approximation of the observable development factors when the approximation errors are given the weights occurring in the representation of the chain ladder factor as a weighted mean.
In what follows we shall study optimality of the chain ladder factors as *predictors* of non-observable development factors. To this end, we first formulate the *prediction problem* and then state the *basic model*:

**Prediction Problem:** For \( k \in \{1, \ldots, n\} \), let \( \mathcal{G}_k \) denote the \( \sigma \)-algebra generated by the family of random variables

\[
\{ S_{i,l} \}_{i \in \{0,1,\ldots,n\}, l \in \{0,1,\ldots,k-1\}}
\]

and let \( \Delta_k \) denote the collection of all random variables \( \delta \) which can be written as

\[
\delta = \sum_{i=0}^{n-k} W_i F_{i,k}
\]

where the weights of the development factors are \( \mathcal{G}_k \)-measurable random variables satisfying

\[
\sum_{i=0}^{n-k} W_i = 1.
\]

For each \( j \in \{n - k + 1, \ldots, n\} \), the problem is to find some \( \delta_j^* \in \Delta_k \) satisfying

\[
E((F_{j,k} - \delta_j)^2 \mid \mathcal{G}_k) = \inf_{\delta \in \Delta_k} E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k).
\]

These quantities can be interpreted as follows:

- The \( \sigma \)-algebra \( \mathcal{G}_k \) represents the information provided by the past preceding development year \( k \).
- The non-observable development factors are to be predicted by a weighted mean of observable development factors from the same development year such that the weights are measurable functions of the aggregate claims in the past. (It is not assumed that the weight are positive.)
- The optimality criterion is conditional expected squared error loss, given the information provided by the aggregate claims in the past.

The conditional loss function instead of the usual unconditional one is reasonable since optimality is desired only with regard to the information provided by the past.

**Basic Model:** For each \( k \in \{1, \ldots, n\} \), there exists a random variable \( F_k \) such that

\[
E(F_{i,k} \mid \mathcal{G}_k) = F_k
\]

\[
\text{cov}(F_{i,k}, F_{j,k} \mid \mathcal{G}_k) = 0
\]

\[
\text{var}(F_{i,k} \mid \mathcal{G}_k) > 0
\]

holds for all \( i, j \in \{0,1,\ldots,n\} \) such that \( i \neq j \).
The following lemma is of interest with regard to the model of Mack which will be studied in Section 5:

**2.1. Lemma** Under the assumptions of the basic model and for each \( k \in \{1, \ldots, n\} \), the following are equivalent:

(a) There exists a real number \( f_k \) such that
\[
F_k = f_k.
\]
(b) The identity
\[
\text{cov}[F_{i,k}, F_{j,k}] = 0
\]
holds for all \( i, j \in \{0, 1, \ldots, n\} \) such that \( i \neq j \).

The prediction problem for the basic model will be studied in Section 4 below.

### 3. Conditional Prediction

In the present section, we study an abstract prediction problem which will later be applied to the prediction of non-observable development factors.

Throughout this section, let \( \{X_i\}_{i \in \{1, \ldots, m+1\}} \) be a family of random variables and let \( \mathcal{G} \) be a sub-\( \sigma \)-algebra of \( \mathcal{F} \). We assume that there exists a random variable \( X \) such that
\[
E(X_i \mid \mathcal{G}) = X
\]
\[
\text{cov}(X_i, X_j \mid \mathcal{G}) = 0
\]
\[
\text{var}(X_i \mid \mathcal{G}) > 0
\]
holds for all \( i, j \in \{1, \ldots, m+1\} \) such that \( i \neq j \). We also assume that the random variables \( X_1, \ldots, X_m \) are observable whereas \( X_{m+1} \) is non-observable.

Let \( \Delta \) denote the collection of all random variables \( \delta \) which can be written as
\[
\delta = \sum_{i=1}^{m} W_i X_i
\]
where the weights are \( \mathcal{G} \)-measurable random variables satisfying
\[
\sum_{i=1}^{m} W_i = 1.
\]

The random variables in \( \Delta \) are called *admissible predictors* of \( X_{m+1} \).

The problem is to find some \( \hat{\delta} \in \Delta \) satisfying
\[
E((X_{m+1} - \delta)^2 \mid \mathcal{G}) = \inf_{\delta \in \Delta} E((X_{m+1} - \delta)^2 \mid \mathcal{G}).
\]
that is, to predict the non-observable random variable \( X_{m+1} \) by a weighted mean of the observable ones such that the weights contain information from outside the sample \( \{ X_1, \ldots, X_m \} \) and such that conditional expected squared error loss in minimized.

**Remark.** The classical case is the case where \( G = \{ \emptyset, \Omega \} \), which means that
- no information from outside the sample is available,
- the random variables \( X_1, \ldots, X_m, X_{m+1} \) are uncorrelated with equal expectations and strictly positive variances,
- the admissible predictors have constant weights, and
- the optimality criterion is unconditional expected squared error loss.

The following lemma is immediate:

3.1. **Lemma.** The identities

\[
E(\delta \mid G) = X
\]

and

\[
E((X_{m+1} - \delta)^2 \mid G) = \text{var}(X_{m+1} \mid G) + \text{var}(\delta \mid G)
\]

hold for all \( \delta \in \Delta \).

