A Journal of the International Actuarial Association

CONTENTS

EDITORIAL AND ANNOUNCEMENTS

Editorial

ARTICLES

A. J. G. Cairns
The Present Value of a Series of Cashflows: Convergence in a Random Environment

W. Neuhaus
Community Rating and Equalisation

O. Hesselager
Modelling of Discretized Claim Numbers in Loss Reserving

R. Schnieper
On the Estimation of the Credibility Factor: A Bayesian Approach

D. C. M. Dickson, A. D. Egídio dos Reis, H. R. Waters
Some Stable Algorithms in Ruin Theory and Their Applications

WORKSHOP

S. Haberman
Pension Funding with Time Delays and the Optimal Spread Period

SHORT CONTRIBUTION

H. U. Gerber
A Teacher's Remark on Exact Credibility

Book Review

Actuarial Vacancy

Ceuterick
EDITORIAL POLICY

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

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 © 1995 Ceuterick
EDITORIAL

WHAT DO THE INITIALS ASTIN ACTUALLY STAND FOR?

This question constituted a competition at a recently held actuarial conference. From the about 150 entries let me mention

Au Secours, mon Tarif est INutilisable

and the winning entry

Accountancy Seems Tremendously Interesting Now.

Though the above competition was obviously a tongue in cheek one, the basic question still remains. Other actuaries have used these pages before in trying to give a sensible answer, see for instance the guest editorial Arne (1990) or the editorials Lemaire (1987) and Reid (1988). I urge the reader to look once more at these interesting publications. The best summary of their arguments is still to be found in the Editorial Policy on the opposite page. ASTIN, as a section of the IAA, and the ASTIN BULLETIN do indeed stand for a forum for any quantitative point of view attacking theoretical and applied problems in any field faced with elements of insurance and risk. Those who wrote these words were wise. Indeed we need such a broad interpretation especially now where deregulation rules, finance and insurance are finding exciting ways in which to combine complementary expert knowledge from both sides and even our own actuarial societies are undergoing basic structural changes more and more blurring once sharp boundaries between countries (in education say) or specific actuarial activities (life versus non-life, asset versus liability modelling for instance).

I personally find it a great challenge to help in shaping the BULLETIN as a top journal in insurance. But let me stress from the start, ASTIN BULLETIN is a society journal. We, the members of ASTIN, shape its content, we are solely responsible for its quality, we all take on the duty in presenting to the world at large a mirror image of the state-of-the-art in actuarial modelling. My answer to the question asked in the title of this Editorial would therefore be: “ASTIN stands for the finest in actuarial modelling”. I personally strongly believe that we should move away from interpreting ASTIN solely as an acronym for Actuarial STudies In Non-life insurance, and indeed continue on the road towards making the ASTIN BULLETIN the finest in the general and much wider field of the mathematics of insurance and its application. As new incoming Editor I very much rely on you, the reader, to make sure that this ultimate goal is reached. As from the next issue, René Schnieper and myself will take over as Co-Editor, respectively Editor in succession of Alois Gisler and Hans Bühlmann. The reason that I am at all able to formulate the high goals of quality set out above is entirely due to their work together with the other Editor, Co-Editor and Editorial Board, and therefore

Many Thanks!

Indeed many thanks in the name of the actuarial profession at large, and the ASTIN membership more in particular to Hans Bühlmann and Alois Gisler for their

invaluable contribution to the well-being of our journal. I find it an honour to have been found fit to replace Hans Bühlmann as Editor. As one member of the ASTIN Committee rightly put it to me: “These are some very big shoes to step into!” That a “Swiss duo” has been proposed to replace Hans Bühlmann and Alois Gisler only stresses once more their excellent handling of affairs.

Finally, I also would like to say thanks to those remaining on board. As much as the shaping of the BULLETIN was teamwork in the past, it will stay like that in the future. As I already stressed above, every society journal thrives on the unselfish input of many. That I am able to start working with such a fine group of colleagues makes me proud. I hope that I will fulfill their expectations.

Many more words of thanks could have been added. I know however that the greatest gift we can give to Hans Bühlmann and Alois Gisler will be the quality of forthcoming ASTIN issues: let us therefore do our utmost to please them!

Paul Embrechts
ETH-Zürich

REFERENCES
THE PRESENT VALUE OF A SERIES OF CASHFLOWS: CONVERGENCE IN A RANDOM ENVIRONMENT

ANDREW J.G. CAIRNS

Heriot-Watt University, Edinburgh, UK

ABSTRACT

The present paper considers the present value, \( Z(t) \), of a series of cashflows up to some time \( t \). More specifically, the cashflows and the interest rate process will often be stochastic and not necessarily independent of one another or through time. We discuss under what circumstances \( Z(t) \) will converge almost surely to some finite value as \( t \to \infty \). This problem has previously been considered by DUFRESNE (1990) who provided a sufficient condition for almost sure convergence of \( Z(t) \) (the Root Test) and then proceeded to consider some specific examples of such processes. Here, we develop Dufresne's work and show that the sufficient condition for convergence can be proved to hold for quite a general class of model which includes the growing number of Office Models with stochastic cashflows.

KEYWORDS

Stochastic discounting; cashflow models; almost sure convergence; office model.

1. INTRODUCTION

Suppose \( \delta(t) \) is the constant force of interest during the period \([t-1, t)\), so that an investment of 1 at time \( t-1 \) will accumulate to \( \exp \delta(t) \) at time \( t \). The present value at time 0 of 1 due at time \( t \) is then

\[
V(t) = \exp \left( - \sum_{s=1}^{t} \delta(s) \right) = \prod_{s=1}^{t} \nu_s
\]

where \( \nu_s = \exp (-\delta(s)) \) is the discount factor for year \( s \).

The present value of a series of cashflows \( C(1) \) at time 1, \( C(2) \) at time 2, \ldots, \( C(t) \) at time \( t \) is therefore

\[
Z(t) = \sum_{s=1}^{t} V(s) C(s) = \sum_{s=1}^{t} \nu_1 \ldots \nu_s C(s)
\]
Such a process has been considered by Dufresne (1990) and Aebi et al. (1994) from the financial point of view and by Vervaat (1979) and Brandt (1986) from the mathematical point of view. All of these works consider the special case where \( \{\delta(t)\}_{t=1}^{\infty} \) and \( \{C(t)\}_{t=1}^{\infty} \) are independent and identically distributed and independent of one another. Dufresne (1990) considers the convergence of \( Z(t) \) as \( t \) tends to infinity and its limiting distribution when the distributions of \( \delta(t) \) and \( C(t) \) are known. Aebi et al. (1994) show how Bootstrap methodology can be used to estimate the limiting distribution when a limited number of past observations of \( \delta(t) \) and \( C(t) \) are available.

Dufresne (1990) also considers more general models and provides sufficient conditions for almost sure convergence of \( Z(t) \). In this paper we consider a number of specific examples for the process \( Z(t) \) and it is demonstrated that Dufresne's conditions hold for quite a wide class of models.

In this paper we will restrict ourselves to discrete time models. However, the results described here also hold for the continuous time models for \( \delta(t) \) described by Parker (1993, 1994d) and Norberg and Möller (1994).

De Schepper, Tuenen and Goovaerts (1994) consider the present value of annuities and of a perpetuity payable continuously. Using Laplace transforms they show that the perpetuity has an inverse Gamma distribution, matching the results of Dufresne (1990).

2. CONDITIONS FOR CONVERGENCE OF \( Z(t) \)

The principal result provided by Dufresne (1990) giving a sufficient condition for the almost sure convergence of \( Z(t) \) is the Root Test:

**Theorem 1** (Root Test, for example, see Dufresne, 1990)

If

\[
\lim_{t \to \infty} \sup_{t} \left| V(t) C(t) \right|^{1/t} < 1 \text{ almost surely}
\]

then \( Z(t) \) converges almost surely to some finite limit as \( t \) tends to infinity.

Now, trivially, this is equivalent to the condition

\[
\lim_{t \to \infty} \sup_{t} \left\{ \log \left| C(t) \right| + \sum_{s=1}^{t} \log \nu_{s} \right\} < 0 \text{ almost surely}
\]

\[
\Rightarrow \lim_{t \to \infty} \sup_{t} \left\{ \log \left| C(t) \right| - \sum_{s=1}^{t} \delta(s) \right\} < 0 \text{ almost surely}
\]

We therefore have the following
Corollary 2

If the force of interest process \( \{ \delta(t) \}_{t=1}^{\infty} \) is ergodic with

\[
\lim_{t \to \infty} \frac{1}{t} \sum_{s=1}^{t} \delta(s) = E[\delta(.)]
\]

almost surely, and if

\[
\limsup_{t \to \infty} \frac{1}{t} \log |C(t)| = \rho \in [-\infty, +\infty)
\]

where \( \rho - \delta < 0 \) then \( Z(t) \) converges almost surely to some finite limit as \( t \) tends to infinity.

[The condition that \( \rho < \delta \) means that the cashflow process, \( C(t) \), must grow more slowly than the accumulation process \( 1/V(t) \). Consider the trivial case of a perpetuity where \( C(t) = \exp \rho t \) and \( V(t) = \exp -\delta t \) are both deterministic. Then \( Z(t) \) converges if and only if \( \rho < \delta \).]

Corollary 2 extends Dufresne's subsequent development by allowing the process \( C(t) \):

- to depend on the force of interest process;
- to be non-ergodic and, in particular, to include inflationary growth and growth in the underlying number of policies.

[A rigorous definition of ergodicity is given by KARLIN and TAYLOR (1975). If a process \( X(t) \) is known to be ergodic then the following results hold:

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} X(t) = \mu
\]

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} [X(t) - \bar{X}_n]^2 = \sigma^2 \text{ where } \bar{X}_n = \frac{1}{n} \sum_{t=1}^{n} X(t)
\]

and if \( I_{a,b}(x) = \begin{cases} 1 & \text{if } a < x \leq b \\ 0 & \text{otherwise} \end{cases} \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} I_{a,b}[X(t)] = F(b) - F(a)
\]

where \( \mu \) and \( \sigma^2 \) are the unconditional mean and variance of \( X(t) \) and \( F(x) \) is the unconditional cumulative distribution function of \( X(t) \).]

3. The Interest Rate Process

Before we concentrate on the cashflow process, it is worth discussing briefly some interest rate processes.

The ergodic condition for the force of interest process is not particularly onerous, and encompasses most of the widely used stochastic investment models.
— Independent and identically distributed returns: for example, WATERS (1978), DUFRESNE (1990), PAPACHRISTOU and WATERS (1991), PARKER (1993, 1994a) and AEBI et al. (1994) give but a few examples.

— Simple autoregressive models for the rate of return, such as the AR(1) time series model, and the Ornstein-Uhlenbeck process: for example, DHAENE (1989, 1992), PARKER (1993, 1994a,b,c,d) and NORBERG and MOLLER (1994).


— Autoregressive Conditionally Heteroscedastic (ARCH) models: for example, see WILKIE (1995, Appendix D).

Some non-ergodic models may still admit convergence but it is worth discussing some special cases which may be considered to be inappropriate.

First, DUFRESNE (1990, Proposition 3.2.1) shows that $Z(t)$ will converge if $\{v_s\}$ is stationary and ergodic and $Pr(v_s = 0) > 0$. This second condition is equivalent to $Pr(\delta(s) = \infty) > 0$ which does not seem appropriate.

Second, some authors (for example, DHAENE, 1989, 1992; DUFRESNE, 1990, and PARKER, 1993) consider non-stationary models for the force of interest process. One of two things may happen.

— The process may have positive drift, so that $\delta(t)$ tends to infinity almost surely (again unrealistic).

— The process may have no drift but unbounded variance. Examples of this are random walk models of the form $\delta(t+1) = \delta(t) + \varepsilon(t+1)$ where the $\varepsilon$ are independent and identically distributed zero mean random variables, and, more generally, $ARIMA(p, d, q)$ models ($d \geq 1$). With such processes it is often easy to show (for example, see DUFRESNE, 1990, Proposition 4.4.4) that

$$\lim_{t \to \infty} \sup_{t} |Z(t)| = \infty.$$

Figures 1 and 2 demonstrate the problem. For many years $\delta(t)$ may remain positive (Figure 1). The process $Z(t)$ (Figure 2) may then give one the impression that it is converging and that it is safe to use $Z(50)$ or $Z(100)$, say, as an approximation to $Z(\infty)$. After a while, though, $\delta(t)$ takes a long excursion below zero and $Z(t)$ shoots off upwards.

4. THE CASHFLOW PROCESS

Corollary 2 provided us with a condition for the behaviour of $C(t)$ as $t$ tends to infinity. We now consider this in more detail and provide the following lemma which will allow us to satisfy the conditions in Corollary 2.
Figure 1. Sample path of a random walk interest rate process, $\delta(t)$.

Figure 2. $Z(t)$ appears to converge initially but eventually diverges.
Lemma 3

Suppose that there exists a deterministic sequence $a(t)$ converging to zero as $t$ tends to infinity such that

$$\sum_{t=1}^{\infty} Pr \left( \frac{1}{t} \log |C(t)| \geq \rho + a(t) \right) < \infty.$$ 

Then

$$\limsup_{t \to \infty} \frac{1}{t} \log |C(t)| \leq \rho \text{ almost surely.}$$

Proof

Let $E_t = \{t^{-1} \log |C(t)| \geq \rho + a(t)\}$. Then

$$\sum_{t=1}^{\infty} Pr(E_t) < \infty$$

$\Rightarrow Pr(\limsup_{t \to \infty} E_t) = 0$ by the first Borel-Cantelli Lemma (for example, see Williams, 1991, Section 2.7)

$\Rightarrow \limsup_{t \to \infty} \frac{1}{t} \log |C(t)| \leq \rho \text{ almost surely.}$

Lemma 3 provides quite a weak condition on the cashflow process: cashflows need not be independent; and the tails of the cashflow distributions can be quite fat. In particular if we suppose that $\mu(t) = E |C(t)|$ then we are able to prove the following theorem which provides us with a relatively easy method for proving the convergence of $t^{-1} \log |C(t)|$.

Theorem 4

Let $\rho_0 = \inf \{ \rho : \limsup_{t \to \infty} e^{-\rho t} \mu(t) < \infty \}$. Then

$$\limsup_{t \to \infty} \frac{1}{t} \log |C(t)| \leq \rho_0 \text{ almost surely.}$$

Proof See Appendix A.

This result covers many cases, some discussed previously by other authors:
- $C(t)$ independent and identically distributed (Dufresne, 1990; Aebi, et al., 1994);
- $C(t)$ an ergodic stochastic process and hence with $\mu(t)$ constant, giving $\rho_0 = 0$ (Dufresne, 1990);
- closed funds (for which $\mu(t)$ tends to zero in finite time) (Papachristou and Waters, 1991; Parker, 1993, 1994a,b,c,d; Frees, 1990);
deterministic processes with exponentially bounded growth (for example, an office model with an assumed new business growth rate).

Theorem 4 gives a stronger condition for convergence than Lemma 3. For example, suppose $C(t)$ has a Log-Pareto distribution with drift: that is,

$$Pr \left[ \log C(t) - \alpha t > x \right] = \frac{\text{constant}}{x^\delta} \quad \text{for } x > x_0 \text{ and } \delta > 0.$$

Then $E \left| C(t) \right| = \infty$ for all $t$ so that $\{C(t)\}$ does not satisfy the conditions for Theorem 4. Nevertheless, $\{C(t)\}$ does satisfy the conditions for Lemma 3, implying that $\lim sup_{t \to \infty} t^{-1} \log \left| C(t) \right|$ is still less than infinity. Theorem 4 does, however, provide us with a condition for convergence which is often easier to verify, as will be demonstrated in the next section.

5. A STOCHASTIC OFFICE MODEL

We now develop the last of these examples to include office models in which cashflows are stochastic. The office's portfolio is assumed to consist only of policies which do not participate in the profits of the company. The model described includes stochastic mortality, stochastic growth of new business volumes, stochastic inflation in the size of individual policies and conditionally independent and identically distributed policy sizes at a given time of inception.

The generality of the model, here, means that the notation may appear to be quite heavy going, but the reader should concentrate on:

— the total premium income at time $t$, $P(t)$;
— the total benefit outgo at time $t$, $B(t)$, which consists of benefits payable on death during the year $(t-1, t]$, and on survival to duration $t$ for $t = 1, 2, \ldots$.

Suppose that $A = \{\lambda(t)\}_{t=-\infty}^{\infty}$, $\lambda(t) = (\lambda_1(t), \lambda_2(t))$, is the process with determines the volume of new business ($\lambda_1(t)$) and the individual policy size index ($\lambda_2(t)$).

Let $(x)_{t,j}$ represent the life corresponding to policy $j$ taken out at time $t$. All policyholders are aged $x$ at entry. Then, using standard notation, we have

$sP_x = \text{probability that an individual now aged } x \text{ survives to age } x + s$
and $s-1 q_x = \text{the probability that an individual now aged } x \text{ dies between ages } x + s - 1 \text{ and } x + s$

First we consider the total premium income at time $t$. This is

$$P(t) = \sum_{s=0}^{\infty} \sum_{j=1}^{N(t-s)} K^p_s A_{t-s}(j) I^S_{t-s,x}(j)$$

$$= P_1(t) + P_2(t)$$

where $P_1(t) = \sum_{s=t}^{\infty} \sum_{j=1}^{N(t-s)} K^p_s A_{t-s}(j) I^S_{t-s,x}(j)$

= premium income in respect of policies issued at or before time $0$
\[ P_2(t) = \sum_{s=0}^{t-1} \sum_{j=1}^{N(t-s)} K_s^P A_{t-s}(j) I_{t-s,s}^S(j) \]

- premium income in respect of policies issued after time 0
- number of new policies taken out at time \( s \)
- \( N(s) \sim Po(\lambda_1(s)), (s > 0) \) and, given \( A, N(1), N(2), \ldots \) are independent
- \( K_s^P \) = premium at duration \( s \) per unit of benefit
- \( A_i(j) \) = number of units of benefit for policy \( j \) taken out at time \( t \)
- \( F_i(x) = Pr(A_i(j) < x) \)

Given these assumptions we can then say that \( P_2(t) \) has the Compound Poisson distribution

\[ P_2(t) \sim CPo(A_P(t), F_i^P) \]

where \( A_P(t) = \sum_{s=0}^{t-1} \lambda_1(t-s)s \)

\[ F_i^P(x) = \frac{1}{A_P(t)} \sum_{s=0}^{t-1} \lambda_1(t-s)s F_i(x/K_s^P) \]

If we assume that the history of the office is known up to and including time 0 then \( P_1(t) \) is subject to much less uncertainty because the numbers and sizes of the existing policies are known. In any event \( P_1(t) \) is equal to zero for all \( t \geq \omega - x \).

Turning now to the benefits process, \( B(t) \), we can proceed in a similar, but slightly more complex, way:

\[ B(t) = \sum_{s=1}^{N(t-s)} \sum_{j=1}^{N(s)} \{ K_s^S A_{t-s}(j) I_{t-s,s}^S(j) + K_s^D A_{t-s}(j) I_{t-s,-s}(j) \} \]

\[ B_1(t) = B_1(t) + B_2(t) \]

where \( B_1(t) \)

\[ B_2(t) = \sum_{s=1}^{N(t-s)} \sum_{j=1}^{N(s)} \{ K_s^S A_{t-s}(j) I_{t-s,s}^S(j) + K_s^D A_{t-s}(j) I_{t-s,,-s}(j) \} \]
THE PRESENT VALUE OF A SERIES OF CASHFLOWS

\( K^s_i \) = amount payable on survival to duration \( s \) per unit benefit

\( K^D_i \) = amount payable on death during the year \( (s-1, s] \) per unit benefit

\( I^s_{t-s, s}(j) = \begin{cases} 
1 & \text{if } (x)_{t-s, j} \text{ dies during the year } (t-1, t] \\
0 & \text{otherwise}
\end{cases} \)

and note that

\( I^s_{t-s, s}(j) + I^D_{t-s, s}(j) = \begin{cases} 
1 & \text{if } (x)_{t-s, j} \text{ is alive at } t-1 \\
0 & \text{otherwise}
\end{cases} \)

As with the premium income we can then see that \( B_2(t) \mid A \) has the Compound Poisson distribution

\( B_2(t) \mid A \sim CPo(A_B(t), F^B_t) \)

where \( A_B(t) = \sum_{s=1}^{t-1} \lambda_1(t-s) p_s \)

\( F^B_t(x) = \frac{1}{A_B(t)} \sum_{s=1}^{t-1} \lambda_1(t-s) \{ p_s F_{t-s}(x/K^s_i) + s q_s F_{t-s}(x/K^D_i) \} \)

We can also say that \( B_1(t) = 0 \) for all \( t > \omega - x \).

Now suppose that \( m_1(t) = E[A_1(j)] \). Then for \( t > \omega - x \)

\[ \mu(t) = E(B(t)) = E[E(B(t) \mid A)] = E \left[ \sum_{s=1}^{t-1} \lambda_1(t-s) m_1(t-s) (p_s K^s_i + s q_s) \right] = \sum_{s=1}^{\omega - x} E[\lambda_1(t-s) m_1(t-s)] k_1(s) \]

where \( k_1(s) = p_s K^s_i + s q_s K^D_i \)

To make further progress we need to make further assumptions about the claim size distributions and in the model for new business growth.

Suppose then we assume that \( F_s(x) = F(x/\lambda_2(t)) \): that is, \( A_1(j)/\lambda_2(t) \) and \( A_1(k)/\lambda_2(s) \) are identically distributed when \( s \neq t \) and where \( \lambda_2(t) \) represents the benefit inflation index. Then \( m_1(t) = \lambda_2(t) m_1 \).

\[ \Rightarrow \mu(t) = \sum_{s=1}^{\omega - x} m_1 k_1(s) E[\lambda_1(t-s) \lambda_2(t-s)] \]

Suppose also that

\[ \lambda_1(t) = \lambda_1 \exp [\rho_1 t + \sigma_1 W_1(t)] \]

\[ \lambda_2(t) = \lambda_2 \exp [\rho_2 t + \sigma_2 W_2(t)] \]
where $W_1(t)$ and $W_2(t)$ are independent standard Brownian motions. In particular, they have the properties for $i = 1, 2$:

- $W_i(0) = 0$;
- if $t_1 < t_2$ then $W_i(t_2) - W_i(t_1) \sim N(0, t_2 - t_1)$;
- if $t_1 < t_2 \leq t_3 < t_4$ then $W_i(t_2) - W_i(t_1)$ and $W_i(t_4) - W_i(t_3)$ are independent.

For $t < 0$, $W_1(t)$ and $W_2(t)$ are known.

Then

$$E[\lambda_1(t) \lambda_2(t)] = E[\lambda_1(t)] E[\lambda_2(t)]$$

$$= \lambda_1 \lambda_2 \exp \left[ \left( \rho_1 + \rho_2 + \frac{1}{2} \sigma_1^2 + \frac{1}{2} \sigma_2^2 \right) t \right]$$

$$= \lambda_1 \lambda_2 \exp \left[ \rho t \right]$$

where $\rho = \rho_1 + \rho_2 + \frac{1}{2} \sigma_1^2 + \frac{1}{2} \sigma_2^2$.

We therefore have

$$\mu(t) = \lambda_1 \lambda_2 \sum_{s=1}^{\infty} m_1 k_1(s) \exp \left[ \rho (t - s) \right]$$

$$= \lambda_1 \lambda_2 \exp (\rho t) \sum_{s=1}^{\infty} m_1 k_1(s) \exp (-\rho s)$$

$$= \mu \exp (\rho t)$$

where $\mu = \lambda_1 \lambda_2 \sum_{s=1}^{\infty} m_1 k_1(s) \exp (-\rho s)$.

Hence, by Theorem 4 we deduce that

$$\limsup_{t \to \infty} - \frac{1}{t} \log |B(t)| \leq \rho \text{ almost surely.}$$

We can prove similarly that

$$\limsup_{t \to \infty} - \frac{1}{t} \log |P(t)| \leq \rho.$$  

Hence, if $\rho - \delta < 0$ then, by Corollary 2

$$Z(t) = \sum_{s=1}^{t} V(s) C(s) = \sum_{s=1}^{t} V(s) P(s) - \sum_{s=1}^{t} V(s) B(s)$$

converges almost surely to some finite limit as $t$ tends to infinity.
6. DISCUSSION

It is possible without great difficulty to relax many of the assumptions made in Section 5.

— Here we assumed that all new entrants will be of the same age in an effort to contain the already complex notation. Relaxing this to include a spread of ages will result in a sum of conditionally independent Compound Poisson processes (given $A$) which is itself a Compound Poisson process.

— Similarly we could allow for more than one policy type, multiple state models (for example, permanent health insurance) and more than one risk group.

— Other forms of distribution for $N(t) | \lambda_1(t)$ would provide similar results. The Poisson assumption was made here for the convenience of its additive properties.

— Inclusion of expenses and reserves. (However, if we discount cashflows at the same rate of interest as that earned on the reserves, then the limiting value of $Z(t)$ will be unchanged.)

It should, therefore, follow that converge can be shown to occur for a wide range of office models, beyond the already general case described here.

Suppose that $N(t) = 0$ for $t \leq 0$. Then, in the context of Section 5, $\lim_{t \to \infty} Z(t)$ represents the present value of profit on future new business. The present paper has shown that, subject to certain conditions, this quantity is well defined and exists almost surely. It is a quantity which is of genuine practical interest since it allows actuaries to assess the underlying value of a company.

It is unlikely that the limiting distribution or the moments of $Z(t)$ will be known under the majority of circumstances. (However, where the cashflows in different years are independent, the methods of Parker, 1993, 1994a,b,c can be used to find the distribution of $Z(t)$ for $t < \infty$. This then provides, for large $t$, an approximation to the limiting distribution of $Z(t)$.) It will often, therefore, be necessary to carry out a Monte-Carlo study, simulating sample paths of $Z(t)$. The results described in this paper indicate that $Z(t)$ will converge to its limit at least as fast as the deterministic annuity function $\bar{a}_{\gamma}$ with force of interest $\delta - \rho$ tends to its limiting value. This gives us a useful guide as to when the difference between $Z(t)$ and its limit falls within the maximum tolerable level of error.

Some idea of the limiting distribution of $Z(t)$ can be obtained by applying the results of Papachristou and Waters (1991) and Frees (1990) for large portfolios. The analogue here is that the distribution of $\lambda_1^{-1} Z(t)$ tends to that of $\lambda_1^{-1} \mathbb{E}[Z(t) | \delta(s), \lambda_1(s), \lambda_2(s), = 1, 2, ..., t]$ as $\lambda_1$ tends to infinity, and similarly the distribution of the limit of $Z(t)$ where this exists almost surely.

ACKNOWLEDGEMENTS

I would particularly like to thank an anonymous referee for helping me to improve Lemma 3 and Theorem 4 which led to a strengthening of the results contained in this paper. I would also like to thank Gary Parker, Angus Macdonald and Howard Waters for the helpful discussions they provided during the preparation of this paper.
APPENDIX A

We make use of the following Corollary to Lemma 3.

**Corollary A.1**

*If there exists a deterministic sequence \( a(t) \) converging to zero as \( t \) tends to infinity such that \( \sum_{t=1}^{\infty} \exp \left[ -pt - ta(t) \right] E \left| c(t) \right| < \infty \) then*

\[
\limsup_{t \to \infty} t^{-1} \log \left| C(t) \right| \leq \rho \text{ almost surely.}
\]

**Proof**

As in Lemma 3 let

\[
E_t = \left\{ \frac{1}{t} \log \left| C(t) \right| \geq \rho + a(t) \right\}
\]

\[
= \left\{ \left| C(t) \right| \geq \exp \left[ pt + ta(t) \right] \right\}
\]

Now \( E \left| C(t) \right| \geq cPr\left[ \left| C(t) \right| \geq c \right] \) for any \( c > 0 \) (for example, see WILLIAMS, 1991, Section 6.4). Hence

\[
E \left| C(t) \right| \geq \exp \left[ pt + ta(t) \right] Pr(E_t)
\]

\[
\Rightarrow \sum_{t=1}^{\infty} Pr(E_t) \leq \sum_{t=1}^{\infty} \exp \left[ -pt - ta(t) \right] E \left| C(t) \right| < \infty.
\]

This is the condition in Lemma 3.

**Proof of Theorem 4**

Take any \( \rho_2 > \rho_0 \) and set \( a(t) = 0 \) in Corollary A.1. Choose any \( \rho_1 \) such that \( \rho_0 < \rho_1 < \rho_2 \) and let \( k = \sup \left\{ \exp \left( -\rho_1 t \right) \mu(t) : t \geq 1 \right\} \). Since \( \rho_1 > \rho_0 \), \( k \) must be finite. Then

\[
\sum_{t=1}^{\infty} \exp \left[ -\rho_2 t - ta(t) \right] E \left| C(t) \right| = \sum_{t=1}^{\infty} \exp \left[ -\rho_1 t \right] E \left| C(t) \right| \exp \left[ -(\rho_2 - \rho_1) t \right]
\]

\[
< \sum_{t=1}^{\infty} k \exp \left[ -(\rho_2 - \rho_1) t \right] < \infty.
\]

Hence \( \limsup_{t \to \infty} t^{-1} \log \left| C(t) \right| \leq \rho_2 \) almost surely.

This is true for all \( \rho_2 > \rho_0 \) so the result follows.
REFERENCES


BOYLE, P.P. (1980) Recent models of the term structure of interest rates with actuarial applications. 21st International Congress of Actuaries, Zurich and Lausanne 4, 95-104.


COMMUNITY RATING AND EQUALISATION

WALThER NEUHAUS

University of Copenhagen and MIRA Consultants Ltd. (Sydney)

ABSTRACT

Several countries have made community rating mandatory for certain lines of insurance, particularly health insurance. This paper offers a theoretical solution to the problem of designing equalisation schemes to support community rating in a market where different insurers are selling different benefit plans. The criterion chosen is that an equalisation scheme should minimise the opportunities for arbitrage between insurers, which community rating otherwise would generate. Several possible measures of arbitrage are presented, and the optimal schemes are compared against data from Australian health insurers. Finally, the approach is extended to partial community rating, for example unisex rating.

KEYWORDS

Arbitrage; Community rating; Credibility theory; Equalisation; Health insurance; Risk adjustment.

