Volume 19, No. 2 November 1989

ASTIN BULLETIN

A Journal of the International Actuarial Association

CONTENTS

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* will publish 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.

ASTIN BULLETIN appears twice a year, each issue consisting of about 80 pages.

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

MEMBERSHIP

ASTIN is a section 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 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 1700.

INDEX TO VOLUMES 1-13

The Cumulative Index to Volumes 1-13 is also published for ASTIN by Ceuterick at the above address and is available for the price of BEF 340.

EDITORIAL AND ANNOUNCEMENTS

EDITORIAL

Actuarial Software Packages: a Chance and a Challenge

Actuaries working in non-life insurance know very well how long is the road from theory to practical application. In other words, how difficult it is to put mathematical methods and models to practical use. Here are just some of the difficulties involved: time pressure; the available data are incomplete and/or inexact; real life problems tend to be complex, "dirty" and difficult to fit into the strict corset of a mathematical model; practical actuaries usually have little time for research and hardly get around even to following passively the new developments in actuarial science by reading the relevant literature. This means that an actuary working in practical insurance has to fulfill high demands. He is constantly required to bridge the considerable gap between correct scientific methodology and the practical needs of the insurance business. He ought to have large practical experience, profound knowledge in risk theory and non-life insurance mathematics, and, last but not least, be an expert in numerical methods and programming. This last point especially is not to be found among the first preferences and interests of an actuary. Although most actuaries are accustomed to using a computer as a technical aid and to writing their own programmes, the programming work for implementing a sophisticated mathematical method is in general very time consuming. Hence new advanced actuarial methods are often rejected for the simple reason that the time required for the programming is considered to be too much. This is where suitable actuarial software packages would bring welcome relief.

Just a few years ago, software packages did not exist in the field of non-life mathematics. Only recently have things started to improve. The first actuarial software packages have come onto the market, especially in the fields of claims reserves, credibility theory and calculation of total claims distributions. Such a development is only to be welcomed. A look across the fence to related fields shows that suitable software packages are likely to have a substantial stimulating impact on applying theoretical results in practice. In classical statistics, for example, methods such as general linear models or time series analysis are nowadays widely used in practice, e.g. in natural science, economics and medecine. The basic theory was already developed in the fifties (linear models) and in the sixties (time series). But the breakthrough in practice only happened some fifteen years later with the availability of statistical software packages. In financial mathematics, the Markowitz approach, one of the bases of modern portfolio theory, goes back to 1952. The famous CAPM-relation (capital asset pricing model) was discovered in the mid sixties. But only in recent years have these theories begun to establish themselves in the practical routines of banks, financial institutions, and insurance companies. One of the reasons for this time-lag is that well-tested computer software with fast and efficient numerical algorithms, carrying out the numerous calculations within the required short period of time, only appeared on the market a relatively short time ago.

It is certain that software packages can only relieve the actuary of a part of his programming. The necessary data have first to be selected, prepared and put into a given format. It is also certain that practical problems in non-life insurance are often individual and specific. It is therefore argued that standard software is of limited use. I agree with this. But is it not equally true for the related fields mentioned above, where software packages are already widely used? In any case, it seems to me, that well tested computer programmes in the field of actuarial mathematics can only be an advantage to the actuarial community. They are a chance for the practitioners to apply more mathematics and to put more sophisticated methods and models to use. Furthermore, certain standards will be set, which should have a positive effect on the overall professional level. One condition of such software being used by a larger group of users is, however, that the input-output-interfaces are well organised and that the programmes are user-friendly. There is also a great danger connected with such software packages: they can be used in the wrong way. A glance at the related fields mentioned above shows what nonsense often results if such packages are used as a magic black box by non-professionals. A profound knowledge of the underlying theories and implications are indispensable to make the best use of such packages for practical purposes. Hence they are also a challenge to the actuary to keep his mathematical knowledge up to date.

ASTIN should be the breeding place for the interaction between sound theoretical thinking and practical application. One of our targets is to support all activities with the aim of putting mathematical models to practical use. In connection with actuarial software, this could mean that ASTIN promotes the spread of knowledge about such products among the actuarial community. A first step in this direction was the decision of the ASTIN Committee in 1987 to establish an actuarial software library (see IAA Bulletin Nr. 6, p. 19). The editors of the ASTIN Bulletin are also prepared to supplement the Book Reviews column with Software Reviews provided they can find persons willing to write such reviews. Should more be done? One could, for example, consider selling advertising space in the ASTIN Bulletin to the suppliers of such software. Would it be an idea for the local ASTIN groups to organise from time to time a demonstration of and discussion on actuarial software? Any suggestions as well as any opinions coming from our readers will be welcomed by the editors.

ALOIS GISLER

OBITUARY

JEAN HAEZENDONCK

1940-1989

Wednesday April 26, 1989 Prof. Dr. JEAN HAEZENDONCK died suddenly. JEAN M. HAEZENDONCK was born on May 8, 1940 in Vilvoorde (Belgium). He studied mathematics at the "Université Libre de Bruxelles" (U.L.B.). He continued his mathematical studies in Paris under the guidance of Prof. Dr. NEVEU and in 1969 he obtained his Ph. D. at the "Vrije Universiteit Brussel" (V.U.B.). He was professor of probability theory at the "Universitaire Instelling Antwerpen" (U.I.A.) and extraordinary professor at the V.U.B. At the U.I.A. he founded an active research group working in risk theory and insurance problems. He organised several international meetings and was one of the founders and thriving forces of Insurance Mathematics and Economics. He was one of the exceptions who didn't have but friends. Many of us will remember him as a perfect gentleman appreciated very much by all of his former mathematical and or actuarial students. We will miss him both as a colleague and as a dear friend. His wife and two children were a genuine support for his scientific work. We wish them strength.

MARC J. GOOVAERTS

ARTICLES

STOCHASTIC INTEREST RATES AND AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES

By Jan Dhaene

Instituut voor Actuariële Wetenschappen, K.U.Leuven, Belgium

Abstract

A practical method is developed for computing moments of insurance functions when interest rates are assumed to follow an autoregressive integrated moving average process.

Keywords

ARIMA (p, d, q)-processes; stochastic interest rates; moments of insurance functions.

1. Introduction

In most of the insurance literature the theory of life contingencies is developed in a deterministic way. This means that mortality happens according to an a priori known mortality table and that the interest rate is assumed to have a constant value. Nevertheless, the traditional theory of life contingencies implicitly deals with the stochastic nature of mortality and interest rates in that conservative assumptions are taken.

A first step forward was to consider the time until decrement as a random variable, while the interest rate was assumed to be constant. This approach is followed in Bowers et al. (1987). This (as one could call) "semi-stochastic" approach contains the traditional theory in that most actuarial functions can be considered as the expected values of certain stochastic functions.

It is only since about 1970 that there has been interest in actuarial models which consider both the time until death and the investment rate of return as random variables.

Boyle (1976) includes the stochastic nature of interest rates in assuming that the force of interest is generated by a white noise series, that is forces of interest in the successive years are normally distributed and uncorrelated.

In the approach of Pollard (1971) the force of interest in a year is related to the force of interest in the preceding years by using an autoregressive process of order two.

Panjer and Bellhouse (1980) and Bellhouse and Panjer (1981) develop a general theory including continuous and discrete models. The theory is further worked out for unconditional and conditional autoregressive processes of order one and two.

GIACCOTTO (1986) develops an algorithm for evaluating present value functions when interest rates are assumed to follow an ARIMA $(p, 0, q)$ or an ARIMA $(p, 1, q)$ process.

The goal of this study is to state a methodology for computing in an efficient manner present value functions when the force of interest evolves according to an autoregressive integrated moving average process of order (p, d, q). As will be seen, the method developed here will require less computing time than Giaccotto's method for autoregressive integrated moving average processes of order $(p, 0, q)$ or $(p, 1, q)$.

It should be remarked that we assume that mortality and interest rates posses a certain stochastic nature and that only accidental fluctuations in this mortality and interest rates are considered. Other fluctuations due to mortality improvement, underwriting practice, the choice of a wrong interest model, investment strategy and so on are not considered here.

2. GENERAL THEORY

The theory developed in this section is mainly based on the work of PANJER and BELLHOUSE (1980) and BELLHOUSE and PANJER (1981).

Let D_t be the stochastic variable denoting the discounted value of one dollar payable in t years ($t = 0, 1, 2, \ldots$). The stochastic variable X_t defined by

$$D_t = \exp(-X_t) \qquad t = 0, 1, 2, \ldots \tag{1}$$

can be interpreted as the force of interest over the first t years.

If δ_i is the force of interest in the i-th year ($i = 1, 2, \ldots$), then

$$X_0 = 0$$

$$X_t = \sum_{i=1}^{t} \delta_i \qquad t = 1, 2, \ldots \tag{2}$$

It is assumed that X_t is normally distributed with mean $\mu(t)$ and variance-covariance function $a(t, s)$. The variance of X_t is equal to $a(t, t)$ and is denoted by $\sigma^2(t)$.

It is immediately seen that $E[D_t^k]$ and $E[D_t^k D_s^l]$ are the moment generating functions of the normal distributed variables kX_t and $(kX_t + lX_s)$ calculated for the value (-1). So one finds that

$$E[D_t^k] = \exp\left[-k\mu(t) + \frac{k^2}{2}\sigma^2(t)\right] \qquad t, k \geq 1 \tag{3}$$

and

$$E[D_t^k D_s^l] = \exp\left[-k\mu(t) - l\mu(s) + \frac{k^2}{2}\sigma^2(t) + \frac{l^2}{2}\sigma^2(s) + kla(t, s)\right] \qquad t, s, k, l \geq 1 \tag{4}$$

PANJER and BELLHOUSE (1980) proved that when the X_t are normally distributed, the moments of and the correlation coefficients between interest, annuity and insurance functions depend upon $E[D_t^k]$ and $E[D_t^k D_s^l]$. For a whole life term insurance, for instance, the moments of the stochastic variable $\underset{\sim}{A}_x$ are given by

$$E[\underset{\sim}{A}_x^k] = \sum_{t=1}^{\infty} {}_{t-1|}q_x \, E[D_t^k] \tag{5}$$

The second moment for the life annuity $\underset{\sim}{a}_x$ is given by

$$E[\underset{\sim}{a}_x^2] = \sum_{t=1}^{\infty} {}_{t|}q_x \sum_{r=1}^{t} \sum_{s=1}^{t} E[D_r D_s] \tag{6}$$

Given a model for the yearly forces of interest δ_t, the problem is to find $\mu(t)$, $\sigma^2(t)$ and $a(t,s)$ for $t, s \geq 1$.

3. AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES

Assume that the stochastic model governing future forces of interest δ_t $(t = 1, 2, \ldots)$ belongs to the class of ARIMA (p, d, q)-processes. Then δ_t is generated by the stochastic difference equation

$$\begin{aligned} \nabla^d \delta_t = {} & \mu + b_1(\nabla^d \delta_{t-1} - \mu) + b_2(\nabla^d \delta_{t-2} - \mu) + \ldots + b_p(\nabla^d \delta_{t-p} - \mu) \\ & + \xi_t - c_1 \xi_{t-1} - c_2 \xi_{t-2} \ldots - c_q \xi_{t-q} \end{aligned} \tag{7}$$

where ∇^d stand for the d-th backward difference operator:

$$\nabla^1 \delta_t \equiv \nabla \delta_t = \delta_t - \delta_{t-1} \tag{8}$$

$$\nabla^d \delta_t = \nabla(\nabla^{d-1} \delta_t) \qquad d = 2, 3, \ldots \tag{9}$$

By convention we set $\nabla^0 \delta_t = \delta_t$. Further ξ_t is a normal white noise series with mean zero and variance σ^2. Equation (7) can also be written as

$$\nabla^d \delta_t = a + b_1 \nabla^d \delta_{t-1} + \ldots + b_p \nabla^d \delta_{t-p} + \xi_t - c_1 \xi_{t-1} - \ldots - c_q \xi_{t-q} \tag{10}$$

with a given by

$$a = \mu(1 - \sum_{i=1}^{p} b_i) \tag{11}$$

Equation (7) indicates that the process describing δ_t will not necessary be stationary. This means that the force of interest δ_t will not necessary have a constant unconditional mean, variance and autocovariance with any δ_{t-k} for $t \neq k$. The d-th difference of δ_t however follows a stationary autoregressive moving average process. This means that the series describing the interest rate exhibits homogeneity in the sense that, apart from local level, or perhaps local level and trend, one part of the series behaves much like any other part.