The following result establishes existence, uniqueness, and the form of the weights of the optimum predictor of \( X_{m+1} \):

3.2. **Theorem.** For

\[
\hat{\delta} = \sum_{i=1}^{m} \hat{W}_i X_i \in \Delta,
\]

the following are equivalent:

(a) There exists a random variable \( \Lambda \) such that

\[
\hat{W}_i = \frac{\Lambda}{\text{var}(X_i \mid G)}
\]

holds for all \( i \in \{1, \ldots, m\} \).

(b) The inequality

\[
E((X_{m+1} - \hat{\delta})^2 \mid G) \leq E((X_{m+1} - \delta)^2 \mid G)
\]

holds for all \( \delta \in \Delta \).
In this case,

\[ \text{var}(\hat{\delta} \mid \mathcal{G}) = \Lambda = \left( \sum_{i=1}^{m} \frac{1}{\text{var}(X_i \mid \mathcal{G})} \right)^{-1}, \]

as well as

\[ E\left( \frac{1}{m-1} \sum_{i=1}^{m} \hat{W}_i (X_i - \hat{\delta})^2 \mid \mathcal{G} \right) = \Lambda \]

when \( m \geq 2 \).

**Proof.** Define

\[ \Lambda := \left( \sum_{i=1}^{m} \frac{1}{\text{var}(X_i \mid \mathcal{G})} \right)^{-1} \]

and let

\[ W_i^* := \frac{1}{\text{var}(X_i \mid \mathcal{G})} \]

for all \( i \in \{1, \ldots, m\} \). For each

\[ \delta = \sum_{i=1}^{m} W_i X_i \in \Delta, \]

we have

\[ \text{var}(\hat{\delta} \mid \mathcal{G}) = \text{var}\left( \sum_{i=1}^{m} W_i X_i \mid \mathcal{G} \right) \]

\[ = \sum_{i=1}^{m} W_i^2 \text{var}(X_i \mid \mathcal{G}) \]

\[ = \sum_{i=1}^{m} (W_i - W_i^*)^2 \text{var}(X_i \mid \mathcal{G}) + 2 \sum_{i=1}^{m} W_i W_i^* \text{var}(X_i \mid \mathcal{G}) - \sum_{i=1}^{m} (W_i^*)^2 \text{var}(X_i \mid \mathcal{G}) \]

\[ = \sum_{i=1}^{m} (W_i - W_i^*)^2 \text{var}(X_i \mid \mathcal{G}) + 2 \Lambda \sum_{i=1}^{m} W_i - \Lambda \sum_{i=1}^{m} W_i^* \]

\[ = \sum_{i=1}^{m} (W_i - W_i^*)^2 \text{var}(X_i \mid \mathcal{G}) + \Lambda. \]

Because of Lemma 3.1, this proves the equivalence of (a) and (b) as well as the identity for \( \text{var}(\hat{\delta} \mid \mathcal{G}) \). The final identity follows by straightforward computation.
**Remark.** In the special case where there exists a random variable $V$ satisfying $\text{var}(X_i | \mathcal{G}) = V$ for all $i \in \{1, \ldots, m\}$, the optimum predictor of $X_{m+1}$ is the sample mean

$$\bar{X} := \frac{1}{m} \sum_{i=1}^{m} X_i$$

and we have

$$\text{var}(\bar{X} | \mathcal{G}) = E\left( \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X})^2 \mid \mathcal{G} \right).$$

In the classical case, this reduces to the well-known fact that

$$\frac{1}{m} \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X})^2$$

is an unbiased estimator of the variance of the sample mean.

### 4. The Results

We now turn to the prediction problem for the basic model. Consider $k \in \{1, \ldots, n\}$.

**4.1. Lemma.** Under the assumptions of the basic model, the identities

$$E(\delta \mid \mathcal{G}_k) = F_k$$

and

$$E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k) = \text{var}(F_{j,k} \mid \mathcal{G}_k) + \text{var}(\delta \mid \mathcal{G}_k)$$

hold for all $\delta \in \Delta_k$ and for all $j \in \{n-k+1, \ldots, n\}$.

This is immediate from Lemma 3.1.

The following result characterizes optimality of the chain ladder factor:

**4.2. Theorem.** Under the assumptions of the basic model, the following are equivalent:

(a) There exists a random variable $V_k$ such that

$$\text{var}(F_{i,k} \mid \mathcal{G}_k) = \frac{V_k}{S_{i,k-1}}$$

holds for all $i \in \{0, \ldots, n-k\}$.

(b) The inequality

$$E((F_{j,k} - \hat{F}_k)^2 \mid \mathcal{G}_k) \leq E((F_{j,k} - \delta)^2 \mid \mathcal{G}_k)$$

holds for all $\delta \in \Delta_k$ and for some $j \in \{n-k+1, \ldots, n\}$. 

(c) The inequality
\[ E((F_{j,k} - \hat{F}_k)^2 \mid G_k) \leq E((F_{j,k} - \delta)^2 \mid G_k) \]
holds for all \( \delta \in \Delta_k \) and for all \( j \in \{n-k+1, \ldots, n\} \).