1. INTRODUCTION

Private insurers in Australia are required by law to practise community rating. They are not allowed to differentiate their contribution rates by the age, sex or any other criteria of the applicant (including ill-health). The only allowable discrimination is between single members and family members, the contribution rate for a family being twice the contribution rate for a single person. While community rating in this way imposes a strong constraint on the rating structure of private health insurers, there is no explicit regulation governing rate levels.

Two tiers of benefit plans are offered by private health insurers in Australia: the basic plan, which is determined by the Government and common to all insurers, and supplementary plans, which may differ between insurers. The basic plan provides for private accommodation in public hospitals as well as some medical expenses, while supplementary plans normally provide for accommodation in private hospitals. A large majority of the privately insured hold both the basic and a supplementary plan.

The reason given for mandatory community rating is one of social equity, i.e. the wish to ensure that private health insurance remains affordable for all who choose.
to buy it. All Australian residents are also covered by the public health care system (Medicare), the introduction of which has reduced the privately insured population significantly.

While it is debatable whether community rating actually achieves its goal (higher enrolment through universal affordability), there is general agreement that community rating can only work if enforced by legislation, or in a monopoly market. For a thorough discussion of community rating in voluntary health insurance, see MacIntyre (1962).

It is well known that mandatory community rating in a competitive market can lead to market instability and adverse selection against some insurers. Insurers with a large proportion of elderly persons are particularly disadvantaged under community rating, as the cost of providing health insurance increases rapidly with a person’s age. In Australia, the average drawing rate of persons aged over 65 is five times that of persons under 65. This is shown in Figure 1.1 (Health insurers refer to the expected benefit cost of a person as his/her drawing rate; actuaries commonly use the term pure premium).

![Cost of Basic Benefits 1991/92](image)

Figure 1.1

Several countries have implemented or are proposing to implement equalisation schemes which support community rating by cross-subsidies between insurers with a young age profile and insurers with an old age profile. De Wit and van Eeghen (1984) as well as Gregorius (1987) describe the situation in the Netherlands. In Australia, an equalisation scheme was introduced in 1989 and modified in 1995. The Department of Health of the Republic of Ireland (1994) has proposed an equalisation scheme to support community rating under the EC Third Directive on Non-Life Insurance.
Broadly speaking, the objective of all equalisation schemes is to level the playing field between insurers with different membership profiles, and to reduce the incentive for insurers to engage in "predatory marketing" or "cherry-picking".

Equalisation is straightforward in a market where all insurers offer just one standard benefit plan: all one has to do is to allocate to each insurer the global average cost per person, times the number of persons it has insured.

This paper provides a theoretical framework for the design of an equalisation scheme in a market where different insurers offer different benefit plans.

The concept of arbitrage is invoked to measure the degree of mismatch between the benefit plans and the community-rated contribution rates of different insurers. The optimal equalisation scheme is the one which minimises the possibility for arbitrage, suitably defined.

In section 2 we discuss the notion of equity (fairness) between insurers and introduce the concept of arbitrage. Section 3 presents a minimum arbitrage scheme which minimises arbitrage opportunities between an insurer and the market average. In section 4 we develop a more elaborate minimum arbitrage scheme, which minimises arbitrage opportunities between all health insurers. Section 5 presents a study of data from Australia; in that section we also compare the optimal schemes with a simplified variant of the scheme used in Australia since 1989.

Section 6 looks at an alternative measure of arbitrage. In section 7 we extend the theory to cover partial community rating, for example unisex rating. Some concluding remarks are given in section 8.

2. EQUITY BETWEEN HEALTH INSURERS

It seems self-evident that while an equalisation should equalise variations in cost which are the result of different membership profiles, it should not equalise variation in cost which can be ascribed to different levels of benefits. In what follows we justify this argument with reference to arbitrage. We will argue that an equalisation scheme provides equity (fairness) between health insurers if it eliminates, or at least minimises the incentive for arbitrage which community rating generates.

An opportunity for arbitrage between two insurers exists when there is a mismatch between the benefit levels they offer and the contribution rates they have to charge to support their benefit levels. Let us consider some examples.

If two insurers differ in their membership profile but offer the same level of benefits, differences in their cost will reflect the difference in membership profiles; the insurer with the highest proportion of elderly members will have the highest cost. If there is no equalisation, the difference in cost will have to be reflected in different contribution rates. This would allow mobile members of the higher-cost insurer to switch to the lower-cost insurer and receive the same level of benefits for a smaller contribution. This is a form of arbitrage.

As an aside we note that the most mobile members tend to be the younger members. Thus, if arbitrage between two insurers occurs along the lines sketched above, then it will tend to increase the incentive for arbitrage as the difference in the age profile grows.
If two insurers have the same membership profile and different levels of benefit, their cost will reflect the differences in benefits paid and will have to be reflected in different contribution rates if there is no equalisation. Thus for a member of one insurer, switching to the other insurer would involve a trade-off between contributions and benefits and there is no opportunity for arbitrage. We conclude that different levels of benefits need not be equalised.

Indeed, equalising different levels of benefits can introduce an arbitrage opportunity, as can be seen from the following argument. If the cost of different benefit levels are equalised, two insurers with different benefit levels may be induced to charge similar contribution rates. This would allow mobile members of the insurer with lower benefits to "upgrade" to higher benefits without a commensurate increase in their contribution; again, a form of arbitrage.

As can be seen from the discussion above, community rating without any form of equalisation generates arbitrage opportunities by preventing insurers from charging new applicants their true pure premium. Even if each insurer at the outset had a perfectly balanced membership, aggressive marketing by some insurers can lead to an imbalance which may be self-reinforcing and destabilise the market.

Thus is can be argued that if a government imposes community rating on a competitive industry (health insurance or otherwise), it has an obligation to support community rating by some form of equalisation. We describe an equalisation scheme as fair if it eliminates, or at least minimises, the opportunities for arbitrage which mandatory community rating otherwise would generate.

The reader should note that in this paper, arbitrage is considered only in terms of the pure benefit cost faced by the insurers. For the insured, there may be arbitrage opportunities generated by other factors, for example different expense levels or investment strategies. There is no need to eliminate those arbitrage opportunities, as they are not a direct result of government policy.

3. MINIMISING ARBITRAGE AGAINST THE MARKET AVERAGE

In this section we develop a class of equalisation schemes which minimises the opportunity for arbitrage between an insurer and the market average. We begin by introducing some notation.

Denote the insurers operating in the market by \( i = 1, \ldots, I \). Let \( X \) denote a finite partition of the insured population into homogeneous risk classes. It is helpful to visualise \( X \) as a collection of age groups, but the partition may reflect other factors that affect cost as well. For an insurer \( i \in \{ 1, \ldots, I \} \) and a risk class \( x \in X \), we define the following quantities:

\[
\begin{align*}
n_i(x) & = \text{the number of person years covered during a given year;} \\
B_i(x) & = \text{the amount of benefits paid or incurred during the same period}.
\end{align*}
\]

Corresponding symbols without the argument \( x \) denote the corresponding quantity summed across all age groups. Likewise, corresponding symbols with the subscript \( i \) omitted denote the corresponding quantity summed across all insurers. Thus, for instance, \( n_l \) denotes the number of persons covered by insurer \( i \), while \( n(x) \) denotes the number of persons in risk class \( x \) in the entire insured population.
Denote by $d_i(x)$ the drawing rate incurred by insurer $i$ in insuring a person in class $x$, that is

\[(3.1) \quad \mathbb{E}B_i(x) = n_i(x) d_i(x). \]

The average drawing rate incurred by all insurers in relation to class $x$ is then

\[(3.2) \quad d(x) = \frac{1}{\mathbb{E}B_i(x)} \sum_{i=1}^{I} n_i(x) d_i(x). \]

The overall drawing rate incurred by insurer $i$ across all classes is

\[(3.3) \quad d_i = n_i^{-1} \sum_{x \in X} n_i(x) d_i(x), \]

and the overall drawing rate incurred by all insurers across all classes is

\[(3.4) \quad d = \sum_{i=1}^{I} \sum_{x \in X} n_i(x) d_i(x). \]

In the absence of equalisation, the contribution rates charged by insurer $i$ are essentially determined by $d_i$, while the average contribution rate of all insurers is essentially determined by $d$. We say that class $x \in X$ has an *arbitrage opportunity* between insurer $i$ and the market, if

\[(3.5) \quad d_i - d = d_i(x) - d(x). \]

Thus an arbitrage opportunity exists whenever there is a mismatch between the difference in contribution rates and the difference in benefits. In order to get an overall measure of arbitrage opportunities present in the market, we define the measure

\[(3.6) \quad Q(d_1, \ldots, d_I) = \sum_{i=1}^{I} \sum_{x \in X} v_i(x) (d_i - d - d_i(x) + d(x))^2, \]

where $\{v_i(x) \mid i = 1, \ldots, I; \ x \in X\}$ is a set of arbitrary, fixed non-negative weights.

Under an equalisation scheme there is a zero-sum reallocation of costs, and the overall drawing rate $d_i$ of insurer $i$ is replaced by a quantity $\tilde{d}_i$, the post-equalisation unit cost.

Thus in order to minimise arbitrage by equalisation, one must solve the constrained minimisation problem

\[(3.8) \quad \text{Minimise } \sum_{i=1}^{I} \sum_{x \in X} v_i(x) (\tilde{d}_i - d - d_i(x) + d(x))^2 \]

with respect to $\tilde{d}_1, \ldots, \tilde{d}_I$, subject to $\sum_{i=1}^{I} n_i \tilde{d}_i = nd$. 

The side condition ensures that the equalisation scheme is balanced, on average. We refer to the side condition as the balancing constraint.

**Theorem 3.1**

The minimum arbitrage scheme against the criterion (3.6) is given by the allocation

$$\tilde{d}_k = d + \Delta_k - \frac{n_k}{\nu_k} \left[ \sum_{i=1}^{l} \frac{n_i^2}{\nu_i} \right]^{-1} \sum_{i=1}^{l} n_i \Delta_i,$$

where $\nu_k = \sum_{s \in X} v_k(s)$ and

$$\Delta_i = \sum_{s \in X} \frac{v_i(s)}{\nu_i} (d_i(s) - d(x)).$$

**Proof:** Lagrange minimisation. One must determine the solution of the $l$ equations

$$\frac{\partial}{\partial \tilde{d}_k} \left[ \frac{1}{2} \sum_{i=1}^{l} \sum_{s \in X} v_i(s) (\tilde{d}_i - d - d_i(s) + d(x))^2 + \lambda \left( \sum_{i=1}^{l} n_i \tilde{d}_i - nd \right) \right] = 0, \forall k,$$

which turns out to be

$$\tilde{d}_k = d + \Delta_k - \frac{n_k}{\nu_k} \lambda.$$

The constraint $\sum_{i=1}^{l} n_i \tilde{d}_i = nd$ is then applied to yield

$$\lambda = \left[ \sum_{i=1}^{l} \frac{n_i^2}{\nu_i} \right]^{-1} \sum_{i=1}^{l} n_i \Delta_i.$$

This proves the theorem. QED

The last term in (3.9) is obviously a balancing correction, ensuring that the balancing constraint is observed.

**Corollary 3.2.**

In the special case where $v_i(s) = c \cdot n_i(s)$ with an arbitrary constant $c > 0$, we obtain the following minimum arbitrage allocation:

$$\tilde{d}_k = d + \sum_{s \in X} p_k(s) (d_k(s) - d(x)).$$
where \( p_k(x) = \frac{n_k(x)}{n_k} \) is the proportion of class \( x \) in the membership of insurer \( k \).

**Proof**: Simple substitution. Note that in this case, \( \lambda = 0 \). QED

**Remark 3.3.**

The choice \( v(x) = c \cdot n_i(x) \) obviously makes sense. It means that the squared arbitrage terms in (3.8) are weighted in proportion to the number of persons exposed to the arbitrage opportunity in question.

**Remark 3.4.**

In our discussion so far, we have discussed arbitrage only in terms of pure premiums, or expected values. In order to develop a fully operational equalisation scheme, one has to replace the drawing rates \( d_i(x) \) by suitable estimates; let us denote the estimates by \( \hat{d}_i^*(x) \). One would normally also want to balance the equalisation against the actual claims cost \( B \) of all insurers, rather than the expected claims cost \( nd \). This will be automatically the case if the estimates balance, i.e.

\[
nd^* = \sum_{i=1}^{I} \sum_{x \in X} n_i(x) \hat{d}_i^*(x) = B.
\]

 Neuhaus (1995) shows how credibility estimators can be corrected to balance.

Alternatively one could estimate each \( d_i(x) \) by the empirical drawing rate

\[
\hat{d}_i(x) := D_i(x) = B_i(x)/n_i(x).
\]

It is easy to verify that the estimators \( \hat{d}_i(x) \) balance. They have the additional advantage of being easy to explain to non-mathematical people. Inserting empirical drawing rates into the scheme defined in (3.14), one obtains the scheme

\[
\tilde{D}_k = D + \sum_{x \in X} p_k(x) (D_k(x) - D(x)),
\]

where \( D(x) = B(x)/n(x) \) is the average drawing rate of class \( x \) and \( D = B/n \) is the overall average drawing rate. The equalisation scheme defined by (3.17) is called a composition-based scheme in MIRA (1993, 1994).

Thomson (1994) suggested the use of robust estimation for the drawing rates. This is entirely possible, but a little more effort will be required to ensure that the estimates balance. Simple grossing up of the robust estimates with a constant factor is one option.

**Remark 3.5.**

The equalisation formula (3.17) can be derived heuristically without any reference to arbitrage. To see this, decompose the drawing rate of insurer \( k \) into three
components as follows:

\[ D_k = D \]

\[ + \sum_{x \in X} p_k(x) (D_k(x) - D(x)) \quad (=: \Delta_k(D)) \]

\[ + \sum_{x \in X} p_k(x) (p_k(x) - p(x)) \quad (=: \Delta_k(p)) \]

The three components can be called the average drawing rate \( D \), the benefit component \( \Delta_k(D) \) of insurer \( k \), and the profile component \( \Delta_k(p) \) of insurer \( k \). In keeping with the intuitively obvious notion that differences in profile should be equalised but not differences in benefit levels, one can eliminate the profile component and equalisation scheme (3.17) appears.

4. MINIMISING ARBITRAGE BETWEEN HEALTH INSURERS

While the measure of arbitrage presented in the previous section leads to a neat and simple formula for the minimum arbitrage scheme, the underlying assumption that arbitrage occurs only between an insurer and the market average may be unrealistic. In reality, people insured with one insurer are able to check out the contribution rates of any number of its competitors, and move their policy if there is a mismatch between the contribution rates and the benefits between their current insurer and any one of its competitors. In this section we develop an equalisation scheme which minimises the opportunity for inter-insurer arbitrage.

Denote by \( \tilde{d}_i \) the post-equalisation unit cost of an insurer \( i \), given any specified equalisation scheme. For two insurers \( i, j \in \{1, ..., l\} \), we say that people in risk class \( x \in X \) have an opportunity for arbitrage if

\[ \tilde{d}_i - \tilde{d}_j \neq d_i(x) - d_j(x). \]

For each \( x \in X \), define a symmetric matrix \( V(x) \) of fixed, non-negative weights,

\[ V(x) = \begin{bmatrix} v_{11}(x) & ... & v_{1l}(x) \\ ... & ... & ... \\ v_{l1}(x) & ... & v_{ll}(x) \end{bmatrix}. \]

In order to get an overall measure of arbitrage opportunities present in the market, we define the measure

\[ Q(\tilde{d}_1, ..., \tilde{d}_l) = \sum_{i=1}^{l} \sum_{j=1}^{l} \sum_{x \in X} v_{ij}(x) (\tilde{d}_i - \tilde{d}_j - d_i(x) + d_j(x))^2. \]

Note that the weighted sum (4.3) involves double counting of the off-diagonal terms, while the diagonal terms are zero. This has been done to simplify the subsequent algebraic bookkeeping. Double counting only changes the measure by a constant factor without affecting the optimal solution.
In order to minimise the opportunity for arbitrage by equalisation, one must solve the constrained minimisation problem

(4.4) Minimise \( \sum_{i=1}^{l} \sum_{j=1}^{l} \sum_{x \in X} v_{ij}(x) (\tilde{d}_i - \tilde{d}_j - d_i(x) + d_j(x))^2 \)

subject to \( \sum_{i=1}^{l} n_i \tilde{d}_i = nd. \)

The side condition ensures that the equalisation scheme is balanced, on average.

**Theorem 4.1**

The minimum arbitrage scheme against the criterion (4.3) is given by the set of linear equations

\[
\begin{align*}
\tilde{v}_k \tilde{d}_k = \sum_{i=1}^{l} v_{ki} \tilde{d}_i + \sum_{x \in X} v_k(x) d_k(x) - \sum_{x \in X} \sum_{i=1}^{l} v_{ki} d_i(x) \\
\text{for } k = 1, \ldots, l, \text{ and }
\end{align*}
\]

\( nd = \sum_{i=1}^{l} n_i \tilde{d}_i, \)

where

\[
\begin{align*}
\tilde{v}_k &= \sum_{i=1}^{l} \sum_{x \in X} v_{ki}(x), \\
v_k &= \sum_{x \in X} v_k(x), \\
v_{ki} &= \sum_{x \in X} v_{ki}(x), \\
v_k(x) &= \sum_{i=1}^{l} v_k(x).
\end{align*}
\]

**Proof:** Lagrange minimisation. Equating the partial derivatives of

\[
\frac{1}{4} Q(\tilde{d}_1, \ldots, \tilde{d}_l) + \lambda \left( \sum_{i=1}^{l} n_i \tilde{d}_i - nd \right)
\]

with respect to \( \tilde{d}_1, \ldots, \tilde{d}_l \) to zero yields

\[
\begin{align*}
\tilde{v}_k \tilde{d}_k = \sum_{i=1}^{l} v_{ki} \tilde{d}_i + \sum_{x \in X} v_k(x) d_k(x) - \sum_{x \in X} \sum_{i=1}^{l} v_{ki} d_i(x) - \lambda n_k,
\end{align*}
\]

for \( k = 1, \ldots, l. \) Summing (4.7) across all values of \( k \) yields \( \lambda = 0. \) Thus the balancing constraint cannot be effectively eliminated and we have to be content with the implicit formula (4.5). QED
Corollary 4.2.

If the weighting matrix $V = \sum_{x \in X} V(x)$ satisfies the condition

\begin{equation}
V_{ij} = c \cdot n_i n_j
\end{equation}

with an arbitrary constant $c > 0$, then the minimum arbitrage scheme against the criterion (4.3) has an explicit solution, which is

\begin{equation}
\tilde{d}_k = d + \sum_{x \in X} \frac{v_k(x)}{v_k} d_k(x) - \sum_{x \in X} \sum_{i=1}^l \frac{v_{ki}(x)}{v_k} d_i(x).
\end{equation}

**Proof:** Under the condition of the corollary, the balancing constraint implies

\begin{equation}
\sum_{i=1}^l v_{ki} \tilde{d}_i = c \cdot n_k \sum_{i=1}^l n_i \tilde{d}_i = c \cdot n_k n d = v_k d.
\end{equation}

QED

The condition of corollary 4.2 is sensible. In using weights of the form (4.8) the aggregate weight one assigns to arbitrage in (4.3) between any two insurers is proportional to the product of their market shares; one could also say that for any given insurer $i$, the seriousness of arbitrage against another insurer $j$ is proportional to the other insurer’s market share. Thus the condition (4.8) formalises the intuitive feeling that arbitrage opportunities should be taken most seriously when the exposed insurers have a large share of the market.

Remark 4.3.

In the general formula (4.5) one can write

\begin{equation}
\tilde{d}_k = \tilde{d}^{(k)} + \sum_{x \in X} \frac{v_k(x)}{v_k} (d_k(x) - d^{(k)}(x)),
\end{equation}

where

\begin{equation}
\tilde{d}^{(k)} = \sum_{i=1}^l \frac{u_{ki}}{v_k} \tilde{d}_i,
\end{equation}

\begin{equation}
d^{(k)}(x) = \sum_{i=1}^l \frac{u_{ki}(x)}{v_k(x)} d_i(x).
\end{equation}

Note the formal similarity between (4.11) and (3.14). The current formula is more sophisticated than (3.14) in that the averages applied to each insurer, depend on the insurer and the weighting of arbitrage opportunities between that insurer and any of its competitors. In contrast, (3.14) is based on the premise that only market averages matter.
Remark 4.4.
The system of equations in (4.5) can be written in matrix form as

\[(4.13) \quad [\text{diag} (V \cdot 1) - V] \bar{d} = \sum_{x \in X} [\text{diag} (V(x) \cdot 1) - V(x)] d(x),\]

\[(4.14) \quad n^T \bar{d} = nd,\]

where \(\bar{d} = [\bar{d}_1, ..., \bar{d}_i]^T\), \(d(x) = [d_1(x), ..., d_i(x)]^T\), \(n = [n_1, ..., n_i]^T\), \(1 = [1, ..., 1]^T\) and \(\text{diag}(x)\) is a diagonal matrix with the elements of vector \(x\) along its main diagonal.

The balancing constraint is essential because \(\text{diag} (V \cdot 1) - V\) is singular (multiply with \(1\) to see this). Barring degeneracy, an invertible system of equations can be constructed by replacing any of the \(i\) equations in (4.13) by the equation (4.14).

Remark 4.5.
The criterion (4.3) is very flexible in that it allows one to incorporate a subjective assessment of the seriousness of arbitrage opportunities between different insurers, or groups of insurers.

Take an example. A number of union-based health funds operate in Australia, each of which recruits its membership exclusively from certain occupational groups. It is arguable that arbitrage opportunities between any of these funds are of no concern, as their members cannot easily transfer (although they can transfer to open funds). If desired, an assessment of which arbitrage opportunities actually matter most can be formalised by adjusting appropriately the weights \(v_j(x)\). In doing so, however, one must take care to ensure that the matrices \(V(x)\) do not become too sparse to allow a unique solution to (4.4).

On a more practical note, care must be taken because an equalisation scheme which bases its allocation on any form of subjective assessment, will be very vulnerable to criticism by the participating insurers, half of which would prefer not to participate in equalisation in the first place.

Remark 4.6.
With regard to replacing the theoretical quantities \(d_i(x)\) with suitable estimates \(d_i^\#(x)\) and balancing the equalisation scheme against the actual rather than the expected claims cost, the same comments as in section 3 apply.

5. A NUMERIC EXAMPLE

This section provides a numeric example to illustrate different equalisation schemes.

Data for 1990-1994 from a random sample of health insurers in Australia was used. The sample was so random, in fact, that the author himself does not know which insurers were chosen; for all intents and purposes the data can be viewed as construed.
The data, consisting of persons covered and basic benefits paid in each of these
years, was split into two age groups, Under 65s and Over 65s. Denote these two
groups by $x_1$ and $x_2$, respectively.

The heterogeneity in benefit plans and thence in the drawing rates $d_i(x)$ was
modelled and estimated by the credibility method set out in the appendix.

We compared the different equalisation schemes for the year $t = 1993$. For insurer
$i$, class $x$, we introduce the following notation:

\[ D_i(x) = \frac{B_i(x)}{n_i(x)} \] = the empirical drawing rate;

\[ d_i^*(x) = \frac{T^*(x)}{\Theta^*_i(x)} \] = the credibility-estimated drawing rate;

We further introduce

\[ D(x) = \frac{B(x)}{n(x)} \] = the empirical drawing rate of class $x$;

\[ d_i^*(x) = \frac{\sum_{i=1}^{t} \frac{n_i(x)}{n(x)} d_i^*(x)} {n(x)} \] = the credibility-estimated drawing rate of class $x$;

\[ D_i = \frac{B_i}{n_i} \] = the empirical drawing rate of insurer $i$;

\[ D = \frac{B}{n} \] = the empirical drawing rate across all classes and

and, finally

\[ E_i \] = the net transfer to equalisation by insurer $i$;

\[ \bar{D}_k = \frac{B_i + E_i}{n_i} \] = the post-equalisation unit cost of insurer $i$.

The following seven equalisation schemes were considered:

**a. Schemes which minimise arbitrage against the average**

a.1 The scheme (3.14). The credibility estimators (A.20) of the theoretical drawing
rates $d_i(x)$ were used. The explicit formula of the scheme is

\[ (5.1) \quad \bar{D}_k = D + \sum_{x \in X} p_k(x) (d_i^*(x) - d^*(x)). \]

a.2 The scheme (3.17). This is essentially the same scheme as in (a.1), except that
empirical drawing rates are used to estimate the theoretical drawing rates. The
explicit formula of the scheme is

\[ (5.2) \quad \bar{D}_k = D + \sum_{x \in X} p_k(x) (D_i(x) - D(x)). \]
b. Schemes which minimise arbitrage between insurers

b.1 The scheme (4.5) with

\[ v_{ij}(x) = \frac{n_i(x) n_j(x)}{n(x)} \]  

(5.3)

The credibility estimators (A.20) of the theoretical drawing rates \( d_i(x) \) were used. A matrix inversion was used to solve (4.5).

b.2 The same scheme as in (b.1), but using empirical drawing rates to estimate the theoretical drawing rates.

c. Miscellaneous schemes

c.1 A simplified version of the 1989 Australian scheme. It is given by the formula

\[ \bar{D}_k = \frac{1}{n_k} \left( B_k(x_1) + \frac{n_k}{n} B(x_2) \right), \]  

(5.4)

c.2 Full equalisation, given by the formula \( \bar{D}_k = D \);

c.3 No equalisation, given by the formula \( \bar{D}_k = D_k \).

\[
\begin{array}{cccccccc}
\text{Equalisation scheme} \\
& i & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
\hline
\text{a.1} & 204.68 & 200.63 & 184.37 & 198.28 & 224.39 & 237.00 & 206.64 & 202.48 & 197.14 & 188.85 & 207.14 & 204.37 & 236.50 \\
\text{a.2} & 185.74 & 206.46 & 171.16 & 200.94 & 226.23 & 238.69 & 238.58 & 199.54 & 195.15 & 169.98 & 198.68 & 204.66 & 236.91 \\
\text{b.1} & 204.37 & 200.38 & 184.71 & 197.94 & 224.38 & 237.17 & 207.04 & 202.53 & 196.99 & 196.57 & 206.81 & 204.66 & \ \\
\text{b.2} & 185.29 & 206.10 & 171.66 & 200.43 & 226.21 & 238.94 & 239.16 & 199.62 & 196.92 & 188.57 & 198.20 & 236.91 & \ \\
\text{c.1} & 204.17 & 219.49 & 195.71 & 219.95 & 219.29 & 220.00 & 229.11 & 205.87 & 208.82 & 183.30 & 218.34 & \ \\
\text{c.3} & 141.17 & 171.31 & 220.06 & 151.22 & 224.00 & 263.05 & 295.90 & 207.20 & 172.86 & 129.87 & 152.01 & 277.36 & \ \\
\end{array}
\]

We then compared the residual arbitrage under each scheme against the measure (4.3), under the assumption that the credibility estimated drawing rates coincide with the theoretical drawing rates.

For each scheme we calculated the quantity

\[ S(\bar{D}_1, ..., \bar{D}_i) = \left( \frac{1}{v} \sum_{i=1}^{l} \sum_{j=1}^{l} \sum_{x \in X} v_{ij}(x) (\bar{D}_i - \bar{D}_j - d^*(x) + d^*(x))^2 \right)^{1/2}, \]
with \( v_j(x) \) defined by (5.3). Scaling (4.3) by \( 1/v \) and the square root transformation were chosen in order to obtain numbers on the same scale as the drawing rates.

Obviously the equalisation scheme (b.1) is optimal against this measure. However, we were interested in just how much additional arbitrage is generated by replacing the credibility estimators \( d^*(x) \) by empirical drawing rates \( D_i(x) \), or by using the simpler schemes (a.1) or (a.2).

The results are displayed in Table 5.2 below. We calculated the residual arbitrage for the years 1990-1993.

Table 5.2 indicates that passing from credibility estimates to empirical drawing rates increases the residual arbitrage. There is, however no significant difference between the residual arbitrage in the optimal scheme (b.1) and the simplified scheme (a.1); the same observation holds true for their respective counterparts (b.2) and (a.2). From this one can conclude that the simple model where arbitrage occurs just against the market average, while unrealistic, produces an optimal scheme which is sufficient for practical purposes, at least against the data used in this study.

The 1989 Australian equalisation scheme (c.1) has residual arbitrage in excess of that achieved by the optimal schemes (a.1) and (b.1) and their counterparts (a.2) and (b.2).

Full equalisation and, of course, no equalisation lead to residual arbitrage which is well in excess of what the optimal schemes achieve.

### Table 5.2

<table>
<thead>
<tr>
<th>Year</th>
<th>a.1</th>
<th>a.2</th>
<th>b.1*</th>
<th>b.2</th>
<th>c.1</th>
<th>c.2</th>
<th>c.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>27.4337</td>
<td>29.4386</td>
<td>27.4327</td>
<td>29.5198</td>
<td>30.9241</td>
<td>35.5532</td>
<td>49.3301</td>
</tr>
<tr>
<td>1993</td>
<td>29.5880</td>
<td>31.9841</td>
<td>29.5869</td>
<td>32.0562</td>
<td>34.7802</td>
<td>38.1619</td>
<td>53.2510</td>
</tr>
</tbody>
</table>

* The scheme b.1 is optimal by assumption.

### 6. ALTERNATIVE MEASUREMENT OF ARBITRAGE

Some people argue, although not exactly in the words used in this paper, that the measure of arbitrage as in (3.6) or (4.3) is too restrictive. In order to understand their point of view, it is instructive to consider the decomposition of the drawing rate (3.18), repeated here for ease of reference:

\[
D_k = D \quad \text{(average drawing rate)}
\]

\[
+ \sum_{x \in X} p_k(x) (D_k(x) - D(x)) \quad \text{(benefit component)}
\]

\[
+ \sum_{x \in X} D(x) (p_k(x) - p(x)) \quad \text{(profile component)}
\]
The scheme (3.17) can be derived by simply eliminating the profile component while letting each insurer pay for its own benefit component.

The argument put forward by critics of this scheme is that the benefit component isn't all benefits. There are two reasons why this could be the case.

The first reason is that the partition of the insured population into supposedly homogeneous groups may not be fine enough. The cost effect of any differences between different insurers' membership profiles which are not reflected in the partition, will make its way into the benefit component and remain unequalised. Figure 1.1 shows, for instance, that there are significant age-related differences in the cost of insuring people aged more than 60 years. In order to make sure that the effect of these differences is equalised, one will have to define very narrow age bands.