In what follows it will implicitly be assumed that the past $(p+d)$ forces of interest $\delta_0, \delta_{-1}, \ldots, \delta_{1-p-d}$ and the past q random disturbances $\xi_0, \ldots, \xi_{1-q}$ are known. Means, variances and covariances will always be considered as conditional on $\delta_0, \delta_{-1}, \ldots, \delta_{1-p-d}, \xi_0, \xi_{-1}, \ldots, \xi_{1-q}$. Remark that if δ_t follows an ARIMA (p, d, q)-process then the X_t given by (2) are normally distributed so that the theory of section 2 can be used.

The variable Y_t is defined as

(12) $$Y_t = \delta_{1-p-d} + \delta_{2-p-d} + \ldots + \delta_t \qquad t \geq 1-p-d$$

Further we set

(13) $$Y_{-p-d} = 0$$

It follows immediately that

(14) $$\delta_t = Y_t - Y_{t-1} \qquad t \geq 1-p-d$$

So if δ_t follows an ARIMA (p, d, q)-process given by (10) with $\delta_0, \ldots, \delta_{1-p-d}, \xi_0, \ldots, \xi_{1-q}$ known then Y_t follows an ARIMA $(p, d+1, q)$-process given by

(15) $$\nabla^{d+1} Y_t = a + b_1 \nabla^{d+1} Y_{t-1} + \ldots + b_p \nabla^{d+1} Y_{t-p} + \xi_t - c_1 \xi_{t-1} - \ldots - c_q \xi_{t-q}$$

with $Y_{-p-d}, Y_{1-p-d}, \ldots, Y_0$ and $\xi_0, \xi_{-1}, \ldots, \xi_{1-q}$ known.

Now it is easy to see that the ARIMA $(p, d+1, q)$-process describing Y_t can be written as an ARIMA $(l, 0, q)$-process with $l = p+d+1$:

(16) $$Y_t = a + \phi_1 Y_{t-1} + \ldots + \phi_l Y_{t-l} + \xi_t - c_1 \xi_{t-1} - \ldots - c_q \xi_{t-q}$$

with $\phi_1, \phi_2, \ldots, \phi_l$ suitable functions of $b_1, \ldots, b_p$.

Examples

(1) If δ_t follows an ARIMA $(p, 0, q)$-process then

(17) $$\delta_t = \mu + b_1(\delta_{t-1} - \mu) + \ldots + b_p(\delta_{t-p} - \mu) + \xi_t - c_1 \xi_{t-1} - \ldots - c_q \xi_{t-q}$$

Y_t can then be written as an ARIMA $(p+1, 0, q)$-process given by

(18) $$Y_t = a + \phi_1 Y_{t-1} + \ldots + \phi_{p+1} Y_{t-p-1} + \xi_t - c_1 \xi_{t-1} - \ldots - c_q \xi_{t-q}$$

with

(19) $$a = \mu(1 - \sum_{i=1}^{p} b_i)$$

and

(20) $$\phi_i = b_i - b_{i-1} \qquad i = 1, \ldots, p+1$$

with $b_0 = -1$ and $b_{p+1} = 0$

(2) If δ_t follows an ARIMA $(p, 1, q)$-process then

(21) $\nabla\delta_t = \mu + b_1(\nabla\delta_{t-1} - \mu) + \ldots + b_p(\nabla\delta_{t-p} - \mu) + \xi_t - c_1\xi_{t-1} - \ldots - c_q\xi_{t-q}$

Y_t can then be written as an ARIMA $(p+2, 0, q)$-process given by

(22) $Y_t = a + \phi_1 Y_{t-1} + \ldots + \phi_{p+2} Y_{t-p-2} + \xi_t - c_1\xi_{t-1} - \ldots - c_q\xi_{t-q} \quad t \geq 1$

with

(23) $$a = \mu(1 - \sum_{i=1}^{p} b_i)$$

and

(24) $$\phi_i = b_i - 2b_{i-1} + b_{i-2} \qquad i = 1, \ldots, p+2$$

with $b_{-1} = b_{p+1} = b_{p+2} = 0$ and $b_0 = -1$

In the next lemma we derive an expression for the Y_t in terms of known values plus a function of future error terms ξ_t.

Lemma 1

Assume that Y_t moves according to an ARIMA $(l, 0, q)$-process given by (16) and with $Y_0, Y_{-1}, \ldots, Y_{1-l}$ and $\xi_0, \xi_{-1}, \ldots, \xi_{1-q}$ known. The Y_t can be written as

(25) $$Y_t = \sum_{i=1}^{l} Y_{i-l} \sum_{j=\max(0,i-t)}^{i-1} \phi_{l-j} a_{j-i+t} - \sum_{i=1}^{q} \xi_{i-q} \sum_{j=\max(0,i-t)}^{i-1} c_{q-j} a_{j-i+t} + a \sum_{i=0}^{t-1} a_i + \sum_{i=0}^{t-1} \beta_i \xi_{t-i} \quad t \geq 1$$

where the coefficients a_i and β_i are given by

(26) $$a_0 = 1, \qquad \beta_0 = 1$$

(27) $$a_i = \sum_{j=1}^{\min(i,l)} \phi_j a_{i-j} \qquad i \geq 1$$

(28) $$\beta_i = a_i - \sum_{j=1}^{\min(i,q)} c_j a_{i-j} \qquad i \geq 1$$

Proof

For arbitrary constants a_i $(i = 0, 1, \ldots, t-1)$ we find for $t \geq 1$

$$\sum_{i=0}^{t-1} a_i Y_{t-i} = \sum_{j=1}^{l} \phi_j \sum_{i=j}^{t+j-1} a_{i-j} Y_{t-i} - \sum_{j=1}^{q} c_j \sum_{i=j}^{t+j-1} a_{i-j}\xi_{t-i} + \sum_{i=0}^{t-1} (a + \xi_{t-i}) a_i$$

By interchanging the order of summation in the second member of this equation and by using the a_i and β_i defined in (26), (27) and (28) we find

$$Y_t = \sum_{i=t}^{t+l-1} Y_{t-i} \sum_{j=i-t+1}^{\min(i,l)} \phi_j a_{i-j} - \sum_{i=t}^{t+q-1} \xi_{t-i} \sum_{j=i-t+1}^{\min(i,q)} c_j a_{i-j}$$

$$+ a \sum_{i=0}^{t-1} a_i + \sum_{i=0}^{t-1} \beta_i \xi_{t-i}$$

After some straightforward calculation (25) is obtained.

Remark that the first, the second and the thirth term in the right member of (25) are constants while the fourth term is stochastic.

In the following theorem expressions are derived for computing $\mu(t)$, $\sigma^2(t)$ and $a(t,s)$.

Theorem 1

If Y_t follows an ARIMA $(l, 0, q)$-process given by (16) then $\mu(t)$, $\sigma^2(t)$ and $a(t,s)$ can be computed by

$$(29) \qquad \mu(t) = a - Y_0\Big(1 - \sum_{i=1}^{l} \phi_i\Big) + \sum_{i=1}^{l} \phi_i \mu(t-i) - \sum_{i=1}^{q} c_i \eta(t-i) \qquad t \geq 1$$

where $\mu(0) = 0$ and $\mu(-i) = -(\delta_0 + \ldots + \delta_{1-i}) \quad i = 1, \ldots, l-1$

$$\text{and} \quad \eta(i) = \begin{cases} \xi_i & i \leq 0 \\ 0 & i > 0 \end{cases}$$

$$(30) \qquad \sigma^2(t) = \sigma^2 \sum_{i=0}^{t-1} \beta_i^2 = \sigma^2(t-1) + \beta_{t-1}^2 \qquad t \geq 1$$

with $\sigma^2(0) = 0$ and the β_i defined in (26), (27) and (28).

$$(31) \qquad a(t,s) = \sigma^2 \sum_{i=1}^{s} \beta_{t-i}\beta_{s-i} \qquad t > s \geq 1$$

Proof

From (2), (12) and (16) we obtain

$$X_t = -Y_0 + a + \phi_1 Y_{t-1} + \ldots + \phi_l Y_{t-l} + \xi_t - c_1 \xi_{t-1} - \ldots - c_q \xi_{t-q} \qquad t \geq 1$$

Taking the expected value of both members gives (29).

(30) and (31) follow immediately from (25).

The results obtained in lemma 1 and theorem 1 become much simpler if Y_t follows an ARIMA $(l, 0, 0)$-process. The expressions to compute $\mu(t)$, $\sigma^2(t)$ and $a(t,s)$ for this case are stated in the following theorem.

Theorem 2

If Y_t follows an ARIMA $(l, 0, 0)$-process given by (16) with $c_1 = c_2 = \ldots = c_q = 0$ then $\mu(t)$, $\sigma^2(t)$ and $a(t, s)$ can be computed by

$$(32) \qquad \mu(t) = a - Y_0(1 - \sum_{i=1}^{l} \phi_i) + \sum_{i=1}^{l} \phi_i \mu(t-i) \qquad t \geq 1$$

where $\mu(0) = 0$ and $\mu(-i) = -(\delta_0 + \ldots \delta_{1-i})$ $\qquad i = 1, \ldots, l-1$

$$(33) \qquad \sigma^2(t) = \sigma^2 \sum_{i=0}^{t-1} a_i^2$$

with $\sigma^2(0) = 0$ and the a_i defined in (26) and (27)

$$(34) \qquad a(t, s) = \sigma^2 \sum_{i=1}^{s} a_{t-i} a_{s-i} \qquad t > s \geq 1$$

The proof follows immediately from theorem 1 by deleting the terms in c_i $(i = 1, \ldots, q)$.

4. REMARKS

The method described by GIACCOTTO (1986) for ARIMA $(p, 0, q)$- and ARIMA $(p, 1, q)$-processes requires for the computation of $\sigma^2(t)$ values of $x_i(t)$ and $y_i(t)$ $(i = 1, \ldots, t)$, which can be computed recursively but that depend on t. In the method developed here for computing $\sigma^2(t)$, the algorithm is written so that the a_i- and β_i-values are independent of t.

We remark from theorem 1 and 2 that $\sigma^2(t)$ and $a(t, s)$ are independent of the past forces of interest $\delta_0, \delta_{-1}, \ldots, \delta_{1-l}$. So it follows that when the same interest rate model is used from year to year with only the past l forces of interest and the past q disturbances changing, the $\sigma^2(t)$ and $a(t, s)$ remain the same. Only the $\mu(t)$ will have to be recomputed every year.

5. EXAMPLE

To use our results the following procedure should be followed:

1) Choose an ARIMA (p, d, q) interest rate model and estimate the parameters involved. (see e.g. BOX and JENKINS (1970)).
2) Write Y_t as an ARIMA $(p+d+1, 0, q)$-process.
3) Compute the a_i's and the β_i's.
4) Compute $\mu(t)$, $\sigma^2(t)$, $a(t, s)$.
5) Compute the moments of actuarial functions.

To illustrate the procedure assume that we have the following model for the interest rate:

$$\delta_t = 0.08 + 0.6(\delta_{t-1} - 0.08) - 0.3(\delta_{t-2} - 0.08) + \xi t \qquad t \geq 1$$

where ξ_t is a white noise series with variance 0.0016 and $\delta_0 = 0.06$ and $\delta_{-1} = 0.07$.

Using (18), (19) and (20) Y_t can be written as

$$Y_t = 0.056 + 1.6\,Y_{t-1} - 0.9\,Y_{t-2} + 0.3\,Y_{t-3} + \xi_t \qquad t \geq 1$$

The a_t, $\mu(t)$, $\sigma^2(t)$ and $a(t, s)$ can then be computed by using theorem 2 and formula (26) and (27).

In table 1 a_t, $\mu(t)$, $\sigma^2(t)$, $E[D_t]$ and $\mathrm{Var}[D_t]$ are given for $t = 0, 1, \ldots, 5$. In the last column the discounted value of 1 $ payable in t years computed with a constant force of interest equal to the unconditional expected value of δ_t is given. In the example described here the stochastic approach leads to higher single premiums. This fact could be expected by observing δ_0 and δ_{-1}.