In this case, \( \delta^* = \hat{F}_k \) holds for each \( \delta^* \in \Delta_k \) such that
\[ E((F_{j,k} - \delta^*)^2 \mid G_k) \leq E((F_{j,k} - \delta)^2 \mid G_k) \]
holds for all \( \delta \in \Delta_k \) and for some \( j \in \{n-k+1, \ldots, n\} \); moreover,
\[ \text{var}(\hat{F}_k \mid G_k) = \frac{V_k}{\sum_{i=0}^{n-k} S_{i,k-1}} = \left( \sum_{i=0}^{n-k} \frac{1}{\text{var}(F_{i,k} \mid G_k)} \right)^{-1}, \]
as well as
\[ E \left( \frac{1}{n-k} \sum_{i=0}^{n-k} \frac{S_{i,k-1}}{\sum_{l=0}^{n-k} S_{l,k-1}} (F_{i,k} - \hat{F}_k)^2 \mid G_k \right) = \text{var}(\hat{F}_k \mid G_k) \]
when \( k \leq n - 1 \).

**Proof.** By Theorem 3.2, the chain ladder factor
\[ \hat{F}_k := \sum_{i=0}^{n-k} \frac{S_{i,k-1}}{\sum_{l=0}^{n-k} S_{l,k-1}} F_{i,k} \]
minimizes conditional expected squared error loss if and only if the identity
\[ \frac{1}{\sum_{i=0}^{n-k} S_{i,k-1}} = \frac{\text{var}(F_{i,k} \mid G_k)}{\sum_{i=0}^{n-k} \frac{1}{\text{var}(F_{i,k} \mid G_k)}} \]
holds for all \( i \in \{0, 1, \ldots, n-k\} \), and this identity is equivalent with
\[ \text{var}(F_{i,k} \mid G_k) = \frac{1}{S_{i,k-1}} \sum_{l=0}^{n-k} \frac{1}{\sum_{i=0}^{n-k} \text{var}(F_{i,k} \mid G_k)} \]
This yields the equivalence of (a) and (b).

The equivalence of (b) and (c) is obvious from Lemma 4.1, and the final assertion follows from Theorem 3.2.

The previous result suggests the definition of the following general model:
**General Model:** For each $k \in \{1, \ldots, n\}$, there exist random variables $F_k$ and $V_k > 0$ such that

\[
E(F_{i,k} \mid G_k) = F_k
\]
\[
\text{cov}(F_{i,k}, F_{j,k} \mid G_k) = 0
\]
\[
\text{var}(F_{i,k} \mid G_k) = \frac{V_k}{S_{i,k-1}}
\]

holds for all $i, j \in \{0, 1, \ldots, n\}$ such that $i \neq j$.

**4.3. Corollary.** Under the assumptions of the general model, the chain ladder factors satisfy

\[
E(\hat{F}_k \mid G_k) = F_k
\]

and

\[
\text{var}(\hat{F}_k \mid G_k) = \frac{V_k}{\sum_{i=0}^{n-k} S_{i,k-1}} = \inf_{\delta \in \Delta_k} \text{var}(\delta \mid G_k)
\]

for all $k \in \{1, \ldots, n\}$ as well as

\[
E((F_{j,k} - \hat{F}_k)^2 \mid G_k) = \inf_{\delta \in \Delta_k} E((F_{j,k} - \delta)^2 \mid G_k)
\]

and

\[
E((F_{j,k} - \hat{F}_k)^2 \mid G_k) = \text{var}(F_{j,k} \mid G_k) + \text{var}(\hat{F}_k \mid G_k) = \frac{V_k}{S_{j,k-1}} + \frac{V_k}{\sum_{i=0}^{n-k} S_{i,k-1}}
\]

for all $k \in \{1, \ldots, n\}$ and for all $j \in \{n-k+1, \ldots, n\}$; moreover, the identity

\[
E\left(\frac{1}{n-k} \sum_{i=0}^{n-k} S_{i,k-1} (F_{i,k} - \hat{F}_k)^2 \mid G_k\right) = V_k
\]

holds for all $k \in \{1, \ldots, n-1\}$.

**Conclusion:** Under the assumptions of the general model, we have, for each $\delta \in \Delta_k$,

\[
E(\delta \mid G_k) = E(F_{j,k} \mid G_k)
\]

and hence

\[
E(S_{n-k+1,k-1} \cdot \delta \mid G_k) = S_{n-k+1,k-1} \cdot E(\delta \mid G_k)
\]
\[
= S_{n-k+1,k-1} \cdot E(F_{j,k} \mid G_k)
\]
\[
= E(S_{n-k+1,k-1} \cdot F_{j,k} \mid G_k)
\]
\[
= E(S_{n-k+1,k-1} \mid G_k),
\]
and this implies that $\delta$ and $S_{n-k+1,k-1}$ are unbiased predictors of $F_{j,k}$ and $S_{n-k+1,k}$, respectively; moreover, we have

$$E((F_{j,k} - \hat{F}_k)^2 \mid G_k) = \inf_{\delta \in \Delta_k} E((F_{j,k} - \delta)^2 \mid G_k)$$

and hence

$$E((S_{n-k+1,k} - \hat{S}_{n-k+1,k})^2 \mid G_k) = \inf_{\delta \in \Delta_k} E((S_{n-k+1,k} - S_{n-k+1,k-1} \cdot \delta)^2 \mid G_k),$$

which means that the chain ladder factor $\hat{F}_k$ and the chain ladder predictor $\hat{S}_{n-k+1,k} = S_{n-k+1,k-1}$ are the optimum predictors of $F_{j,k}$ and $S_{n-k+1,k}$, respectively.

This solves completely the prediction problem for the first non-observable year $n+1$.