The other reason cited is the insidious effect of self-selection. This means that insurers with generous benefit plans are likely to attract a less healthy membership than insurers with no-frills benefit plans. The fact that some insurers aggressively market exclusion plans which are unsuitable for elderly people—excluding hip replacements or bypass operations, for example—even encourages self-selection. It is said in the industry that most people are fairly good judges of their own health care need in the medium term. As a result of self-selection, it is argued, two insurers with identical membership profiles may still have differences in cost which exceed those that can be directly ascribed to differences in their benefit plans.

While the first problem is essentially a statistical one and must be solved in a statistical framework, the problem of self-selection can only be addressed by reviewing our measurement of arbitrage.

Denote the set of all possible treatments which are insurable, by G. The treatments in G could be more or less aggregated; at a very fine level of classification, G could consist of all Diagnostically Related Groups (DRGs).

For an insurer \( i \in \{1, ..., l\} \), a risk class \( x \in X \), and a treatment \( g \in G \), we define the following quantities:

\[
H_i(x, g) = \text{the number of hospital episodes paid for;} \\
B_i(x, g) = \text{the amount of benefits paid.}
\]

As before, omission of any argument indicates summation.

Now denote by \( a_i(x, g) \) the expected benefit cost incurred by insurer \( i \) in insuring a person in class \( x \) to receive treatment \( g \), that is

\[
(6.4) \quad E[B_i(x, g) \mid H_i(x, g)] = H_i(x, g) a_i(x, g).
\]

The assumption underlying self-selection is that in choosing an insurer, people compare the overall price with the specific benefits they are likely to receive in return; and, accepting that people are good judges of their own health care need, one may surmise that many self-select quite cynically. We therefore assert that an arbitrage opportunity between insurers \( i \) and \( j \) exists for people in class \( x \) wishing to receive treatment \( g \), if

\[
(6.5) \quad \tilde{d}_i - \tilde{d}_j = a_i(x, g) - a_j(x, g).
\]

This means that the comparison between different insurers is based on intended usage, rather than the statistical average of benefits received.
In support of this measure of arbitrage one could also argue (Dubey, 1994) that since the insured person is in no position to know his/her drawing rate, he/she is likely to compare competing insurers by the benefits they offer for specific treatments.

An overall measure of the arbitrage opportunities present in the market would then be

\[ R(\tilde{d}_1, ..., \tilde{d}_l) = \sum_{i=1}^l \sum_{j=1}^l \sum_{x \in X} \sum_{g \in G} w_{ij}(x, g)(\tilde{d}_i - \tilde{d}_j - a_i(x, g) + a_j(x, g))^2, \]

where

\[ W(x, g) = \begin{bmatrix} w_{11}(x, g) & ... & w_{ll}(x, g) \\ \vdots & \ddots & \vdots \\ w_{11}(x, g) & ... & w_{ll}(x, g) \end{bmatrix} \]

are symmetric matrices of fixed, non-negative weights.

Not wanting to entirely abandon the criterion (4.3), we propose to blend the two criteria (4.3) and (6.6) and find an equalisation scheme to solve the constrained minimisation problem

\[ \text{Minimise } Q(\tilde{d}_1, ..., \tilde{d}_l) + R(\tilde{d}_1, ..., \tilde{d}_l) \]

\[ \text{subject to } \sum_{i=1}^l n_i \tilde{d}_i = nd. \]

By exactly the same technique as in theorem 4.1 we can prove

**Theorem 6.1.**

The solution of (6.8) is given by the set of equations

\[ (v_k + w_k)\tilde{d}_k = \sum_{i=1}^l (v_{ki} + w_{ki})\tilde{d}_i \]

\[ + \sum_{x \in X} v_k(x) d_k(x) - \sum_{x \in X} \sum_{i=1}^l v_{ki}(x) d_i(x) \]

\[ + \sum_{x \in X} \sum_{g \in G} w_k(x, g) a_k(x, g) - \sum_{x \in X} \sum_{g \in G} \sum_{i=1}^l w_{ki}(x, g) a_i(x, g) \]

for \( k = 1, ..., l \), and

\[ nd = \sum_{i=1}^l n_i \tilde{d}_i. \]
Remark 6.2.
The minimum arbitrage scheme can be written in the form

\[
\tilde{d}_k = \tilde{d}^{(k)} + \frac{v_k}{v_k + w_k} \Delta_k(d) + \frac{w_k}{v_k + w_k} \Delta_k(a),
\]

where

\[
\tilde{d}^{(k)} = \sum_{i=1}^{I} \frac{v_{ki} + w_{ki}}{v_k + w_k} \tilde{d}_i
\]
is a weighted average of all insurers’ post-equalisation unit cost,

\[
\Delta_k(d) = \sum_{x \in X} \frac{v_k(x)}{v_k} \left( d_k(x) - \sum_{i=1}^{I} \frac{v_{ki}(x)}{v_k(x)} d_i(x) \right)
\]
is a weighted average of the drawing rate differential between insurer \(k\) and the suitably calculated average, and

\[
\Delta_k(a) = \sum_{x \in X} \sum_{g \in G} \frac{w_k(x, g)}{w_k} \left( a_k(x, g) - \sum_{i=1}^{I} \frac{w_{ki}(x, g)}{w_k(x, g)} a_i(x, g) \right)
\]
is a weighted average of the benefit differential between insurer \(k\) and a suitably calculated average.

Schemes of the form (6.10) can be called mixed schemes, as they minimise arbitrage opportunities with respect to a mixed criterion. In the extreme case where \(v_k = 0 \ (\forall k)\), the resulting equalisation scheme can be called a usage scheme, as it minimises arbitrage opportunities based on intended usage only. The scheme proposed for Ireland is a usage scheme, although it does not differentiate between different treatments.

Corollary 6.3.

If the matrix \(V + W = \sum_{x \in X} V(x) + \sum_{x \in X} \sum_{g \in G} W(x, g)\) satisfies the condition

\[
v_{ij} + w_{ij} = cn_i n_j
\]
with an arbitrary constant \(c > 0\), then optimal scheme against the criterion (6.8) has an explicit solution.

Remark 6.4.
With regard to replacing the theoretical quantities \(d_i(x)\) with suitable estimates \(d_i^*(x)\) and balancing the equalisation scheme against the actual rather than the expected claims cost, the same comments as in section 3 apply. However, to implement the equalisation scheme (6.10), one also needs estimates of the quantities \(a_i(x, g)\); if benefits are paid in the form of predetermined case payments, such estimates are readily available.
7. PARTIAL COMMUNITY RATING

In sections 3 and 4 we developed equalisation schemes which were designed to support total community rating, i.e. just one flat contribution rate for any person/policy.

Some countries/states require partial community rating. As an example of partial community rating we will use unisex rating in motor vehicle insurance.

By the argument developed in section 2, mandatory partial community rating can imply an obligation to implement an equalisation scheme, just as total community rating. A pragmatic test as to whether equalisation is warranted would be to consider the difference in cost between hypothetical insurers located at the extremes of membership profile: if for instance an insurer which has recruited only males has significantly higher cost than one which has recruited only females, other things being equal, then equalisation is warranted. Following this argument further, one can conjecture that equalisation normally will be warranted whenever partial community rating is mandatory; for unless there were significant differences in cost, there would be no need to make partial community rating mandatory.

In this section we develop an equalisation scheme that supports partial community rating.

As before, assume that there exists a partition $X$ of the insured population and that contribution rates may not depend on $x \in X$. The partition could be $X = \{\text{male}, \text{female}\}$, for instance. Now assume that there exists another partition $Y$, and that contribution rates may differ with $y \in Y$; this partition may be made up of several other rating variables, like the type of car, district, usage, etc. Thus we assume that each insurer $i$ is allowed to have a vector of (net) contribution rates,

$$
(7.1) \quad \tilde{d}_i = \text{col}(\tilde{d}_i(y)).
$$

Denote by $d_i(x, y)$ the pure premium of insurer $i$ in insuring a person belonging to the class $(x, y) \in X \times Y$.

We say that an opportunity for arbitrage exists if for insurers $i, j \in \{1, \ldots, l\}$ and risk classes $x \in X$, $y, y' \in Y$ we have

$$
(7.2) \quad \tilde{d}_i(y) - \tilde{d}_j(y') \neq d_i(x, y) - d_j(x, y')
$$

This measure implies that

a. for a person in class $(x, y) \in X \times Y$, an arbitrage opportunity exists when there is a mismatch between the contributions and benefits between the two insurers; and

b. for a person in class $x \in X$, an arbitrage opportunity exists if there is a mismatch between contributions and benefits for two values $y, y' \in Y$, be it with the same insurer or a different insurer.

Arbitrage of type (a) is of the form we have already seen previously, between insurers. The motivation for including (b) is that in transferring between classes $y$ and $y'$ (say, changing cars), the change in contribution rate for a person $x$ should
reflect the change in pure premium for class $x$. The change in contribution rate should not reflect the fact that car make $y$ appeals more to males (females) than car $y'$. If contributions for $y'$ were much higher than those for $y$ on account of a different mix of $x$, the purpose of partial community rating—equality between classes of $X$—would be defeated. Barred from discriminating on the basis of $x$, insurers would surely find proxy variables to include in $y$.

Having motivated what we mean by arbitrage, we can define an overall measure

$Q(\tilde{d}_1, \ldots, \tilde{d}_I) = \sum_{i,j=1}^{I} \sum_{x \in X} \sum_{y, y' \in Y} v_{ij}(x; y, y') (\tilde{d}_i(y) - \tilde{d}_j(y') - d_i(x, y) + d_j(x, y'))^2,$

where the $v_{ij}(x; y, y')$ are non-negative weights obeying the symmetry condition $v_{ij}(x; y, y') = v_{ji}(x; y', y)$.

An optimal equalisation scheme against the measure (7.3) is given by the set of values $\{d_i(y) : i = 1, \ldots, I; y \in Y\}$ which minimises (7.3) under the constraint

$\sum_{i=1}^{I} \sum_{x \in X} \sum_{y \in Y} n_i(x, y) d_i(x, y) =: nd.$

Using Lagrange minimisation we obtain

**Theorem 7.1.**

The minimum arbitrage scheme against the criterion (7.3) is given by the set of linear equations

$\forall_{k = 1, \ldots, I} \mathcal{X} \mathcal{Y} \mathcal{Z}$

\begin{align*}
\left(7.5\right) & \sum_{i=1}^{I} \sum_{x \in X} \sum_{y \in Y} n_{i}(x, y) d_{i}(x, y) = \sum_{k=1}^{I} \sum_{z \in Y} n_{k}(z) \tilde{d}_{k}(z) = nd,
\end{align*}

where a dot in place of an argument indicates summation.

**Proof:** Lagrange minimisation,

QED
Remark 7.2.

The system of equations (7.5)-(7.6) has an explicit solution if the following condition holds:

\[ u_{ki}(\cdot ; z, y) = c \cdot n_k(z) n_i(y), \text{ with } c \text{ an arbitrary constant.} \]

8. CONCLUDING REMARKS

This paper provides a theoretical framework for the development of equalisation schemes, be it in health insurance or other areas. Broadly, it argues that community rating creates arbitrage opportunities which are self-reinforcing and can destabilise a market. As a consequence, an equalisation scheme which is meant to support community rating, must be designed to minimise opportunities for arbitrage.

The use of theoretically sound equalisation schemes may help to bridge the gap between the advocates of unrestricted risk rating and the advocates of (partial) community rating, which is lucidly described in Jewell (1980).

The general approach advocated in this paper still leaves some degree of freedom to the designer. In particular, the measure used to quantify arbitrage could be varied, although the measure proposed here (weighted sum of squares) has the great advantage of being mathematically tractable. Even when retaining a measure based on weighted sum of squares, one can vary the weights which indicate the "seriousness" of different arbitrage opportunities.

The weighted sum of squares measure leads to equalisation formulae which are very tractable and, to the trained eye, intuitively obvious. It is the author's hope that this paper will contribute towards a more disciplined approach to the construction of equalisation schemes.

With effect from 1995 the Australian equalisation scheme was modified following the recommendations in MIRA (1993, 1994). The new scheme combines features of the composition-based scheme (3.17) and the 1989 Australian scheme (5.4).

ACKNOWLEDGEMENTS

This paper developed out of a study commissioned by the Australian Department of Health, Housing and Community Services in 1992. Many persons within the Department in the health insurance industry have contributed to the ideas presented above. All contributions are gratefully acknowledged, in particular those of Malcolm Murray (Health Insurance Commission) and Russell Schneider (Australian Health Insurance Association) and lan Heppell (then MIRA Consultants Ltd).
The heterogeneity in benefit plans and thence in the pure premiums $d_i(x)$ was modelled as follows: It was assumed that the relative level of benefits paid by insurer $i$ is characterised by a latent vector

\[(A.1) \quad \Theta_i = \text{col} \left( \theta_i(x) \right)_{x \in X} \]

which is independent of time. The notation col denotes a column vector.

We then assumed that the pure premium vector for insurer $i$ in year $t$ is

\[(A.2) \quad d_{i}^{(t)} = \text{col} \left( d_i^{(t)}(x) \right) = \text{col} \left( T^{(i)}(x) \theta_i(x) \right) = T^{(i)} \Theta_i, \]

where $T^{(i)} = \text{diag} \left( T_i(x) \right)$ and $T^{(i)}(x)$ adjusts benefits paid to class $x$ in year $t$ to reflect inflation.

As an aside, note that the adjustment $T^{(i)}$ could be chosen dependent on the insurer $i$, thereby incorporating in the model any prior knowledge about the relative benefit levels provided by the different insurers. We have preferred not to use this option and rather let the latent vector $\Theta_i$ reflect all the variation in benefits between insurers.

Of actual benefits paid by insurer $i$ in year $t$ we assumed that

\[(A.3) \quad \text{E}[B_i^{(t)}] = \text{col} \left( \text{E}B_i^{(t)}(x) \right) = \text{col} \left( n_i^{(t)}(x) d_i^{(t)}(x) \right) = N_i^{(t)} T^{(i)} \Theta_i, \]

where $N_i^{(t)} = \text{diag} \left( n_i^{(t)}(x) \right)$. Of the variance we assumed

\[(A.4) \quad \text{Var}[B_i^{(t)}] = \text{diag} \left( T^{(i)}(x) n_i^{(t)}(x) \Phi(x) \right) = T^{(i)} \Phi, \]

independent of $\Theta_i$, where $\Phi = \text{diag} \left( \Phi(x) \right)$ is a fixed coefficient matrix.

Finally we assumed that $\Theta_1, \ldots, \Theta_I$ are independent random vectors with mean

\[(A.5) \quad \text{E}\Theta_i = \eta \]

and variance

\[(A.6) \quad \text{Var}\Theta_i = \Lambda. \]
The best linear estimator of $\Theta_i$, homogeneous in the $B^{(i)}$, is

$$\Theta_i = Z_i \hat{\Theta}_i + (1 - Z_i) \bar{\eta},$$

where

$$\hat{\Theta}_i = \left( \sum_i N_i^{(i)} \right)^{-1} \sum_i N_i^{(i)} \hat{\Theta}_i^{(i)},$$

and

$$\hat{\Theta}_i^{(i)} = [N_i^{(i)} T^{(i)}]^{-1} B_i^{(i)}.$$

are estimators of $\Theta_i$.

$$Z_i = \Lambda (N_i^{(i)-1} \Phi + \Lambda)^{-1},$$

is the credibility matrix (the dot denoting summation), and finally

$$\hat{\eta} = \left( \sum_{i=1}^I Z_i \right)^{-1} \sum_{i=1}^I Z_i \hat{\Theta}_i$$

is the best linear unbiased estimator of $\eta$.

It now remains to estimate the parameters of the model. The parameters $T^{(i)}(x)$ were estimated by

$$T^{(i)*}(x) = \frac{B^{(i)}(x)/n^{(i)}(x)}{B^{(1994)}(x)/n^{(1994)}(x)},$$

and the resulting estimate inserted in (A.9).

The structural parameters were estimated by the method proposed by de Vylder (1981) and analysed by Hesselager (1988).

The covariance matrix $\Phi$ was estimated by the unbiased estimator

$$\Phi^* = \text{diag} \left( \frac{1}{t} \sum_{i=1}^I \sum_{r=1}^t h_i^{(r)}(x) (\hat{\Theta}_i^{(r)}(x) - \hat{\Theta}_i(x))^2 \right).$$

In this expression, $\tau_i$ represents the number of years that the insurer $i$ has been under observation.

The covariance matrix $\Lambda$ was estimated by the limit of a convergent sequence of the form

$$\Lambda^*(k+1) = \frac{1}{t-1} \sum_{i=1}^I Z_i^*(k) (\hat{\Theta}_i - \bar{\eta}(k)) (\hat{\Theta}_i - \bar{\eta}(k))^T,$$

with

$$Z_i^*(k) = \Lambda^*(k) (N_i^{(i)*} \Phi^* + \Lambda^*(k))^{-1}. $$
and

\[(A.16) \quad \hat{\eta}(k) = \left( \sum_{i=1}^{I} Z_{*}^{i}(k) \right)^{-1} \sum_{i=1}^{I} Z_{*}^{i}(k) \hat{\Theta}_{i}.\]

Denoting the limit by \(\Lambda^{*}(\infty)\), the final estimator used was the symmetrised

\[(A.17) \quad \Lambda^{**} = \frac{1}{2} [\Lambda^{*}(\infty) + \Lambda^{*}(\infty)^{T}].\]

Using \(\Lambda = I\) (identity matrix) as starting value, this procedure worked extremely well.

The resulting estimates were

\[(A.18) \quad \pi^{1990\ldots1994^{*}} = \begin{bmatrix} 0.92 & 0.92 & 0.98 & 0.99 & 1.00 \\ 0.76 & 0.86 & 0.92 & 0.95 & 1.00 \end{bmatrix}.\]

and

\[(A.19) \quad \hat{\eta} = \begin{bmatrix} 143.08 \\ 730.37 \end{bmatrix}, \quad \Phi^{*} = \begin{bmatrix} 3185^2 & 0 \\ 0 & 6156^2 \end{bmatrix}, \quad \Lambda^{**} = \begin{bmatrix} 14.38 & 49.31 \\ 49.31 & 644.18 \end{bmatrix}.\]

We denote the credibility estimator of the form (A.7), with the parameter values (A.19) inserted, by

\[\hat{\Theta}_{i} = \text{col} \left( \hat{\Theta}_{*}^{f}(x) \right).\]

Note in passing that inflation apparently has been significantly stronger for the Over 65s than the Under 65s, see (A.18). There are two possible explanations for this. One explanation is that the 1989 equalisation scheme essentially has removed any incentive for insurers to control benefits paid to that group. The other possible explanation is that there has been an ageing in that group; a glance at Figure 1.1 reveals that increasing the average age with just a few years would lead to a significant increase in cost. We did not have the data available to decide the extent to which each of the two explanations outlined above explain the observed difference in inflation.

The estimates \(\hat{\Theta}_{i}^{*}\) are shown in Table A.1. Based on these estimates one can estimate the pure premium vector of insurer \(i\) in year \(t\) as

\[(A.20) \quad d_{i}^{(t)*} = T^{(t)*} \hat{\Theta}_{i}^{*} = \text{col} \left( T^{(t)*}(x) \hat{\Theta}_{i}^{*}(x) \right).\]
## TABLE A.1
CREDIBILITY ESTIMATES

<table>
<thead>
<tr>
<th>i</th>
<th>( n_{1i}^{(1)}(x_i) ) in 1000</th>
<th>( n_{1i}^{(2)}(x_i) ) in 1000</th>
<th>( \hat{\Theta}_{i}(x_i) )</th>
<th>( \hat{\Theta}<em>{j}(x</em>{j}) )</th>
<th>( \Theta^{*}(x_i) )</th>
<th>( \Theta^{*}(x_{j}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95</td>
<td>4</td>
<td>127.56</td>
<td>583.65</td>
<td>140.71</td>
<td>716.46</td>
</tr>
<tr>
<td>2</td>
<td>247</td>
<td>13</td>
<td>128.86</td>
<td>616.84</td>
<td>138.35</td>
<td>700.57</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>31</td>
<td>160.78</td>
<td>501.55</td>
<td>140.84</td>
<td>660.28</td>
</tr>
<tr>
<td>4</td>
<td>781</td>
<td>20</td>
<td>131.05</td>
<td>515.07</td>
<td>134.94</td>
<td>665.50</td>
</tr>
<tr>
<td>5</td>
<td>1,996</td>
<td>232</td>
<td>157.11</td>
<td>851.67</td>
<td>155.06</td>
<td>831.16</td>
</tr>
<tr>
<td>6</td>
<td>2,653</td>
<td>416</td>
<td>160.95</td>
<td>904.53</td>
<td>159.36</td>
<td>885.49</td>
</tr>
<tr>
<td>7</td>
<td>314</td>
<td>73</td>
<td>172.72</td>
<td>724.58</td>
<td>151.06</td>
<td>741.92</td>
</tr>
<tr>
<td>8</td>
<td>2,223</td>
<td>284</td>
<td>143.58</td>
<td>725.78</td>
<td>143.35</td>
<td>726.99</td>
</tr>
<tr>
<td>9</td>
<td>1,611</td>
<td>115</td>
<td>134.38</td>
<td>689.03</td>
<td>136.67</td>
<td>696.97</td>
</tr>
<tr>
<td>10</td>
<td>578</td>
<td>26</td>
<td>105.52</td>
<td>686.52</td>
<td>126.35</td>
<td>676.33</td>
</tr>
<tr>
<td>11</td>
<td>55</td>
<td>2</td>
<td>143.22</td>
<td>559.64</td>
<td>142.73</td>
<td>725.48</td>
</tr>
<tr>
<td>12</td>
<td>126</td>
<td>24</td>
<td>177.17</td>
<td>709.42</td>
<td>147.51</td>
<td>737.24</td>
</tr>
</tbody>
</table>

## REFERENCES


WALther NEUHAUS

Laboratory of Actuarial Mathematics, University of Copenhagen
Universitetsparken 5, DK-2100 Copenhagen Ø
MODELLING OF DISCRETIZED CLAIM NUMBERS IN LOSS RESERVING

BY OLE HESSELAGER

University of Copenhagen
Laboratory of Actuarial Mathematics

ABSTRACT

We investigate the usual method of discretizing loss reserving data by calendar year and show how this procedure may introduce fluctuations in the delay probabilities. These fluctuations, when treated as random fluctuations, possess a special correlation structure and we present a simple credibility method accounting for these fluctuations. The results are illustrated by a numerical example.

KEYWORDS

Loss reserving; Claim numbers; Poisson process; Discretizing by calendar year; Variations in delay probabilities; Credibility adjustments.

1. INTRODUCTION

Starting from a continuous time model, with claims occurring according to an inhomogeneous Poisson process and with the waiting times until notification being iid real-valued variables, it is shown how the traditional way of discretizing the observations according to calendar year can introduce a time dependence in the delay probabilities. Thus, variations between occurrence years in the delay probabilities may occur simply as a consequence of the way the data is discretized, even when the distribution of the actual delays is independent of time.

In recent years a number of papers on loss reserving have appeared where the delay probabilities for observations in discrete time are assumed to vary between occurrence years. In these papers (HESSELAGER and WITTING, 1988; NEUHAUS, 1992; LAWLESS, 1994; HESS and SCHMIDT, 1994a, b) the variations are treated as random fluctuations between occurrence years and are modelled by a Dirichlet distribution which is a mathematically very convenient construct. In the present paper it is shown that fluctuations induced in the discretizing process possess a special structure and that these fluctuations can not be described by a Dirichlet distribution in a reasonable manner. In fact, in typical cases the probability that a claim is reported in the year of occurrence is negatively correlated with all the other delay probabilities, and these are positively correlated. In a Dirichlet distribution, all probabilities are negatively correlated.

It should be noted that this paper deals only with the fluctuations which are introduced in the discretizing process, and we are of course aware that there may be other sources of variation.

2. FROM CONTINUOUS TO DISCRETE TIME MODELS

We start by considering the continuous time process \((T_n, W_n), n = 1, 2, \ldots\), where \(0 < T_1 < T_2 < \ldots\) denote the occurrence times, and \(W_1, W_2, \ldots\) are the corresponding real-valued waiting times from occurrence until notification. Most methods for loss reserving assume that the observations have been discretized according to calendar year as illustrated in Figure 1.

The time axis is divided into intervals \((i-1, i], i = 1, 2, \ldots\) and the time interval \((i-1, i]\) is called year \(i\). A claim incurred in year \(i\) is according to this method reported with a delay of \(j\) years if it is reported in year \(i+j\), and Figure 1 shows the combinations of values \((T_n, W_n)\) which fall into cell \((i, j)\). It should be noted that the cell \((i, 0)\) is differently shaped and only half the size of the cells \((i, j), j > 1\). Note also that the actual waiting time corresponding to observations in \((i, j)\) is in the interval \([j-1, j+1], j \geq 1\).

For \(i = 1, 2, \ldots\) and \(j = 0, 1, \ldots\), we denote by

\[
N_{ij} = \sum_{n \geq 1} 1 \{T_n \in (i-1, i], T_n + W_n \in (i+j-1, i+j]\},
\]
the number of claims incurred in year \( i \) and reported with a delay of \( j \) years. The total number of claims incurred in year \( i \) is

\[
N_i = \sum_{j=0}^{\infty} N_{ij} = \sum_{n \geq 1} I\{T_n \in (i-1, i]\}.
\]

We now make the following model assumptions pertaining to the continuous time setting

(a) The claims occurrences \( \{T_n\}_n \) are generated by a Poisson process with parameter \( \Lambda(\cdot) \) on \([0, \infty)\), where \( \Lambda(t) \) is a non-negative, non-decreasing and right-continuous function representing the risk exposure for the interval \([0, t]\).

(b) The delays \( W_1, W_2, \ldots \) are independent of \( \{T_n\}_n \) and iid with common distribution \( F \).

For a Poisson process with parameter \( \Lambda(\cdot) \) the number of events occurring in disjoint time intervals are stochastically independent, and the number of events occurring in \((s, t]\) is Poisson distributed with parameter \( \lambda(s, t) = \Lambda(t) - \Lambda(s) \). A Poisson process with intensity \( \lambda(\cdot) \) has a parameter \( \Lambda(t) = \int_0^t \lambda(x) \, dx \), such that \( \Lambda(\cdot) \) in this case is the measure with density \( \lambda(\cdot) \).

While it is customary in collective risk theory to assume the existence of an intensity, we choose here to work with the more general setting primarily because we shall consider mixed Poisson models where \( \Lambda(\cdot) \) is considered random, and where \( t \to \Lambda(t) \) is not necessarily continuous (with probability one). Apart from this technical reason, the reader should note that the case where \( \Lambda(\cdot) \) is a non-random measure having a mass point at \( t_0 \), say, represents the situation where some event is known to take place at time \( t_0 \) generating a Poisson distributed number of claims with parameter \( \Lambda(t_0) \) \( - \) \( \Lambda(t_0^-) \). When in addition \( \Lambda(\cdot) \) is taken to be random (a stochastic process), one will also cover the cases where the time epochs for such multiple claims (or catastrophes) are not known in advance.

For the moment we consider \( \Lambda(\cdot) \) as non-random and introduce for \( i = 1, 2, \ldots \)

\[
(2.1) \quad \Lambda_i = \Lambda(i-1, i), \quad U_i(x) = \Lambda(i-1, i-1+x)/\Lambda_i, \quad x \in (0, 1],
\]

which is the total exposure for year \( i \) and the relative distribution of the exposure over year \( i \), respectively. It then follows from assumptions (a), (b) that the claim numbers \( N_{ij} \) are mutually independent and

\[
N_{ij} \sim \text{Poisson}(\Lambda_i p_{ij}),
\]

(2.2) where

\[
p_{ij} = \int_0^1 [F(j+1-x) - F(j-x)] U_i(dx), \quad j \geq 0,
\]

and it is understood that \( F(z) = 0 \) for \( z < 0 \). The result (2.2), (2.3) is easily verified by standard calculations making use of the fact that the occurrence times within year \( i \), conditionally given \( N_i \), are distributed as the order statistics of \( N_i \) iid variables with distribution \( U_i(\cdot) \) on \((0, 1]\). Alternatively, the result can be shown using arguments similar to those given in the proof of Theorem 2 in Norberg (1993).
The delay probabilities $p_{ij}$ will according to (2.3) in general depend on time $i$ through the distribution $U_i$ of exposure over year $i$. This is very unfortunate from a practical point of view, since it makes statistical estimation of the delay probabilities difficult. However, if the Poisson process has an intensity $\lambda(t)$ which is piecewise constant over the occurrence years, 

$$\lambda(t) = \lambda_{[t]+1},$$

where $[t]$ denotes the integer part of $i$, then

$$U_i(x) = \frac{\int_0^x \lambda(i-1+s)ds}{\int_0^1 \lambda(i-1+s)ds} = x, x \in (0, 1],$$

is the uniform distribution and is independent of time $i$. The delay probabilities $p_{ij}$ are therefore also independent of $i$, and can now be estimated by statistical methods in a straightforward manner. A more general situation where $p_{ij}$ becomes independent of $i$ is that where

$$\Lambda(i-1, i-1+x) = \lambda_i \Lambda_0(x), x \in (0, 1], i = 1, 2, ...$$

This covers the situation where there exists a fixed measure $\Lambda_0(x)$ on $(0,1]$ representing seasonal variation in the exposure within occurrence years, whereas the factors $\lambda_i$ represent the variation between the occurrence years. We obtain in this case from (2.1) that

$$U_i(x) = \frac{\Lambda_0(x)}{\Lambda_0(1)}, x \in (0, 1],$$

which is again independent of $i$.