TABLE 1

MEAN AND VARIANCE OF A PAYMENT OF 1 $ DUE IN t YEARS

t	a_t	$\mu(t)$	$\sigma^2(t)$	$E[D_t]$	$\mathrm{Var}[D_t]$	$\exp(-0.08\,t)$
0	1	0	0	1	0	1
1	1.6000	0.0710	0.0016	0.9322	0.0014	0.9231
2	1.6600	0.1516	0.0057	0.8618	0.0042	0.8521
3	1.5160	0.2347	0.0101	0.7948	0.0064	0.7866
4	1.4116	0.3163	0.0138	0.7339	0.0075	0.7261
5		0.3964	0.0170	0.6784	0.0080	0.6703

ACKNOWLEDGEMENT

The author wishes to thank the anonymous referees for their helpful comments.

REFERENCES

BELLHOUSE, D. R. and PANJER, H. H. (1981) Stochastic modelling of interest rates with applications to life contingencies — part II. *Journal of Risk and Insurance*, **vol. XLVII no. 4**, 628–637.

BOWERS, N. L., GERBER, H. U., HICKMAN, J. C., JONES, D. A. and NESBIT, C. J. (1987) *Actuarial Mathematics*, Society of Actuaries.

BOX, G. E. P. and JENKINS, G. M. (1970) *Time Series Analysis*. Holden-Day, San Francisco.

BOYLE, P. P. (1976) Rates of return as random variables. *Journal of Risk and Insurance*, **vol. XLIII no. 4**, 693–713.

GIACCOTTO, C. (1986) Stochastic modelling of interest rates: actuarial vs. equilibrium approach. *Journal of Risk and Insurance*, **vol. LIII no. 3**, 435–453.

PANJER, H. H. and BELLHOUSE, D. R. (1980) Stochastic modelling of interest rates with applications to life contingencies. *Journal of Risk and Insurance*, **vol. XLVII no. 1**, 91–110.

POLLARD, J. H. (1971) On fluctuating interest rates. *Bulletin van de Koninklijke Vereniging van Belgische Actuarissen* **nr. 66**, 68–97.

JAN DHAENE

Instituut voor Actuariële Wetenschappen, K.U.Leuven, Dekenstraat 2, 3000 Leuven, Belgium.

THE CLAIMS RESERVING PROBLEM IN NON-LIFE INSURANCE: SOME STRUCTURAL IDEAS

By Elja Arjas

University of Oulu, Finland

Abstract

We present some relatively simple structural ideas about how probabilistic modeling, and in particular, the modern theory of point processes and martingales, can be used in the estimation of claims reserves.

1. Introduction

The claims reserving problem, or the run off problem, has been studied rather extensively. The monograph by Taylor (1986) covers most of the developments so far, and, interestingly enough, creates a taxonomy to the models introduced. The booklet of van Eeghen (1981) has a somewhat similar aim. Because of these recent surveys we do not intend to describe "the state of the art" in this area but confine ourselves to a few remarks.

There has been a clear tendency away from deterministic "accounting methods" into more descriptive probabilistic models. Early works in this direction were Bühlmann et al. (1980), Hachemeister (1980), Linnemann (1980) and Reid (1981). Of more recent contributions we would like to mention particularly Pentikäinen and Rantala (1986), and three papers dealing with unreported (IBNR) claims: Norberg (1986), Robbin (1986) and Jewell (1987).

Most authors today tend to agree that there are important benefits from using structurally descriptive probabilistic models in insurance. However, there appears to be a new problem: With the increased realism of such models, many papers introduce, very early on, a long list of special assumptions and a correspondingly complicated notation. A reader may then not be able to see what ideas are really important and characteristic to the entire claims reserving problem, and what are less so, only serving to make the calculations more explicit. It would be more pleasant if the modeling could be started virtually without any assumptions, and then only adding assumptions as it becomes clear that advancing otherwise is difficult. We think that the modern theory of stochastic processes comes here to aid, and try to illustrate this in the following. We are mainly using "the martingale approach to point processes", as discussed e.g. in Brémaud (1981) and Karr (1986). However, apart from some calculations towards the end, no previous knowledge of this theory is really needed to understand the paper.

The emphasis of this paper is in the conceptual analysis of Section 2 and the structural results of Section 3. Section 4 provides an illustration of how the actual stochastic calculus, in a simple form, can be applied to obtain more explicit results.

We want to stress that this paper contains very little that could be called "new results": it is more important to us here how we arrive at them.

2. CLAIMS, INFORMATION AND SETTLEMENT AS MARKED POINT PROCESSES

Considering a fixed accident year, say the unit interval (0, 1], let the exact occurrence times of the accidents be $T_1^* \leq T_2^* \leq \ldots$ An accident which occurs at time T_i^* is reported to the company after a random delay D_i so that its reporting time is $R_i^* = T_i^* + D_i$. We denote the ordered reporting times (order statistics) by $T_1 < T_2 < \ldots$, assuming for simplicity that they are all different.

We follow the convention that the accidents are indexed according to the order in which they are reported to the company, i.e., the accident reported at T_i is called "the i^{th} accident". Because of the random delay in reporting this indexing is often different from the one that refers to the occurrence times.

In practice the number of accidents in a given year of occurrence is of course finite. We denote this (random) number by N. As a convention, we let the sequence (T_i) be infinite but define $T_{N+1} = T_{N+2} = \ldots = \infty$.

Let us then assume that every time a new accident is reported to the company, this will be followed by a sequence of "handling times". These handlings could be times at which claim payments are paid, but also times at which the file concerning the accident is updated because of some arriving new information. Supposing that the i^{th} accident has altogether N_i handling times following its reporting, we denote them by

$$T_i = T_{i0} < T_{i1} < T_{i2} < \ldots < T_{i,N_i}. \tag{2.1}$$

Again, we let $T_{i,N_i+1} = T_{i,N_i+2} = \ldots = \infty$.

Next we need to specify the event that takes place at T_{ij}. If a payment is made then, we denote the amount paid by X_{ij}. If nothing is paid at T_{ij} we simply let $X_{ij} = 0$. Similarly, it is convenient to have a notation for the information which is used for updating the accident file. Let I_{i0} be the information which becomes available when the accident is reported, and let I_{ij} be the new information which arrives at handling time T_{ij}. If there is no such information, we set $I_{ij} = \emptyset$, signalling "no new information". In particular we set $X_{ij} = 0$ and $I_{ij} = \emptyset$ whenever $T_{ij} = \infty$.

Our analysis will not depend on what explicit form the variables I_{ij} are thought to have. They could well be strings of letters and numbers, reflecting, for example, how the accident is classified by the company at time T_{ij}. I_{i0} will often determine what was the delay in reporting the i^{th} accident. If further payments are made after the case was thought in the company to be closed, it is probably convenient to consider the arrival of the first such claim as a new reporting time, also initiating a new sequence of handlings.

The above definitions give rise to a number of stochastic processes which are of interest in the claims reserving problem. The first definitions will be accident specific, after which we obtain the corresponding collective processes by simple summation.

We start by assigning the payments X_{ij} to the handling times T_{ij}. In this way we arrive, for each i, at a sequence $(T_{ij}, X_{ij})_{j \geq 0}$, where $T_i = T_{i0} \leq T_{i1} \leq \ldots$ (with strict inequalities if the variables are finite) and $X_{ij} \geq 0$. Thus $(T_{ij}, X_{ij})_{j \geq 0}$ can be viewed as a *marked point process* (MPP) on the real line, with non-negative real "marks" X_{ij}. We call it *the payment process*. Equivalently, of course, we can consider the cumulative payment process $(X_i(t))$ defined by

$$X_i(t) = \sum_{\{j:\, T_{ij} \leq t\}} X_{ij}. \tag{2.2}$$

Clearly, $X_i(t)$ represents the total amount of payments (arising from accident i) made before time t. The function $t \mapsto X_i(t)$ is an increasing step function, with $X_i(t) = 0$ for $t < T_i$ (= reporting time) and $X_i(t)$ approaching, as $t \to \infty$, the limit

$$X_i(\infty) = \sum_{j \geq 0} X_{ij}, \tag{2.3}$$

which is the total compensation paid for the i^{th} accident. Similarly,

$$\begin{aligned} U_i(t) &= X_i(\infty) - X_i(t) \\ &= \sum_{\{j:\, T_{ij} > t\}} X_{ij} \end{aligned} \tag{2.4}$$

represents the total liability at t coming from future payments, with $U_i(t) = X_i(\infty)$ for $t < T_i$ and $U_i(t)$ decreasing stepwise to 0 as $t \to \infty$.

We remark here that, in order to keep this simple structure, we do not consider explicitly the effects of interest rate or inflation. This means, among other things, that the future claims must be expressed in standardized (deflated) currency.

Second, we can consider the sequence $(T_{ij}, I_{ij})_{j \geq 0}$ and call it *the information process* for the i^{th} accident. This, too, is an MPP, with mark I_{ij} taking values in some conveniently defined set. As mentioned earlier the form of the marks is not restricted in any real way: It will suffice, for example, that there is a countable number of possible marks.

Our third MPP is obtained by combining the marks of the other two, into pairs (X_{ij}, I_{ij}). We call $(T_{ij}, (X_{ij}, I_{ij}))_{j \geq 0}$ *the settlement process* of the i^{th} accident.

Considering finally all accidents collectively, we obtain the corresponding collective payment process, information process and settlement process by a simple summation (superposition) over the index i. However, we do not need a separate notation for these MPP's and will therefore confine ourselves to *the cumulative payment process*

$$X.(t) = \sum_i X_i(t) \tag{2.5}$$

and *the liability process*

$$U.(t) \sum_i U_i(t). \tag{2.6}$$

Observe that it is not necessary to restrict the summation to indices i satisfying $i \leq N$ because, unless this is satisfied, $X_i(t) = U_i(t) = 0$ for all t.

3. CLAIMS RESERVES AS A PREDICTION PROBLEM

The estimation of the claims reserves can now be viewed as a prediction problem where, at a given time t representing "the present", an assessment of the future payments is made on the basis of the available information. Most of our mathematical considerations do not depend on whether the assessment concerns the payments from an individual i^{th} accident, or all accidents during the considered year of occurrence. Because of this we will often simply drop the subscript ("i" or "$\cdot$") from the notation. Thus, for example, $U(t)$ can be taken to be either the accident specific liability $U_i(t)$ or their sum $U.(t)$.

The role of the information process above is to provide a formal basis for the assessments made. This is done most conveniently in terms of histories, i.e., families of σ-fields in the considered probability space, which correspond to the knowledge of the values of the random variables generating them. In particular, we let the σ-field

$$\mathcal{F}_t^N = \sigma\{(T_{ij}, I_{ij}, X_{ij})_{i \geq 1, j \geq 0} : T_{ij} \leq t\} \tag{3.1}$$

respresent the information carried by the pre-t settlement process arising from all claims. (For background, see e.g. KARR (1986), Section 2.1). For completeness, we also allow for the possibility of having information which is exogenous to the settlements. Writing $\mathcal{E}_t$ for such pre-t information, we shall base the estimation of the future payments on the history $(\mathcal{F}_t)$, with

$$\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{E}_t. \tag{3.2}$$

In an obvious sense, the most complete assessment at time t concerning $X(\infty)$, the total of paid claims, is provided by the conditional distribution

$$\mu_t(\cdot) = P(X(\infty) \in \cdot \mid \mathcal{F}_t).$$

When t varies, these conditional distributions form a so called *prediction process* (μ_t) (see e.g. NORROS (1985)). Here, however, we restrict our attention to the first two moments of μ_t. Assuming square integrability throughout this paper, we write

$$M_t = E^{\mathcal{F}_t}(X(\infty)) \qquad \left(= \int x\mu_t(dx) \right) \tag{3.3}$$

and

$$V_t = \mathrm{Var}^{\mathcal{F}_t}(X(\infty)) \qquad \left(= \int x^2\mu_t(dx) - M_t^2 \right). \tag{3.4}$$

We now derive some fundamental properties of (M_t) and (V_t). From now on we also write X_t and U_t instead of $X(t)$ and $U(t)$.

Having introduced the idea that $\mathcal{F}_t$ represents "information which the company has at time t", it is of course the case that the payments already made are, at least in principle, included in such knowledge. Formally this corresponds to the decomposition of X_∞ into X_t and U_t (see (2.4)), i.e.,

$$X_\infty = X_t + U_t, \tag{3.5}$$

where X_t is determined from $\mathcal{F}_t$ (i.e., $\mathcal{F}_t$-measurable). Therefore, the $(\mathcal{F}_t)$-based prediction of X_∞ is equivalent to predicting U_t.