5. THE MODEL OF MACK

For all $i, k \in \{0, 1, \ldots, n\}$, define

$$S_{i,k} := \sigma(S_{i,j})_{i \in \{0, 1, \ldots, k\}}.$$

These $\sigma$-algebras are needed to formulate the model of Mack:

**Model of Mack:** The family of $\sigma$-algebras $\{S_{i,n}\}_{i \in \{0, 1, \ldots, n\}}$ is independent and, for each $k \in \{1, \ldots, n\}$, there exist real numbers $f_k$ and $v_k > 0$ such that

$$E(F_{i,k} \mid S_{i,k-1}) = f_k$$

$$\text{var}(F_{i,k} \mid S_{i,k-1}) = \frac{v_k}{S_{i,k-1}}$$

holds for all $i \in \{0, 1, \ldots, n\}$.

The main problem when comparing the model of Mack with the general model consists in the fact that (unconditional) independence does not imply conditional independence (and vice versa). Nevertheless, we have the following result:

5.1. **Theorem.** The model of Mack is a special case of the general model.

**Proof.** Consider $k \in \{1, \ldots, n\}$. Since the family $\{S_{i,n}\}_{i \in \{0,1,\ldots,n\}}$ is independent, the family $\{S_{i,k-1}\}_{i \in \{0,1,\ldots,n\}}$ is independent as well. Also, for all $i \in \{0, 1, \ldots, n\}$, we have $S_{i,k-1} \subseteq G_k$. This yields

$$E(F_{i,k} \mid G_k) = E(F_{i,k} \mid S_{i,k-1}) = f_k$$

and

$$\text{var}(F_{i,k} \mid G_k) = \text{var}(F_{i,k} \mid S_{i,k-1}) = \frac{v_k}{S_{i,k-1}}.$$
Furthermore, using independence repeatedly and in a similar manner as before, we obtain, for all $i, j \in \{0, 1, \ldots, n\}$ such that $i \neq j$,

$$E(F_{i,k}F_{j,k} \mid \mathcal{G}_k) = E\left(E\left(F_{i,k}F_{j,k} \mid \sigma(S_{i,k-1} \cup S_{j,k-1})\right) \mid \sigma(S_{i,k-1} \cup S_{j,k-1})\right)$$

$$= E\left(F_{i,k} \cdot E\left(F_{j,k} \mid \sigma(S_{i,k-1} \cup S_{j,k-1})\right) \right) \cdot E\left(F_{j,k} \mid \sigma(S_{i,k-1} \cup S_{j,k-1})\right)$$

$$= E\left(F_{i,k} \mid \sigma(S_{i,k-1} \cup S_{j,k-1})\right) \cdot E\left(F_{j,k} \mid S_{j,k-1}\right)$$

$$= E\left(F_{i,k} \mid S_{i,k-1}\right) \cdot E\left(F_{j,k} \mid S_{j,k-1}\right)$$

and thus

$$\text{cov}(F_{i,k}, F_{j,k} \mid \mathcal{G}_k) = 0$$

The assertion follows.

Because of Lemma 2.1, the model of Mack is even properly contained in the general model; this is also true when the random variables $F_k$ and $V_k$ of the general model are assumed to be constant.

6. COMPLEMENT: UNBIASED PREDICTION OF ULTIMATE AGGREGATE CLAIMS IN A MODIFIED MODEL

In the general model, the chain ladder predictor $\hat{S}_{i,n-i+1}$ is the optimum predictor of the aggregate claims $S_{i,n-i+1}$ in the first non-observable calendar year $n + 1$. By contrast, optimum prediction of the ultimate aggregate claims $S_{i,n}$ remains an open problem (except for the case $i = 1$).

Mack proved, in this model, that the chain ladder predictor of ultimate aggregate claims is unbiased. We now formulate a modification of the general model in which every predictor of the form

$$S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} \delta_l$$

with $\delta_l \in \Delta_l$ for all $l \in \{n - i + 1, \ldots, n\}$ turns out to be an unbiased predictor of the ultimate aggregate claims.

**Modified Model:** For each $k \in \{1, \ldots, n\}$, there exists a random variable $F_k$ such that

$$E(F_{i,k} \mid \mathcal{G}_k) = F_k$$
and the identity

$$\text{cov}\left( F_{i,k}, \prod_{l=k+1}^{n} F_l \mid G_k \right) = 0$$

holds for all $k \in \{1, \ldots, n\}$ and $i \in \{0, 1, \ldots, n\}$.

The general model and the modified model can be combined without any problem. Moreover, if in the general model the random variables $F_i$ are assumed to be constant, then the assumptions of the modified model are automatically fulfilled; in particular, the model of Mack is a special case of the modified model.

6.1. Lemma. Under the assumptions of the modified model, the identities

$$E(\delta_k \mid G_k) = F_k$$

and

$$E\left( \prod_{l=k}^{m} \delta_l \cdot \prod_{l=m+1}^{n} F_l \mid G_k \right) = E\left( \prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m}^{n} F_l \mid G_k \right)$$

hold for all $k \in \{1, \ldots, n\}$ and $m \in \{k, \ldots, n\}$ and for every choice of $\delta_i \in \Delta_i$ for all $l \in \{k, \ldots, m\}$.