**Remark 1.** In the case (2.4) with a uniform distribution of exposure we may express the delay probabilities (2.3) in terms of the stop-loss transform

$$\Pi(x) = \int_x^{\infty} (1 - F(y)) dy$$

for the continuous time delay distribution. This yields the expression

$$p_j = \begin{cases} 
1 + \Pi(1) - \Pi(0), & j = 0 \\
\Pi(j-1) + \Pi(j+1) - 2\Pi(j), & j \geq 1.
\end{cases}$$

**Remark 2.** Had the occurrence times $T_n$ and the waiting times $W_n$ been discretized separately, we would have that

$$N_{ij} = \sum_{n=1}^{\infty} I\{T_n \in (i-1, i], W_n \in (j, j+1]\},$$

\[\square\]
and it would then hold that the claim numbers \( N_{ij} \) are mutually independent and Poisson distributed as

\[
N_{ij} \sim \text{Poisson}(\Lambda_i p_j),
\]

with \( \Lambda_i \) as in (2.1) and now with delay probabilities

\[
p_j = F(j + 1) - F(j), j = 0, 1, \ldots,
\]

which are always time-independent. The problem with this approach is of course that the statistics \( N_{ij} \) with \( i + j = r \) can not be constructed (observed) at the end of year \( r \) (at time \( r \)), since they involve reportings in the interval \((r-1, r+1]\).

From these remarks we are now able to conclude that when discretizing the observations according to calendar year, or some other time unit, it is important to choose the time unit in such a way that the risk can be regarded as constant over these time periods, except possibly for the same seasonal variation within the time periods. If this is not the case, one will introduce fluctuations in the delay probabilities between the occurrence years. In particular, when seasonal variations may occur within calendar years, one should be cautious about discretizing on a quarterly basis, as is often done in practice.

3. THE STRUCTURE OF THE INDUCED VARIATION

We investigate here the structure of the fluctuations in the delay probabilities \( p_{ij} \) induced by variations in the distribution \( U_i \) of exposure over year \( i \). Considering only a fixed year \( i \), we shall drop the subscript \( i \) in this section. For reasons of simplicity we also assume that the continuous time delay distribution \( F \) has a density \( f \).

With \( P_j = p_0 + \ldots + p_j \) being the cdf. for the discretized delay distribution, we obtain from (2.3) that

\[
P_j = \int_0^1 F(j + 1 - x) U(dx)
\]

\[
(3.1) = F(j) + \int_0^1 f(j + 1 - x) U(x) dx,
\]

where the latter equality follows by partial integration.

A pair of probability distributions with cumulative distribution functions \( G \) and \( G^* \) are ordered in stochastic order, written as \( G \leq_s G^* \), if \((1 - G)(x) \leq (1 - G^*)(x)\) for all \( x \). In the actuarial literature one also says that \( G^* \) is more dangerous than \( G \), written as \( G \leq_d G^* \), if there exists a \( c \) such that

\[
G(x) \leq G^*(x), x < c
\]

\[
G(x) \geq G^*(x), x \geq c,
\]

and \( \int xG(dx) \leq \int xG^*(dx) \). In addition to the sign change condition (3.2) for the cumulative distribution function we shall also work with a sign change condition
for the discretized delay probabilities. In this case we write $P \leq_{st} P^*$ if

$$
p_0 \leq p_0^*, \quad p_j \leq p_j^*, \quad j > 0
$$

(3.3)

It is well-known (e.g. KAAS et al., 1994, Th. III.1.3) that $G \leq_d G^*$ implies that $G$ is smaller than $G^*$ in stop-loss order, and also that $P \leq_{st} P^*$ implies that $P \leq_{st} P^*$ (e.g. KAAS et al., 1994, Th. II.1.3).

**Lemma 1**

(a) $U \leq_{st} U^* \Rightarrow P \leq_{st} P^*$

(b) If the density $f$ is decreasing, then $U \leq_{st} U^* \Rightarrow P \leq_{st} P^*$

(c) If the density $f$ is decreasing and convex, then $U \leq_{d} U^* \Rightarrow P \leq_{o} P^*$

**Proof.** Assertion (a) follows immediately from (3.1) since

$$
P_j^* - p_j = \int_0^1 f(j + 1 - x)(U^*(x) - U(x)) \, dx \leq 0.
$$

Since $p_0 = p_0^*$, this also proves (3.3) for $j = 0$. For $j \geq 1$, when $f$ is decreasing, we similarly find that

$$
p_j^* - p_j = \int_0^1 (f(j + 1 - x) - f(j - x))(U^*(x) - U(x)) \, dx \geq 0.
$$

which verifies (3.3) and hence assertion (b).

To verify (c) we write

$$
p_0 - p_0^* = \int_0^1 f(1 - x)(U(x) - U^*(x)) \, dx
$$

$$
= \int_0^1 [f(1 - x) - f(1 - c)](U(x) - U^*(x)) + f(1 - c)(m^* - m),
$$

where $m = \int (1 - U(x)) \, dx$ is the mean of the distribution $U$, and $m^*$ is the mean of $U^*$. Since $f(1 - x)$ is increasing it holds that $f(1 - x) - f(1 - c)$, and by assumption also $U(x) - U^*(x)$, is positive for $x > c$ and negative for $x < c$, and therefore

$$
p_0 - p_0^* \geq f(1 - c)(m^* - m) \geq 0.
$$
For \( j \leq 1 \) we note that \( f \) being convex implies that \( h(x) = f(j + 1 - x) - f(j - x) \) is decreasing. It then holds that \( h(x) - h(c) \) and \( U(x) - U^*(x) \) have opposite signs, and

\[
p_j - p^*_j = \int_0^1 [h(x) - h(c)](U(x) - U^*(x)) dx + h(c) \int_0^1 [U(x) - U^*(x)] dx
\]

\[
\leq h(c) \int_0^1 [U(x) - U^*(x)] dx
\]

\[= h(c)(m^* - m) \leq 0,\]

since \( h(c) \leq 0 \). QED

Assertion (a) allows us to obtain lower and upper bounds for the discrete time delay distribution when the distribution \( U \) of the exposure is completely unknown. It follows that the lower and upper bounds are obtained by putting all the exposure at the beginning and at the end of the year, respectively, which gives

\begin{align*}
(3.4) & \quad P^- = (F(1), F(2), \ldots), \\
(3.5) & \quad P^+ = (0, F(1), F(2), \ldots).
\end{align*}

(Note that the \( - \) signifies the smallest distribution in stochastic order, which is the largest cdf, and similarly for the \( + \).)

If the continuous time delay distribution has a decreasing density it can be concluded that a stochastic increase in the distribution of exposure results in a special type of stochastic increase in the discrete time delay distribution; namely where mass is taken from the reportings with delay \( j = 0 \) and transferred to the reportings with delay \( j \geq 1 \). The same type of stochastic increase occurs if the distribution of exposure becomes "more dangerous" (which is a weaker requirement), provided that the delay density is also convex.

We have seen that at least certain types of changes in the exposure distribution lead to the special type of change in the delay probabilities where the probability of immediate reporting varies inversely with the remaining delay probabilities. Thus, if variations in the exposure distribution are considered as random, and only variations satisfying the conditions of Lemma 1 are followed for, we will have that the random probability of immediate reporting is negatively correlated with the rest of the delay probabilities, and that these are positively correlated.

4. INDUCED RANDOM FLUCTUATIONS

In this section we consider the situation where the parameter \( \Lambda(\cdot) \) itself is a stochastic process. Note that there is a one to one correspondence between \( \Lambda(\cdot) \) and the pairs \( (\Lambda_i, U_i) \), \( i = 1, 2, \ldots \), and in the stochastic models considered here these
pairs are iid for \( i = 1, 2, \ldots \), and \( \Lambda_i \) is independent of \( U_i \). The delay probabilities (see (2.3))

\[
(4.1) \quad p_{ij} = \int_0^1 h_j(x) U_i(dx),
\]

\[
(4.2) \quad h_j(x) = F(j + 1 - x) - F(j - x), \quad j = 0, 1, \ldots,
\]

are in this case random variables with

\[
(4.3) \quad \pi_j = \mathbb{E} p_{ij} = \int_0^1 h_j(x) u(dx),
\]

\[
(4.4) \quad c_{jl} = \text{Cov}(p_{ij}, p_{il}) = \int_0^1 \int_0^1 h_j(x) h_l(y) \varphi(dx, dy),
\]

where

\[
(4.5) \quad u(x) = \mathbb{E} U_i(x), \quad \varphi(dx, dy) = \text{Cov}(U_i(dx), U_i(dy)).
\]

Conditionally given \((\Lambda_i, U_i)\), the claim numbers \( N_{ij}, j = 0, 1, \ldots \), are mutually independent and Poisson distributed with parameters \( \Lambda_i p_{ij} \) (see (2.2)), which gives the (unconditional) moments

\[
(4.6) \quad \mathbb{E} N_{ij} = \nu \pi_j,
\]

\[
(4.7) \quad \text{Cov}(N_{ij}, N_{il}) = \delta_{jl} \nu \pi_j + \pi_j \pi_l \tau^2 + c_{jl}(\tau^2 + \nu^2),
\]

where

\[
(4.8) \quad \nu = \mathbb{E} \Lambda_i, \quad \tau^2 = \text{Var} \Lambda_i.
\]

At the end of calendar year \( l \), the observed claim numbers in respect of occurrence year \( i \) are \( N_{i\leq} = (N_{i0}, \ldots, N_{i,i-i})' \), and for \( j > l - i \) we may predict \( N_{ij} \) using credibility estimation as (see e.g. SUNDT, 1993, p. 34)

\[
(4.9) \quad \tilde{N}_{ij} = \mathbb{E} N_{ij} + \text{Cov}(N_{ij}, N_{i\leq}) (\text{Var} N_{i\leq})^{-1} (N_{i\leq} - \mathbb{E} N_{i\leq}).
\]

The second order moments \( c_{jl} \) may be written as \( c_{jl} = \sqrt{c_{jj} c_{ll}} \chi_{jl} \), where \( \chi_{jl} \) denotes the coefficient of correlation. We shall see that these for all practical purposes can be approximated by

\[
(4.10) \quad \{\chi_{jl}\}_{jl} = \begin{bmatrix}
1 & -1 & -1 & -1 & \cdots \\
-1 & 1 & 1 & 1 & \cdots \\
-1 & 1 & 1 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix},
\]

and for this situation we have the following
Lemma 2

With a moment structure (4.6), (4.7), and a correlation matrix (4.10), the credibility estimator (4.9) is given by

\[
\tilde{N}_{ij} = \sum_{l \leq l-i} N_{ij}/\pi_i,
\]

where

\[
\bar{\pi}_i = \frac{\bar{\pi}^2}{\bar{\pi} + \bar{\pi}^2}, \quad \bar{\lambda}_i = \sum_{l \leq l-i} N_{ij}/\pi_i,
\]

\[
\hat{\Omega}_i = \frac{1}{m_i \pi_i} \sum_{l \leq l-i} v_i (N_{ij} - z_i \bar{\lambda}_i \pi_i), \quad E \hat{\Omega}_i = v (1 - z_i) \bar{v}/m_i,
\]

\[
\eta_i = \frac{(\bar{\pi}^2 + \bar{\pi}^2) m_i \pi_i}{v + (\bar{\pi}^2 + \bar{\pi}^2) m_i \pi_i}, \quad m_i = \frac{1}{\pi_i} \sum_{l \leq l-i} \pi_i v_i^2 - z_i \bar{v}^2,
\]

and

\[
\pi_i = \sum_{l \leq l-i} \pi_i, \quad \bar{v} = \frac{1}{\pi_i} \sum_{l \leq l-i} \pi_i v_i,
\]

with \(v_0 = -\sqrt{c_{00}/\pi_0}\) and \(v_j = \sqrt{c_{jj}/\pi_j}\) for \(j \geq 1\).

Proof. Letting \(a_j = v_j \pi_j\), and assuming a correlation matrix (4.10), we may according to (4.7) write the covariance matrix \(\text{Var}(N_{i \infty})\) as

\[
\text{Var}(N_{i \infty}) = v \text{Diag}(\pi) + \bar{\pi}^2 \pi \pi^t + (\bar{\pi}^2 + \bar{\pi}^2) \pi^t \pi,
\]

where \(\pi = (\pi_0, \ldots, \pi_{l-i})^t\), and Diag \((\pi)\) is the diagonal matrix with the elements of \(\pi\) placed in the diagonal. Using the inversion lemma (e.g. SUNDT, 1984, Lemma 6.1)

\[
(Q + f g')^{-1} = Q^{-1} - Q^{-1} f g' Q^{-1} \frac{1}{1 + g' Q^{-1} \bar{g}},
\]

one first calculates the inverse of \(P = v \text{Diag}(\pi) + \bar{\pi}^2 \pi \pi^t\) and subsequently uses the same lemma to invert the matrix \(P + (\bar{\pi}^2 + \bar{\pi}^2) \pi^t \pi\). The result then follows by straightforward calculations from (4.9).

Remark 3. The term \(\pi_j [(1 - z_i) v + z_i \bar{\lambda}_i]\) appearing in Lemma 2 is the credibility estimator for \(N_{ij}\) in the discrete time mixed Poisson model

\[
N_{ij} \mid \Lambda_i \overset{\text{indep}}{\sim} \text{Poisson}(\Lambda_i, \pi_i),
\]

(4.11)

\[E \Lambda_i = v, \quad \text{Var} \Lambda_i = \tau^2,\]

with fixed delay probabilities \(\pi_j\) and a random parameter \(\Lambda_i\) as considered by NORBERG (1986). The second term can therefore be viewed as a correction term,
taking account of the random fluctuations in the delay probabilities induced in the discretizing process.

4.1. The exponential delay distribution

Consider the exponential continuous time delay distribution

$$F(x) = 1 - e^{-\lambda \mu}, \ x \geq 0.$$  

From (2.3) we obtain that

(4.12) \hspace{1cm} p_{i0} = 1 - e^{-\lambda \mu} \Psi_i, \\
(4.13) \hspace{1cm} p_{ij} = e^{-\lambda \mu} (1 - e^{-\lambda \mu}) \Psi_i, \ j = 1, 2, ..., \ \\

where

(4.14) \hspace{1cm} \Psi_i = \int_0^1 e^{\xi \mu} U_i(dx). \\

It is seen from (4.12), (4.13) that the probabilities $p_{ij}$ are linearly dependent, and that the correlation structure (4.10) is exact in this case.

Because (4.10) holds true, we may then use the credibility approach described in Lemma 2, and for the constants $\nu_j$ introduced in Lemma 2 we furthermore observe from (4.12), (4.13) that

(4.15) \hspace{1cm} \nu_j = \frac{\sqrt{\text{Var} \Psi_i}}{\mathbb{E} \Psi_i}, \ j \geq 1, \ \ \nu_0 = -\frac{\sqrt{\text{Var} \Psi_i}}{e^{\lambda \mu} - \mathbb{E} \Psi_i}. \\

Thus, if the parameters $\pi_j$ and $\nu, \tau^2$ of the mixed Poisson model (4.11) are known, it is also possible to calculate the credibility adjustment accounting for random fluctuations in the delay probabilities $p_{ij}$, when the mean $\mathbb{E} \Psi_i$ and the coefficient of variation $\sqrt{\text{Var} \Psi_i}/\mathbb{E} \Psi_i$ are known.

Consider the simple model where

(4.16) \hspace{1cm} \frac{dU_i(x)}{dx} = (1 - b_i)1_{[0, 1]}(x) + (1 + b_i)1_{[1, \infty]}(x), \\

and $b_i$ is a random variable with mean zero. This means that the exposure is (a priori) expected to be constant over the year, but that variations occur such that the actual exposure is $(1 - b_i) \times 100\%$ during the first half of the year and $(1 + b_i) \times 100\%$ during the second half. In this case we find that

$$\Psi_i = \int_0^1 e^{\xi \mu} dx + b_i \left[ \int_{1/2}^1 e^{\lambda \mu} dx - \int_0^{1/2} e^{\xi \mu} dx \right].$$
4.1. The exponential delay distribution

Consider the exponential continuous time delay distribution

\[ F(x) = 1 - e^{-\lambda x}, \quad x \geq 0. \]

From (2.3) we obtain that

\[ p_{\lambda} = 1 - e^{-1/\lambda} \psi_{j}, \quad p_{\lambda} = e^{-\lambda j} (1 - e^{-1/\lambda}) \psi_{j}, \quad j = 1, 2, \ldots. \]

where

\[ \psi_{j} = \int_{0}^{1} e^{\lambda w} U_{j}(dx). \]

It is seen from (4.12), (4.13) that the probabilities \( p_{\lambda} \) are linearly dependent, and that the correlation structure (4.10) is exact in this case.

Because (4.10) holds true, we may then use the credibility approach described in Lemma 2, and for the constants \( \psi_{j} \) introduced in Lemma 2 we furthermore observe from (4.12), (4.13) that

\[ \psi_{j} = \frac{\sqrt{\text{Var} \psi_{j}}}{E \psi_{j}}, \quad j = 1, \ldots, n, \quad \psi_{0} = -\frac{\sqrt{\text{Var} \psi_{0}}}{e^{1/\lambda} - E \psi_{0}}. \]

Thus, if the parameters \( \tau_{j} \) and \( \nu, \nu^{2} \) of the mixed Poisson model (4.11) are known, it is also possible to calculate the credibility adjustment accounting for random fluctuations in the delay probabilities \( p_{\lambda} \), when the mean \( E \psi_{j} \) and the coefficient of variation \( \sqrt{\text{Var} \psi_{j}}/E \psi_{j} \) are known.

Consider the simple model where

\[ \frac{dU_{j}(x)}{dx} = (1 - b_{j}) l_{0}(x) l_{1}(x) + (1 + b_{j}) l_{0}(x) l_{1}(x), \]

and \( b_{j} \) is a random variable with mean zero. This means that the exposure is (a priori) expected to be constant over the year, but that variations occur such that the actual exposure is \( (1 - b_{j}) \times 100\% \) during the first half of the year and \( (1 + b_{j}) \times 100\% \) during the second half. In this case we find that

\[ \psi_{j} = \int_{0}^{1/2} e^{\lambda w} dx + b_{j} \left[ \int_{1/2}^{1} e^{\lambda w} dx - \int_{0}^{1/2} e^{\lambda w} dx \right]. \]

and

\[ E \psi_{j} = \mu (e^{1/\lambda} - 1), \]

\[ \sqrt{\text{Var} \psi_{j}} = \frac{\sqrt{\text{Var} \psi_{0}}}{e^{1/\lambda} - E \psi_{0}}. \]

Remark 4. We may view the above as a quick and dirty method of performing the credibility adjustment in Lemma 2, based on the assumption of an exponential delay distribution. Note that we do not necessarily suggest that the parameters \( \tau_{j} \) are obtained from (4.3) with an exponential delay distribution, since these can be estimated in a straightforward manner from the observed claim numbers \( N_{j} \) in the run-off triangle. The exponential distribution is only used to generate the correlation structure (4.10) as assumed in Lemma 2, and to reduce the number of second order parameters \( c_{ij} \) (or \( \psi_{j} \)) via (4.15). In Section 5 it will be demonstrated by example that the correlation structure (4.10) can safely be assumed even when the delay distribution is not exponential.

The model (4.16) is not essential to the simple approach presented here, and the coefficient of variation \( \sqrt{\text{Var} \psi_{j}}/E \psi_{j} \) can easily be obtained using other models for \( U_{j} \). Adopting (4.16) we suggest to fix the standard deviation \( \sqrt{\text{Var} b_{j}} \) on a purely subjective basis.

4.2. The time continuous Poisson/Gamma model

Recall that the claim occurrences in the continuous time setting are assumed to be generated by a Poisson process with parameter \( \Lambda (\cdot) \). In this section we consider the time continuous Poisson/Gamma model, where \( \Lambda (\cdot) \) is viewed as the realization of a gamma process with parameters \( \gamma (\cdot), \beta \).

The gamma process with parameters \( \gamma (\cdot), \beta \) has independent increments and \( \Lambda (x, \cdot) \) is gamma distributed with shape parameter \( \gamma (s, t) = \gamma (t) - \gamma (s) \) and scale parameter \( \beta \). Since the sample paths of a gamma process are not continuous, the (conditional) Poisson with parameter \( \Lambda (\cdot) \) does not have an intensity.

Remark 5. The Poisson/Gamma process has independent increments since the Poisson process as well as the mixing gamma process have independent increments. With

\[ N(s, t) = \# \{ \tau \mid T_{\tau} \in (s, t) \} \]

denoting the number of occurrences in \( (s, t) \), it follows that the distribution of \( N(s, t) \) is a gamma mixture of Poisson distributions, which yields a negative binomial distribution,

\[ P(N(s, t) = k) = \binom{\gamma (s, t) + k - 1}{k} \beta^{(1 - \beta)\gamma (s, t)} \]

with \( q = (1 + \beta)^{-1} \).
From the distribution of $\Lambda(\cdot)$ we want to derive the distribution of $(\Lambda_i, U_i)$ for $i = 1, 2, \ldots$, and use this to investigate the distribution of the (random) delay probabilities.

Since $\Lambda_i$ and $U_i$ depend only on the increments of $\Lambda(\cdot)$ over year $i$ according to (2.1), we immediately conclude that $(\Lambda_i, U_i)$, $i = 1, 2, \ldots$, are stochastically independent, and since $\Lambda_i$ is the increment of $\Lambda(\cdot)$ over year $i$ we also have that $\Lambda_i \sim \text{Gamma}(\gamma(i-1, i), \beta)$. The distribution $U_i$ is obtained from (2.1) by normalizing the increments of $\Lambda(\cdot)$ over year $i$, and so it follows from FERGUSON (1973, Section 4) that $U_i(\cdot)$ is a Dirichlet process on $(0,1]$ with parameter $\gamma(i-1, i-1+x), x \in [0,1]$, and furthermore that $U_i(\cdot)$ is stochastically independent of $\Lambda_i$.

In particular, $U_i(x)$ has a beta distribution with mean

$$u(x) = \mathbb{E} U_i(x) = \gamma(i-1, i-1+x)/\gamma(i-1, i),$$

and (FERGUSON, 1973, Th. 4)

$$\text{Cov}(U_i(dx), U_t(dy)) = \frac{1}{\gamma(i-1, i) + 1} \{ \delta_{x,y} u(dx) - u(dx) u(dy) \},$$

where $\delta_{x,y}$ equals 1 if $x=y$ and zero otherwise.

In order for the pairs $(\Lambda_i, U_i)$ to be iid (which was previously assumed) we then need to require that $\gamma(i-1, i-1+x)$ and $\gamma(i-1, i)$ are independent of $i$, and expressed in terms of the parameters (4.8) we have that $\gamma(i-1, i) = \nu^2/\tau^2$ and $\beta = \nu/\tau^2$. From (4.3)-(4.5) together with (4.19) we then find the first and second order moments of the discrete time delay probabilities, which become

$$\pi_j = p_{ij} = \int_0^1 h_j(x) u(dx),$$

$$c_{jl}^2 = \frac{\tau^2}{\nu^2 + \tau^2} \left\{ \int_0^1 h_j(x) h_l(x) u(dx) - \pi_j \pi_l \right\},$$

where $h_j(x) = F(j+1-x) - F(j-x)$ was introduced in (4.2).

Notice that the bracket in (4.21) can be written as $\text{Cov}(h_j(Y), h_l(Y))$, where $Y$ is a random variable with distribution $u(\cdot)$ on $(0,1]$. The delay probabilities $p_{ij}$ and $p_{il}$ are therefore positively correlated if both $h_j$ and $h_l$ are either increasing or decreasing, and are negatively correlated if one is increasing while the other is decreasing. It is seen from (4.2) that $h_0(x)$ is always decreasing, and $h_j(x)$ for $j \geq 1$ is certainly increasing if $F$ has a decreasing density. This result matches assertion (b) in Lemma 1.

**Example 1.** The (American) Pareto distribution with parameters $(\eta, \alpha)$ has cdf (see e.g. HOGG and KLUGMAN, 1984)

$$F(x) = 1 - \left( \frac{\eta}{\eta + x} \right)^\alpha, \quad x \geq 0,$$
a (decreasing) density

\[ f(x) = \frac{\alpha \eta^x}{(\eta + x)^{\alpha+1}}, \quad x \geq 0, \]

and a mean \( \mu = \frac{\eta}{\alpha - 1} \). It is a noteworthy property of this distribution that when \( X \) is Pareto distributed with parameters \((\eta, \alpha)\), then the conditional distribution of \( X - y \) given that \( X > y \) is again Pareto with parameters \((\eta + y, \alpha)\). The mean residual lifetime (MRL) then becomes

\[
m(y) = E(X - y | X > y) = \frac{\eta + y}{\alpha - 1} = \mu + \frac{1}{\alpha - 1} y.
\]

Benktander and Segerdahl (1960) suggested to use the MRL as a useful tool for investigating the tail of a severity distribution.

With a Pareto \((\eta, \alpha)\) delay distribution, and \( u(x) = x, \ x \in [0, 1] \), the average delay distribution is calculated from (2.7) with the stop-loss transform given by

\[
\Pi(x) = (1 - F(x)) m(x) = \mu \left( \frac{\eta}{\eta + x} \right)^{\alpha - 1},
\]

whereas the parameters \( \pi_{ij} \) appearing in (4.21) have to be calculated numerically.

5. NUMERICAL RESULTS

For a portfolio of accident policies we have observed all claim occurrences between 1/1/82 and 31/12/90, which have been reported before 3/3/92. The data have been discretized according to calendar period, using an interval length of 3 month. Table 1 shows the run-off triangle by 31/12/90 containing the numbers of reported disability claims, and since the portfolio has been observed until 3/3/92 we are also able to construct the claim numbers which were eventually reported in this case.

<table>
<thead>
<tr>
<th>Run-off triangle</th>
<th>Actual reportings</th>
</tr>
</thead>
<tbody>
<tr>
<td>delay ( j )</td>
<td>delay ( j )</td>
</tr>
<tr>
<td>0    1    2    3    4    5</td>
<td>0    1    2    3    4    5</td>
</tr>
<tr>
<td>0</td>
<td>72    35    7    4    3    0</td>
</tr>
<tr>
<td>1</td>
<td>71    35    6    3    2    ---</td>
</tr>
<tr>
<td>2</td>
<td>69    42    4    4    ---    ---</td>
</tr>
<tr>
<td>3</td>
<td>70    31    9    ---    ---    ---</td>
</tr>
<tr>
<td>4</td>
<td>67    31    ---    ---    ---    ---</td>
</tr>
<tr>
<td>5</td>
<td>55    ---    ---    ---    ---    ---</td>
</tr>
</tbody>
</table>

TABLE 1

Disability claims reported by 31/12/90, and later reportings

These are shown in Table 1 as well. For the period 1/1/82-31/12/90 there has been reported a total of 4015 disability claims, and the average reporting delay for these claims was \( \mu = 0.91 \) (with a time unit of 3 month).
In the mixed Poisson model (4.11) we have estimated the delay probabilities \( \pi_j \) on the basis of all the observations for the period 1/1/82-31/12/90, and the estimated probabilities \( \pi_j \) are shown for \( j = 0, ..., 5 \) in Table 2. Finally, using the method of moments, we have estimated the parameters (4.8) obtaining the results
\[
\nu = 110.5, \quad \tau^2 = 164.0
\]

In this mixed Poisson model, the credibility predictions for the outstanding claim numbers are calculated in accordance with Remark 4, and the result is shown in Table 3 as the first set of predictions.

### Table 2
**Discrete Time Delay Probabilities in %**

<table>
<thead>
<tr>
<th>( j )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_j )</td>
<td>58.03</td>
<td>29.27</td>
<td>4.72</td>
<td>2.38</td>
<td>1.57</td>
<td>0.69</td>
</tr>
</tbody>
</table>

### Table 3
**Credibility Predictions**

<table>
<thead>
<tr>
<th>( )</th>
<th>&quot;Ordinary&quot; credibility predictions</th>
<th>Credibility predictions based on Lemma 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>delay ( j )</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>30</td>
</tr>
</tbody>
</table>

By inspection of the data in Table 1 it is seen that the observed number of immediate reportings in respect of period 6 is significantly below average. In the mixed Poisson model this will be interpreted as a result of a low-risk period, and the predictions will be correspondingly low. An alternative explanation would be that the claims in this period have occurred later than expected by the model, and that a smaller proportion (than expected) of the incurred claims have therefore been reported already in the occurrence period. Comparing with the actual reportings in Table 1 it is seen that this was probably the case, since the number of reported claims with delay \( j = 1 \) is above average for that period.

In order to account for fluctuations due to the discretization we apply the credibility method in Lemma 2 using the simple approach described in Section 4.1. With a standard deviation \( \sqrt{\text{Var} b_j} = 0.2 \) we obtain the predictions as shown in Table 3. Comparing with Table 1 it is seen that the credibility method based on Lemma 2 performs better in this case, even though both methods underestimate the number \( N_{6,1} \).
Lemma 2 is based on the assumption of a correlation structure (4.10), which from Section 4.2 is known to be correct in the case of an exponential delay distribution. For the data considered here we have access to the continuous time delays \( W_n, n = 1, \ldots, 4015 \), and in Figure 2 we have shown the empirical MRL for these observations, the MRL for the Pareto distribution considered in Example 1, and we have also fitted the curve

\[
m_1(x) = \mu (1 + x/a)^{1-c},
\]

obtaining the values \( \hat{\mu} = 0.409 \) and \( \hat{c} = 0.226 \). It is seen that this gives an almost perfect fit, and the corresponding distribution, a shifted version of Benktander's type II distribution (see Beard et al., p. 82), has cdf

\[
F(x) = 1 - \left( \frac{a}{a + x} \right)^{1-c} e^{-\frac{a}{\hat{\mu}} (1 + x/a)^{-1}}.
\]

In Figure 3 we have also plotted the empirical MRL for reporting delays for dental claims from the same portfolio. It is seen that the Pareto distribution from Example 1 gives a adequate description in this case, and the estimated parameters are \( \mu = 1.14 \) (with a time unit of 1 month in this case) and \( \hat{\alpha} = 1.343 \).

For these two delay distributions we have calculated the coefficients of correlation assuming the time continuous Poisson/Gamma model treated in Section 4.2. This
gives the results shown in Table 4, and it is seen that even though the continuous
time delay distributions in these cases are far from exponential, the correlation
matrix is well approximated by (4.10) as assumed in Lemma 2.

**TABLE 4**

Correlation matrix. Benktander type II delay distribution for disability claims (upper table)
and Pareto delay distribution for dental claims (lower table)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.999</td>
<td>-0.991</td>
<td>-0.982</td>
<td>-0.977</td>
<td>-0.973</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.984</td>
<td>0.974</td>
<td>0.967</td>
<td>0.963</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.999</td>
<td>0.997</td>
<td>0.995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.999</td>
<td>-0.982</td>
<td>-0.969</td>
<td>-0.960</td>
<td>-0.955</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.976</td>
<td>0.960</td>
<td>0.951</td>
<td>0.945</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.998</td>
<td>0.995</td>
<td>0.993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.** The MRL for the continuous time delay distribution, dental claims.
ACKNOWLEDGEMENT

I am grateful to WALTHER NEUHAUS for giving several useful comments and for correcting errors in an earlier version of the manuscript.