CONDITIONAL EXPECTATIONS. Let us first consider the expected values M_t. As a stochastic process, (M_t) is easily seen to have the *martingale-property*: For any $t < u$,

$$E^{\mathcal{F}_t}(M_u) = M_t. \tag{3.6}$$

Thus, since M_t is an estimate of X_∞ at time t and M_u is a corresponding updated estimate at a later time u, (3.6) expresses the simple consistency principle:

(P1) "Current estimate of a later estimate, which is based on more information, is the same as the current estimate".

Another way to express the martingale property is to say that the estimates (M_t) have no trend with respect to t.

Since X_t is determined from $\mathcal{F}_t$ we clearly have

$$M_t = X_t + E^{\mathcal{F}_t}(U_t) \underset{\text{def}}{=} X_t + m_t.$$

Here, the estimated liability at t,

$$m_t = E^{\mathcal{F}_t}(U_t), \tag{3.7}$$

is a *supermartingale*, with the "decreasing trend property"

$$E^{\mathcal{F}_t}(m_u) \le m_t \qquad \text{for} \qquad t < u .$$

This follows readily from the fact that the true liability U_t is decreasing in time, as more and more of the claims are paid. Unfortunately such a monotonicity property is of little direct practical use because the process (U_t) is unobservable: Only the differences $U_u - U_t = X_t - X_u$ can be observed, but not the actual values of U_u or U_t.

The trend properties of (M_t) and (m_t) lead to a crude idea about how the reserve estimates should behave as functions of time. Considering them as a time series may therefore be useful. On the other hand, one has to remember that the (super)martingale property is quite weak and only concerns the $(\mathcal{F}_t)$-conditional expected values. Thus an apparently downward trend in an observed time series could be balanced by a rare but big jump upwards.

For a more refined analysis, it would be interesting to study (M_t) in terms of its martingale integral representation (see e.g. BRÉMAUD (1981)). The key ingredient in that representation is the innovation gains process which determines how (M_t) is updated in time when $(\mathcal{F}_t)$ is observed. This theory is well understood. Unfortunately, however, actuaries seem to have very little idea about what properties the updating mechanism should realistically possess, and presently there is no detailed enough data to study the question statistically. Therefore, a more systematic research effort must wait.

It is instructive to still consider the differences

$$M(t, u) = M_u - M_t, \quad t < u. \tag{3.8}$$

By the martingale property (3.6) we clearly have $E^{\mathcal{F}_t}(M(t, u)) = 0$. Now, using the analogous notation $X(t, u) = X_u - X_t$ for the cumulative payments we easily find that

$$M(t, u) = [X(t, u) - E^{\mathcal{F}_t}(X(t, u))] + [E^{\mathcal{F}_u}(U_u) - E^{\mathcal{F}_t}(U_u)].$$

The first term on the right is the error in the estimate concerning payments in the time interval $(t, u]$. The second term, then, is the updating correction which is made to the estimated liability when the time of estimation changes from t to u. Both terms have $\mathcal{F}_t$-conditional expected value 0. This suggests that it might be beneficial in practice to split the estmate into two parts: one that covers the time interval to the next update (typically a year) and another for times thereafter.

CONDITIONAL VARIANCES. The variances V_t give rise to somewhat similar considerations. First observe that, since X_t is determined by $\mathcal{F}_t$, the variance V_t defined in (3.4) satisfies

$$V_t = \mathrm{Var}^{\mathcal{F}_t}(U_t) = \mathrm{Var}^{\mathcal{F}_t}(M(t, \infty)). \tag{3.9}$$

Thus, if the used estimation method produces also estimates of V_t, the observed oscillations in (U_t) can be compared with the square root of V_t. (Warning: Do not expect normality in short time series!) Second, it is interesting to note that (V_t) is a *supermartingale* as well, i.e.,

(3.10) $$E^{\mathcal{F}_t}(V_u) \le V_t \qquad \text{for} \qquad t < u.$$

This expresses the following intuitively plausible principle:

(P2) "Measured by the conditional variance, the estimates M_t tend to become more accurate as time increases and more information becomes available".

To show that (3.10) holds, we first find that

$$E^{\mathcal{F}_t}(M(t,u)\,X(u,\infty)) = E^{\mathcal{F}_t}(M(t,u)\,E^{\mathcal{F}_u}M(u,\infty)) = 0$$

so that $M(t,u)$ and $M(u,\infty)$ are uncorrelated. This implies the well known additivity property ("Hattendorf's formula", e.g. GERBER (1979))

(3.11) $$\mathrm{Var}^{\mathcal{F}_t}(M(t,\infty)) = \mathrm{Var}^{\mathcal{F}_t}(M(t,u)) + \mathrm{Var}^{\mathcal{F}_t}(M(u,\infty)).$$

On the other hand,

(3.12) $$\mathrm{Var}^{\mathcal{F}_t}(M(u,\infty)) = E^{\mathcal{F}_t}(\mathrm{Var}^{\mathcal{F}_u}(X_\infty)) = E^{\mathcal{F}_t}(V_u),$$

so that (3.10) follows by combining (3.9), (3.11) and (3.12).

REMARK. Recall the following well-known result which complements this picture: with respect to a quadratic loss function, the conditional expectation M_t is the optimal estimate of $X(\infty)$. More precisely, for any estimate $\tilde{M}_t$ of X_∞ which can be determined from $\mathcal{F}_t$ (i.e., $\tilde{M}_t$ is $\mathcal{F}_t$-measurable), the following inequality is satisfied:

(3.13) $$E^{\mathcal{F}_t}((X_\infty - \tilde{M}_t)^2) \ge E^{\mathcal{F}_t}((X_\infty - M_t)^2) \;\; (= V_t).$$

KNOWN AND UNKNOWN ACCIDENTS. Finally in this section we divide the collective estimate $m_{.t} = E^{\mathcal{F}_t}(U_{.t})$ into two parts depending on whether the considered accidents are at time t *known* (= reported, IBNER) or *unknown* (= not reported, IBNR).

Let the number of *known (= reported) accidents* at time t be

(3.14) $$N_t = \sum_i 1_{\{T_i \le t\}}.$$

The corresponding liability from future payments is then $\sum_{i \le N_t} U_{it}$. Since the events $\{T_i \le t\}$ are determined by $\mathcal{F}_t$, the corresponding $\mathcal{F}_t$-conditional estimate is simply given by

(3.15) $$E^{\mathcal{F}_t}\left(\sum_{i \le N_t} U_{it}\right) = \sum_{i \le N_t} E^{\mathcal{F}_t}(U_{it}) = \sum_{i \le N_t} m_{it}.$$

This formula expresses the intuitively obvious fact that the reserves corresponding to reported accidents could, at least in principle, be assessed individually.

If we are willing to make the assumption, which may not be completely realistic, that the liabilities U_{it} are uncorrelated across accidents given $\mathcal{F}_t$, we also have a corresponding equality for variances:

$$\text{(3.16)} \qquad \mathrm{Var}^{\mathcal{F}_t}\left(\sum_{i \le N_t} U_{it}\right) = \sum_{i \le N_t} \mathrm{Var}^{\mathcal{F}_t} U_{it} = \sum_{i \le N_t} V_{it}.$$

Note that although the processes (m_{it}) and (V_{it}) were above found to be supermartingales, the processes defined by (3.15) and (3.16) do not have this property. This is because N_t is increasing.

Considering then the *unknown (IBNR) accidents*, it is obvious that also their number $N - N_t$ is unknown (i.e., not determined by $\mathcal{F}_t$) and therefore the liability estimate $E^{\mathcal{F}_t}\left(\sum_{i > N_t} U_{it}\right)$ cannot be determined "termwise" as was done in (3.15). Therefore the estimate needs to be determined collectively for all IBNR-accidents, a task which we consider in the next section. The only qualitative property which we note here is that the process $\left(E^{\mathcal{F}_t}\left(\sum_{i > N_t} U_{it}\right)\right)$ is again a *supermartingale*. This is an easy consequence of the supermartingale property of (m_{it}), which was established above, and the fact that N_t is increasing.

4. AN ILLUSTRATION: THE ESTIMATION OF IBNR CLAIMS RESERVES

We now illustrate, considering the IBNR claims reserves, how the mathematical apparatus of the stochastic calculus can be used to derive explicit estimates. But we are also forced to introduce some more assumptions in order to reach this goal.

For known accidents, the delays in the reporting times T_i are only important in so far as they are thought to influence the distribution of the corresponding payment process. For unknown accidents the situation is completely different: For unknown accidents the only thing which is known is that if an i^{th} accident occurred during the considered year and it is still unknown at time t, its reporting time T_i exceeds t. (Recall the convention that $T_i = \infty$ for $i > N$). Therefore, it is impossible to estimate the IBNR reserves individually. A natural idea in this situation is to use the information which has been collected about other (i.e., known) accidents and hope that they would have enough in common with those still unknown. The problem resembles closely those in software reliability, where the aim is to estimate the unknown number of "bugs" remaining in the program. More generally, it is a *state estimation* or *filtering* problem.

It is most convenient to formulate the "common elements" in terms of unobservable (latent) variables whose distribution is updated according to the information $\mathcal{F}_t$. $\mathcal{F}_t$ has thereby an indirect effect on the behaviour of IBNR claims. In the following we study the expected value and the variance of the IBNR liability. The presentation has much in common with JEWELL (1980, 1987), and ROBBIN (1986), and in particular NORBERG (1988).

Since the marked points belonging to the settlement process of an unknown accident are all "in the future", most considerations concerning the reserves will not change if the payments are assigned directly to the reporting time T_i. This is possible because we, as stated before, don't consider the effects of interest rate or inflation. This will simplify the notation to some extent. We therefore consider the MPP (T_i, X_i), where $X_i = X_i(\infty)$ is the size of the claim caused by the i^{th} accident. The corresponding counting process is $\{N_t(A);\ t \geq 0, A \subset R^1\}$, where

$$N_t(A) = \sum_i 1_{\{T_i \leq t,\, X_i \in A\}} \tag{4.1}$$

counts the number of accidents reported before t and such that their liability X_i is in the set A. (Note that $N_t(A)$ cannot in general be determined from $\mathcal{F}_t$ since the X_i's counted before t may also include payments made after time t. Also observe the connection to (3.14): $N_t = N_t(R^1)$).

For the purpose of using the apparatus of the stochastic calculus we start by writing the total liability from IBNR claims as an integral (pathwise):

$$\sum_{i>N_t} U_{it} = \sum_{\{i:T_i>t\}} X_i = \int_{s=t}^{\infty} \int_{x=0}^{\infty} x\, dN_s(dx)\,. \tag{4.2}$$

We also let

$$\tilde{U}(t, u; A) = \sum_{\{i:t<T_i\leq u,\, X_i\in A\}} X_i = \int_{s=t}^{u} \int_{x\in A} x\, dN_s(dx)\,,$$

so that $\sum_{i>N_t} U_{it} = \tilde{U}(t, \infty; R^1)$.

Adapting the idea from NORBERG (1986) we now suppose that the above mentioned latent variables form a pair (Φ, Θ) and are such that Φ can be viewed as a parameter of the distribution of the process (N_t), formed by the reporting times, whereas Θ parametrizes the distribution of the claim sizes (X_i). (Note that this simple model is "static" in the sense that the latent variables do not depend on time. This assumption could be relaxed, for example, by introducing an autoregressive scheme of state equations, as in the Kalman filter). There are no restrictions on the dimension of (Φ, Θ). On the other hand, these parameters are assumed to be *sufficient* in the sense that if Φ and Θ, together with some initial information $\mathcal{F}_0$, were known, no information from $\mathcal{F}_t$ would change the prediction concerning the IBNR claims after t. Thus the estimates of Φ and Θ which are obtained from $\mathcal{F}_t$, or more exactly, their conditional distribution given $\mathcal{F}_t$, can be said to include "that part of $\mathcal{F}_t$-information which is relevant in the IBNR-problem".

The formal expression of this idea is as follows. Fixing t ("the present") we consider times $u \geq t$ and define

$$\tilde{\mathcal{F}}_u = \mathcal{F}_0 \vee \sigma\{\Phi, \Theta\} \vee \sigma\{(T_i, X_i);\ t < T_i \leq u\}. \tag{4.3}$$

Thus $\tilde{\mathcal{F}}_\infty$ represents the information contained collectively in $\mathcal{F}_0$, the parameters Φ and Θ, and all post-t payments, cf. KARR (1986), Section 2.1. We then assume the conditional independence property

$$\tilde{\mathcal{F}}_\infty \coprod_{\mathcal{F}_0 \vee \sigma\{\Phi, \Theta\}} \mathcal{F}_t, \tag{4.4}$$

stating that $\mathcal{F}_t$ is irrelevent for predicting the post-t payments provided that $\mathcal{F}_0$ and (Φ, Θ) are known.