Proof. The first identity is obvious. Furthermore, we have

$$E\left( \delta_m \cdot \prod_{l=m+1}^{n} F_l \mid G_m \right) = E\left( \prod_{l=m}^{n} F_l \mid G_m \right)$$

and hence

$$E\left( \prod_{l=k}^{m} \delta_l \cdot \prod_{l=m+1}^{n} F_l \mid G_k \right) = E\left( \prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m+1}^{n} F_l \mid G_k \right)$$

$$= E\left( \prod_{l=k}^{m-1} \delta_l \cdot E\left( \delta_m \cdot \prod_{l=m+1}^{n} F_l \mid G_m \right) \mid G_k \right)$$

$$= E\left( \prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m}^{n} F_l \mid G_m \right) \mid G_k$$

$$= E\left( \prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m}^{n} F_l \mid G_m \right) \mid G_k$$

$$= E\left( \prod_{l=k}^{m-1} \delta_l \cdot \prod_{l=m}^{n} F_l \mid G_m \right) \mid G_k$$

which proves the second identity.
6.2. Theorem. Under the assumptions of the modified model, the identity

\[ E\left(S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} \delta_l \mid \mathcal{G}_{n-i+1}\right) = E(S_{i,n} \mid \mathcal{G}_{n-i+1}) \]

holds for all \( i \in \{0, 1, \ldots, n\} \) and for every choice of \( \delta_l \in \Delta_i \) for all \( l \in \{n-i+1\} \).

Proof. By Lemma 6.1, we have

\[ E\left(\prod_{l=n-i+1}^{n} \delta_l \mid \mathcal{G}_{n-i+1}\right) = E\left(\prod_{l=n-i+1}^{n} F_{i,l} \mid \mathcal{G}_{n-i+1}\right) \]

and

\[ E\left(\prod_{l=n-i+1}^{n} F_{i,l} \mid \mathcal{G}_{n-i+1}\right) = E\left(\prod_{l=n-i+1}^{n} F_{i} \mid \mathcal{G}_{n-i+1}\right), \]

and hence

\[ E\left(S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} \delta_l \mid \mathcal{G}_{n-i+1}\right) = S_{i,n-i} \cdot E\left(\prod_{l=n-i+1}^{n} \delta_l \mid \mathcal{G}_{n-i+1}\right) \]

\[ = S_{i,n-i} \cdot E\left(\prod_{l=n-i+1}^{n} F_{i,l} \mid \mathcal{G}_{n-i+1}\right) \]

\[ = S_{i,n-i} \cdot E\left(\prod_{l=n-i+1}^{n} F_{i,l} \mid \mathcal{G}_{n-i+1}\right) \]

\[ = E\left(S_{i,n-i} \cdot \prod_{l=n-i+1}^{n} F_{i,l} \mid \mathcal{G}_{n-i+1}\right) \]

\[ = E(S_{i,n} \mid \mathcal{G}_{n-i+1}) \]

as was to be shown.

Conclusion: Under the assumptions of the modified model, the chain ladder predictor is an unbiased predictor of the ultimate aggregate claims, but many other predictors are unbiased as well.

In order to establish optimality, and not only unbiasedness, of the chain ladder predictor, the modified model should be restricted by additional assumptions which are in the spirit of the general model. These additional assumptions should concern products of development factors instead of single ones.
7. REMARKS

At the first glance, it may appear to be somewhat strange that $\sigma$-algebras $G_k$ which are used for conditioning, include (except for the case $k = 1$) non-observable information. However, non-observable information drops out automatically in the formulas for the optimum predictors of non-observable development factors. Moreover, all results remain valid when the $\sigma$-algebras $G_k$ are replaced by the $\sigma$-algebras $E_k := \sigma \left( \left\{ S_{i,k-1} \right\}_{i \in \{0,1,\ldots,n-k\}} \right)$ or by any $\sigma$-algebras $\mathcal{F}_k$ satisfying $E_k \subseteq \mathcal{F}_k \subseteq G_k$; a natural choice would be to take $\mathcal{F}_k := G_k \cap D$, where $D$ denotes the $\sigma$-algebra generated by the run-off triangle. The choice of the $\sigma$-algebras $G_k$ considered here allows to capture the model of Mack, which also uses conditioning with respect to $\sigma$-algebras including non-observable information.

In the modified model, it is easily seen that the additional assumption

$$E \left[ \prod_{i=k}^{n} F_i \right] = \prod_{i=k}^{n} E \left[ F_i \right]$$

implies

$$E \left[ \prod_{i=k}^{n} \hat{F}_i \right] = \prod_{i=k}^{n} E \left[ \hat{F}_i \right],$$

which means that successive chain ladder factors are uncorrelated. This assumption is automatically fulfilled if in the general model the random variables $F_k$ are assumed to be constant; in particular, the assumption is fulfilled in the model of Mack. To the present authors, however, uncorrelatedness of chain ladder factors seems to be of minor importance when compared with unbiasedness of the chain ladder predictor, and assumptions on unconditional expectations appear to be a bit strange in the general setting of conditional prediction considered in this paper.

ACKNOWLEDGEMENT

The authors gratefully acknowledge various discussions on the subject with Thomas Mack and Lothar Partzsch.

REFERENCES


KLAUS D. SCHMIDT
Lehrstuhl für Versicherungsmathematik
Technische Universität Dresden
D-01062 Dresden
Germany

ANJA SCHNAUS
Kölner Rück
Postfach 10 22 44
D-50642 Köln
Germany
BOOK REVIEW


The new book by Jean Lemaire will be welcomed and appreciated by actuaries working in Automobile Insurance. The presentation of different perspectives of the subject makes it also a valuable reading for actuaries outside Automobile Insurance, who are interested in experience rating and or optimal deductibles. Although it gives a broad and deep introduction to the subject, it is not a boring academic textbook guiding the reader through the matter with lemmas and theorems. The mathematical developments are sufficient to be followed without having to reinvent the wheel, but not so extended to make the reading boring. The student will learn that the development of a bonus-malus system is not only an academic exercise, but embedded in a social and political context, in which several opinions and interests are to be considered. Actuaries will find a complete introduction to the subject, including basics of risk theory and statistics. The book contains many examples and ideas, but no recipe giving simple answers: The reader is not told which is best bonus-malus system, but the reader finds many ideas to approach his specific problem, and many different sources to analyse a given bonus-malus system, to identify its main characteristics, weaknesses and strengths, and to modify some keys according to given targets.