REFERENCES


OLE HESSELAGER
University of Copenhagen
Laboratory of Actuarial Mathematics
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
ON THE ESTIMATION OF THE CREDIBILITY FACTOR:  
A BAYESIAN APPROACH*

BY RENÉ SCHNIEPER  
"Zürich " Versicherungs-Gesellschaft, Zürich

ABSTRACT

In practical applications of Credibility Theory the structure parameters usually have 
to be estimated from the data. This leads to an estimator of the a posteriori mean 
which is often biased and where the credibility factor depends on the data. A more 
coherent approach to the problem would be to also treat the unknown parameters as 
random variables and to simultaneously estimate the a posteriori mean and the 
structure parameters. Different statistical models are proposed which allow for such 
a solution. These models all lead to an estimation of the posterior mean which is a 
weighted average of the prior mean and of the observed mean, the weights 
depending on the observations.

KEYWORDS

Credibility theory; Structure parameters; Statistical estimation; exponential famil-
ies; Pricing of risks.

1. INTRODUCTION

We have k different risks with a claims record over a certain number of years

\[ X_{11}, X_{12}, ..., X_{1n_1} \]
\[ X_{21}, X_{22}, ..., X_{2n_2} \]
\[ \vdots \]
\[ X_{k1}, X_{k2}, ..., X_{kn_k} \]

Depending on the specific problem, the data are numbers of claims from different 
insurance policies, loss ratios from insurance portfolios for instance in fire, liability 
or workmen's compensation insurance and burning costs from different reinsurance 
treaties. 

With each claims record \( X_{ij} \) there is an associated measure of risk exposure \( p_{ij} \) 
which is a number of risk years, a sum insured, a turnover, a total amount of wages 
or a premium income depending on the specific problem.

* Paper presented at the XXIVth ASTIN Colloquium in Cambridge.
Within the framework of credibility theory it is assumed that
1. Each risk $i$ is characterized by an individual unknown risk parameter $\theta_i$; the risk parameters $\theta_1$, $\theta_2$, ..., $\theta_k$ are i.i.d. random variables. The random vectors $(\theta_i, X_{i1}, ..., X_{in})$ $i = 1$, ..., $k$ are independent.
2. Given $\theta_i$ the observations

$$X_{i1}, X_{i2}, ..., X_{in},$$

are independent with finite second moments

$$E(X_{ij} \mid \theta_i) = \mu(\theta_i)$$

$$\mathrm{Var}(X_{ij} \mid \theta_i) = \frac{\sigma^2(\theta_i)}{p_{ij}}$$

The individual premium $\mu(\theta_i)$ is to be approximated by a premium which is linear in the observations

$$\hat{\mu}(\theta_i) = \alpha_{i0} + \sum_{j=1}^{n_i} \alpha_{ij} X_{ij}$$

and which minimizes the expected squared error

$$E[(\mu(\theta_i) - \hat{\mu}(\theta_i))^2]$$

It is shown that the optimal linear premium is a weighted average of the individual mean and of the a priori mean

$$\hat{\mu}(\theta_i) = z_i \bar{x}_i + (1 - z_i) m$$

where

$$m = E[\mu(\theta)]$$

$$\bar{x}_i = \left( \frac{\sum_{j=1}^{n_i} p_{ij} x_{ij}}{\sum_{j=1}^{n_i} p_{ij}} \right)$$

The weight $z_i$ given to the individual mean $\bar{x}_i$ is called the credibility factor. It is equal to

$$z_i = \frac{p_i b}{p_i b + w}$$

where

$$p_i = \sum_{j=1}^{n_i} p_{ij}$$

$$b = \mathrm{Var}[\mu(\theta)]$$

$$w = E[\sigma^2(\theta)]$$

$b$ and $w$ are the between risks and within risks variance components respectively. To practically compute the credibility premium we need to estimate the structure parameters $m$, $b$ and $w$. 
m is either known a priori e.g. from some nationwide statistics or is estimated from the collective. The best linear unbiased estimator for m is

\[ \hat{m} = \sum_{i=1}^{k} \frac{Z_i}{Z} \bar{x}_i \]

with

\[ Z = \sum_{i=1}^{k} z_i \]

In what follows we shall assume that m is known a priori and focus on the estimation of \( \theta \) and \( \omega \). [In Klugman (1986) m is assumed to be a random variable and included in the Bayesian analysis].

In their pioneering paper Bühlm and Straub (1970) propose the following "natural" estimators.

\[ \hat{\omega} = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{n_i - 1} \sum_{j} p_{ij} (x_{ij} - \bar{x}_i)^2 \]

\[ \hat{\theta} = \frac{1}{c} \left( \sum_{i=1}^{k} \frac{p_i}{\bar{x}_i - \bar{x}}^2 - (k - 1) \frac{\hat{\omega}}{\bar{p}} \right) \]

where

\[ p = \sum_{i} p_i, \]

\[ \bar{x} = \left( \sum_{i} p_i \bar{x}_i \right) / p \]

\[ c = \sum_{i} \frac{p_i}{p} \left( 1 - \frac{p_i}{p} \right) \]

Since the estimator of the between risks variance component may be negative, it is replaced by \( \max (\hat{\theta}, 0) \) in practical applications.

Other estimators have also been proposed; for a review of the subject see for instance Dubey and Gisler (1981). A common property of these estimators is that they usually lead to a biased estimator of \( \mu (\theta) \).

We shall adopt a different approach. Since \( \theta \) and \( \omega \) are unknown quantities we shall treat them as random variables and we shall propose statistical models which allow for a simultaneous estimation of \( \theta \), \( \omega \) and \( \mu (\theta) \). The credibility factor will depend upon the observations, but this is also the case for the Bühlmann-Straub estimator once \( \theta \) and \( \omega \) have been replaced by \( \hat{\theta} \) and \( \hat{\omega} \).

A general discussion of Bayesian inference on variance components in a normal model as well as in a generalisation of the normal model can be found in Box and Tiao (1973).

The present paper focuses on applications to credibility theory.
2. THE BASIC MODEL

We assume that the data described at the outset is generated by the following model:
- the individual premiums $\mu_i$ stem from a normal distribution with unknown precision $\pi$;
- the claims record of risk $i$ $X_{i1}, X_{i2}, \ldots, X_{in_i}$ stems from a normal distribution with mean $\mu_i$ and unknown precision $\rho$.

Note that instead of emphasizing the variance of a random variable, we emphasize its precision. This model is best illustrated by the following picture where the realizations of random variables are represented by drawings from an urn.
Conventions and notation

— Densities are indexed by their arguments; thus $p(\pi)$ and $p(\rho)$ do not necessarily represent the same density.

— Distribution laws of random variables are symbolized by $p(\cdot)$. However we do not necessarily assume that the density of a random variable exists.

— $\varphi(m, \sigma^2)$ denotes the normal density with mean $m$ and variance $\sigma^2$.

The formal model assumptions are as follows
1) The precision $\pi$ of the individual premiums and the precision $\rho$ of the observations are random variables whose distribution function will be specified later.
2) Each risk is characterized by an individual premium $\mu_i$ and a common precision $\rho$.
   i) $\mu_1, \mu_2, ..., \mu_k$ are independent random variables given $\pi$; their common density is $\varphi(m, \pi^{-1})$.
   ii) $(\mu_i, x_{i1}, ..., x_{in})$ are independent random vectors given $\pi$ and $\rho$.
   iii) $\mu_1, \mu_2, ..., \mu_k$, are independent random variables given $\pi$.
3) Given $(\mu_1, \rho), x_{ij} (j = 1, ..., n_i)$ are independent random variables with common density $\varphi(\mu_i, \rho^{-1})$.

Remarks
1) If the distribution of $\pi$ and $\varrho$ were degenerated [i.e. if the probability mass of the common distribution of $\pi$ and $\varrho$ is concentrated in some possibly unknown points $(\pi_0, \varrho_0)$] the above model would be a special case of the credibility model.
2) A similar model, but with individual precisions $\rho_{ij}$ for each observation, has been proposed by Jewell and Schnieper (1983) for the treatment of outliers.

We now turn to the problem of the estimation of the individual premium. Let $D$ denote the set of the claims records from all individual risks; the best estimator of the individual premium (best in the sense that it minimizes the expected squared error) is the posterior mean.

$$E(\mu_i | D) = \int E(\mu_i | D, \rho, \pi) p(\rho, \pi | D) d\rho \cdot d\pi$$

Because of assumptions 2) and 3) we have

$$E(\mu_i | D, \rho, \pi) = \frac{\pi m + \rho n_i \bar{x}_i}{\pi + \rho n_i} = x_i^*$$

because the credibility formula is exact in the case of a normal likelihood (with known variance) and a normal prior.

$\bar{x}_i$ is as in section 1 but with all measures of exposure equal to one.
Therefore we have

$$E(\mu_1 | D) = E \left( \frac{\rho n_i}{\rho n_i + \pi} \bigg| D \right) \cdot \bar{x}_i + E \left( \frac{\pi}{\rho n_i + \pi} \bigg| D \right) \cdot m$$

$$E(\mu_1 | D) = z_1(D) \bar{x}_i + [1 - z_1(D)] \cdot m$$

The posterior mean is given by a credibility type formula where the credibility factor depends on the data. To determine the credibility factor and the posterior mean, we must determine the posterior distribution of $\pi$ and $\rho$ given the data.

Using Bayes' Theorem we have

$$p(\rho, \pi | D) \propto p(D | \rho, \pi)p(\rho, \pi)$$

The common density of $\rho$ and $\pi$ will be specified later; for the likelihood we obtain

$$p(D | \rho, \pi) = \int p(D | \mu, \rho, \pi)p(\mu | \rho, \pi)d\mu_1 \ldots d\mu_k$$

where

$$\mu = (\mu_1, \mu_2, \ldots, \mu_k).$$

From assumption 2) ii) and 3) it follows

$$p(D | \mu, \rho, \pi) \propto \rho^{\frac{k}{2}} \cdot a_i^{n_i} \cdot e^{-\frac{1}{2} \rho \sum_{i} (x_i - \mu_i)^2}$$

and from 2) i)

$$p(\mu | \rho, \pi) \propto \frac{k}{\pi^2} \cdot e^{-\frac{1}{2} \pi \sum_{i} (\mu_i - m_i)^2}$$

therefore the likelihood becomes

$$p(D | \rho, \pi) \propto \pi^2 \cdot \rho^{\frac{k}{2}} \cdot \frac{1}{\pi^2} \cdot \frac{1}{\pi^2} \cdot e^{- \frac{1}{2} \pi \sum_{i} (x_i - (\pi + \rho n_i) \mu_i + \pi m_i)^2} \cdot \prod_{i} \left( \pi + \rho n_i \right)$$

and upon integration we obtain

$$p(D | \rho, \pi) \propto \frac{k}{\pi^2} \rho^{\frac{k}{2}} \cdot e^{-\frac{1}{2} \rho \sum_{i} (x_i - (\pi + \rho n_i) \mu_i + \pi m_i)^2}$$

from which it is seen that

$$\sum_j x_{ij} \text{ and } \sum_j x_{ij}^2 \quad i = 1, \ldots, k$$
are sufficient statistics.

After some straightforward but tedious algebraic transformations the likelihood can be written in an intuitively more appealing form

\[ p(D \mid \rho, \pi) \propto \frac{1}{\pi^2 \rho^2} \frac{1}{2} \sum_{i=1}^{k} \frac{1}{\pi^2 (\pi - m)^2 + \pi^2 (\pi - \pi_i)^2} \prod_{i}^{k} (\pi + \rho \pi_i) \]

from which we can compute the posterior density of \( \rho \) and \( \pi \) (once the prior density has been specified) and the posterior mean. The reason why the problem remains tractable is because \( p(D \mid \rho, \pi) \) is in analytical form; this in turn is due to the fact that for given \( \pi \) and \( \rho \) \( p(D \mid \mu, \rho, \pi) \) and \( p(\mu \mid \rho, \pi) \) are conjugate priors.

Remark

The following relation is true in general

\[ E[\mu_i \mid D] = \int E[\mu_i \mid D, \rho, \pi] p(\rho, \pi \mid D) d\rho, d\pi, \]

and from credibility theory we obtain

\[ E[\mu_i \mid D, \rho, \pi] = \frac{\pi \pi_i + \rho \pi_i}{\pi + \rho} \]

independently of the above distributional assumptions.

Therefore the form of the optimal estimator

\[ E[\mu_i \mid D] = z(D) \bar{x}_i + (1 - z(D)) m \]

is independent of the distributional assumptions; these are only needed to compute the credibility factor

\[ z(D) = \int \frac{\rho \pi_i}{\pi + \rho \pi_i} p(\rho, \pi \mid D) d\rho d\pi. \]

3. A NUMERICAL ILLUSTRATION

We illustrate the results of section 2 with the following very simple numerical example

\[ m = 0, \quad \pi = \frac{1}{3} \]

\[ \rho = \begin{cases} 1 & \text{with probability 0.333} \\ 0.001 & \text{with probability 0.667} \end{cases} \]

\[ E(\rho) = 0.334 \]
There is only one risk and let us also assume that there is only one observation $x$.

$$
\rho \mid D = \begin{cases} 
1 & \text{with probability } p(D \mid \rho = 1) \cdot 0.333 \\
0.001 & \text{with probability } p(D \mid \rho = 0.001) \cdot 0.667 
\end{cases}
$$

with

$$
p(D \mid \rho) \propto \left( \frac{\rho}{0.333 + \rho} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \left( 0.333 x^*(\rho)^2 + \rho | x^*(\rho)| \right)}
$$

$$
x^*(\rho) = \frac{\rho x}{0.333 + \rho} = \begin{cases} 
0.750 x & \text{for } \rho = 1 \\
0.003 x & \text{for } \rho = 0.001 
\end{cases}
$$

As $x$ becomes "large", i.e. deviates strongly from $m$ the whole probability mass is shifted towards $\rho = 0.001$ and $E(\mu \mid D)$ converges towards $0.003 x$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(Q = 0.001 \mid D)$</th>
<th>$z(D)$</th>
<th>$E(\mu \mid D)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.112</td>
<td>0.666</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.125</td>
<td>0.657</td>
<td>0.657</td>
</tr>
<tr>
<td>2</td>
<td>0.172</td>
<td>0.621</td>
<td>1.243</td>
</tr>
<tr>
<td>3</td>
<td>0.279</td>
<td>0.541</td>
<td>1.624</td>
</tr>
<tr>
<td>4</td>
<td>0.481</td>
<td>0.391</td>
<td>1.564</td>
</tr>
<tr>
<td>5</td>
<td>0.739</td>
<td>0.198</td>
<td>0.988</td>
</tr>
<tr>
<td>6</td>
<td>0.918</td>
<td>0.064</td>
<td>0.387</td>
</tr>
<tr>
<td>7</td>
<td>0.983</td>
<td>0.016</td>
<td>0.112</td>
</tr>
<tr>
<td>8</td>
<td>0.997</td>
<td>0.005</td>
<td>0.040</td>
</tr>
<tr>
<td>9</td>
<td>1.000</td>
<td>0.003</td>
<td>0.029</td>
</tr>
<tr>
<td>10</td>
<td>1.000</td>
<td>0.003</td>
<td>0.030</td>
</tr>
</tbody>
</table>

where

$$
z(D) = \left( \frac{Q}{0.333 + Q} \right) \mid D
$$

$$
E(\mu \mid D) = z(D) \cdot x + [1 - z(D)] \cdot m = z(D) \cdot x
$$

If we take two observations instead of one a more interesting picture emerges. Two "large" observations are given a high credibility factor because the within variance component is low, whereas one "small" and one "large" observation are given a low factor. The result is best illustrated by the contour plot of the posterior mean where $X_1$ and $X_2$ are the two observations and the "estimate" is the posterior mean. (See Appendix 1).

4. THE SIMPLE EXPONENTIAL FAMILY

We now show that the method used to derive simultaneous estimations of the individual premium and of the hyperparameters can be applied to the whole simple exponential family. We use some of the results of Jewell (1974).
Before defining the new model, we simplify our parameterization. In the basic model, the a posteriori individual premium for risk $i$, given the precisions, is

$$E(\mu_i \mid D, \rho, \pi) = \frac{\pi \cdot m + \rho n_i \bar{x}_i}{\pi + \rho n_i}$$

which can be rewritten in the following way

$$E(\mu_i \mid D, \rho, \pi) = \frac{vm + n_i \bar{x}_i}{v + n_i}$$

with

$$v = \frac{\pi}{\theta}$$

and it is apparent that only the ratio of $\pi$ over $\rho$ is relevant, and not the two variance components separately.

We shall refer to $v$ as to the time constant. It is equal to the number of individual claims records necessary for the credibility factor to be equal to one half.

This new parametrisation is more "natural" than the one introduced in section 2 since in practical applications one often has a priori information on the credibility factor and therefore on the time constant but not necessarily on the two variance components separately. The reason for using both $\pi$ and $\rho$ in section 2 is mathematical tractability.

The exponential family likelihoods with the sample mean as sufficient statistics is

$$p(x \mid \theta) = a(x)c(\theta)^{-1} e^{-\theta x}$$

and their natural conjugate priors are

$$p(\theta) = d(n_0, x_0)^{-1} c(\theta)^{-n_0} e^{-\theta x_0}.$$  

Jewell (1974) shows that under certain regularity conditions

$$E[\mu(\theta)] = \frac{x_0}{n_0} = m$$

The above family is closed under sampling, so that if we observe $x_{i1}, ..., x_{in}$, for fixed $\theta_i$ then $p(\theta_i \mid D)$ is of the same form with new parameters

$$n'_0 = n_0 + n_i$$

$$x'_0 = x_0 + \sum_j x_{ij}$$

It follows that

$$E[\mu(\theta_i) \mid D] = \frac{x_0 + \sum_j x_{ij}}{n_0 + n_i} = \frac{n_0 m + n_i \bar{x}_i}{n_0 + n_i}$$
We now generalize the model by introducing a random time constant. The model assumptions are the following:

1) The time constant $\nu$ is a random variable whose distribution function will be specified later.

2) Each risk is characterized by an individual parameter $\theta_i$.
   i) $\theta_1, \theta_2, \ldots, \theta_k$ are independent random variables given $\nu$; their common density is $p(\theta | \nu) = d(\nu, \nu m)^{-1} c(\theta)^{-1} e^{-\theta \nu m}$
   ii) $(\theta_i, X_{i1}, \ldots, X_{in_i}) i = 1, \ldots, k$ are independent random vectors given $\nu$.

3) Given $\theta_i, X_{ij} (j = 1, \ldots, n_i)$ are independent with common density $p(x | \theta) = a(x) c(\theta)^{-1} e^{-\theta x}$

The posterior mean now is

$$E(\mu(\theta_i | D) = \int E(\mu(\theta_i | D, \nu) p(\nu | D) d\nu$$

where

$$E[\mu(\theta_i | D, \nu) = \frac{\nu m + n_i \bar{x}_i}{\nu + n_i}$$

and therefore

$$E[\mu(\theta_i | D) = E\left(\frac{n_i}{\nu + n} \mid D\right) \bar{x}_i + E\left(\frac{\nu}{\nu + n_i} \mid D\right) m$$

$$= z(D) \cdot \bar{x}_i + [1 - z(D)] \cdot m$$

On the other hand

$$p(\nu | D) = p(D | \nu) p(\nu) \cdot p(\nu)^{-1}$$

and

$$p(D | \nu) = \int p(D | \nu, \theta) p(\theta | \nu) d\theta_1 \cdot d\theta_2 \cdots d\theta_k$$

$$\propto \prod_{i=1}^k \int c(\theta_i)^{-n_i} e^{-\theta_i \sum x_i} d^{-1}(\nu, \nu m) c(\theta_i)^{-\nu} e^{-\theta_i \nu m} d\theta_i$$

$$\propto d^{-k}(\nu, \nu m) \prod_{i=1}^k \int c(\theta_i)^{-(\nu + n_i)} e^{-\theta_i (\nu m + \sum x_i)} d\theta_i$$

$$\propto \prod_{i=1}^k \frac{d\left(\nu + n_i, \nu m + \sum x_{ij}\right)}{d(\nu, \nu m)}$$
As in our basic model, \( p(D \mid v) \) is in analytic form. This is so because for given \( v \)
\( p(D \mid \theta, v) \) and \( p(\theta \mid v) \) are conjugate priors.
We illustrate our general result with the following.

**Example**

We assume a Poisson likelihood and gamma prior.
\[
p(x \mid \theta) = e^{-\theta} \frac{\theta^x}{x!} \quad x = 0, 1, \ldots
\]
\[
p(\theta \mid v) = \frac{v^m}{\Gamma(mv)} \theta^{vm-1} e^{-\theta v} \quad \theta > 0
\]

and we have
\[
p(D \mid v) = \prod_{i, j} \frac{1}{x_{ij}!} \left( \frac{v^m}{\Gamma(vm)} \right)^k \prod_{i=1}^k \frac{\Gamma(vm + \sum_j x_{ij})}{\Gamma(v + n_i) \Gamma(vm + \sum_j x_{ij})}
\]

We have five risks, each with recorded number of claims for two years
\[
D = \begin{bmatrix}
2 & 1 \\
1 & 1 \\
0 & 1 \\
0 & 0 \\
0 & 0
\end{bmatrix}
\]

using the estimators given in section 1, we have
\[
\hat{\omega} = 0.2
\]
\[
b = 0.325
\]

and the time constant is
\[
\hat{\nu} = \frac{\hat{\omega}}{b} = 0.615
\]

If we compute the likelihood function \( p(d \mid v) \) another picture emerges
\[
\hat{\nu}_{MLE} = 9
\]

and the likelihood is very flat for \( v \) larger than 2, as is apparent from the graph in appendix 2 where we have assumed \( m = 0.6 \). From the shape of the likelihood it is obvious that \( p(v \mid D) \) and therefore \( z(D) \) will heavily depend on the choice of the prior \( p(v) \).
5. ARBITRARY MEASURES OF EXPOSURE – THE BÜHLMANN-STRAUB EXAMPLE

Our basic model can be generalized to the case where the measures of exposure are arbitrary. We make the same assumptions as in section 2 except for assumption 3) which now reads: given \((\mu, \rho, X_j)\) \((j = 1, \ldots, n_i)\) are independent random variables with common density \(\varphi(\mu, (\rho_j \rho)^{-1})\). \(p_j\) are known measures of exposure.

The posterior mean is computed as in section 2.

\[
E(\mu_j | D) = \int E(\mu_j | D, \rho, \pi) \rho(\rho, \pi | D) d\rho d\pi
\]

but now we have

\[
E(\mu_j | D, \rho, \pi) = \frac{\pi m + \rho \bar{p}_j \bar{x}_i}{\pi + \rho \bar{p}_j} = x_i^*\]

with \(p_i\) and \(\bar{x}_i\) as in section 1. The posterior distribution of \(\rho\) and \(\pi\) is formally as in section 2.

\[
p(\rho, \pi | D) \propto p(D | \rho, \pi) p(\rho, \pi)
\]

but the likelihood now is

\[
p(D | \rho, \pi) \propto \frac{1}{\pi^2 \rho^2} \left( \frac{1}{\pi \rho} \right)^{n_i} \cdot \exp \left\{ -\frac{1}{2} \left[ \pi \left( \sum \Sigma (x_i - m)^2 + \rho \sum \Sigma p_j (x_i - x_i^*)^2 \right) \right] \right\} \prod \left( \pi + \rho \bar{p}_j \right)
\]

The proofs are as in section 2. We now reanalyze the data of BÜHLMAN-STRAUB (1970). In their paper the authors give as-if burning costs of seven different excess of loss treaties. For each treaty we have the burning costs from five different years and each treaty is characterized by a measure of exposure and the gross premium income.

The burning costs (in percent of the gross premium income) are as follows:

<table>
<thead>
<tr>
<th>Treaty</th>
<th>Year</th>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>4.2</td>
<td>0.0</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.3</td>
<td>25.0</td>
<td>18.5</td>
<td>14.3</td>
<td>30.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8.0</td>
<td>1.9</td>
<td>7.0</td>
<td>3.1</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5.4</td>
<td>5.9</td>
<td>7.1</td>
<td>7.2</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9.7</td>
<td>8.9</td>
<td>6.7</td>
<td>10.3</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>9.7</td>
<td>14.5</td>
<td>10.8</td>
<td>12.0</td>
<td>13.1</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>9.0</td>
<td>9.6</td>
<td>8.7</td>
<td>11.7</td>
<td>7.0</td>
</tr>
</tbody>
</table>
ON THE ESTIMATION OF THE CREDIBILITY FACTOR: A BAYESIAN APPROACH

and the gross premium income (in some monetary unit) are

<table>
<thead>
<tr>
<th></th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>70</td>
</tr>
</tbody>
</table>

Bühlmann-Straub compute \( \hat{w} \) and \( \hat{b} \) using the formula given in section 1; they obtain

\[ \hat{w} = 209.0 \cdot 10^{-4} \]
\[ \hat{b} = 12.1 \cdot 10^{-4} \]

which gives the following estimates for the precisions

\[ \rho = \hat{w}^{-1} = 48 \]
\[ \hat{\pi} = \hat{b}^{-1} = 826 \]

the time constant being

\[ \hat{\nu} = \frac{\hat{\pi}}{\hat{\rho}} = 17 \]

Instead of computing a point estimate we look at \( p(D | \pi, \rho) \). The contour plot of the likelihood is given in appendix 3.

It is seen that the maximum likelihood estimator is approximately

\[ \hat{\rho}_{ml} \approx 50 \]
\[ \hat{\pi}_{ml} \approx 500 \]

thus giving a much smaller time constant

\[ \hat{\nu} \approx 10 \]

Thus the impact of the variance estimates on the credibility premium of a small treaty can be quite important. A full bayesian analysis would entail specifying a joint prior distribution for \( \pi \) and \( \rho \) and computing the posterior mean through numerical integration. Since this is straightforward it is omitted here.
REFERENCES


RENÉ SCHNIEPER
“Zürich” Versicherungs-Gesellschaft
Postfach, CH-8022 Zürich

Appendix 1
Plot of estimate
Appendix 2
Plot of I_khd vs \( \nu \)

Appendix 3
Plot of I_khd
SOME STABLE ALGORITHMS IN RUIN THEORY
AND THEIR APPLICATIONS

BY

DAVID C. M. DICKSON
The University of Melbourne

ALFREDO D. EGÍDIO DOS REIS
ISEG, Lisbon

HOWARD R. WATERS
Heriot-Watt University, Edinburgh

ABSTRACT

In this paper we present a stable recursive algorithm for the calculation of the probability of ultimate ruin in the classical risk model. We also present stable recursive algorithms for the calculation of the joint and marginal distributions of the surplus prior to ruin and the severity of ruin. In addition we present bounds for these distributions.

KEYWORDS

Probability of ruin; Severity of ruin; Surplus prior to ruin; Recursive calculation; Stable algorithm; Compound binomial model.

1. INTRODUCTION

In this paper we present recursive algorithms for the (approximate) numerical calculation of various quantities for a classical surplus process. These quantities are the probability of ultimate ruin, the distribution of the severity of ruin, the moments of the severity of ruin, the distribution of the surplus immediately prior to ruin and the joint distribution of the surplus immediately prior to ruin and the severity of ruin. Recursive algorithms for the calculation of some of these quantities have already appeared in the actuarial literature, particularly for the probability of ultimate ruin. However, not all of these algorithms are numerically stable. The stability of recursive algorithms has been discussed by Panjer and Wang (1993) and, in their words, “For unstable recursions, alternative methods of evaluation merit further research”. The main purpose of this paper is to present stable
algorithms. In addition we present bounds and approximations to the (defective) distributions of the severity of ruin and the surplus immediately prior to ruin, and for the (defective) joint distribution of these two quantities.

Our general method for producing algorithms is to approximate the classical surplus process by a discrete process (discrete time and discrete claim amount distribution) and then to derive an algorithm for the appropriate quantity for the discrete model. The discrete model we will be using is an example of a compound binomial model, studies of which have already appeared in the actuarial literature [Gerber (1988), Shiu (1989), Willmot (1992) and Dickson (1994)]. Hence, although in this paper we will regard our algorithms as providing approximations to, for example, the probability of ultimate ruin for a (continuous time) classical surplus process, we could have chosen to regard them as providing exact values for a compound binomial model.

In the next section we introduce the basic continuous time surplus model, the discrete approximation to the basic model and some notation. In Section 3 we discuss the probability of ultimate ruin. In particular, we discuss the stability of some algorithms which have appeared in the actuarial literature, present a new stable algorithm and discuss numerical examples. In Section 4 we consider the calculation of the distribution of the severity of ruin. In Section 5 we use the algorithm presented in Section 4 to derive an algorithm for the calculation of the moments of the severity of ruin. Both the probability of ruin and, perhaps to a lesser extent, the severity of ruin are of obvious interest. Our reasons for considering also the moments of the severity of ruin are that these moments are of some interest in their own right and that these moments can be used to calculate the moments of durations of negative surplus, as shown by dos Reis (1993). Finally, in Section 6 we discuss the distribution of the surplus prior to ruin and the joint distribution of the surplus immediately prior to ruin and the severity of ruin.

2. MODELS AND NOTATION

Let \( \{U(u, t)\}_{t \geq 0} \) be a classical continuous time surplus process, so that

\[
U(u, t) = u + ct - \sum_{i=1}^{N(t)} X_i
\]

where:

- \( u \) is the insurer's initial surplus,
- \( c \) is the insurer's rate of premium income per unit time,
- \( N(t) \) is the number of claims in the time interval \((0, t]\) and has a Poisson \((\lambda t)\) distribution, and,
- \( \{X_i\}_{i=1}^\infty \) is a sequence of i.i.d. random variables representing the individual claim amounts.

Throughout this paper we adopt the convention that \( \sum_{i=1}^{0} X_i = 0 \).

We denote by \( P(x) \) the distribution function of \( X_i \). We assume that \( P(x) = 0 \) for \( x < 0 \), so that all claim amounts are non-negative. We assume that the mean of \( X_i \),
which we denote $p_1$, is finite and that any other moments of $X_i$ which we require are also finite. We assume that $c > \lambda p_1$.