Let the $(\tilde{\mathcal{F}}_u)$-intensity of counting process $(N_u(A))_{u \geq t}$, be $(\tilde{\lambda}_u(A))_{u \geq t}$, with $A \subset R^1$. The probabilistic interpretation of $\tilde{\lambda}_u(A)$ is that

$$\tilde{\lambda}_u(A)\, du = P(dN_u(A) = 1 \mid \tilde{\mathcal{F}}_{u-}) = P(T_i \in du,\ X_i \in A \mid \tilde{\mathcal{F}}_{u-}) \tag{4.5}$$

on the interval $T_{i-1} < u \leq T_i$. On the other hand, $\tilde{\lambda}_u(A)$ can obviously be expressed as the product

$$\tilde{\lambda}_u(A) = \bar{\lambda}_u\, \varphi_u(A), \tag{4.6}$$

where $\bar{\lambda}_u = \tilde{\lambda}_u(R^1)$ and $\varphi_u(A)/\bar{\lambda}_u$ (cf. KARR (1986), Example 2.24). Here $(\bar{\lambda}_u)$ is the $(\tilde{\mathcal{F}}_u)$-intensity of the counting process (N_u), i.e., $\bar{\lambda}_u du = P(dN_u = 1 \mid \tilde{\mathcal{F}}_{u-}) = P(T_i \in du \mid \tilde{\mathcal{F}}_{u-})$ for $T_{i-1} < u \leq T_i$, whereas $\varphi_u(A)$ can be interpreted as the conditional probability of $\{X_i \in A\}$ given $\tilde{\mathcal{F}}_{u-}$ and that $T_i = u$.

It follows from (4.4) that the intensity $(\tilde{\lambda}_u(\cdot))_{u>t}$ can be chosen to be $\mathcal{F}_0$-measurable and parametrized by (Φ, Θ). According to "the division of roles of Φ and Θ" we now assume that in fact $\bar{\lambda}_u$ in (4.6) is parametrized by Φ, and $\varphi_u(\cdot)$ by Θ. $\bar{\lambda}_u$ can then be expressed in the form $\bar{\lambda}_u = h(u; \Phi)$,where, for fixed Φ, $u \mapsto h(u; \Phi)$ is $\mathcal{F}_0$-measurable. This is only another way of saying that the reporting process (N_u) is assumed to be *a doubly stochastic (non-homogeneous) Poisson process* (or Cox process) with random parameter Φ.

Similarly, we assume that the claim size distributions $\varphi_u(\cdot)$ can be written as $\varphi_u(A) = F_u(A; \Theta)$, where, for fixed u and Θ, $F_u(\cdot\,; \Theta)$ is a distribution function on R^1_+. This, then, amounts to saying that, given Θ and the (unobserved) IBNR reporting times, the claim sizes X_i are independent.

We now derive an expression for *the expected IBNR-liability* $E^{\mathscr{F}_t}\left(\sum_{i>N_t} U_{it}\right)$.

First note that $\tilde{\mathscr{F}}_t = \mathscr{F}_0 \vee \sigma(\Phi, \Theta)$. By a straightforward calculation we get that

$$E^{\tilde{\mathscr{F}}_t}\left(\sum_{i>N_t} U_{it}\right) = E^{\tilde{\mathscr{F}}_t}\left(\int_{u=t}^{\infty}\int_{x=0}^{\infty} x\, dN_u(dx)\right)$$

$$\underset{(*)}{=} E^{\tilde{\mathscr{F}}_t}\left(\int_{u=t}^{\infty}\int_{x=0}^{\infty} x\, \tilde{\lambda}_u(dx)\, du\right)$$

$$= \int_{u=t}^{\infty} h(u; \Phi) \int_{x=0}^{\infty} x\, F_u(dx; \Theta)\, du$$

$$= \int_{u=t}^{\infty} h(u; \Phi)\, m_u(\Theta)\, du\,,$$

where $m_u(\Theta)$ is the mean

$$m_u(\Theta) = \int_{x=0}^{\infty} x\, F_u(dx; \Theta)\,. \tag{4.7}$$

(The equality (*) here is a simple consequence of the definition of $(\tilde{\lambda}_u)$; for a general result see e.g. KARR (1986, Theorem 2.22). On the other hand, because of the conditional independence (4.4), we have that

$$E^{\tilde{\mathscr{F}}_t}\left(\sum_{i>N_t} U_{it}\right) = E^{\mathscr{F}_t \vee \tilde{\mathscr{F}}_t}\left(\sum_{i>N_t} U_{it}\right),$$

and therefore finaly

$$E^{\mathscr{F}_t}\left(\sum_{i>N_t} U_{it}\right) = \int_{u=t}^{\infty} E^{\mathscr{F}_t}(h(u; \Phi)\, m_u(\Theta))\, du\,. \tag{4.8}$$

We consider some special cases at the end of this section.

Let us then go over to calculating the corresponding *conditional variance* expression $\mathrm{Var}^{\mathscr{F}_t}\left(\sum_{i>N_t} U_{it}\right)$. The calculation goes as follows.

$$\mathrm{Var}^{\tilde{\mathcal{F}}_t}\left(\sum_{i>N_t} U_{it}\right)$$

$$= \mathrm{Var}^{\tilde{\mathcal{F}}_t}\left(\int_{u=t}^{\infty}\int_{x=0}^{\infty} x\, dN_u(dx)\right) \underset{(*)}{=} E^{\tilde{\mathcal{F}}_t}(\langle \tilde{U}(t,\cdot\,;R^1)\rangle_\infty)$$

$$= E^{\tilde{\mathcal{F}}_t}\left(\int_{x=0}^{\infty} \langle \tilde{U}(t,\cdot\,;dx)\rangle_\infty\right) \underset{(**)}{=} E^{\tilde{\mathcal{F}}_t}\left(\int_{x=0}^{\infty}\int_{u=t}^{\infty} x^2\, d\langle N_{(\cdot)}(dx)\rangle_u\right)$$

$$= E^{\tilde{\mathcal{F}}_t}\left(\int_{x=0}^{\infty}\int_{u=t}^{\infty} x^2\, \tilde{\lambda}_u(dx)\, du\right) = \int_{u=t}^{\infty} h(u,\Phi)\int_{x=0}^{\infty} x^2 F_u(dx;\Theta)\, du$$

$$= \int_{u=t}^{\infty} h(u;\Phi)\, m_u^{(2)}(\Theta)\, du\,,$$

where $m_u^{(2)}(\Theta)$ is the second moment

$$m_u^{(2)}(\Theta) = \int_{x=0}^{\infty} x^2 F_u(dx;\Theta)\,. \tag{4.9}$$

(Here $(\langle\cdot\rangle_u)$ is the predictable variation process, see e.g. KARR (1986), Appendix B, (*) is a direct consequence of the definition of this process, and (**) follows from Theorem B.12 in KARR (1986)). Therefore, and again using the conditional independence (4.4),

$$\mathrm{Var}^{\mathcal{F}_t}\left(\sum_{i>N_t} U_{it}\right) = E^{\mathcal{F}_t}\,\mathrm{Var}^{\tilde{\mathcal{F}}_t}\left(\sum_{i>N_t} U_{it}\right) + \mathrm{Var}^{\mathcal{F}_t}\, E^{\tilde{\mathcal{F}}_t}\left(\sum_{i>N_t} U_{it}\right) \tag{4.10}$$

$$= \int_{u=t}^{\infty} E^{\mathcal{F}_t}(h(u;\Phi)\, m_u^{(2)}(\Theta))\, du + \mathrm{Var}^{\mathcal{F}_t}\left(\int_{u=t}^{\infty} h(u;\Phi)\, m_u(\Theta)\, du\right).$$

The formulas (4.8) and (4.10) can be briefly summarized by saying that the conditional expectation and the conditional variance of the IBNR liability $\sum_{i>N_t} U_{it}$ can be obtained if the following are known:

(i) the intensities $h(\cdot\,;\Phi)$;
(ii) the first two moments of the distributions $F(\cdot\,;\Theta)$, and
(iii) the conditional distribution of the latent variables (Φ,Θ) given $\mathcal{F}_t$.

Concerning (i), the common expression for $h(\cdot\,;\Phi)$ (e.g. RANTALA (1984)) is obtained by assuming that during the considered year (= unit interval (0, 1]) accidents occur according to the Poisson(Φ)-process, and that the reporting delays D_i are i.i.d. and distributed according to some known distribution $G(\cdot)$. Then it is easily seen that

$$h(u;\Phi) = \Phi[G(u) - G((u-1)^+)]. \tag{4.11}$$

More generally, Φ can parametrize both the occurrence process and the distribution of the delays in the reporting, cf. JEWELL (1987).

The simplest case in (ii) is of course when only the number $N - N_t$ of future claims is considered, instead of the liability they cause. Then we can make the obvious convention that every $X_i = 1$, giving $m_u(\Theta) = m_u^{(2)}(\Theta) = 1$.

Requirement (iii), finally, strongly supports the use of the Bayesian paradigm. It is particularly appealing to use the Poisson-gamma conjugate distributions for the pair (N_t, Φ) since this makes the updating extremely simple (see GERBER (1979) and NORBERG (1986)). Since deciding on claims reserves is a management decision, rather than a problem in science in which some physical constant needs to be determined, Bayesian arguments should not be a great deterrent to a practitioner. Choosing a reasonable prior for (Φ, Θ) could be viewed as a good opportunity for an actuary to use, in a quantitative fashion, his experience and best hunches.

ACKNOWLEDGEMENT

Most of what I today know about insurance mathematics I have learned, during many long conversations, from JARMO JACOBSSON, of the Statistical Center of the Finnish Insurance Companies. I also appreciate the referees' careful reading of the paper.

This paper was read at the ASTIN Annual Meeting which was held in Helsinki in July 1988, in connection with the 23rd International Congress of Actuaries.

REFERENCES

BRÉMAUD, P. (1981) *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, Berlin.

BÜHLMANN, H., SCHNIEPER, R. and STRAUB, E. (1980) Claims reserves in casualty insurance based on a probabilistic model. *Bulletin of the Association of Swiss Actuaries*, 21–45.

GERBER, H. U. (1979) *An Introduction to Mathematical Risk Theory*. S.S. Huebner Foundation of Insurance Education, University of Pennsylvania, Philadelphia.

HACHEMEISTER, C. A. (1980) A stochastic model for loss reserving. *Transactions of the 21st International Congress of Actuaries* **1**, 185–194.

JEWELL, W. S. (1980) Generalized models of insurance business. Report of Introduction. *Transactions of the 21st International Congress of Actuaries* S, 87–141.

JEWELL, W. S. (1987) Predicting IBNYR events and delays I. Continuous time. Department of Industrial Engineering and Operations Research, University of California at Berkeley, Research Report.

KARR, A. F. (1986) *Point Processes and Their Statistical Inference*. Dekker, New York.

LINNEMANN, P. (1980) A multiplicative model for loss reserves: A stochastic process approach. *University of Copenhagen, Laboratory of Actuarial Mathematics, Working Paper No. 32.*

NORBERG, R. (1986) A contribution to modelling of IBNR claims. *Scandinavian Actuarial Journal 1986*, 155–203.

NORROS, I. (1985) Systems weakened by failures. *Stochastic Processes Appl.* **20**, 181–196.

PENTIKÄINEN, T. and RANTALA, J. (1986) Run-off risk as a part of claims fluctuation. *ASTIN Bulletin* **16**, 113–147.

RANTALA, J. (1984) An application of stochastic control theory to insurance business. *Acta Universitatis Tamperensis* **A, 164** (Academic dissertation).

REID, D. H. (1981) A method of estimating outstanding claims in motor insurance with applications to experience rating. Colloque: Les mathématiques en Sciences Actuarielles-1981. Institut des Hautes Etudes de Belgique, Brussels, Belgium, 275–289.

ROBBIN, I. (1986) A Bayesian credibility formula for IBNR counts. *Proceedings of the Cas. Actuarial Society* **Vol. LXXIII** (1986), No. 139–140, 129–167 (with discussion).

TAYLOR, G. C. (1986) *Claims Reserving in Non-Life Insurance.* North-Holland, Amsterdam.