For obvious reasons, I read it from A to Z. I think that it is the best approach to the book, as the different concepts are introduced where they are needed. The book can also serve as reference. The preface contains as detailed overview over the content of the 16 chapters and two appendices forming the book, and therefore allows for a quick orientation. In that preface, the author mentions that he is definitely biased towards simplicity, and therefore wrote a book that can be reads by the entire ASTIN membership, not only its academic subset. "An unavoidable consequence of this approach is that the level of mathematics will tend to be somewhat uneven." This is the main strength of the book and makes it a pleasant and interesting reading.

Part I (Chapters 1-3) defines a bonus-malus system, illustrates the development of the Belgian bonus-malus system over the last 30 years, and introduces to different models for the claim number distribution. The later chapter also discusses several goodness of fit tests. In Appendix A these tests are compared by applying them to some examples in other disciplines.

Part II (Chapter 4-9) develops four tools for analyzing bonus-malus systems: The relative stationary average level, the coefficient of variation of the insured's premium, the elasticity of a bonus-malus system and the average optimal retention. In each of the chapters, the defined key is computed for thirty different systems from twenty-two countries. All these bonus-malus systems are described in

Appendix B. The last chapter of Part II compares the four measures with factor analysis. The method is introduced, and can be understood also by a reader who is not familiar with this part of statistics.

Parts III and IV of the book are more sophisticated than the preceding ones. Nevertheless, more sophisticated techniques are not developed for their own merit, but for answering natural questions arising when developing new bonus-malus systems and, possibly, negotiating them with involved third parties. Part III (Chapters 10-14) deals with the construction of an optimal system using the expected value principle with different loss functions. The concept is extended to other premium calculation principles. Moreover, a possible way of allowing for the severity of claims as factor of the bonus or malus is introduced. Finally, this part discusses the influence of an additional expense loading to the pure risk premium.

Let us cite the author: Except in life insurance, where there are specific cost models for sales commissions (in many cases of regulated form), there seems to be no further modelling principles used, except multiplying the risk premium by a factor 1+A. This lacuna in the literature is all the more surprising, as it is in sharp contrast to the fields of engineering and business management, where extensive and sophisticated cost allocation and modelling are the order of the day. Are these activities outside the realm of the actuary?

In Part IV (Chapters 15-16), ways of replacing a bonus-malus system by a high deductible are studied.

The book contains a long list of references, an author and a subject index.

Benedetto Conti
ANNOUNCEMENT FOR THE
7th INTERNATIONAL AFIR COLLOQUIUM
Cairns, Australia, 13-15 August 1997

The 7th International AFIR Colloquium will take place in Cairns, Australia at the Cairns International Hotel from 13-15 August 1997. For the first time both AFIR and ASTIN Colloquia will be held in the same week with ASTIN being held at the same venue from 10-13 August 1997. Joint AFIR/ASTIN sessions will be held on 13 August 1997.

The actuarial profession in Australia celebrates its 100th anniversary in 1997 and The Institute of Actuaries of Australia is holding its Centenary Convention at Sheraton Mirage, Gold Coast, Queensland from 17-20 August, 1997 and ASTIN/AFIR members are invited to attend this Convention.

Cairns is located in North Queensland and August is the best time of year to visit the Great Barrier Reef and the beautiful tropical Queensland Coast.

Papers can be submitted on any financial or investment topic of interest to AFIR. Papers are welcome that present innovative ideas, address significant practical issues or provide surveys of recent developments. Topic areas (both theory and application) could include: asset-liability modelling, stochastic investment models, asset allocation, portfolio management, options, interest rate derivatives, term structure models, risk management and assessment, balance sheet management, markets and instruments. The program will be constructed on the basis of papers submitted. Papers can also be submitted on topics suitable for the joint ASTIN/AFIR day. Topic areas could include: links between insurance and financial markets, capital and risk management in insurance, financial economic approach to insurance capitalisation, integration of actuarial and financial economic approaches to insurance. AFIR papers will be subject to review and the scheduling and determination of suitability will be decided by the Scientific Committee.

A number of internationally recognised speakers have been invited to present talks on topics of interest to AFIR participants on both the joint ASTIN/AFIR day and for the AFIR sessions.

The 2nd AFIR best paper prize will be awarded to the paper that best meets the following: “An applied or theoretical paper that provides an innovative solution to a practical investment or financial problem of interest to actuaries.”