We define $\theta$ to be such that

$$c = (1 + \theta) \lambda p_1$$

so that $\theta$ is the insurer's premium loading factor.

Without loss of generality, we make the following two assumptions

$$c = 1 \text{ and } p_1 = 1$$

We will refer to the process described above as our "basic process".

We want to produce a discrete approximation to this basic process but before doing so it is convenient to rescale the basic process by multiplying all monetary amounts by some positive number $\beta$ and taking a new time unit to be $\beta^{-1}$ times the original time unit so that the premium income per unit time for the rescaled process is still 1. In all our numerical examples $\beta$ will be 100.

Now let $\{X_{d, i}\}_{i=1}^\infty$ be a sequence of i.i.d. random variables whose (common) distribution is approximately the same as that of $\beta X_i$ and which are distributed on the non-negative integers. We denote the probability function of $X_{d, i}$ by $f(k)$ so that

$$f(k) = \Pr(X_{d, i} = k), \quad k = 0, 1, 2, \ldots$$

Let $N_d(t)$ be defined to be $N(\beta^{-1} t)$ so that $\{N_d(t)\}_{t=0}^\infty$ is a Poisson process with parameter $\lambda \beta^{-1}$. Now consider the discrete time surplus process $\{U_d(u, n)\}_{n=0}^\infty$ defined as

$$U_d(u, n) = u + n - \sum_{i=1}^{N_d(1)} X_{d, i}$$

so that the premium income per unit time is 1 and the initial surplus is $u$. The implied premium loading factor for this discrete surplus process will be denoted $\theta_d$ and is given by the formula

$$l = (1 + \theta_d) \lambda \beta^{-1} E[X_{d, i}]$$

Note that if $E[X_{d, i}] = \beta$ then $\theta_d = \theta$. We will always choose $\beta$ and the distribution of $X_{d, i}$ to be such that $\theta_d$ is positive. Let $S_d$ denote the aggregate claims over the first time period for the discrete model. We will denote by $H_d(k)$ and $h_d(k)$ the distribution function and probability function, respectively, of $S_d$, so that

$$H_d(k) = \sum_{j=0}^k h_d(j) = \Pr(S_d \leq k) = \Pr\left(\sum_{i=1}^{N_d(1)} X_{d, i} \leq k\right) \text{ for } k = 0, 1, 2, \ldots$$

Then it is clear that for any integer $n$, $U_d(\beta u, \beta n)$ will have approximately the same distribution as $U(u, n)$. It should also be clear that by increasing the value of $\beta$ we ought to be able to improve this approximation.
3. THE PROBABILITY OF ULTIMATE RUIN

Let $T$ be the time to ruin for the basic process, starting from initial surplus $u$, so that

$$T = \begin{cases} \inf \{t: U(u, t) < 0\} & \text{if } U(u, t) \geq 0 \text{ for all } t > 0 \\ \infty & \text{if } U(u, t) < 0 \text{ for some } t > 0 \end{cases}$$

The probability of ultimate ruin for the basic process, $\psi(u)$, and its complement, the probability of ultimate survival, $\delta(u)$, are defined as follows

$$\psi(u) = 1 - \delta(u) = Pr(T < \infty)$$

We are interested in the probability of ruin for our discrete process. However, since we will always take the initial surplus for the discrete process to be an integer we need to define "ruin" carefully. We will use two definitions of ruin for our discrete process, depending on whether or not a surplus of zero, other than at time zero, is regarded as ruin. Accordingly we define

$$T_d = \begin{cases} \min \{n: U_d(u, n)<0\} & \text{if } U_d(u, n)>0 \text{ for all } n \\ \infty & \text{if } U_d(u, n)<0 \text{ for some } n \end{cases}$$

$$T_{d*} = \begin{cases} \min \{n: U_d(u, n)>0\} & \text{if } U_d(u, n)<0 \text{ for all } n \\ \infty & \text{if } U_d(u, n)<0 \text{ for some } n \end{cases}$$

with $\delta_d(u) = 1 - \psi_d(u)$ and $\delta_{d*}(u) = 1 - \psi_{d*}(u)$ denoting the corresponding probabilities of ultimate survival. We need to define the defective probability function of the severity of ruin for the discrete model. For $u = 0, 1, 2, \ldots$ and $y = 1, 2, 3, \ldots$, we define

$$g_d(u, y) = Pr(T_d < \infty \text{ and } U_d(u, T_d) = -y)$$

For $u = 0, 1, 2, \ldots$ and $y = 0, 1, 2, \ldots$, we define

$$g_{d*}(u, y) = Pr(T_{d*} < \infty \text{ and } U_{d*}(u, T_{d*}) = -y)$$

It is immediate that

$$(3.1) \quad \psi_{d*} = \psi_d(u - 1) \text{ for } u = 1, 2, 3, \ldots$$

and that

$$(3.2) \quad g_{d*}(u, y) = g_d(u - 1, y + 1) \text{ for } u = 1, 2, 3, \ldots, \text{ and } y = 0, 1, 2, \ldots$$

It is well known that

$$(3.3) \quad \delta_d(0) = \theta_d / [(1 + \theta_d) h_d(0)]$$

$$(3.4) \quad \delta_{d*}(0) = \theta_d / (1 + \theta_d)$$
See, for example, DICKSON and WATERS (1991, Section 7.1). We need the following formula for \( g_d(0, y) \)

(3.5) \[ g_d(0, y) = (1 - H_d(y))/h_d(0) \]

for \( y = 1, 2, 3, \ldots \)

This can be proved by noting first that

(3.6) \[ g_d^*(0, y) = 1 - H_d(y) \]

(this follows from DICKSON and WATERS [1992, formula (3.5)]) and then conditioning on the aggregate claims in the first time period

\[ g_d^*(0, y) = h_d(0) g_d^*(1, y) + h_d(y+1) \]

Using (3.6) and (3.2), and rearranging, gives (3.5).

DICKSON and WATERS (1991, formula (7.2)) presented the following formula for the calculation of \( \delta_d(u) \) for positive integer values of \( u \)

(3.7) \[ \delta_d(u) = \frac{1}{h_d(0)} \left( \delta_d(u-1) - \sum_{i=1}^{u} h_d(i) \delta_d(u-i) \right) \]

This formula can be used recursively starting from formula (3.3). We can then use (3.1) to calculate \( \delta_d^*(u) \), with \( \delta_d^*(0) \) given by (3.4).

In the context of a compound binomial model, this formula has been put forward by GERBER (1988, formulae (6) and (7)), WILLMOT (1993, see the remark following formula (3.3)) and DICKSON (1994, formulae (5.1) and (5.2)). Unfortunately, the recursive scheme based on this formula is not stable. See DICKSON and WATERS (1991, Sections 7.2 and 7.3) and PANJER and WANG (1993, Section 11.5).

As an alternative to formula (3.7) we propose the following formula:

(3.8) \[ \delta_d(u) = \delta_d(0) + \sum_{k=1}^{u} g_d(0, k) \delta_d(u-k) \]

Formula (3.8) can be used to calculate \( \delta_d(u) \) recursively for \( u = 1, 2, 3, \ldots \), starting from (3.3), and using (3.5).

The derivation of (3.8) is elementary. Starting from surplus \( u \), ruin does not occur if either the surplus never falls below \( u \) (\( \delta_d(0) \)) or falls below \( u \) for the first time to \( u-k \), where \( k = 1, 2, \ldots, u, (g_d(0, k)) \) but ruin does not occur subsequently from this new level (\( \delta_d(u-k) \)). The important feature of (3.8) is that it is stable. In fact, Theorem 7 of PANJER and WANG (1993) shows that it is, in their terminology, strongly stable.

By choosing a distribution for \( X_d, i \), that is, in some sense, a good approximation to that of \( \beta X_i \), we can use (3.8) to provide a good approximation to \( \delta(u) \). For reasons explained by DICKSON and WATERS (1991, Section 8), \( \delta_d^*(\beta u) \) is usually a better approximation to \( \delta(u) \) than is \( \delta_d(\beta u) \). However, we can also use the discrete model to provide upper and lower bounds for \( \delta(u) \).

**Result 1**

Let \( X_d, i \) be defined as follows:

\[ X_d, i = k \text{ if } k-1 \leq \beta X_i < k \text{ for } k = 1, 2, \ldots \]
Then for any $u > 0$

$$\delta^*_\beta([\beta u]) \leq \delta(u)$$

where $[\beta u]$ is the integer part of $\beta u$.

**Proof**

Suppose ruin occurs for the basic process at time $t$, where $(n - 1)/\beta < t \leq n/\beta$, for some positive integer $n$. Then for the basic process

$$u + n/\beta - \sum_{i=1}^{N(n/\beta)} X_i < 1/\beta$$

Hence

$$\beta u + n - \beta (X_1 + X_2 + \ldots + X_{N(n/\beta)}) < 1$$

Hence for the discrete process

$$[\beta n] + n - (X_{d, 1} + X_{d, 2} + \ldots + X_{d, N_d(n)}) < 1$$

and so $U_d([\beta u], n) \leq 0$. Hence

$$\psi^*_\beta([\beta u]) \geq \psi(u)$$

and the result follows.

**Result 2**

Let $X_{d, i}$ be defined as follows

$$X_{d, i} = k - 1 \text{ if } k - 1 \leq \beta X_i < k \text{ for } k = 1, 2, \ldots, K$$

$$= K \text{ if } \beta X_i \geq K$$

for some positive integer $K$, which could be $\infty$. Then for any $u > 0$

$$\delta^*_\beta([\beta u]) \geq \delta(u)$$

where $[\beta u]$ is the least integer greater than or equal to $\beta u$.

**Proof**

Suppose ruin occurs for the discrete process at time $n$, regarding hitting zero as ruin, starting from initial surplus $[\beta u]$. Then

$$[\beta u] + n - (X_{d, 1} + X_{d, 2} + \ldots + X_{d, N_d(n)}) \leq 0$$

Hence

$$\beta u + n - \beta (X_1 + X_2 + \ldots + X_{N(n/\beta)}) \leq 0$$

Hence

$$u + n/\beta - (X_1 + X_2 + \ldots + X_{N(n/\beta)}) \leq 0$$
and so the basic process is ruined at or before time $n/\beta$ starting from initial surplus $u$. Hence

$$\psi^n_\beta (\{\beta u\}) \leq \psi(u)$$

and the result follows.

**Comment**

The use of the recursive scheme based on formula (3.8) requires knowledge of the premium loading factor for the discrete model, $\theta_d$. This is equivalent to knowing $E[X_d, i]$. When applying Result 2 it may be possible to calculate $E[X_d, i]$ for $K = \infty$, i.e. it may be possible to sum the appropriate infinite series. If not, $K$ can be chosen to be suitably large, but still finite, and $E[X_d, i]$ can be calculated by direct summation. The calculation of $E[X_d, i]$, and hence of $\theta_d$, for the lower bound in Result 1 may not be quite as simple. If the appropriate infinite series cannot be summed, we can use the fact that for the discrete model

$$E[X_d, i] \leq 1 + \beta E[X_i] = 1 + \beta$$

and hence

$$\theta_d \geq \frac{\beta \theta - 1}{1 + \beta}$$

and hence

$$\delta_d(0) \geq 1 - \frac{1 + \beta}{\beta (1 + \theta) h_d(0)} \quad (3.9)$$

Now note that since $g_d(0, k)$ is known for all $k$, the values of $\delta_d(u)$, for positive integer values of $u$ are all proportional to $\delta_d(0)$. Hence, using the right hand side of (3.9) as an approximation to $\delta_d(0)$ in formula (3.8) will produce approximations to $\delta_d(\{\beta u\})$ which are lower bounds (and which are lower than the correct values of $\delta_d(\{\beta u\})$ by the same factor for all $u$) and hence lower bounds for $\delta(u)$.

**3.1. Examples**

In the numerical examples at the end of this section we will compare numerical results produced by formulae (3.3) and (3.8) (and the relationship between $\delta_d(u)$ and $\delta_d^\#(u)$) with those produced by a different recursive algorithm. This alternative algorithm is called “Method 1” by DUFRESNE and GERBER (1989) and attributed by them to GOOVAERTS and DE VYLDERS (1984) and PANJER (1986). “Method 1” is a stable recursive scheme since it is based on Panjer’s recursion for a compound geometric distribution, which PANJER and WANG (1993, Section 9) show to be stable. It also has the advantage that it produces upper and lower bounds for $\psi(u)$. It requires an interval of discretisation to be chosen. In our examples we will take this to be the unit interval for the rescaled basic process, which is equivalent to an interval of length $\beta^{-1}$ for the basic process. Recall that $\beta = 100$ in all our examples.
3.1.1. Example 1

We assume that individual claim amounts for the basic process have an exponential distribution (with mean 1) and that $\theta = 0.1$. In this case we can calculate the exact value of $\delta(u)$, which is given by

$$
\delta(u) = 1 - \frac{1}{1 + \theta} \exp \left\{ - \frac{\theta u}{1 + \theta} \right\}
$$

The columns of Table 1 show for the values of $u$ indicated:

(1) A lower bound for $\delta(u)$ calculated as in Result 1. In this example it is easy to show that the premium loading factor for the discrete process, $\theta_d$, is $(1 + \theta)\beta (1 - e^{-0.01}) - 1 = 0.094518$.

(2) An approximation to $\delta(u)$ based on formula (3.8). The discretisation of the rescaled individual claim amounts for this approximation uses the method of De Vylder and Goovaerts (1988). This method preserves the mean of the distribution so that $\theta_d = \theta = 0.1$.

(3) An upper bound for $\delta(u)$ calculated as in Result 2 with $K = \infty$. The value of $\theta_d$ can be shown to be $(1 + \theta)\beta (e^{0.01} - 1) - 1 = 0.105518$.

(4) The relative percentage difference between the approximation in (2) and the correct value for $\delta(u)$, i.e. $100 \times \frac{\text{approximation} - \text{correct value}}{\text{correct value}}$.

(5) A lower bound for $\delta(u)$ calculated using “Method 1”.

(6) An upper bound for $\delta(u)$ calculated using “Method 1”.

(7) The relative percentage difference between the average of the values in (5) and (6) and the correct value for $\delta(u)$.

3.1.2. Example 2

Now assume individual claim amounts have a Pareto(2,1) distribution (so that its mean is 1).
The columns of Table 2 show for the values of $u$ indicated:

1. A lower bound for $\delta(u)$ calculated as in Result 1. The value of $\theta_d$ has been bounded below as described in the Comment following Result 2, so that its value has been taken to be 0.089109.

2. An approximation to $\delta(u)$ based on formula (3.8). The discretisation of the rescaled individual claim amounts for this approximation uses the method of De Vylder and Goovaerts (1988), so that $\theta_d = 0.1$.

3. An upper bound for $\delta(u)$ calculated as in Result 2 with $K = 35,000$. The value of $\theta_u$ can be shown to be 0.108683.

4. A lower bound for $\delta(u)$ calculated using "Method 1".

5. An upper bound for $\delta(u)$ calculated using "Method 1".

6. The average of the values in (4) and (5).

### Table 2

<table>
<thead>
<tr>
<th>$u$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.08182</td>
<td>0.09091</td>
<td>0.09803</td>
<td>0.09091</td>
<td>0.09091</td>
<td>0.09091</td>
</tr>
<tr>
<td>2</td>
<td>0.17874</td>
<td>0.18977</td>
<td>0.20337</td>
<td>0.18952</td>
<td>0.19003</td>
<td>0.18978</td>
</tr>
<tr>
<td>4</td>
<td>0.22726</td>
<td>0.25024</td>
<td>0.26744</td>
<td>0.24992</td>
<td>0.25057</td>
<td>0.25024</td>
</tr>
<tr>
<td>6</td>
<td>0.27107</td>
<td>0.29785</td>
<td>0.31767</td>
<td>0.29750</td>
<td>0.29821</td>
<td>0.29785</td>
</tr>
<tr>
<td>8</td>
<td>0.30810</td>
<td>0.33795</td>
<td>0.35983</td>
<td>0.33758</td>
<td>0.33833</td>
<td>0.33796</td>
</tr>
<tr>
<td>10</td>
<td>0.34045</td>
<td>0.37287</td>
<td>0.39642</td>
<td>0.37249</td>
<td>0.37325</td>
<td>0.37287</td>
</tr>
<tr>
<td>20</td>
<td>0.46090</td>
<td>0.50186</td>
<td>0.53055</td>
<td>0.50148</td>
<td>0.50224</td>
<td>0.50186</td>
</tr>
<tr>
<td>40</td>
<td>0.60339</td>
<td>0.65211</td>
<td>0.68446</td>
<td>0.65179</td>
<td>0.65242</td>
<td>0.65211</td>
</tr>
<tr>
<td>60</td>
<td>0.68750</td>
<td>0.73935</td>
<td>0.77244</td>
<td>0.73911</td>
<td>0.73960</td>
<td>0.73935</td>
</tr>
<tr>
<td>80</td>
<td>0.74276</td>
<td>0.79598</td>
<td>0.82888</td>
<td>0.79579</td>
<td>0.79617</td>
<td>0.79598</td>
</tr>
<tr>
<td>100</td>
<td>0.78135</td>
<td>0.83514</td>
<td>0.86755</td>
<td>0.83499</td>
<td>0.83529</td>
<td>0.83514</td>
</tr>
</tbody>
</table>

3.1.3. Comments on Examples 1 and 2

From Table 1 it can be seen that the numerical results provided by (3.8) and by "Method 1" are very accurate—recall that columns (4) and (7) give the percentage relative errors—at least for exponential claim amounts. For Pareto individual claim amounts we cannot check the accuracy of the two methods, since the exact values are not known, but we can see from Table 2 that the two methods give remarkably similar answers, agreeing to 4 decimal places in all cases and 5 in most. The bounds produced by "Method 1" are closer than those produced by Results 1 and 2.

4. THE PROBABILITY AND SEVERITY OF RUIN

In this section we present a stable recursive algorithm for the approximate numerical calculation of the probability and severity of ruin for our basic process. Using a different approach, we also derive lower and upper bounds for this quantity.

Let $G(u, y)$ denote the probability that ruin occurs for our basic process, given initial surplus $u$, and that the deficit at the time of ruin is less than $y$, so that for
Using the discrete approximation to our basic process, an approximation to \( G(u, y) \) is \( G_d(\beta u, \beta y) \) where

\[
G_d(\beta u, \beta y) = \Pr(T_\beta < \infty \quad \text{and} \quad U_\beta(u, T_\beta) > -y)
\]

Dickson and Waters (1992, Section 3) presented the following formula for the calculation of \( G_d(u, y) \) for \( u = 0, 1, 2, ... \) and \( y = 1, 2, 3, ... \)

\[
G_d(u + 1, y) = \frac{1}{h_d(0)} \left( G_d(u, y) - \sum_{j=1}^{u} h_d(j) G_d(u + 1 - j, y) + H_d(u) - H_d(u + y) \right)
\]

This formula allows recursive calculation of \( G_d(u, y) \) starting from

\[
G_d(0, y) = \sum_{j=0}^{y-1} (1 - H_d(j))
\]

Although this algorithm provides very good approximations for moderate values of \( u \), it is unstable. An alternative approach to calculating \( G_d(u, y) \) is as follows. Define, for \( u = 0, 1, 2, ... \) and \( y = 1, 2, 3, ... \)

\[
G_d(u, y) = \Pr(T_d < \infty \quad \text{and} \quad U_d(u, T_d) > -y)
\]

Then for \( u = 0, 1, 2, ... \)

\[
G_d(u, y) = G_d(0, u + y) - G_d(0, u) + \sum_{k=1}^{u} g_d(0, k) G_d(u - k, y)
\]

This formula follows by considering the level of the surplus process on the first occasion that the surplus falls below its initial level (if this ever occurs). We can calculate \( G_d(0, y) \) in a recursive manner from (3.5) and hence can also calculate \( G_d(u, y) \) recursively. Once again, by Theorem 7 of Panjer and Wang (1993), this is a stable recursive algorithm.

We will give an example to illustrate the use of this algorithm at the end of this section. Before doing so we show how to derive lower and upper bounds for \( G(u, y) \). The method does not involve the discrete approximation to the basic process. For the remainder of this section we will make the additional assumption that \( P(x) \) is absolutely continuous and we will denote its density function by \( p(x) \).

Let \( g(u, y) \) denote the derivative of \( G(u, y) \) with respect to \( y \). It is well known that

\[
g(0, y) = \frac{\lambda}{c} (1 - P(y))
\]
(see, for example, GERBER et al. (1987)). We can write

\[ 4.4 \quad g(u, y) = \frac{1}{\delta(0)} \left( \frac{\lambda}{c} \int_0^u p(y + z) \psi(u - z) \, dz + g(0, u + y) - \psi(u)g(0, y) \right) \]

(see PANJER and WILLMOT (1992) or DICKSON and DOS REIS (1994)) and it follows by integrating (4.4) over \( y \) that

\[ 4.5 \quad G(u, y) = \frac{1}{\delta(0)} \left( \int_0^u \psi(u - z) [g(0, z) - g(0, z + y)] \, dz ight) 
+ G(0, u + y) - G(0, u) - \psi(u)G(0, y) \]

\[ = \frac{1}{\delta(0)} \left( \sum_{r=0}^{u-1} \int_r^{r+1} \psi(u - z) [g(0, z) - g(0, z + y)] \, dz 
+ G(0, u + y) - G(0, u) - \psi(u)G(0, y) \right) \]

Now let \( \psi^l(u) \) and \( \psi^h(u) \) denote lower and upper bounds respectively for \( \psi(u) \), calculated, for example, by one of the methods in the previous section.

From (4.5), a lower bound for \( G(u, y) \) is \( G^l(u, y) \) where

\[ G^l(u, y) = \frac{1}{\delta(0)} \left( \sum_{r=0}^{u-1} \psi^l(u - r) \int_r^{r+1} [g(0, z) - g(0, z + y)] \, dz 
+ G(0, u + y) - G(0, u) - \psi^h(u)G(0, y) \right) \]

\[ = \frac{1}{\delta(0)} \left( \sum_{r=0}^{u-1} \psi^l(u - r) [G(0, r + 1) - G(0, r)] 
- \sum_{r=0}^{u-1} \psi^l(u - r) [G(0, r + y + 1) - G(0, r + y)] 
+ G(0, u + y) - G(0, u) - \psi^h(u)G(0, y) \right) \]

and an upper bound is \( G^h(u, y) \) where

\[ G^h(u, y) = \frac{1}{\delta(0)} \left( \sum_{r=0}^{u-1} \psi^h(u - r - 1) [G(0, r + 1) - G(0, r)] 
- \sum_{r=0}^{u-1} \psi^h(u - r - 1) [G(0, r + y + 1) - G(0, r + y)] 
+ G(0, u + y) - G(0, u) - \psi^l(u)G(0, y) \right) \]
Since it is always possible to compute $G(0, y)$, either because (4.3) can be integrated in closed form or because we can integrate (4.3) numerically to any degree of accuracy we choose, we can always compute these bounds for $G(u, y)$. In our examples we will calculate bounds for the rescaled basic process, with $\beta = 100$.

### 4.1. Examples

#### 4.1.1. Example 3

Table 3 shows bounds and exact values for, and approximations to, $G(u, y)$ when the individual claim amount distribution is exponential with mean $1$, $\theta = 0.1$ and the bounds for $\psi(u)$ have been calculated using “Method 1” as described in the previous section. The key to Table 3 is as follows:

(1) gives the value of $G^l(u, y)$,
(2) gives the exact value of $G(u, y)$ (see, for example, DICKSON (1992)),
(3) gives the approximation to $G(u, y)$, calculated from the recursive algorithm,
(4) gives the approximation to $G(u, y)$, calculated by averaging the lower and upper bounds, and
(5) gives the value of $G^h(u, y)$.

<table>
<thead>
<tr>
<th></th>
<th>$v = 1$</th>
<th>$v = 3$</th>
<th>$v = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u = 20$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.077091</td>
<td>0.115884</td>
<td>0.121134</td>
</tr>
<tr>
<td>(2)</td>
<td>0.093278</td>
<td>0.140217</td>
<td>0.146570</td>
</tr>
<tr>
<td>(3)</td>
<td>0.093034</td>
<td>0.140119</td>
<td>0.146549</td>
</tr>
<tr>
<td>(4)</td>
<td>0.093279</td>
<td>0.140219</td>
<td>0.146572</td>
</tr>
<tr>
<td>(5)</td>
<td>0.109468</td>
<td>0.164554</td>
<td>0.172009</td>
</tr>
<tr>
<td>$u = 60$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.001179</td>
<td>0.001772</td>
<td>0.001852</td>
</tr>
<tr>
<td>(2)</td>
<td>0.002458</td>
<td>0.003694</td>
<td>0.003862</td>
</tr>
<tr>
<td>(3)</td>
<td>0.002451</td>
<td>0.003692</td>
<td>0.003861</td>
</tr>
<tr>
<td>(4)</td>
<td>0.002458</td>
<td>0.003695</td>
<td>0.003863</td>
</tr>
<tr>
<td>(5)</td>
<td>0.003738</td>
<td>0.005619</td>
<td>0.005874</td>
</tr>
<tr>
<td>$u = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.000009</td>
<td>0.000013</td>
<td>0.000014</td>
</tr>
<tr>
<td>(2)</td>
<td>0.000065</td>
<td>0.000097</td>
<td>0.000102</td>
</tr>
<tr>
<td>(3)</td>
<td>0.000065</td>
<td>0.000097</td>
<td>0.000102</td>
</tr>
<tr>
<td>(4)</td>
<td>0.000065</td>
<td>0.000097</td>
<td>0.000102</td>
</tr>
<tr>
<td>(5)</td>
<td>0.000121</td>
<td>0.000182</td>
<td>0.000190</td>
</tr>
</tbody>
</table>

#### 4.1.2. Example 4

Table 4 shows bounds for, and approximations to, $G(u, y)$ when the individual claim amount distribution is Pareto $(2, 1)$, $\theta = 0.1$ and the bounds for $\psi(u)$ have been calculated using “Method 1” as described in the previous section. The key to Table 4 is as follows:

(1) gives the value of $G^l(u, y)$,
(2) gives the approximation to $G(u, y)$, calculated from the recursive algorithm,
(3) gives the approximation to $G(u, y)$, calculated by averaging the lower and upper bounds, and
(4) gives the value of $G^b(u, y)$.

\[
\text{TABLE 4}
\text{SEE EXAMPLE 4, SECTION 4, FOR DETAILS}
\]

<table>
<thead>
<tr>
<th>$u = 20$</th>
<th>$y = 1$</th>
<th>$y = 5$</th>
<th>$y = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.075914</td>
<td>0.204566</td>
<td>0.274804</td>
</tr>
<tr>
<td>(2)</td>
<td>0.079821</td>
<td>0.211242</td>
<td>0.282126</td>
</tr>
<tr>
<td>(3)</td>
<td>0.079990</td>
<td>0.211347</td>
<td>0.282184</td>
</tr>
<tr>
<td>(4)</td>
<td>0.084065</td>
<td>0.218128</td>
<td>0.289563</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$u = 100$</th>
<th>$y = 1$</th>
<th>$y = 5$</th>
<th>$y = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.011382</td>
<td>0.033331</td>
<td>0.047841</td>
</tr>
<tr>
<td>(2)</td>
<td>0.012918</td>
<td>0.035929</td>
<td>0.050693</td>
</tr>
<tr>
<td>(3)</td>
<td>0.012945</td>
<td>0.035948</td>
<td>0.050705</td>
</tr>
<tr>
<td>(4)</td>
<td>0.014509</td>
<td>0.038566</td>
<td>0.053569</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$u = 200$</th>
<th>$y = 1$</th>
<th>$y = 5$</th>
<th>$y = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.003056</td>
<td>0.009230</td>
<td>0.013560</td>
</tr>
<tr>
<td>(2)</td>
<td>0.003593</td>
<td>0.010137</td>
<td>0.014554</td>
</tr>
<tr>
<td>(3)</td>
<td>0.003601</td>
<td>0.010142</td>
<td>0.014558</td>
</tr>
<tr>
<td>(4)</td>
<td>0.004146</td>
<td>0.011054</td>
<td>0.015555</td>
</tr>
</tbody>
</table>

4.1.3. Comments on Examples 3 and 4

(i) In each example, the two approximations to $G(u, y)$ are close to each other. We can see in Example 3 that for smaller values of $u$, the approximation based on the bounds is slightly superior, but for large values of $u$ both approximations give values very close to the true value.

(ii) The calculation of $G^l(u, y)$ and $G^b(u, y)$ is not recursive so that separate calculations are required for each combination of $u$ and $y$. The calculation of $G_d(u, y)$ using (4.2) is recursive in $u$, and so is more convenient if values are required for several values of $u$.

5. MOMENTS OF THE SEVERITY OF RUIN

In this section we are interested in the moments of the severity of ruin for the basic process. For this process, let $Y$ be a defective random variable denoting the severity of ruin. The $k$-th unconditional moment of $Y$ is given by

\[
E(Y^k) = \int_0^\infty y^k g(u, y) \, dy
\]

and the conditional moment is found by dividing this quantity by $\psi(u)$.

We can use results from the previous two sections to obtain approximations to these moments.

Let $Y_d$ and $Y^e_d$ denote the deficit at the time of ruin for the discrete process, the distinction being that, for the latter, a surplus of zero, other than at time zero, is
regarded as ruin. The unconditional moments of these defective random variables are given by

\[ E(Y_d^k | u) = \sum_{y=0}^{\infty} y^k g_d(u, y) \quad \text{and} \quad E(Y_d^{\# k} | u) = \sum_{y=0}^{\infty} y^k g_d^{\#}(u, y) \]

We will approximate (5.1) by \( \beta^{-k} E(Y_d^{\# k} | \beta u) \). We will derive a recursive algorithm for \( E(Y_d^{\# k} | u) \) and then use this to calculate \( E(Y_d^{\# k} | u) \). In our examples we will consider only the first three moments.