VAN EEGHEN, J. (1981) Loss Reserving Methods. *Surveys of Actuarial Studies* **No. 1**, Nationale Nederlanden N.V.

ELJA ARJAS
Department of Applied Mathematics and Statistics, University of Oulu,
90570 Oulu, Finland.

ON EXPERIENCE RATING AND OPTIMAL REINSURANCE

BY JUKKA RANTALA

The Ministry of Social Affairs and Health
Helsinki, Finland

ABSTRACT

This paper presents applications of stochastic control theory in determining an insurer's optimal reinsurance and rating policy. Optimality is defined by means of variances of such variables as underwriting result of the insurer, solvency margins of the insurer and reinsurer and the premiums paid by policyholders.

KEYWORDS

Optimal reinsurance; control theory; Kalman filter.

INTRODUCTION

The problem of optimal reinsurance has been widely discussed in risk-theoretical literature. This problem has several answers depending on the optimality criteria used and assumptions on random variables involved. However, from the theoretical point of view a marked simplification is possible. It has been shown e.g. by BORCH (see GERBER 1979) p. 95) that for every pair of concave utility functions of the cedant and reinsurer the optimal reinsurance arrangement can be found among those where the reinsurer's share of the claims s a function of the total claims amount only; dependence on individual risks or claim sizes is not needed. In PESONEN (1984), Theorem 10.5, a method for constructing an optimal reinsurance form is also presented when the utility functions are known but arbitrary. Usually the problem of optimal reinsurance is treated as a static one; i.e. the problem is to divide the total claims amount of a fixed time period, e.g. one year, into cedant's and reinsurer's components in an optimal way. In this paper a longer perspective is taken by assuming that

a) a reinsurance contract between two insurance companies (the cedant and reinsurer) has been made for a fairly long period and both parties will look for an arrangement which would be optimal (under some criterion) over a longer term.

This assumption justifies among other things the use of asymptotic methods.

Moreover, we assume that

b) the reinsurer's annual share of the total claims amount is a function of present and past annual total claim amounts only (i.e. reinsurance does not depend on individual risks);

and

c) the reinsurer's share is a linear function.

Assumption (b) is motivated by the above-mentioned theorem of BORCH. The linearity assumption (c) allows us to use the methods of linear stochastic control theory. It has been shown by PESONEN (1984), Theorem 10.13, that linear functions are optimal if the utility functions of the cedant and the reinsurer are linear functions of each other.

It is obvious that the three parties involved, the policy-holders, the cedant and the reinsurer, have conflicting interests. Each of them desires to have as small a share as possible of the total variation emerging from claims occurrences. It is in the interest of policy-holders that fluctuation in the premium rates be only moderate. The cedant and the reinsurer put value on smooth flows of underwriting results and solvency margins. In this paper we attempt to find a balance between these different interests by stating the optimality criteria in terms of the variances of the main variables. Examples are minimization of the variance of the total claims amount retained, subject to a constraint on the variance of the reinsurer's accumulated profit; or minimization of the variance of the premiums collected by the cedant, subject to a constraint on the sum of the variances of cedant's and reinsurer's accumulated profits.

The basic model is introduced in Section 1. Section 2 studies a simple case where both cedant's and reinsurer's premiums are assumed to be constants. In that section we use a technique of BOX-JENKINS (1976), Section 13.2; see also RANTALA (1984). In Section 3 a more general case is considered. It is then assumed that the premiums paid by policy-holders to the cedant company are also a controllable variable. This introduces an experience rating aspect into the model. The numerical solutions are relatively easy to find with the aid of the Kalman filter technique (see also RANTALA (1986)).

The main purpose of this paper is more to show a feasible way to attack the problems of reinsurance than to give explicit results directly applicable in practice. Related works are among others those by BOHMAN (1986), (who also considers the reinsurance contract on a long-term basis), GERBER (1984) and LEMAIRE-QUAIRIERE (1986) (who consider reinsurance chains).

1. The Basic Model

Consider two insurance companies. The variables relating to company $j (i = 1, 2)$ are labelled with the subscript j. Company 1 is called the *cedant* and company 2 the *reinsurer*. All variables are measured as proportions of a joint basic volume measure $V(t)$. This may be taken as e.g. the sum of insurance sums, payroll, a suitable monetary index multiplied by the number of policies, or it may be some measure which is a basis for tariffication. Thus the variables may be termed *rates* (claims rate, premium rate etc.). Moreover, all variables refer to that part of the portfolio which is covered by the reinsurance agreement in question.

We assume that $V(t)$ progresses according to equation

$$V(t) = r_g(t)\, r_x(t)\, V(t-1). \tag{1.1}$$

In equation (1.1) the total growth of the volume $V(t)$ is attributed to two factors: the growth in number of policies or risks units described by $r_g(t)$ and the growth due to inflation described by $r_x(t)$.

Now the accumulated profit (rate) $u_j(t)$ of company j satisfies equation (see BEARD-PENTIKÄINEN-PESONEN (1984), Section 6.5)

$$u_j(t) = r_j(t)\, u_j(t-1) + p_j(t) - x_j(t), \tag{1.2}$$

where $p_j(t)$ is the rate of the premiums and $x_j(t)$ the rate of the total claims amount retained by company j, $r_j(t) = r_{ij}(t)/r_g(t)\, r_x(t)$ and $r_{ij}(t)$ is the interest coefficient of company j and $r_j(t)$ may be called the relative interest rate of company j. The nature of $r_j(t)$'s is stochastic, but for simplicity they are in the following taken as time-independent non-random constants $r_j (j = 1, 2)$.

Note that even if there is variation in $r_{ij}(t)$ and $r_x(t)$, coefficient $r_j(t)$ will be fairly stable if $r_{ij}(t)/r_x(t)$ and $r_g(t)$ are stable as can often be assumed. In general, values of r_j:*s* around 1.0 are perhaps the most usual.

In addition, $x_j(t)$'s and $p_j(t)$'s must satisfy the equations

$$\begin{cases} p(t) = p_1(t) + P_2(t) \\ x(t) = x_1(t) + x_2(t), \end{cases} \tag{1.3}$$

where $p(t)$ is the total premium rate paid by the policy-holders and $x(t)$ is the total claims rate.

Another form of (1.2) and (1.3) which better brings out the control-theoretic aspects is

$$\begin{cases} u_1(t) = r_1 u_1(t-1) + y_1(t) \\ u_2(t) = r_2 u_2(t-1) + p(t) - x(t) - y_1(t), \end{cases} \tag{1.4}$$

where $y_1(t) = p_1(t) - x_1(t)$ is the cedant's underwriting result in the year t. The controllable variables in (1.4) are $y_1(t)$ (both through $p_1(t)$ and $x_1(t)$) and $p(t)$.

We study first in Section 2 a simpler case where premium rates $p(t)$, $p_1(t)$ and $p_2(t)$ are kept as constants and the problem is only do divide $x(t)$ into cedant's and reinsurer's shares.

2. The case of constant premium rates

Assume that $Ex(t)$ is known and both the total premium rate $p(t)$ and the reinsurer's premium rate $p_2(t)$ are constants. *In order to prevent* $u_j(t)$:*s from* *unlimited asymptotic behaviour it has to be assumed that* $r_j < 1$ (which has generally been the case in many countries due to rapid growth in business volume and high inflation). This assumption can be relaxed when premium control is also introduced in Section 3. *Moreover, to simplify notation we consider only deviations from corresponding expectations and thus take* $Ex(t) = 0$. Hence the premium rates are in fact the corresponding safety loadings. Determination of their rational magnitude can be based on the variances of $u_j(t)$'s but is omitted here (see however Example in Section 2.1).

Thus the accumulated profits are governed by the equations

$$\begin{cases} u_1(t) = r_1 u_1(t-1) + p_1 - x_1(t) \\ u_2(t) = r_2 u_2(t-1) + p_2 - (x(t) - x_1(t)) . \end{cases} \tag{2.1}$$

In the following we briefly sketch the method for finding the optimal linear reinsurance policy

$$x_1(t) = a_0\, x(t) + a_1\, x(t-1) + \ldots, \tag{2.2}$$

when optimality is defined to mean

(a) minimization of Dx_1 when Du_2 is restricted to a given value (or vice versa)
(b) minimization of $D(\Delta x_1)$ when Du_2 is restricted to a given value (or vice versa),

where D denotes standard deviation (i.e. D^2 is the variance operator) and Δ is the difference operator: $\Delta x(t) = x(t) - x(t-1)$.

The former criterion aims at restricting the variation range (i.e. minimums and maximums) of the cedant's annual profit, whereas the latter stresses more its smooth flow from year to year. Variation in the reinsurer's accumulated profit can be controlled by the choice of the admissible value for Du_2. If the safety margin p_2 in ceded premiums is an increasing function of Du_2, criteria (a) and (b) also give the answers to the problem: minimize loading p_2 for given Dx_1 or $D\Delta x_1$.

In what follows the derivation of the optimal coefficients $a_0, a_1, \ldots$ in (2.2) is limited in case (a) to autoregressive claims rates $x(t)$ of at most order two (abbreviated as AR(2) processes and in case (b) for AR(1) claims rates. An important special case of these, usually considered in traditional risk theory, is the white noise process of identically and independently distributed (abbreviated i.i.d.) random variables. The motivation for considering AR claims processes is the empirical observation (see BEARD-PENTIKÄINEN-PESONEN (1984), PENTIKÄINEN-RANTALA (1982), RANTALA (1988)) that claims processes are at least in some cases subject to cyclical variations. Such variations can be generated by AR(2) processes by a suitable choice of parameters. AR (or more generally ARMA processes) are also used in KREMER (1982) to find credibility premiums. A natural way to introduce the AR component into the claims process is to assume that the structure variation (see BEARD-PENTIKÄINEN-PESONEN (1984), Section 2.7) of the claims process is of autoregressive character and the process has also the usual Poisson "random noise". However, this decomposition is not used in this paper so as not to overcomplicate the model-structure and the better to extract the relevant features of the control problems.

In both cases (a) and (b) a modification of the method presented in BOX-JENKINS (1976), Section 13.2 is used to find the optimal rules. Also the Kalman filter technique to be presented in Section 3 could be used in Section 2.1, but not in Section 2.2.

2.1. *Minimization of $Dx_1(t)$ subject to a constraint on $Du_2(t)$*

The problem is (a): i.e. to minimize Dx_1 when $Du_2(t)$ is given. As stated above we restrict our considerations to autoregressive processes of at most order two. Solutions for more general processes could be found by solving the general difference equations (A1.12)-(A1.13) in Appendix 1. Thus the claims rate process is assumed to obey the difference equation

(2.1.1) $$x(t) = \phi_1 x(t-1) + \phi_2 x(t-2) + \varepsilon(t),$$

where $\varepsilon(t)$'s are uncorrelated random variables with mean zero and with variance σ_ε^2. To have finite variance for $x(t)$ coefficients ϕ_1 and ϕ_2 must satisfy the stationarity conditions

(2.1.2) $$\begin{cases} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1 . \end{cases}$$

The formulas become more handy if the so-called backward shift operator B (e.g. $Bx(t) = x(t-1)$) is taken into use. With this notation (2.1.1) can be rewritten as

(2.1.3) $$\Phi(B)\,x(t) = \varepsilon(t),$$

where

(2.1.4) $$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2.$$

It is shown in Appendix 1 that for this claims process the solution to problem (a) is (see equations (A1.25)-(A1.26) in Appendix 1)

(2.1.5) $$x_1(t) = [-(1-r_2 B)\,\mu(B)\,\Phi(B) + 1]\,x(t)$$

or equivalently

(2.1.6) $$x_1(t) = [-(1-r_2 B)\,\mu(B) + \Phi^{-1}(B)]\,\varepsilon(t),$$

where $^{-1}$ denotes the inverse operator and

(2.1.7) $$\mu(B) = A(1-z_0 B)^{-1} + (W_1 + W_2 B)\,\Phi^{-1}(B)$$

and coefficients A, W_1, and W_2 are given by equations (A1.14), (A1.21)-(A1.24) in Appendix 1 and z_0 is that solution of (A1.16) for which $|z_0| < 1$. Note that the formulas do not depend on σ_ε^2. The relevant parameters are ϕ_1, ϕ_2, r_2 and the parameter ν in (A1.14) defining the ratio Du_2/Dx_1.