Abstracts are required by 15 December 1996 and papers in their final form must be submitted by 1 March 1997. A copy of the first announcement and call for papers can be obtained from:

ASTIN/AFIR Colloquia Secretariat
Conference Action
PO Box 1231
North Sydney, NSW, 2059, Australia
Tel.: +61 2 9956 8333
Fax: +61 2 9956 5154
email : confact@real.net.au

Details about the Scientific Program can be obtained from
  Michael Sherris
  School of Economic and Financial Studies
  Macquarie University
  Sydney, NSW, Australia, 2109
  Tel.: +61 2 9850 8572
  Fax: +61 2 9850 9481
  email: msherris@efs.mq.edu.au

Details about AFIR will be placed on the WWW at url
  http://www.ocs.mq.edu.au/~msherris/afir97.html
XXVIIIth ASTIN COLLOQUIUM
Cairns, Australia — August 1997

The XXVIIIth International ASTIN Colloquium will be held back to back with the 7th International AFIR Colloquium in Cairns, Australia on the beautiful tropical Queensland coast and opposite the Great Barrier Reef. The relevant dates are:

**ASTIN**
10-13 August 1997

**AFIR**
13-15 August 1997

The Cairns International Hotel is the venue for both Colloquia. Capturing the essence of the tropics with commanding views of the harbour and surrounding hills, whilst being conveniently located in the heart of Cairns, giving you the best of Far North Queensland.

Apart from the Cairns International Hotel, a wide range of hotel prices is available within convenient distance of the conference venue.

**Program**

Topics for ASTIN are not circumscribed other than that they be relevant to non-life insurance. The program will be constructed on the basis of papers submitted. It is intended that a theme of natural catastrophes will run through the Colloquium, and contributions on this subject are especially encouraged.

A stimulating cultural program is also being arranged. In addition, Australia offers a wide range of pre- and post-Colloquium tourist options.

**Invited speakers**

A number of internationally recognised speakers have been invited to present talks on topics of interest to participants in both the ASTIN and AFIR Colloquia. Several of these are expected to relate to the catastrophe theme.

Enquiries concerning the ASTIN Colloquium can be directed to:

Greg Taylor
ASTIN Scientific Committee
Tillinghast - Towers Perrin
GPO Box 3279
Sydney NSW 2001
Telephone: 61 2 9229 5513
Facsimile: 61 2 9229 5588
e-mail: taylorg@towers.com

Institute of Actuaries of Australia Centenary Convention

The Institute of Actuaries of Australia celebrates its 100th anniversary in 1997 and is holding its Centenary Convention in the week immediately following the ASTIN and AFIR Colloquia on Queensland’s popular Gold Coast. ASTIN participants are
Welcome to attend the Institute of Actuaries of Australia Centenary Convention which will deal with a range of issues of interest to international attendees. The theme of the Convention is "Shaping the Next Century".

Macquarie University Research Conference

There will be a research conference held at Macquarie University in Sydney on Friday 22 August, 1997 to honour Professor Alfred Hurlstone Pollard. A programme of invited talks will make this a very interesting conference. Enquiries concerning this conference can be directed to Michael Sherris (see AFIR announcement for address).
CALL FOR PAPERS

AMERICAN RISK AND INSURANCE ASSOCIATION
1997 ANNUAL MEETING

August 10-13, 1997
Doubletree Hotel
San Diego, California

You are encouraged to submit a proposal to present research findings at the 1997 meeting of the American Risk and Insurance Association. Papers on any risk- or insurance-related topic are welcome. Specific subject areas include, but are not limited to, insurance law or regulation, public policy, economics, finance, health care, international issues, employees benefits or risk management. Executive summaries (not exceeding three pages) that focus on the purpose, expected results and importance of the research are preferred. The names and affiliations of all co-authors, with telephone and fax numbers and e-mail address (if available) of the designated contact person, should be provided on a separate cover page attached to the proposal. Proposals from doctoral students are encouraged.

The deadline for submission is February 15, 1997.
This deadline will not be extended.

Four copies of your proposal should be mailed to ARIA Vice-President and 1997 Program Chair:
Stephen P. D'Arcy
Department of Finance
University of Illinois
340 Commerce West
1206 S. Sixth Street
Champaign, IL 61820-6271

Phone: 217-333-0772
Fax: 217-244-3102
E-mail: sdarcy@commerce.cba.uiuc.edu

Submissions may be sent via e-mail provided that the author information requested above is also included. E-mail submissions should be sent as the text of an e-mail message rather than as an attached file.

Questions or suggestions concerning the program can be directed to Professor D'Arcy. Other questions about the San Diego meeting should be directed to the ARIA office by phone at 800-951-2020, fax at 914-699-2025 or e-mail at ARIA@PIPELINE.COM. For more information about ARIA, visit our website at http://www.aria.org.
ASTIN DONATES MAJOR ACTUARIAL BOOKS TO 120 UNIVERSITIES IN EMERGING COUNTRIES

Article 2 of the ASTIN rules state that ASTIN has as its objective the promotion of actuarial research, particularly in non-life insurance. To this end it organises the ASTIN Colloquia and publishes the ASTIN Bulletin. In addition to these two basic activities it may engage in other activities serving to advance the above objective.

During its September meeting in Copenhagen, the ASTIN Committee unanimously approved a proposal to donate 18 major actuarial books to 120 universities and actuarial associations in emerging countries.

The selection process took place as follows: in September 1995, in Brussels, the Committee created a task force, with the assignment to find ways to spend part of ASTIN’s assets for the benefit of actuarial science in emerging countries. The task force wrote to the entire ASTIN Committee, the Editorial Board of the ASTIN Bulletin, and other respected ASTIN members (45 members in total), and asked them to recommend fifteen books. The selections were quite convergent, and 18 books clearly emerged as the best actuarial books recently published. The following table lists the selected books, by decreasing order of votes.