Since \( g_d^{\#}(u, y) = g_d(u - 1, y + 1) \) for \( u = 1, 2, 3, \ldots \) and \( y = 0, 1, 2, \ldots \), it follows that for \( u = 1, 2, 3, \ldots \)

\[
\begin{align*}
E(Y_d^{\# 1} | u) &= E(Y_d^{\# 1} | u - 1) - \psi_d(u - 1) \\
E(Y_d^{\# 2} | u) &= E(Y_d^{\# 2} | u - 1) - 2E(Y_d^{\# 1} | u - 1) + \psi_d(u - 1) \\
E(Y_d^{\# 3} | u) &= E(Y_d^{\# 3} | u - 1) - 3E(Y_d^{\# 2} | u - 1) + 3E(Y_d^{\# 1} | u - 1) - \psi_d(u - 1)
\end{align*}
\]

For \( u = 0 \) we have

\[
E(Y_d^{\# 0} | 0) = \sum_{y=0}^{\infty} yg_d^{\#}(0, y) = \sum_{y=0}^{\infty} y(1 - H_d(y)) = \frac{1}{2}(E(S_d^3) - E(S_d))
\]

Similarly

\[
E(Y_d^{\# 2} | 0) = \frac{1}{3}E(S_d^3) - \frac{1}{2}E(S_d^2) + \frac{1}{3}E(S_d)
\]

and

\[
E(Y_d^{\# 3} | 0) = \frac{1}{4}E(S_d^4) - \frac{1}{2}E(S_d^3) + \frac{1}{4}E(S_d^2)
\]

with

\[
E(Y_d^{\# k} | 0) = h_d(0)^{-1} E(Y_d^{\# k} | 0) \quad \text{for} \ k = 1, 2, 3, \ldots
\]

From (4.2) it is easy to see that for \( u = 1, 2, 3, \ldots \) and \( y = 1, 2, 3, \ldots \)

\[
g_d(u, y) = g_d(0, u + y) + \sum_{k=1}^{u} g_d(0, k) g_d(u - k, y)
\]

and so

\[
E(Y_d^{\# k} | u) = \sum_{y=1}^{\infty} y^k g_d(0, u + y) + \sum_{y=1}^{\infty} y^k \sum_{k=1}^{u} g_d(0, k) g_d(u - k, y) = \Sigma_k(u) + \sum_{k=1}^{u} g_d(0, k) E(Y_d^{\# k} | u - k)
\]
where

\[ \Sigma_k(u) = \sum_{y=1}^{\infty} y^k g_d(0, u+y) \]

Assuming \( \Sigma_k(u) \) is known, (5.2) is a stable recursion formula for \( E(Y^k \mid u) \). Note that \( \Sigma_k(0) = E(Y^k \mid 0) \), which is known. We can calculate \( \Sigma_k(u) \) recursively for \( u = 1, 2, 3, ... \) as follows:

\[ \Sigma_1(u+1) = \Sigma_1(u) - \psi_d(0) + G_d(0, u) \]

\[ \Sigma_2(u+1) = \Sigma_2(u) - 2 \Sigma_1(u+1) - \psi_d(0) + G_d(0, u) \]

\[ \Sigma_3(u+1) = \Sigma_3(u) - 3 \Sigma_2(u+1) - 3 \Sigma_1(u+1) - \psi_d(0) + G_d(0, u) \]

Unfortunately, these recursion formulae are unstable. In our examples we have applied these formulae but have constrained them to satisfy the following inequalities:

\[ 0 \leq \Sigma_k(u+1) \leq \Sigma_k(u) \quad \text{for } k = 1, 2, 3 \quad \text{and } u = 0, 1, 2, ... \]

\[ \Sigma_k(u) \leq \Sigma_{k+1}(u) \quad \text{for } k = 1, 2 \quad \text{and } u = 0, 1, 2, ... \]

5.1. Examples

We have used the method of this section to calculate the conditional moments of the severity of ruin in two cases: firstly, when individual claim amounts have an exponential distribution with mean 1 and, secondly, when they have a Pareto(4,3) distribution. Thus, we have calculated \( \beta^{-k} E(Y^k \mid \beta u) / \psi(\beta u) \) and we regard this as an approximation to the conditional moment \( E(Y^k \mid u) / \psi(u) \) for the basic process.

The calculation of \( E(Y^k \mid 0) \) requires \( E(S_d^{k+1}) \) to be finite. For this reason, we have calculated just the first two conditional moments of the severity of ruin for the Pareto(4,3) distribution. For the exponential distribution, where \( E(S_d^{k+1}) \) is finite for all \( k \), we have calculated the first three moments.

PANJER and LUTEK (1983) describe a method which may provide a discretisation of the rescaled individual claim amount distribution that preserves the moments of the original distribution. Because we need values of \( E(S_d^{k+1}) \) we have adopted this discretisation method for this section only. PANJER and LUTEK (1983) mention the possibility of obtaining negative values for probabilities under this method. In the examples below we used the software Mathematica and specified a high numerical precision for all calculations in the discretisation procedure. In this way we obtained positive values for all probabilities in the discretised distribution.

5.1.1. Example 5

When the individual claim amount distribution is exponential, so too is the distribution of the severity of ruin given that ruin occurs. In particular, \( E(Y^k \mid u) / \psi(u) \)
\( \psi(u) \) is independent of \( u \). Hence, when the individual claim amount is exponential,
\[
E(Y|u)\psi(u) = 1
\]
\[
E(Y^2|u)\psi(u) = 2
\]
\[
E(Y^3|u)\psi(u) = 6
\]
The method of this section gives the following results for \( \theta = 0.1 \):
\[
\beta^{-1}E(Y^\#|\beta u)\psi^\#(\beta u) = \begin{cases} 
0.9995 & \text{for } u = 0, 2, 3, 4, ..., 100 \\
0.9996 & \text{for } u = 1 
\end{cases}
\]
\[
\beta^{-2}E(Y^\#^2|\beta u)\psi^\#(\beta u) = 2.0082 & \text{for } u = 0, 1, 2, ..., 100 \\
\beta^{-2}E(Y^\#^3|\beta u)\psi^\#(\beta u) = 6.0518 & \text{for } u = 0, 1, 2, ..., 100 
\]
In this example it was necessary to apply the constraints described above in the calculation of the functions \( \Sigma_k(u) \), for \( k = 1, 2, 3 \).

5.1.2. Example 6

Now suppose that the individual claim amounts have a Pareto \((4,3)\) distribution. The method of this section gives the results in Table 5 for \( \theta = 0.1 \).

| \( u \) | \( \beta^{-1}E(Y^\#|\beta u)\psi^\#(\beta u) \) | \( \beta^{-2}E(Y^\#^2|\beta u)\psi^\#(\beta u) \) |
|---|---|---|
| 0 | 1.4995 | 9.0123 |
| 40 | 3.7585 | 111.85 |
| 80 | 5.8098 | 432.11 |
| 120 | 11.670 | 1,868.1 |
| 160 | 27.067 | 7,098.8 |
| 200 | 52.985 | 18,892 |

Using formula (4.3) and (5.1) and the fact that \( \psi(0) = 1/(1 + \theta) \), it is easy to show that \( E(Y|0)\psi(0) = 1.5 \) and \( E(Y^2|0)\psi(0) = 9 \) in this example. It is not possible to check the accuracy of the results in this example, other than when \( u = 0 \). It is, however, interesting to note that the conditional moments of \( Y^\# \) increase with \( u \).

In this example there was no need to apply the constraints described above in the calculation of the functions \( \Sigma_k(u) \), for \( k = 1, 2, 3 \).

6. DISTRIBUTIONS FOR THE SURPLUS PRIOR TO RUIN

In this section we present stable recursive algorithms for the approximate numerical calculation of the (defective) distribution of the surplus immediately prior to ruin, and for the (defective) joint distribution of the surplus immediately prior to ruin and the severity of ruin for our basic process. We will also apply the ideas introduced in Section 4 to derive bounds for these distributions.
Define \( U(u, \hat{T}) \) to be the surplus immediately prior to ruin for our basic process and for \( u \geq 0 \) and \( x > 0 \) define

\[
F(u, x) = \Pr(T < \infty \text{ and } U(u, \hat{T}) < x)
\]

so that \( F(u, x) \) is the probability that ruin occurs (from initial surplus \( u \)) and that the surplus immediately prior to ruin is less than \( x \).

Using the discrete approximation to our basic process, an approximation to \( F(u, x) \) is \( F_d^* (\beta u, \beta x) \) where for \( u = 0, 1, 2, \ldots \) and \( x = 1, 2, 3, \ldots \)

\[
F_d^* (u, x) = \Pr(T_d^* < \infty \text{ and } U_d(u, T_d^* - 1) < x)
\]

DICKSON (1992) presents the following formulae for the calculation of \( F_d^* (u, x) \) for \( x = 1, 2, 3, \ldots \):

\[
\begin{aligned}
F_d^* (u, x) &= \frac{1}{h_d(0)} \left( F_d^* (u - 1, x) - \sum_{j=0}^{u-1} h_d(j) - \sum_{j=1}^{u-1} h_d(j) F_d^* (u-j, x) \right)
\end{aligned}
\]

for \( u = 0, 1, 2, \ldots, x \) and

\[
\begin{aligned}
F_d^* (u, x) &= \frac{1}{h_d(0)} \left( F_d^* (u - 1, x) - \sum_{j=1}^{u-1} h_d(j) F_d^* (u-j, x) \right)
\end{aligned}
\]

for \( u = x+1, x+2, x+3, \ldots \) We can use these formulae to calculate \( F_d^* (u, x) \) recursively starting from

\[
(6.1) \quad F_d^* (0, x) = \sum_{j=0}^{x-1} (1 - H_d(j))
\]

This algorithm provides good approximations for moderate values of \( u \) but is unstable. To provide an alternative method of calculating \( F_d^* (u, x) \) we require the following definitions. For \( u = 0, 1, 2, \ldots \) and \( x = 1, 2, 3, \ldots \), define

\[
F_d(u, x) = \Pr(T_d < \infty \text{ and } U_d(u, T_d - 1) < x)
\]

and for \( u = 0, 1, 2, \ldots, x = 0, 1, 2, \ldots \) and \( y = 1, 2, 3, \ldots \), define

\[
f_d(u, x, y) = \Pr(T_d < \infty, U_d(u, T_d) = -y \text{ and } U_d(u, T_d - 1) = x)
\]

Now for \( u = 1, 2, 3, \ldots \) and \( x = 1, 2, 3, \ldots \)

\[
(6.2) \quad F_d^*(u, x) = F_d(u - 1, x - 1)
\]

We can find \( F_d(0, x) \) by conditioning on the aggregate claim amount in the first time period. We have

\[
F_d^*(0, x) = h_d(0) F_d^*(1, x) + 1 - h_d(0)
\]
Substituting (6.1) for $F_d^+(0, x)$ and $F_d(0, x - 1)$ for $F_d^+(1, x)$, and rearranging leads to

\[(6.3)\quad F_d(0, x) = \frac{1}{h_d(0)} \sum_{j=1}^{x} (1 - H_d(j))\]

We can calculate $F_d(u, x)$ for $x = 1, 2, 3, \ldots$ from the following formulae. For $u = 1, 2, 3, \ldots, x - 1$

\[(6.4)\quad F_d(u, x) = \sum_{j=1}^{u} g_d(0, j) F_d(u-j, x) + \sum_{s=0}^{x-u-1} \sum_{y=u+1}^{x} f_d(0, s, y)\]

and for $u = x, x + 1, x + 2, \ldots$

\[(6.5)\quad F_d(u, x) = \sum_{j=1}^{u} g_d(0, j) F_d(u-j, x)\]

Formula (6.5) follows by considering the first occasion on which the surplus falls below its initial level (if this ever occurs). The first term of (6.4) comes from the same consideration. The second term in (6.4) comes from considering the situation when ruin occurs on the first occasion that the surplus falls below its initial level. In this case the surplus must be no more than $x - u - 1$ above its initial level at time $T_d - 1$ in order for the surplus at that time to be less than $x$. From GERBER (1988, equation (35)), it follows that $f_d(0, x, y) = h_d(x+y+1)/h_d(0)$. Substituting this expression in (6.4) we find that for $u = 1, 2, 3, \ldots, x - 1$

\[(6.6)\quad F_d(u, x) = \sum_{j=1}^{u} g_d(0, j) F_d(u-j, x) + \sum_{j=u+1}^{x} g_d(0, j)\]

Formulae (6.6) and (6.5), used in this order, provide a stable recursive algorithm for calculating $F_d(u, x)$ with the initial value $F_d(0, x)$ given by (6.3). We will illustrate the use of this algorithm later in this section.

Let us now consider how to calculate bounds for $F(u, x)$. DICKSON (1992) shows that

\[F(u, x) = \frac{1 - G(0, x)}{1 - \psi(0)} \psi(u) = \frac{\psi(0) - G(0, x)}{1 - \psi(0)} \quad \text{for } 0 \leq u \leq x\]

and

\[F(u, x) = G(u-x, x) - \frac{1 - G(0, x)}{1 - \psi(0)} (\psi(u-x) - \psi(u)) \quad \text{for } u \geq x\]

Then for $0 \leq u \leq x$, a lower bound for $F(u, x)$ is $F^l(u, x)$, where

\[F^l(u, x) = \frac{1 - G(0, x)}{1 - \psi(u)} \psi(u) - \frac{\psi(0) - G(0, x)}{1 - \psi(0)}\]
and an upper bound is \( F^h(u, x) \) where

\[
F^h(u, x) = \frac{1 - G(0, x)}{1 - \psi(0)} \psi^h(u) - \frac{\psi(0) - G(0, x)}{1 - \psi(0)}
\]

For \( u \geq x \), a lower bound for \( F(u, x) \) is \( F^l(u, x) \) where

\[
F^l(u, x) = G^l(u - x, x) - \frac{1 - G(0, x)}{1 - \psi(0)} \left( \psi^l(u - x) - \psi^l(0) \right)
\]

and an upper bound is \( F^h(u, x) \) where

\[
F^h(u, x) = \frac{1 - G(0, x)}{1 - \psi(0)} \psi^h(u - x, x) - \psi^h(0) \psi^h(0)
\]

These bounds are easily calculated by the methods described in Sections 3 and 4.

Now define

\[
F(u, x, y) = \Pr(T < \infty, U(u, T) > -y \text{ and } U(u, T - 1) < x)
\]

so that \( F(u, x, y) \) gives the (defective) joint distribution of the severity of ruin and the surplus immediately prior to ruin for our basic process. Using the discrete approximation to our basic process, an approximation to \( F(u, x, y) \) is \( F^\beta(u, \beta x, \beta y) \) where

\[
F^\beta(u, x, y) = \Pr(T^\beta < \infty, U^\beta(u, T^\beta) > -y \text{ and } U^\beta(u, T^\beta - 1) < x)
\]

for \( u = 0, 1, 2, \ldots, x = 1, 2, 3, \ldots \text{ and } y = 1, 2, 3, \ldots \). We can compute values of \( F^\beta(u, x, y) \) by first computing values of \( F_d(u, x, y) \) where

\[
F_d(u, x, y) = \Pr(T_d < \infty, U_d(u, T_d + 1) > -y \text{ and } U_d(u, T_d - 1) < x)
\]

since

\[
F_d(u, x, y) = F(u - 1, x - 1, y)
\]

for \( u = 1, 2, 3, \ldots, x = 1, 2, 3, \ldots \text{ and } y = 1, 2, 3, \ldots \). We can calculate \( F_d(u, x, y) \) through a stable recursive algorithm. The starting value for the algorithm is

\[
F_d(0, x, y) = \sum_{j=0}^{x-1} \sum_{s=1}^{y} f_d(0, j, s)
\]

For computational purposes we can write this as

\[
F_d(0, x, y) = \frac{1}{h_d(0)} \sum_{j=1}^{x} \left( H_d(y+j) - H_d(j) \right)
\]

An alternative way of writing \((6.8)\) is

\[
F_d(0, x, y) = F_d(0, x) + G_d(0, y) - G_d(0, x+y)
\]
which corresponds to the expression for $F(0, x, y)$ for our basic process given by Dickson and Dos Reis (1994, equation (2.1)).

For $u = 1, 2, 3, ..., x - 1$ we can use the same reasoning that we used to write formula (6.4) to write

\[(6.9) \quad F_d(u, x, y) = \sum_{j=1}^{u} g_d(0, j) F_d(u-j, x, y) + \sum_{s=0}^{x-u-1} \sum_{j=u+1}^{u+y} f_d(0, s, j)\]

\[= \sum_{j=1}^{u} g_d(0, j) F_d(u-j, x, y) + \sum_{j=u+1}^{x} (g_d(0, j) - g_d(0, j+y))\]

Similarly to (6.5), for $u = x, x+1, x+2, ...$

\[(6.10) \quad F_d(u, x, y) = \sum_{j=1}^{u} g_d(0, j) F_d(u-j, x, y)\]

Formulae (6.9) and (6.10), with (6.8) as a starting value, give a stable recursive algorithm for calculating $F_d(u, x, y)$. An application of this algorithm is given at the end of this section.

Finally, let us consider bounds for $F(u, x, y)$. Dickson and Dos Reis (1994) show that

\[F(u, x, y) = G(u, y) + \frac{\delta(u)}{\delta(0)} (G(0, x) - G(0, x+y)) \quad \text{for} \quad 0 \leq u \leq x\]

and

\[F(u, x, y) = G(u, y) - G(u-x, x+y) + G(u-x, x)\]

\[+ \frac{\psi(u-x) - \psi(u)}{\delta(0)} (G(0, x) - G(0, x+y)) \quad \text{for} \quad u \geq x\]

Then for $0 \leq u \leq x$, a lower bound for $F(u, x, y)$ is $F^l(u, x, y)$, where

\[F^l(u, x, y) = G^l(u, y) + \frac{\delta^l(u)}{\delta(0)} (G(0, x) - G(0, x+y))\]

and an upper bound is $F^h(u, x, y)$, where

\[F^h(u, x, y) = G^h(u, y) + \frac{\delta^h(u)}{\delta(0)} (G(0, x) - G(0, x+y))\]

For $u \geq x$, a lower bound for $F(u, x, y)$ is $F^l(u, x, y)$, where

\[F^l(u, x, y) = G^l(u, y) - G^h(u-x, x+y) + G^l(u-x, x)\]

\[+ G(0, x)(\psi^l(u-x) - \psi^h(u))/\delta(0)\]

\[+ G(0, x+y)(\psi^l(u) - \psi^h(u-x))/\delta(0)\]
and an upper bound is \( F^h(u, x, y) \), where

\[
F^h(u, x, y) = G^h(u, y) - G^j(u - x, x + y) + G^h(u - x, x) + G(0, x)(\psi^h(u - x) - \psi^j(u))/\delta(0) + G(0, x + y)(\psi^h(u) - \psi^j(u - x))/\delta(0)
\]

### 6.1. Examples

**6.1.1. Example 7**

Table 6 shows some bounds and approximations to \( F(u, x) \) when the individual claim amount distribution is Pareto(2,1), the premium loading factor, \( \theta \), is 0.1 and the bounds for \( \psi(u) \) and \( G(u, y) \) have been calculated as in Sections 3 (using "Method 1") and 4. The key to Table 6 is as follows:

1. gives the value of \( F^j(u, x) \),
2. gives the approximation to \( F(u, x) \) calculated from the recursive algorithm for \( F^d(u, x) \),
3. gives the approximation to \( F(u, x) \) calculated by averaging \( F^j(u, x) \) and \( F^h(u, x) \), and
4. gives the value of \( F^h(u, x) \).

**TABLE 6**

See Example 7, Section 6, for details.

<table>
<thead>
<tr>
<th></th>
<th>( x = 5 )</th>
<th>( x = 10 )</th>
<th>( x = 15 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u = 10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.161668</td>
<td>0.287422</td>
<td>0.393461</td>
</tr>
<tr>
<td>(2)</td>
<td>0.169434</td>
<td>0.287847</td>
<td>0.393936</td>
</tr>
<tr>
<td>(3)</td>
<td>0.169869</td>
<td>0.288154</td>
<td>0.394084</td>
</tr>
<tr>
<td>(4)</td>
<td>0.178070</td>
<td>0.288886</td>
<td>0.394706</td>
</tr>
<tr>
<td>( u = 30 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.064448</td>
<td>0.107883</td>
<td>0.139130</td>
</tr>
<tr>
<td>(2)</td>
<td>0.072663</td>
<td>0.116525</td>
<td>0.148012</td>
</tr>
<tr>
<td>(3)</td>
<td>0.072851</td>
<td>0.116654</td>
<td>0.148110</td>
</tr>
<tr>
<td>(4)</td>
<td>0.081254</td>
<td>0.125426</td>
<td>0.157089</td>
</tr>
<tr>
<td>( u = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.035739</td>
<td>0.060270</td>
<td>0.077271</td>
</tr>
<tr>
<td>(2)</td>
<td>0.042324</td>
<td>0.067322</td>
<td>0.084765</td>
</tr>
<tr>
<td>(3)</td>
<td>0.042434</td>
<td>0.067398</td>
<td>0.084823</td>
</tr>
<tr>
<td>(4)</td>
<td>0.049130</td>
<td>0.074526</td>
<td>0.092375</td>
</tr>
</tbody>
</table>

**6.1.3. Example 8**

Table 7 shows some bounds, approximations and exact values for \( F(u, x, y) \) when the individual claim amount distribution is exponential with mean 1, the premium loading factor, \( \theta \), is 0.1 and the bounds for \( \psi(u) \) and \( G(u, y) \) have been calculated as in Sections 3 (using "Method 1") and 4. The key to Table 7 is as follows:

1. gives the value of \( F^j(u, x, y) \),
2. gives the exact value of \( F(u, x, y) \),
3. gives the approximation to \( F(u, x, y) \) calculated from the recursive algorithm for \( F^d(u, x, y) \),
(4) gives the approximation to $F(u, x, y)$, calculated by averaging $F_t(u, x, y)$ and $F_s'(u, x, y)$, and
(5) gives the value of $F^h(u, x, y)$.

### TABLE 7

See Example 8, Section 6, for details

<table>
<thead>
<tr>
<th></th>
<th>$x = y = 1$</th>
<th>$x = y = 3$</th>
<th>$x = y = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u = 20$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>-0.070275</td>
<td>-0.021141</td>
<td>0.000549</td>
</tr>
<tr>
<td>(2)</td>
<td>0.023040</td>
<td>0.109159</td>
<td>0.139331</td>
</tr>
<tr>
<td>(3)</td>
<td>0.022529</td>
<td>0.108601</td>
<td>0.139137</td>
</tr>
<tr>
<td>(4)</td>
<td>0.023041</td>
<td>0.109161</td>
<td>0.139333</td>
</tr>
<tr>
<td>(5)</td>
<td>0.116357</td>
<td>0.239464</td>
<td>0.278116</td>
</tr>
<tr>
<td>$u = 60$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>-0.006932</td>
<td>-0.008188</td>
<td>-0.008862</td>
</tr>
<tr>
<td>(2)</td>
<td>0.000607</td>
<td>0.002876</td>
<td>0.003671</td>
</tr>
<tr>
<td>(3)</td>
<td>0.000594</td>
<td>0.002862</td>
<td>0.003666</td>
</tr>
<tr>
<td>(4)</td>
<td>0.000607</td>
<td>0.002877</td>
<td>0.003672</td>
</tr>
<tr>
<td>(5)</td>
<td>0.008147</td>
<td>0.013942</td>
<td>0.016206</td>
</tr>
<tr>
<td>$u = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>-0.000317</td>
<td>-0.000417</td>
<td>-0.000467</td>
</tr>
<tr>
<td>(2)</td>
<td>0.000016</td>
<td>0.000076</td>
<td>0.000097</td>
</tr>
<tr>
<td>(3)</td>
<td>0.000016</td>
<td>0.000075</td>
<td>0.000097</td>
</tr>
<tr>
<td>(4)</td>
<td>0.000016</td>
<td>0.000076</td>
<td>0.000097</td>
</tr>
<tr>
<td>(5)</td>
<td>0.000349</td>
<td>0.000569</td>
<td>0.000661</td>
</tr>
</tbody>
</table>

6.1.3. Comments on Examples 7 and 8

(i) In each case, the approximations are close together, and we can see from Example 8 that the approximations are close to the true values. As with the approximations to $G(u, y)$ when the individual claim amount distribution is exponential, approximations to $F(u, x, y)$ based on the bounds are slightly better for small values of $u$.

(ii) Example 8 illustrates that the lower bound for $F(u, x, y)$ can be negative, as can be the lower bound for $F(u, x)$. Thus, the bounds themselves may be of little practical value. However, averaging the bounds produces reasonable approximations since this process simply averages bounds for the functions $\psi(\cdot)$ and $G(\cdot, \cdot)$, and the bounds for $F(u, x)$ and $F(u, x, y)$ depend on the bounds for these functions.

(iii) We have seen in Example 1 that averaging bounds for $\psi(u)$ gives an excellent approximation to $\psi(u)$. Hence, when $u \leq x$ the average of the bounds for $F(u, x)$ should be a very good approximation to $F(u, x)$ since these bounds are linear functions of the bounds on $\psi(u)$.

### REFERENCES


DAVID C. M. DICKSON
Centre for Actuarial Studies, Faculty of Economics and Commerce
University of Melbourne
Parkville, Victoria 3052, Australia

ALFREDO D. EGÍDIO DOS REIS
Departamento de Matemática, Instituto Superior de Economia e Gestão
Rua Miguel Lupi 20, P-1200 Lisbon, Portugal

HOWARD R. WATERS
Department of Actuarial Mathematics and Statistics
Heriot-Watt University, Edinburgh
Midlothian EH14 4AS, Great Britain
WORKSHOP

PENSION FUNDING WITH TIME DELAYS AND THE OPTIMAL SPREAD PERIOD

BY STEVEN HABERMAN

Department of Actuarial Science and Statistics
The City University, London, UK

ABSTRACT

The paper extends earlier results by demonstrating that there is an optimal range of values for the period for amortizing valuation surpluses or deficiencies, in the case when there is a one year time delay between fixing a contribution rate and the accounting information about current fund levels. The optimal range is compared for the cases where there is no time delay and there is a one year time delay.

KEYWORDS

Pension funding; time delays; optimal spread period.

INTRODUCTION

We shall consider the financial structure of a defined benefit pension scheme, as represented by a simple mathematical model, which can be regarded as an extension to that originally proposed by TROWBRIDGE (1952). We focus on the effect of varying investment returns on the contribution rate and fund level for the scheme and consider possible choices of two important control parameters: the spread period and the delay in fixing contribution rates.

We consider defined benefit pension schemes where the benefits promised in the event of various contingencies are defined by a formula while the contributions are to be determined by the actuary by means of the valuation process. The funding method then represents the means by which the contribution rate is fixed at each valuation. We shall consider the case of annual valuations: at which the actuary values the prospective liabilities, allowing for future contributions to be paid, and compares this result with the value of the assets currently held.

The paper provides a natural follow-up to the earlier work of DUFRESNE (1988) and HABERMAN (1992) and gives a comparison with these earlier results.

As in these earlier papers, we shall consider the funding methods described by the following pairs of equations:

\[ C(t) = NC(t) + ADJ(t) \]

where \( C(t) \) is the contribution rate at time \( t \), \( NC(t) \) is the normal cost at time \( t \) and \( ADJ(t) \) is an adjustment to the contribution rate at time \( t \), represented by the liquidation of the unfunded liability, \( UL(t) \). \( UL(t) \) is defined by:

\[ UL(t) = AL(t) - F(t) \]

where $AL(t)$ is the total actuarial liability in terms of all members at time $t$ and $F(t)$ is the fund level at time $t$, measured in terms of the market value of the underlying assets.

We are using a discrete time approach with $t$ taking integer values 0, 1, 2 and so on.

At each time $t$, a valuation is carried out to estimate $C(t)$ and $F(t)$ based on the membership of the scheme at that time. As $t$ changes, however, we allow for new entrants to the membership so that the population remains stationary. See the assumptions listed below.

In the ensuing mathematical discussion, we make the following assumptions:

1. All actuarial assumptions are consistently borne out by experience, except for investment returns.
2. The population is stationary in size and structure from the start.
3. Salaries increase at a deterministic rate of inflation. For simplicity, each active member's annual salary is set at 1 unit at the minimum age at entry. There is no promotional salary scale. We allow for salary inflation by considering the real rate of investment return i.e. the rate in excess of salary inflation. In parallel, we assume that benefits in payment increase at the same rate of salary inflation.
4. The real interest rate assumption for valuation purposes is fixed.
5. It is assumed that the contribution income and benefit outgo occur at the start of each scheme year.

It is straightforward to relax some of these assumptions e.g. replace 2 by allowing the population to grow at a fixed compound rate (i.e. be stable in the sense of KEYFITZ (1985)); include a promotion salary scale in 3; use a different timing assumption in 5.

Assumptions 1.-4. imply that the following are constants with respect to time, $t$ (after rescaling to allow for the predetermined growth in line with salary inflation):

- $NC$: the total normal cost.
- $AL$: the total actuarial liability.
- $B$: the overall benefit outgo per unit time.

Also, assumptions 1., 2., 4. and 5. imply that the following equation of equilibrium holds:

$$AL = (1 + i) (AL + NC - B)$$

or equivalently

$$B = d \cdot AL + NC$$

where $d = i (1 + i)^{-1}$, the compound interest discount rate.

This equation of equilibrium can be also found in the earlier papers of TROWBRIDGE (1952) and BOWERS et al. (1976).

We make the following further assumptions regarding the real interest rate earned on the fund and the stochastic nature of $F(i)$:

6. The real interest rate earned on the fund during the period $(t, t+1)$ is $i (t+1)$, where $Ei(t+1) = i$, the real valuation rate of interest. Thus, the valuation rate is correct “on average”. This assumption is not essential mathematically but is in agreement with classical ideas on pension fund valuation. We further define $\sigma^2 = \text{Var} i(t+1)$. 


7. It is assumed that the earned real rates of return \( i(t) \) for \( t > 1 \) are independent, identically distributed random variables (with \( i(t) > -1 \) with probability 1).

8. \( \text{Prob}[F(0) = F_0] = 1 \) for some \( F_0 \).

Given these assumptions, the random variable \( i(t) \) leads to \( F(t) \) being a random variable and hence \( UL(t) \), \( ADJ(t) \) and \( C(t) \) being random variables.

A continuous time formulation would be possible, in which case stochastic differential equations would be utilised in the mathematical discussion rather than difference equations.