The reinsurance scheme (2.1.5) leads to the following equations for u_1 and u_2:

(2.1.8) $$(1-r_1 B)\, u_1(t) = -[-(1-r_2 B)\, \mu(B)\, \Phi(B)+1]\, x(t)+p_1$$

and

(2.1.9) $$\Phi^{-1}(B)\, u_2(t) = -\mu(B)\, x(t)+p_2/(1-\phi_1-\phi_2)\,(1-r_2).$$

The variances connected with these equations are fairly easy to calculate from the ARMA presentations containing $\varepsilon(t)$'s, which result when $x(t)$ is replaced by $\Phi^{-1}(B)\,\varepsilon(t)$ in (2.1.8) and in (2.1.9). The details are omitted here (see e.g. BOX-JENKINS (1976) Section 3.4.2).

EXAMPLE. Take the classical case of risk theory that $x(t)$:s are i.i.d. random variables: $\phi_j = 0$ for $j = 1, 2$. Then $K = D_j = W_j = 0$ $(j = 1, 2)$ in equations (A1.24), and thus

(2.1.10) $$\mu(B) = r_2^{-1} z_0 (1-z_0 B)^{-1},$$

where z_0 is that root of $r_2 z^2-(1+r_2^2+\nu)z+r_2 = 0$ whose modulus is less than one. Here ν is the parameter fixing the ratio Du_2/Dx_1. The optimal reinsurance scheme is from (2.1.5) and (2.1.7)

(2.1.11) $$x_1(t) = (1-z_0 B)^{-1}(1-r_2^{-1} z_0)\, x(t)$$

or equivalently

(2.1.12) $$x_1(t) = z_0 x_1(t-1)+(1-r_2^{-1} z_0)\, x(t),$$

i.e. $x_1(t)$ is calculated according to the classical exponential smoothing formula of experience rating theory. The corresponding variance is

(2.1.13) $$D^2 x_1 = D^2 x \cdot (1-r_2^{-1} z_0)^2/(1-z_0^2).$$

The resulting solvency rate of the cedant is, from (2.1.8),

(2.1.14) $$(1-r_1 B)(1-z_0 B)\, u_1(t) = -(1-r_2^{-1} z_0)\, x(t)+p_1 (1-z_0)$$

with variance

(2.1.15) $$D^2 u_1 = \frac{(1+z_0 r_1)(1-r_2^{-1} z_0)^2}{(1-z_0 r_1)(1-r_1^2)(1-z_0^2)}\, D^2 x$$

The solvency rate of the reinsurer is

(2.1.16) $$u_2(t) = z_0 u_2(t-1)-r_2^{-1} z_0 x(t)+p_2 \cdot \frac{1-z_0}{1-r_2}.$$

and hence $u_2(t)$ is an AR(1) process with variance

$$D^2 u_2 = D_2 x (r_2^{-2} z_0^2/(1-z_0^2)). \tag{2.1.17}$$

The following figure gives the optimal combinations of Du_1, Du_2, Dx_1 and the long-term safety loadings defined by $\lambda_1 = 3(1-r_1)\,Du_1$, $\lambda_2 = 3(1-r_2)\,Du_2$ and $\lambda = \lambda_1 + \lambda_2$ as multiples of Dx when $r_1 = r_2 = 0.95$.

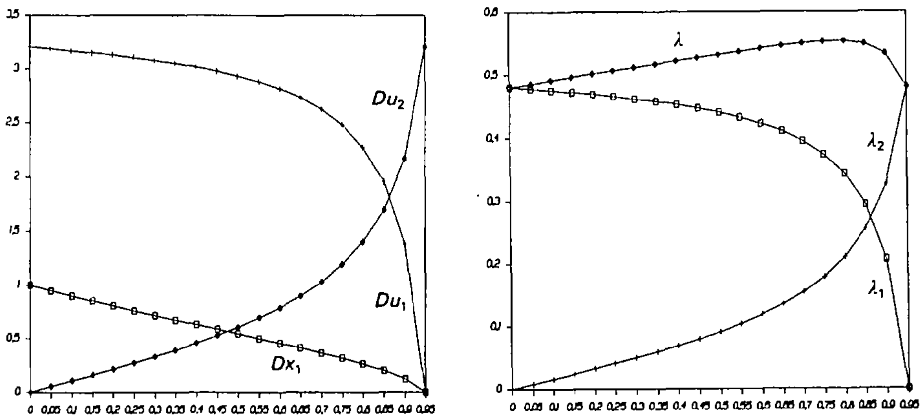

FIGURE 2.1.1. Optimal combinations of the main variables as multiples of D_x in Example 1 when $r_1 = r_2 = 0.95$.

Since an increase in z_0 means that the ceded share of the business increases it is quite natural that Dx_1 and Du_1 decrease and Du_2 increases when z_0 gets larger. Intuitively it is not so obvious that the sum of the safety loadings has its minimum when the whole risk is carried by one insurer only; i.e. if the risk is shared by two companies the safety loading is higher than without risk sharing. The reason is that in the case with reinsurance the total safety loading must maintain two solvency margins, both of which have with high probability to be positive: it is not sufficient that their sum is positive, as is in fact required in the case of no risk-sharing.

2.2. *Minimization of $D(\Delta x_1(t))$ subject to a constraint on $Du_2(t)$*

Now the problem is to minimize $D(\Delta x_1(t))$ when $Du_2(t)$ is given.

To simplify the formulas we restrict ourselves to AR(1) claims rate processes; i.e. coefficient ϕ_2 is zero in (2.1.1). Thus

$$x(t) = \phi x(t-1) + \varepsilon(t), \tag{2.2.1}$$

where $|\phi| < 1$ and $\varepsilon(t)$'s are a series of uncorrelated random variables with mean zero and with variance σ_ε^2. Moreover, let $Eu_1(t) = Eu_2(t) = 0$.

As is shown in Appendix 2 (formulas A2.18-A2.21), the solution is

$$x_1(t) = [-(1-r_2 B)(1-\phi B)\,\mu(B) + 1]\,x(t) \tag{2.2.2}$$

or

$$(2.2.3) \qquad x_1(t) = [-(1-r_2 B)\mu(B) + (1-\Phi B)^{-1}]\varepsilon(t),$$

$$(2.2.4) \qquad (1-r_1 B)u_1(t) = -x_1(t),$$

$$(2.2.5) \qquad u_2(t) = -\mu(B)\varepsilon(t),$$

where $\mu(B)$ is given by (A2.15) in Appendix 2. Thus processes $u_1(t)$, $u_2(t)$ and $x_1(t)$ are ARMA processes, whose variances are easy to compute from the presentations containing $\varepsilon(t)$'s (see BOX-JENKINS (1976), Section 3.4.2).

As a limiting case when ϕ approaches 1 we obtain from (2.2.1) a random walk process. This process also follows as a special case of an ARIMA (0, 1, 1) process:

$$(2.2.6) \qquad \Delta x(t) = (1-\theta B)\varepsilon(t)$$

with $\varepsilon(t)$'s uncorrelated and with $0 \leq \theta \leq 1$.

Equation (2.2.6) has the interpretation that every year a shock $\varepsilon(t)$ is added to the current "level" of the claims rate to produce a value $x(t)$. However, only a proportion $1-\theta$ of the shock is actually absorbed into the level to have lasting influence (see BOX-JENKINS (1976) Chapter 4).

In practice perhaps not every new shock changes the level; possible changes occur only occasionally. Thus (2.2.6) may be regarded as a cautious "upper limit)" for actual claims processes. Such changes in the claims level are to be expected e.g. due to changed policy conditions or changes in claims settlement practice. When $\theta \to 0$ we obtain a random walk process; i.e. every new shock is totally absorbed into the level, this being the most dangerous alternative. When θ is put to one we arrive at the traditional white noise claims process.

WHITE NOISE CASE $\theta = 1$. As is shown in Appendix 2 (see equation (A2.27)), the optimal reinsurance scheme is now

$$(2.2.7) \qquad (1-k_0 B + k_1 B^2)x_1(t) = (1-r_2^{-1}k_0 + r_2^{-2}k_1)x(t) \stackrel{\text{def}}{=} b_0 x(t),$$

where k_0 and k_1 are given by the procedure I-III in Appendix 2. The variance of $x_1(t)$ is

$$(2.2.8) \qquad D^2 x_1 = \frac{(1+k_1)(b_0^2+b_1^2)+2b_0 b_1 k_0}{(1-k_1)[(1+k_1)^2-k_0^2]} D^2 x.$$

with $b_1 = 0$.

The accumulated process $u_1(t)$ is an ARMA process

$$(2.2.9)\quad (1-k_0B+k_1B^2)(1-r_1B)\,u_1(t) = -(1-r_2^{-1}k_0+r_2^{-2}k_1)\,x(t),$$

whose variance is readily calculable. Moreover, $u_2(t)$ is an ARMA (2, 1) process

$$(2.2.10)\quad (1-k_0B+k_1B^2)\,u_2(t) = -[-r_2^{-2}k_1+r_2^{-1}k_0-r_2^{-1}k_1B]\,x(t) \overset{\text{def}}{=} (c_0+c_1B)\,x(t),$$

whose variance is given by (2.2.8) when b's are replaced by c's.

The following Figure 2.2.1 shows Dx_1, Du, and Du_2 for different values of parameter v, when $r_1 = r_2 = 0.95$. The curves should be compared to those of figure 2.2.1. An increase in Dx_1 is reflected as an increase in Du_1 and as a decrease in Du_2. When $v \to \infty$ the total variation is shifted to u_1, the cedant then taking the whole risk. Naturally the minimum for Dx_1 and $D\Delta x_1$ is zero, which is achieved when $v = 0$. Then Du_2 has its maximum.

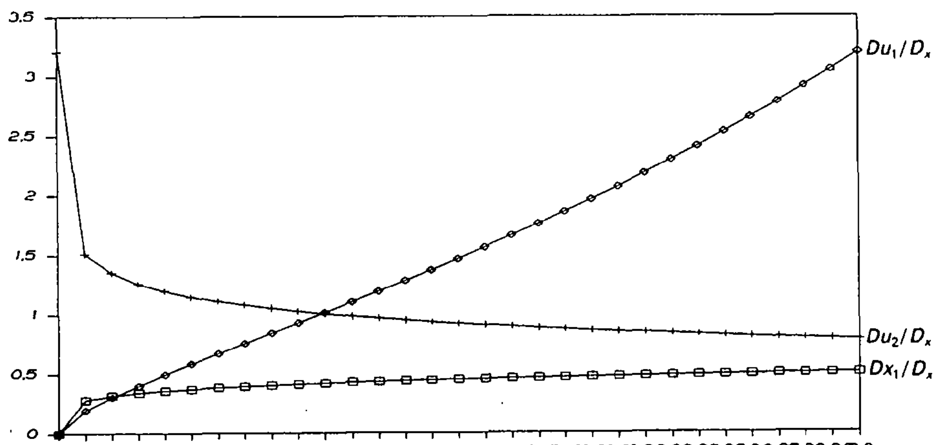

FIGURE 2.2.1. Dx_1, Du_1 and Du_2 as a functions of parameter v, when $r_1 = r_2 = 0.95$, $x(t)$ is a white noise process and $D\Delta x_1$ is minimized for given Du_2.

RANDOM WALK CASE $\theta = 0$. As is shown in Appendix 2, $u_2(t)$ corresponding to the optimal scheme is now an AR (2) process with variance (see (A2.27))

$$(2.2.11)\quad D^2u_2 = \frac{(1+k_1)(r_2^{-1}k_1)^2}{(1-k_1)[(1+k_1)^2-k_0^2]}\,\sigma_\varepsilon^2 .$$

The optimal reinsurance scheme itself is

$$(2.2.12)\quad (1-k_0B+k_1B^2)\,x_1(t) = [(1-r_2^{-1}k_1)+(r_2^{-1}k_1+k_1-k_0)B]\,x(t).$$

Thus $x_1(t)$ is a non-stationary process with infinite variance since the "driving" process $x(t)$ on the r.h.s. of (2.2.12) is such. The variance of Δx_1 is

$$(2.2.13)\qquad D^2(\Delta x_1) = \frac{(1+k_1)(w_0^2+w_1^2)+2k_0w_0w_1}{(1-k_1)[(1+k_1)^2-k_0^2]}\,\sigma_\varepsilon^2,$$

where $w_0 = (1-r_2^{-1}k_1)$ and $w_1 = r_2^{-1}k_1+k_1-k_0$.