1. Daykin, Pentikainen, Pesonen.
   "Practical Risk Theory for Actuaries"
2. Gerber
   "An Introduction to Mathematical Risk Theory"
3. Bühlmann
   "Mathematical Models in Risk Theory"
4. Lemaire
   "Automobile Insurance: Actuarial models"
5. Bowers, Gerber, Jones, Hickman, Nesbitt
   "Actuarial Mathematics"
6. Panjer and Willmot
   "Insurance Risk Models"
7. Hogg and Klugman
   "Loss Distributions"
8. Gerber
   "Life Insurance Mathematics"
9. Lemaire
   "Bonus-malus systems in Automobile Insurance"
10. Casualty Actuarial Society
    "Foundations of Casualty Actuarial Science"
11. Institute of Actuaries
    "Claims Reserving Manual"

ASTIN DONATES MAJOR ACTUARIAL BOOKS

12. Straub
   "Non-life Insurance Mathematics"
13. Taylor
   "Claims Reserving in Non-life Insurance"
14. Van Eeghen
   "Rate-Making"
15. Hossack, Pollard, Zehnwirth
   "Introductory Statistics with Applications in General Insurance"
16. Van Eeghen
   "Loss Reserving Methods"
17. Hart, Buchanan, Howe
   "Actuarial Practice of General Insurance"
18. Kastelijn
   "Solvency"

The authors and publishers of these books were contacted and responded enthusiastically to our project. All authors agreed to give up their royalties. Publishers provided discounts as high as 75%. The books ranked 2, 3, 4, 14, 16, and 18, that were out of print, are going to be specially reprinted, as an unexpected byproduct of our action.

The task force is in the process of putting together a list of recipient universities and national actuarial associations. While the list is not yet complete, recipients have already been identified in Albania, Belarus, Bulgaria, China, Croatia, the Czech Republic, Estonia, Egypt, Hungary, Georgia, Ghana, India, Jamaica, Korea, Latvia, Lithuania, Malaysia, Mexico, Poland, Romania, Russia, Slovakia, Slovenia, South Africa, Taiwan, Thailand, Turkey, the Ukraine, and Vietnam. It is expected that the books will be forwarded early in 1997.

The ASTIN Committee wishes to express its most sincere thanks to all members who contributed to this exciting project, with special thanks to Chris Daykin.

JEAN LEMAIRE
The College of Business Administration, Georgia State University, is seeking to hire a Director of its Actuarial Science Program. The position, which resides in the Department of Risk Management & Insurance, is available at any rank beginning in either June or September 1997. Candidates should have an earned doctorate or a master’s degree and a fellowship designation from one of the leading actuarial societies. Candidates nearing fellowship will also be considered.

The program director must have a strong commitment to students, teaching, program management and development, and external relations with the professional actuarial community. Salary and rank are dependent upon qualifications and experience.

Send current resume and three references to Dr. Bruce A. Palmer, Chair, Department of Risk Management and Insurance, College of Business Administration, Georgia State University, P.O. Box 4036, Atlanta, GA 30302-4036.

Preference will be given to applications received by JANUARY 15, 1997.

AN EQUAL EDUCATIONAL AND EMPLOYMENT OPPORTUNITY EMPLOYER.
GUIDELINES TO AUTHORS

1. Papers for publication should be sent in quadruplicate to one of the Editors:
   Paul Embrechts,
   Department of Mathematics, ETH-Zentrum,
   CH-8092 Zurich, Switzerland.
   or to one of the Co-Editors:
   René Schnieper,
   Zurich Insurance Company,
   P.O. Box, CH-8022 Zurich, Switzerland.
   D. Harry Reid,
   Eagle Star Insurance Company Ltd,
   The Grange, Bishop's Cleeve
   Cheltenham Glos GL52 4XX, United Kingdom.
   David Wilkie
   Watson Wyatt
   Watson House, London Rd.,
   Reigate, Surrey RH2 9PQ, United Kingdom.

   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. Manuscripts should be typewritten on one side of the paper, double-spaced with wide margins. 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. Upon acceptance of a paper, any figures should be drawn in black ink on white paper in a form suitable for photographic reproduction with lettering of uniform size and sufficiently large to be legible when reduced to the final size.

7. References should be arranged alphabetically, and for the same author chronological. 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:

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

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

*Continued overleaf.*
Committee of ASTIN

Hans Bühlmann Switzerland Honorary Chairman
James N. Stanard USA Chairman
Jukka Rantala Finland Vice-Chairman
Bouke Posthuma Netherlands Secretary
Jean Lemaire Belgium/USA Treasurer/IAA-Delegate
Paul Embrechts Switzerland Editor
D. Harry Reid United Kingdom Editor
James MacGinnitie USA Member/IAA-Delegate
Björn Ajne Sweden Member
Edward J. Levay Israel Member
Charles Levi France Member
Thomas Mack Germany Member
Ermanno Pitacco Italy Member
Gregory C. Taylor Australia Member
René Schnieper Switzerland Co-Editor
David Wilkie United Kingdom Co-Editor

Neither the Committee of ASTIN nor Ceuterick 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. 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.)

10. Electronic Typesetting using Word Perfect 5.1 is available. Authors who wish to use this possibility should ask one of the editors for detailed instructions.

11. Authors will receive from the publisher two sets of page proofs together with the manuscript. One corrected set of proofs plus the manuscript should be returned to the publisher within one week. Authors may be charged for alterations to the original manuscript.

12. 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.