We are not suggesting (through assumption 7) that the rates of return actually achieved by pension funds form an independent and identically distributed sequence. Indeed, rates of return are more generally viewed as autoregressive-moving average processes (for example, PANJER and BELLHOUSE (1980)). In parallel work, HABERMAN (1991, 1993) has investigated the effect of using dependent investment return models, in particular autoregressive models of low order. These more sophisticated models are not pursued here. It is only because it keeps the mathematical discussion tractable that assumption 7 is imposed here.

**CHOICE OF ADJ: SPREAD PERIOD AND DELAY**

We consider a particular method for defining the contribution adjustment term \( ADJ(t) \) which is an approach widely used in the U.K. and involves putting \( ADJ(t) \) equal to the overall unfunded liability divided by the present value of an annuity for a term of \( M \) years, calculated at the valuation rate of interest \( i \). It is common practice to use values of \( M \) in the range 20-25 years, on the grounds that this would represent the average remaining active lifetime within the scheme of the current membership.

As in HABERMAN (1992), we shall allow for delays in the collection and processing of data and the preparation of the accounts, and assume that the adjustment term at time \( t \) depends on \( UL(t-1) \). Thus, with \( k = 1/\overline{a}_M \),

\[
ADJ(t) = k \cdot UL(t-q) \text{ where } q = 0 \text{ or } q = 1.
\]

(4)

\( q = 0 \) corresponds to the analysis of DUFRESNE (1988) and \( q = 1 \) corresponds to HABERMAN (1992). Then:

\[
C(t) = NC + k \cdot (AL - F(t-q)) \text{ where } q = 0 \text{ or } q = 1.
\]

(5)

We shall now view \( k \) (and hence \( M \)) and \( q \) as being control parameters which the actuary may choose with the objective of meeting certain specified criteria (see later) connected with controlling the behaviour of \( C(t) \) or \( F(t) \) over time. The choice of \( M \) would not be completely free: \( M \) would probably have a lower bound to limit the income tax deductibility of contributions and an upper bound to prevent large increases in \( UL(t) \).

Equation (5) includes a negative feedback component, whereby the current status is compared with a target and corrective action is taken to deal with the discrepancy.

With \( q = 1 \) in equation (4) we see that an element of delay is introduced into the way that changes in \( F(.) \) feed back into changes in \( C(.) \).
Other values of \( q \) are considered in detail in ZIMBIDIS and HABERMAN (1993) and are not discussed here.

We note that this choice of \( ADJ(t) \) uses the same fraction of the unfunded liability regardless of the latter's sign, so surpluses and deficiencies are treated in the same way, which would not always be the case in practice.

**MOMENTS OF \( F(t) \) and \( C(t) \)**

In the case of \( q = 0 \), we repeat from DUFRESNE (1988) the recurrence relation for \( F(t) \):

\[
F(t+1) = \frac{u(t+1)}{u}(pF(t)+r)
\]

and from HABERMAN (1992) the corresponding relation when \( q = 1 \)

\[
F(t+1) = \frac{u(t+1)}{u}(uF(t)-ukF(t-1)+r)
\]

where we have introduced the subsidiary parameters

\[
u = (1+i), \quad p = (1+i)(1-k), \quad r = (1+i)(NC-B+kAL)
\]

and

\[
u(t+1) = 1 + i(t+1).
\]

Using conditional expectation and variance based methods, DUFRESNE (1988) obtains explicit equations for the expectation and variance of \( F(t) \) and \( C(t) \) for finite \( t \) when \( q = 0 \). In the limit as \( t \to \infty \), he demonstrates that, providing that \( M > 1 \),

\[
\lim_{t \to \infty} E \ F(t) = AL \\
\lim_{t \to \infty} E \ C(t) = NC
\]

and that providing that \( y(1-k)^2 < 1 \)

\[
\lim_{t \to \infty} Var \ F(t) = \frac{\sigma^2 AL^2}{u^2(1-y(1-k)^2)}
\]

where \( y = \sigma^2 + u^2 \), and

\[
\lim_{t \to \infty} Var \ C(t) = \frac{\sigma^2 k^2 AL^2}{u^2(1-y(1-k)^2)}
\]

In this discussion, we exclude pay-as-you-go funding and terminal funding for which \( AL = 0 \) and initial funding.
Using conditional expectation and generating function based methods, Haberman (1992) similarly obtain equations for the first two moments of $F(t)$ and $C(t)$ when $q = 1$. In the limit as $t \to \infty$, he demonstrates that

$$\lim_{t \to \infty} EF(t) = AL$$
$$\lim_{t \to \infty} EC(t) = NC$$

providing that $M \geq 2$ and $\alpha_M > 1$.

Under more complex restrictions on the parameters (Appendix I), Haberman (1992) obtains, for the case $q = 1$, that

$$\lim_{t \to \infty} \text{Var} F(t) = \frac{\sigma^2 AL^2 (1 + uk)}{u^2 (1 + ku - y(1 - uk + k^2 + uk^3))}$$

and

$$\lim_{t \to \infty} \text{Var} C(t) = \frac{\sigma^2 k^2 AL^2 (1 + uk)}{u^2 (1 + ku - y(1 - uk + k^2 + uk^3))}$$

[Note that there is a typographical error in equations (14), (15) and (B.6) in Haberman (1992)].

**Trade off in variances**

We introduce the following notation for the scaled variances

$$\alpha_i(M) = \frac{\lim \text{Var} F(t)}{(\lim EF(t))^2} \quad \text{for} \quad q = i \quad \text{where} \quad i = 0 \lor 1$$

and

$$\beta_i(M) = \frac{\lim \text{Var} C(t)}{(\lim EC(t))^2}$$

Then Dufresne has shown that, if $y > 1$, then there exists $M^*$ such that

i) for $M \leq M^*$, $\alpha_0(M)$ increases and $\beta_0(M)$ decreases with increasing $M$,

ii) for $M > M^*$, both $\alpha_0(M)$ and $\beta_0(M)$ increase with increasing $M$ and that

$$\hat{a}_{M^*} = \frac{1}{k^*} \quad \text{where} \quad k^* = 1 - \frac{1}{y}.$$ 

In a sense, the choice of $M$ in the range of $(1, M^*)$ is "optimal". If our objective in choosing $M$ is to reduce uncertainty and to keep the limiting variances of $F(t)$ and $C(t)$ to a minimum, then any $M > M^*$ is to be discarded since clearly some
other $M < M^*$ would at least reduce $\alpha_0(M)$ while keeping $\beta_0(M)$ the same. For most pension funds, it is likely that the variance of $C(t)$ will be the principal criterion of interest.

We shall now consider the extent to which similar properties hold for the case $q = 1$.

We firstly note that, for $M > 1$,

\[
\begin{align*}
\alpha_0(M) &< \alpha_1(M) \\
\text{and} \\
\beta_0(M) &< \beta_1(M),
\end{align*}
\]

(as demonstrated in the numerical examples in HABERMAN (1992)).

The proof is straightforward. We consider

\[
\frac{\alpha_1(M)}{\alpha_0(M)} = \frac{\beta_1(M)}{\beta_0(M)} = \frac{(1 + uk)(1 - y(1 - k)^2)}{1 + uk - y(1 - ku + k^2 + uk^3)}
\]

\[
= \frac{1 - y(1 - k)^2}{1 - y(1 - ku + k^2 + uk^3)/(1 + uk)}
\]

\[
= \frac{1 - y(1 + k^2 - 2k)}{1 - y(1 + k^2 - 2uk/(1 + uk))}
\]

The difference between the terms in the numerator and denominator is the coefficient of "2k". Now, if $M > 1$,

\[
\frac{u}{1 + uk} < 1 \text{ because } 1 + uk - u = \frac{1}{\alpha_M} - i > 0.
\]

Hence $\frac{\alpha_1(M)}{\alpha_0(M)} = \frac{\beta_1(M)}{\beta_0(M)} > 1$.

(HABERMAN and ZIMBIDIS (1993) show that these inequalities hold for higher values of $q$). This result is intuitive: the introduction of a one year time delay means that we have lost information about the fund since time $t - 1$ and we would expect the resulting variances to be increased.

We consider the behaviour of $\alpha_1(M)$ as $M$ varies. It is convenient to view $\alpha_1(\ )$ as a function of $k$ and then use the 1-1 correspondence between values of $k$ and $M$.

We can show that

\[
\frac{d}{dk} \alpha_1(k) = \frac{\sigma^2 y (k + ku)^2 - u}{u^2 (1 + ku - y(1 - ak + k^2 + ak^3))^2}.
\]

We are interested in the turning points of $\alpha_1(k)$ in the range for $k$ of $(d, 1)$, corresponding to values of $M$ in the range $(1, \infty)$.
The cubic equation $p(k) = k(1 + ku)^2 - u$ has only one real root since

$$D = \frac{1}{u^4} (27u^2 + 4) > 0$$

(see Appendix II). We let $k_1$ be this real root.

We note that $p(d) = u(i - 1) < 0$ (if $|i| < 100\%$) and $p(1) = 1 + u + u^2 > 0$.

Then $p(k) < 0$ for $d < k < k_1$ and $p(k) > 0$ for $k_1 < k < 1$.

$k_1$ depends on the value of $u$ and numerical experiments indicate the following values:

<table>
<thead>
<tr>
<th>$i$</th>
<th>$u$</th>
<th>$k_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.4656</td>
</tr>
<tr>
<td>1%</td>
<td>1.01</td>
<td>0.4666</td>
</tr>
<tr>
<td>5%</td>
<td>1.05</td>
<td>0.4704</td>
</tr>
<tr>
<td>10%</td>
<td>1.10</td>
<td>0.4707</td>
</tr>
<tr>
<td>20%</td>
<td>1.20</td>
<td>0.4818</td>
</tr>
</tbody>
</table>

In each of the cases, $k_1$ approximately corresponds to $M_1 = 2$.

Hence

$$\frac{d}{dk} \alpha_1(k) < 0 \text{ for } d < k < k_1$$

and

$$\frac{d}{dk} \alpha_1(k) > 0 \text{ for } k_1 < k < 1$$

(subject to $k$ satisfying the constraints implied by Appendix I) which are equivalent to

$$\frac{d}{dM} \alpha_1(M) < 0 \text{ for } 1 < M < M_1 = 2$$

and

$$\frac{d}{dM} \alpha_1(M) > 0 \text{ for } M_1 < M < \infty.$$ 

We now consider the behaviour of $\beta_1(k)$, viewed initially as a function of $k$. We can show that

$$\frac{d}{dk} \beta_1(k) = \frac{\sigma^2 k(1 - y + ku(2 - y) + u^2 k^2 (1 + y))}{u^2 (1 + ku - y(1 - uk + k^2 + uk^3))^2}$$

(14)
We are interested in the turning points of $\beta_1(k)$ in the feasible range $k$ of $(d, 1)$, again subject to the constraints implied by Appendix I, corresponding to values of $M$ in the range $(1, \infty)$.

The quadratic equation $s(k) = 1 - y + ku(2 - y) + u^2k^2(1 + y)$ has one real root in the specified range since $y > 0.8$. We let $k_2$ be this real root.

We note that
\[
s(0) = 1 - y < 0
\]
and
\[
s(1) = 1 - y + u(2 - y) + u^2(1 + y) = u^3(u - 1) + 2u + 1 + \sigma^2(u^2 - u - 1)
\]
so
\[
s(1) > 0 \text{ if } i > 61.8\%, \text{ given that } u > 0
\]
or
\[
s(1) > 0 \text{ if } \sigma^2 < \frac{u^3(u - 1) + 2u + 1}{1 + u - u^2} = \sigma_1^2, \text{ say.}
\]

[This restriction on $\sigma^2$ is not too onerous! It would correspond to the following values:]

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\sigma_1^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0$</td>
<td>$300%$</td>
</tr>
<tr>
<td>$1%$</td>
<td>$306%$</td>
</tr>
<tr>
<td>$5%$</td>
<td>$333%$</td>
</tr>
<tr>
<td>$10%$</td>
<td>$375%$</td>
</tr>
<tr>
<td>$20%$</td>
<td>$493%$</td>
</tr>
</tbody>
</table>

Clearly $k_2$ corresponds to a minimum value of $\beta(k)$.

The explicit value of $k_2$ is
\[
(15) \quad k_2 = \frac{- (2 - y) + \sqrt{y(5y - 4)}}{2u(1 + y)}
\]

Thus
\[
\frac{d}{dk} \beta_1(k) < 0 \text{ for } d < k < k_2
\]
and
\[
\frac{d}{dk} \beta_1(k) > 0 \text{ for } k_2 < k < 1.
\]

If $k_2$ corresponds to $M_2$ we can translate this statement into
\[
\frac{d}{dM} \beta_1(M) < 0 \text{ for } 1 < M < M_2
\]
and

\[ \frac{d}{dM} \beta_1(M) > 0 \text{ for } M_2 < M < \infty. \]

Given the restrictions on the parameters mentioned in the above discussion, we see that for the case of \( q = 1 \) that is also a trade off between variability in \( C(t) \) and variability in \( F(t) \) and that there is an optimal choice of \( M \), and hence of \( k \), if our objective is to keep the limiting variances to a minimum. The optimal spread period in this case is \((1, M_2)\).

A comparison of the optimal periods defined by \( M^* \) and \( M_2 \) (for the cases \( q = 0 \) and \( q = 1 \)) would be useful. Again it is convenient to examine the corresponding annuity values, as represented by \( k \).

We note that

\[ s'(k^*) = u(2 - y) + 2u^2(1 + y) \frac{(y - 1)}{y} \]

\[ = \frac{u}{y} [2(\sigma^2 + u^2 - u) + y^2(2u - 1)]. \]

The sign of \( s'(k^*) \) depends on the values of \( u \) and \( \sigma \). Clearly, if \( u \geq 1 \) (i.e. \( i > 0 \)) then \( s'(k^*) > 0 \). Since \( s(k) \) is a quadratic, with minimum at \( k = k_2 \), this implies that \( k^* > k_2 \) and hence that \( M_2 > M^* \) (for \( i > 0 \)).

This is confirmed by the numerical example given in Haberman (1992). As Table 1 illustrates, the values of \( M_2 \) and \( M^* \) are numerically close.

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>0</th>
<th>0.01</th>
<th>0.03</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M^* ) and ( M_2 )</td>
<td>401</td>
<td>60</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>( M^* ) and ( M_2 )</td>
<td>101</td>
<td>42</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>( M^* ) and ( M_2 )</td>
<td>45</td>
<td>28</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>( M^* ) and ( M_2 )</td>
<td>26</td>
<td>19</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>( M^* ) and ( M_2 )</td>
<td>17</td>
<td>14</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

### CONCLUSIONS

A simple stochastic model is used to represent the real investment returns for a defined benefit pension scheme.
The paper shows that in the presence of a one year time delay between fixing a contribution rate and the information about current fund levels, it is possible to set up formulae for studying the limiting behaviour of the expected values and variances of the contributions and fund levels. The paper demonstrates that, as with the case when there is no time delay, there is an optimum range of values for the spread period, $M$ (for amortizing valuation surpluses or deficiencies). The relationship between the optimum range of values of $M$ in the case of no time delay ($q = 0$) and with a one year time delay ($q = 1$) is investigated.

ACKNOWLEDGEMENT

This research work was performed under EEC contract SPES-CT91-0063.

Appendix I

Based on Haberman (1992) Appendix C, we require the following conditions for the convergence of equations (10):

With $b = \sigma^2 + u^2 - uk$

\[
c = (\sigma^2 + u^2)k(u-k)
\]

\[
e = (\sigma^2 + u^2)uk^3
\]

\[
D = 27e^2 + 4c^3 - 18bce - b^2c^2 + 4b^3e,
\]

we require

i) $1 + c > |b + e|

ii) (a) if $D > 0$, $e^3 - be + c - 1 < 0$

(b) if $D < 0$, $|b| < \frac{1}{2}(3 + c)$.

i) and ii) can be considered to provide restrictions on $k$ (and hence $M$) or on $\sigma^2$.

Appendix II – Roots of a cubic equation

In general, the roots of the cubic equation

\[
p(x) = x^3 - bx^2 + cx - e
\]

are

\[
x_1 = \frac{1}{3} b + U + V
\]

\[
x_2, x_3 = \frac{1}{3} b - \frac{1}{2} (U + V) \pm \frac{\sqrt{3}}{2} i(U - V)
\]
where

\[ U, V = \left[ \frac{b^3}{27} - \frac{bc}{6} + \frac{e}{2} \pm \frac{1}{6\sqrt{3}} \sqrt{D} \right] \]

and

\[ D = 27e^2 + 4c^3 - 18bce - b^2c^2 + 4b^3e. \]

REFERENCES


Steven Haberman

Department of Actuarial Science and Statistics

The City University

Northampton Square

London ECIV GHB

UK.
In the classical Bayesian approach to credibility the claims are conditionally independent and identically distributed random variables, with common density \( f(x, \vartheta) \). The unknown parameter \( \vartheta \) is a realization of a random variable \( \Theta \) having initial (prior) density \( u(\vartheta) \). Let

\[
\mu(\vartheta) = \int x f(x, \vartheta) \, dx.
\]

The initial pure premium is

\[
E[X_1] = E[\mu(\Theta)].
\]

The premium for \( X_{t+1} \), given \( X_1, \ldots, X_t \), is the conditional expectation

\[
E[\mu(\Theta) \mid X_1, \ldots, X_t] = \int \mu(\vartheta) u(\vartheta \mid X_1, \ldots, X_t) \, d\vartheta.
\]

A central question is for which pairs \( f(x, \vartheta) \) and \( u(\vartheta) \) this expression is linear, i.e. of the form

\[
Z \cdot \bar{X} + (1 - Z) \cdot E[\mu(\Theta)]
\]

where \( \bar{X} = (X_1 + \ldots + X_t)/t \) is the observed average. This is indeed the case for about half a dozen famous examples. JEWELL (1974) has found an elegant and general approach to unify these examples, see also GOOVAERTS and HOOGSTAD (1987, chapter 2). The classical examples can be retrieved as special cases; however a preliminary reparameterization has to be performed on a case by case basis. The purpose of this note is to propose an alternative (but of course strongly related) formulation of the general model, from which the classical examples can be retrieved in a straightforward way.

The common density of the claims is supposed to be of the form

\[
f(x, \vartheta) = \frac{a(x) \cdot b(\vartheta)^x}{c(\vartheta)}, \quad x \in A.
\]

Here \( A \) is the set of possible values of the claims (discrete or continuous), and \( c(\vartheta) \) is the normalizing constant:

\[
c(\vartheta) = \int_A a(x) \cdot b(\vartheta)^x \, dx.
\]
Then
\[ \mu(\theta) = \frac{b(\theta) c'(\theta)}{b'(\theta) c(\theta)} . \]

As a prior density we choose
\[ u(\theta) = \frac{c(\theta)^{-n_0} b(\theta)^{x_0} b'(\theta) d(n_0, x_0)}{d(n_0, x_0)} , \]

where \( \theta \) varies in some interval,
\[ d(n_0, x_0) = \int c(\theta)^{-n_0} b(\theta)^{x_0} b'(\theta) d\theta \]

is the normalizing constant and \( n_0 \) and \( x_0 \) are two parameters. Then it is easy to see that the posterior density
\[ u(\theta | X_1, ..., X_t) \]

is a member of the same family, with updated parameter values:
\[ n_t = n_0 + t, \quad x_t = x_0 + X_1 + ... + X_t . \]

Hence, if we have an expression for \( E[\mu(\Theta)] \), it suffices to replace \( n_0 \) by \( n_t \) and \( x_0 \) by \( x_t \) to obtain \( E[\mu(\Theta) | X_1, ..., X_t] \).

By definition,
\[ E[\mu(\Theta)] = \frac{1}{d(n_0, x_0)} \int c(\theta)^{-n_0 - 1} c'(\theta) b(\theta)^{x_0 + 1} d\theta \]

Now we perform a partial integration and assume that the function
\[ c(\theta)^{-n_0} b(\theta)^{x_0 + 1} \]

vanishes at the integration limits.

Then we obtain
\[ E[\mu(\Theta)] = \frac{x_0 + 1}{n_0 d(n_0, x_0)} \int c(\theta)^{-n_0} b(\theta)^{x_0} b'(\theta) d\theta = \frac{x_0 + 1}{n_0} . \]

Hence the premium for \( X_{t+1} \) is
\[ \frac{x_t + 1}{n_t} = \frac{x_0 + t \cdot \bar{X} + 1}{n_0 + t} = Z \cdot \bar{X} + (1 - Z) \cdot E[\mu(\Theta)] \text{ with } Z = \frac{t}{n_0 + t} . \]
The classical examples can be retrieved directly as follows:

a. POISSON-GAMMA

\[ A = \{0, 1, 2, \ldots\}, \quad 0 < \vartheta < \infty \]

\[ a(x) = \frac{1}{x!}, \quad b(\vartheta) = \vartheta, \quad c(\vartheta) = e^\vartheta \]

\[ u(\vartheta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \vartheta^{\alpha-1} e^{-\beta \vartheta} \quad \text{with} \quad \alpha = x_0 + 1, \quad \beta = n_0 \]

b. GEOMETRIC-BETA

\[ A = \{0, 1, 2, \ldots\}, \quad 0 < \vartheta < 1 \]

\[ a(x) = 1, \quad b(\vartheta) = 1 - \vartheta, \quad c(\vartheta) = \frac{1}{\vartheta} \]

\[ u(\vartheta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \vartheta^{\alpha-1} (1 - \vartheta)^{\beta-1} \quad \text{with} \quad \alpha = n_0 + 1, \quad \beta = x_0 + 1 \]

c. EXPONENTIAL-GAMMA

\[ A = (0, \infty), \quad 0 < \vartheta < \infty \]

\[ a(x) = 1, \quad b(\vartheta) = e^{-\vartheta}, \quad c(\vartheta) = \frac{1}{\vartheta} \]

\[ u(\vartheta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \vartheta^{\alpha-1} e^{-\beta \vartheta} \quad \text{with} \quad \alpha = n_0 + 1, \quad \beta = x_0 + 1 \]

d. NORMAL-NORMAL

\[ A = (-\infty, \infty), \quad -\infty < \vartheta < \infty \]

\[ a(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad b(\vartheta) = \exp\left(\frac{\vartheta}{\sigma^2}\right), \quad c(\vartheta) = \sqrt{2\pi} \sigma \exp\left(\frac{\vartheta^2}{2\sigma^2}\right) \]

\[ u(\vartheta) = \frac{1}{\sqrt{2\pi} v} e^{-\left(\frac{(\vartheta - \mu)^2}{2v^2}\right)} \quad \text{with} \quad \mu = \frac{x_0 + 1}{n_0}, \quad v^2 = \frac{\sigma^2}{n_0} \]
e. BERNOLLI-BETA

\[ A = \{0, 1\}, \ 0 < \theta < 1 \]

\[ a(x) = 1, \ b(\theta) = \frac{\theta}{1 - \theta}, \ c(\theta) = \frac{1}{1 - \theta} \]

\[ u(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \theta^{\alpha - 1} (1 - \theta)^{\beta - 1} \quad \text{with} \quad \alpha = x_0 + 1, \ \beta = n_0 - x_0 - 1 \]

ACKNOWLEDGMENT

At the ASTIN Colloquium in Leuven the author had fruitful and joyful discussions with Hans Bühlmann and William Jewell. They contributed to an improvement of this note.

REFERENCES


BOOK REVIEW


The all embracing nature of the title of this published thesis is perhaps somewhat misleading given the content. In the thesis, aspects of three non-life actuarial techniques viz the ordering of risks, credibility theory and portfolio models are applied in a number of actuarial life insurance settings. The material is well organised into three parts with a brief well balanced discussion of the issues at the end of each part. Some of the work is relatively new and sections of the thesis have been published elsewhere in scientific journals. The list of references is informative and potentially useful.

The first part of the thesis is devoted to the ordering of risks in the analysis of life insurances or annuities. It begins with a restatement of basic definitions and results for moment ordering, stochastic ordering and stop-loss ordering in terms of well defined random variables and their distribution functions. There then follow sections dealing with both single life and multi-state applications. The former contains a study of the effects of changes in the remaining lifetime distribution on single premiums, ordering by variance through the optimisation of loss functions together with generalisations to stop-loss ordering. The multi-state application is based on various orderings applied to two dependent lives.

The second part of the thesis is concerned with aspects of credibility theory, in particular, the application of regression methods in credibility theory to parametric graduation at elderly ages. While the author begins by documenting references to credibility applications for portfolios of grouped life insurances, the main emphasis is on a review of the technical detail of both the Hachemeister regression model and the De Vylder non-linear regression model in credibility theory. Adjustments to the detail of these models are discussed and credibility regression methodology applied to the graduation of mortality data, at elderly ages, by estimating the parameters of the Makeham mortality law. While the concept of using credibility theory to smooth mortality data at elderly ages in this way, where the data are sparse, has an appealing ring, it is perhaps questionable whether the method will attract general applicability given both the wide range of well established flexible graduation techniques available and the small reported discontinuity in the graduations induced by the method.

The third and final substantive part of the thesis is concerned with two, largely theoretical, aspects of portfolio models. The first of these deals mainly with computational aspects of the total claims distribution over a fixed limited time period. While a brief review of the potential for using the individual model in a life insurance setting is given, the great bulk of this work focuses on numerical recursive schemes based on the collective model. The detailed implementation of a new numerical procedure for a class of compound generalised distributions and

which generalises existing methods is described within this context. This is accompanied by only a partial review of numerical portfolio models since a comprehensive review is already to be found in Panjar & Willmot (1992). Given both the title of the thesis and the non-life insurance background underpinning the more recent developments in numerical portfolio models, the emphasis on the specific application to life insurance is perhaps somewhat lightweight here. The second of the two aspects looks at the multi-period approach and deals mainly with computational aspects of the survival probability in finite time ruin theory. A recursive procedure for calculating finite time ruin probabilities for the compound Poisson process is advanced and a comparison made with other methods due to Wikstad, DeVylder & Goovaerts and Dickson & Waters. The procedure is extended to more general claim number distributions using a renewal equation approach. This part of the thesis concludes with some brief comments on the practical implement of the methods described.

REFERENCES


Arthur E. Renshaw
UBS RESEARCH FELLOWSHIPS
IN FINANCIAL MATHEMATICS

Applications are invited for two three-year post-doctoral research fellowships in financial mathematics, to be instituted in collaboration with the Union Bank of Switzerland (UBS). The first position is affiliated to the Statistical Laboratory of the University of Cambridge, the second to the Department of Mathematics of the Eidgenössische Technische Hochschule (ETH) in Zürich. Besides the establishment of research contacts between Cambridge and the ETH, the research fellows will be expected to interact closely with derivatives professionals at UBS, both in London and Zürich. In order to facilitate this collaboration, the research fellows will be expected to spend some time at both the London and Zürich offices of UBS.

It is expected that the persons appointed will be recent Ph.Ds with outstanding records and direct research experience in financial mathematics or expertise in an appropriate area of applicable mathematics. The mathematical backgrounds of the persons appointed are likely to be in one, or more, of the following areas:

(i) probability theory and stochastic calculus;
(ii) mathematical modelling and applied probability;
(iii) optimization and stochastic control theory;
(iv) numerical methods for partial or stochastic differential equations;
(v) the statistical analysis of financial time series.

Applications, including a curriculum vitae and the names of three referees (at least two of which should be academic) should be sent by 29th February 1996 to The Director, Statistical Laboratory, 16 Mill Lane, Cambridge CB2 1SB, UK (for the Cambridge position), and to Professor P. Embrechts, Department of Mathematics, ETH Zentrum, CH-8092 Zürich, Switzerland (for the ETH position). If not otherwise indicated, candidates will be considered as applying for either position.

Both the University of Cambridge and the ETH Zürich aim to achieve the highest quality in teaching and research.

The University of Cambridge and ETH Zürich are Equal Opportunities Employers.
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.

   or to one of the Co-Editors:
   Harry Reid,
   D. Eagle Star Insurance Company Ltd
   The Grange, Bishop’s Cleeve
   Cheltenham Glos GL52 4XX, United Kingdom.

2. Manuscripts should be typewritten on one side of the paper, double-spaced with wide margins.
   Submission of a paper is held to imply that it contains original unpublished work and is not being
   submitted for publication elsewhere.

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 chronologically. Use a, b, c,
   etc. to separate publications of the same author in the same year. For journal references give
   author(s), year, title (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:

   JEWELL, W.S. (1975a) Model variations in credibility theory. In Credibility: Theory and

   JEWELL, W.S. (1975b) Regularity conditions for exact credibility. ASTIN Bulletin 8,
   336–341.

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

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

   Continued overleaf.
COMMITTEE OF ASTIN

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hans Bühlmann</td>
<td>Switzerland</td>
<td>Honorary Chairman</td>
</tr>
<tr>
<td>James N. Stanard</td>
<td>USA</td>
<td>Chairman</td>
</tr>
<tr>
<td>Jukka Rantala</td>
<td>Finland</td>
<td>Vice-Chairman</td>
</tr>
<tr>
<td>Bouke Posthuma</td>
<td>Netherlands</td>
<td>Secretary</td>
</tr>
<tr>
<td>Jean Lemaire</td>
<td>Belgium/USA</td>
<td>Treasurer/IAA-Delegate</td>
</tr>
<tr>
<td>Paul Embrechts</td>
<td>Switzerland</td>
<td>Editor</td>
</tr>
<tr>
<td>D. Harry Reid</td>
<td>United Kingdom</td>
<td>Editor</td>
</tr>
<tr>
<td>James MacGinitie</td>
<td>USA</td>
<td>Member/IAA-Delegate</td>
</tr>
<tr>
<td>Björn Aine</td>
<td>Sweden</td>
<td>Member</td>
</tr>
<tr>
<td>Edward J. Levay</td>
<td>Israel</td>
<td>Member</td>
</tr>
<tr>
<td>Charles Levi</td>
<td>France</td>
<td>Member</td>
</tr>
<tr>
<td>Thomas Mack</td>
<td>Germany</td>
<td>Member</td>
</tr>
<tr>
<td>Ermanno Pitacco</td>
<td>Italy</td>
<td>Member</td>
</tr>
<tr>
<td>Gregory C. Taylor</td>
<td>Australia</td>
<td>Member</td>
</tr>
<tr>
<td>René Schnieper</td>
<td>Switzerland</td>
<td>Co-Editor</td>
</tr>
<tr>
<td>David Wilkie</td>
<td>United Kingdom</td>
<td>Co-Editor</td>
</tr>
</tbody>
</table>

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.