The corresponding $u_1(t)$ process obeys equation

$$(2.2.14)\quad (1-r_1B)(1-k_0B+k_1B^2)\,u_1(t) = [1-r_2^{-1}k_1+(r_2^{-1}k_1+k_1-k_0)B]\,x(t)$$

and is thus non-stationary, since $x(t)$ is such a process.

Hence in the case of a random walk claims process the procedure produces finite $D(\Delta x_1)$ and Du_2 but with constant $p_1(t)$ Du_1 will be infinite. A finite Du_1 can be achieved if $p_1(t)$ is allowed to be non-stationary.

Although the cases considered in this section may be of some practical interest, their applicability may be rather limited since the premium rate $p(t)$ is unrealistically kept as a constant. In reality premiums are obviously also adjusted according to the observed claims experience. To obtain a more realistic model the variable premium rates should be incorporated into equations and the variation of the premium rate should also be regarded in optimality criteria.

Another limitation to the model above is that the relative interest rates r_j have to satisfy $|r_j| < 1$ in order not to have infinite variances for $u_j(t)$'s. If premium rate control is also introduced this assumption is not necessary.

3. The case where the premium rate may also vary

The technique of Box-Jenkins used in the preceding section becomes rather messy when the number of the control variables or the complexity of the claims process increases. In the following the well-known Kalman filter is used instead. However, we then obtain only numerical solutions, not analytic expressions like (2.1.5) and (2.2.2). In addition, loss function (3.7) is not suitable for such optimization as envisaged in Section 2.2, since the order of the difference of $p(t)$ which occurs in (3.7) is the same as the smallest difference parameter d for the claims process (3.2) at which $\Delta^d x(t)$ is stationary.

Since the premiums are usually charged at the beginning of the insurance period, the optimal premium rate control scheme cannot utilize the most recent $x(t)$ to determine $p(t)$; i.e. $p(t)$ is a function $x(t-1)$, $x(t-2)$, ... In order to keep the formulas as simple as possible, we then assume that the same set of data is used to determine also the retained part $x_1(t)$ of the claims. In many cases it would also be more realistic to let the time delay be even longer.

RANTALA (1986) illustrates the incorporation of a time delay in a simple case.

Take the model in the form (1.4); i.e.

$$(3.1)\qquad \begin{cases} u_1(t) = r_1 u_1(t-1) + y_1(t) \\ u_2(t) = r_2 u_2(t-1) + p(t) - y_1(t) - x(t)\,. \end{cases}$$

The control variables are the underwriting result $y_1(t)$ of the cedant and the total premiums $p(t)$. It is clear that the optimality criterion must include each of $u_1(t)$ (or alternatively $y_1(t)$), $u_2(t)$ and $p(t)$ if a solution is sought where none of these variables is identically constant: if the variation of only two variables is restricted the total variation produced by $x(t)$ can be directed to the remaining third variable by letting the other variables be constant.

We make the general assumption that the claims rate is an ARIMA (s, d, q) process

$$(3.2)\qquad \Phi(B)\,\Delta^d x(t) = \Theta(B)\,\varepsilon(t),$$

where

$$(3.3)\qquad \begin{cases} \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_s B^s \\ \Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots \theta_q B^q \\ \varepsilon(t) \;\; = \text{a sequence of uncorrelated random variables with mean zero and with variance } \sigma_\varepsilon^2. \end{cases}$$

If $d > 0$, then the $x(t)$ process defined by (3.2) is non-stationary, but if the roots of equation

$$(3.3)\qquad \Phi(B) = 0$$

lie outside the unit circle the d-th difference $\Delta^d x(t)$ of $x(t)$ is stationary. Note that for $d > 0$ the variances of $\Delta^i u_j(t)$ and $\Delta^i p(t)$ for $i < d$ and $j = 1, 2$ cannot all be finite. A natural demand is that $Du_j(t)$ $(j = 1, 2)$ and $D\Delta^d p(t)$ should be finite, i.e. the accumulated profits have finite variances and the "stationarity order" of the premium process is the same as that of the claims process.

Next (3.1) and (3.2) are transformed to a state-space model. Equations (3.1) can be rewritten as

$$(3.4)\qquad \begin{cases} (1 - r_1 B)\,\Delta^d u_1(t) = \Delta^d y_1(t) \\ (1 - r_2 B)\,\Phi(B)\,\Delta^d u_2(t) = \Phi(B)\,[\Delta^d p(t) - \Delta^d y_1(t)] - \Theta(B)\,\varepsilon(t)\,. \end{cases}$$

Let $n_1 = d+1$, $n_2 = \max\{s+d+1,\ q+1\}$ and $n = n_1 + n_2$.

Introduce n state variables $Z(i, t)$ $(i = 1, 2, \ldots, N)$ obeying equation

$$(3.5)\qquad Z(t+1) = AZ(t) + G\begin{pmatrix} \Delta^d y_1(t) \\ \Delta^d p(t) \end{pmatrix} - M\,\varepsilon(t),$$

where

(3.6)
$$
\left\{
\begin{aligned}
A &= \left[\begin{array}{c|c}
\begin{array}{c|c} a_1 & \\ \vdots & I_{n_1}-1 \\ \hline a_{n_1} & 0\;\;0\ldots0 \end{array} & O_{n_2} \\ \hline
O_{n_1} & \begin{array}{c|c} \beta_1 & \\ \vdots & I_{n_2}-1 \\ \hline \beta_{n_2} & 0\;\;0\ldots0 \end{array}
\end{array}\right] \\
I_n &= \text{identity matrix of order } n, \\
O_n &= nxn \text{ matrix of zeroes,} \\
G &= \left(\begin{array}{cc|ccc} 1 & \overbrace{0\ldots0}^{n_1-1} & -1, & \phi_1,\ldots, & \phi_{n_2} \\ 0 & 0\ldots0 & 1, & -\phi_1,\ldots, & -\phi_{n_2} \end{array}\right)', \\
M &= (\underbrace{0\ldots0}_{n_1} \mid 1, \quad -\theta_1,\ldots,-\theta_{n_2})', \\
a(B) &= (1-r_1B)\Delta^d \overset{\text{def}}{=} 1-a_1B-a_2B^2-\ldots-a_{n_1}B^{n_1}, \\
\beta(B) &= (1-r_2B)\Delta^d\Phi(B) \overset{\text{def}}{=} 1-\beta_1B-\beta_2B^2-\ldots-\beta_{n_2}B^{n_2}
\end{aligned}
\right.
$$

with $\phi_i = 0$ for $i > s$ and $\theta_i = 0$ for $i > q$ and $'$ denoting transpose.

The accumulated profits $u_1(t)$ and $u_2(t)$ are given by $Z(1, t+1)$ and $Z(n_1+1, t+1)$.

Let the loss function to be minimized be

$$E\left\{Z(N)'\,Q_0\,Z(N) + \sum_{j=1}^{N} (Z(j)'\,Q_1\,Z(j)) + Y(j)'\,Q_2\,Y(j)\right\}, \tag{3.7}$$

where Q_0, Q_1 and Q_2 are symmetric positive definite matrices, $Y(j) = (\Delta^d y_1(j),\ \Delta^d p(j))'$ and $\{1,\ldots,N\}$ is the planning horizon (a suitable choice for which is the duration of the reinsurance agreement). According to our assumption at the beginning of this section $Y(t)$ can depend on $Z(t)$, $Z(t-1),\ldots$ but not on $Z(t+1)$.

The optimal linear control rule giving the minimum for this loss function is (see e.g. ÅSTRÖM (1970): Theorem 4.1 in Section 8.4):

$$Y(t) = -L(t)\,Z(t), \tag{3.8}$$

where $Y(t)$ is the vector of the cedant's optimal profit and premium setting to be applied at time t. $L(t)$ is a $(2 \times n)$ matrix of constants given by

$$L(t) = [Q_2 + G'\,S(t+1)\,G]^{-1}\,G'\,S(t+1)\,A, \tag{3.9}$$

where $S(t+1)$ is obtained from

$$S(t) = A'\,S(t+1)\,A + Q_1 - A'\,S(t+1)\,GL(t) \tag{3.10}$$

with the initial condition

$$S(N) = Q_0. \tag{3.11}$$

Thus the optimal procedure is quite easy to reach from recurrence equations (3.8)-(3.11). However, it depends on the initial values of the state vector Z; i.e. on the immediate past of the accumulated profits $u_j(t)$. It can be shown that as the planning horizon $N \to \infty$, matrix $S(t)$ will converge to a unique steady-state positive definite value S. Denote the corresponding limit of $L(t)$ by L. Numerical calculation by computer of this *steady-state solution* is quite easy from equations (3.9) and (3.10) by successive iteration. (Note also that the results of Section 2 are in fact steady-state solutions.) The steady-state feedback rating and ceding formula is

$$Y(t) = -LZ(t). \tag{3.12}$$

This equation is quite easy to translate into a more traditional form involving only past $p(t)$'s and $u_j(t)$'s or $x(t)$'s. An example is given later.

The corresponding steady-state covariance matrix C_Z of the state vector $Z(t)$ can be obtained by iteration from equation

$$C_Z = (A - GL)\,C_Z(A - GL)' + \sigma_\varepsilon^2\,MM'. \tag{3.13}$$

The corresponding variance of $Y(t)$ is

$$\operatorname{Var} Y(t) = C_Y = LC_Z L'. \tag{3.14}$$

The steady-state variances of the accumulated profits and $\Delta^d y_1$ and $\Delta^d p$ can be found as the appropriate elements of matrices C_Z and C_Y.

Note that when $d > 0$ the variance of the premiums (as that of $x(t)$) is infinite but the variances of the accumulated profits and cedant's profit $y_1(t)$ are finite. Note also that the KALMAN filter technique can easily be extended to more than one reinsurer.

EXAMPLE 1. Take first the white noise $x(t)$ process of traditional risk theory. This case was considered in the examples of Sections 2.1 and 2.2. Now the state-space equation (3.5) is simply

$$(3.15)\quad \left\{ \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} \begin{pmatrix} u_1(t-1) \\ u_2(t-1) \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} y_1(t) \\ p(t) \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} x(t) \right.$$

and $MM' = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

Choose the matrices Q_0, Q_1 and Q_2 in loss function (3.7) as

$$(3.16)\qquad Q_0 = 0_2, Q_1 = \begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}, \qquad Q_2 = \begin{pmatrix} w_3 & 0 \\ 0 & w_4 \end{pmatrix}.$$

By varying w_i's different optimum combinations can be produced. As an example we take $r_1 = r_2 = 1.0$, $w_1 = 0.1$, $w_2 = 0.025$, $w_3 = 0.0001$ and $w_4 = 1$. Since w_3 is negligible this in fact means that the variance of premiums is minimized subject to $w_1 D^2 u_1 + w_2 D^2 u_2 =$ a given value. Furthermore, an increase in $D^2 p$ is ten times "worse" than in $D^2 u_1$ and forty times "worse" than in $D^2 u_2$ and an increase in $D^2 u_1$ four times "worse" than in $D^2 u_2$. This choice of weights reflects the thinking that the reinsurer should carry most of the fluctuations and the policy-holder the least.

With these parameters the steady-state optimal scheme turns out to be

$$(3.17)\qquad \begin{cases} y_1(t) = -0.826 \cdot u_1(t-1) + 0.173 \cdot u_2(t-1) \\ p(t) = -0.132 \cdot u_1(t-1) - 0.132 \cdot u_2(t-1) \end{cases}$$

with corresponding variances

$$(3.18)\qquad \begin{cases} D^2 y_1 = 0.0322\,\sigma_\varepsilon^2 \\ D^2 u_1 = 0.122\,\sigma_\varepsilon^2 \\ D^2 p = 0.0705\,\sigma_\varepsilon^2 \\ D^2 u_2 = 2.96\,\sigma_\varepsilon^2 . \end{cases}$$

Using equations (3.1) it can be shown that (3.17) is equivalent to

$$(3.19)\qquad \begin{cases} (1-2.652\,B+1.652\,B^2)\,y_1(t) = (0.173-0.173\,B)\,B(p(t)-x(t)) \\ (1-1.868\,B+0.868\,B^2)\,p(t) = (0.264-0.264\,B)\,By_1(t) + \\ \qquad + (0.132-0.132\,B)\,Bx(t) . \end{cases}$$

Figures 3.1 and 3.2 show the steady-state standard deviations of the main variables in the optimal schemes as a function of w_1, where loss matrices (3.16) are used with $w_3 = 0.0001$, $w_4 = 1$ and with two constant ratios $w_1/w_2 = 4$ and $w_1/w_2 = 1$.

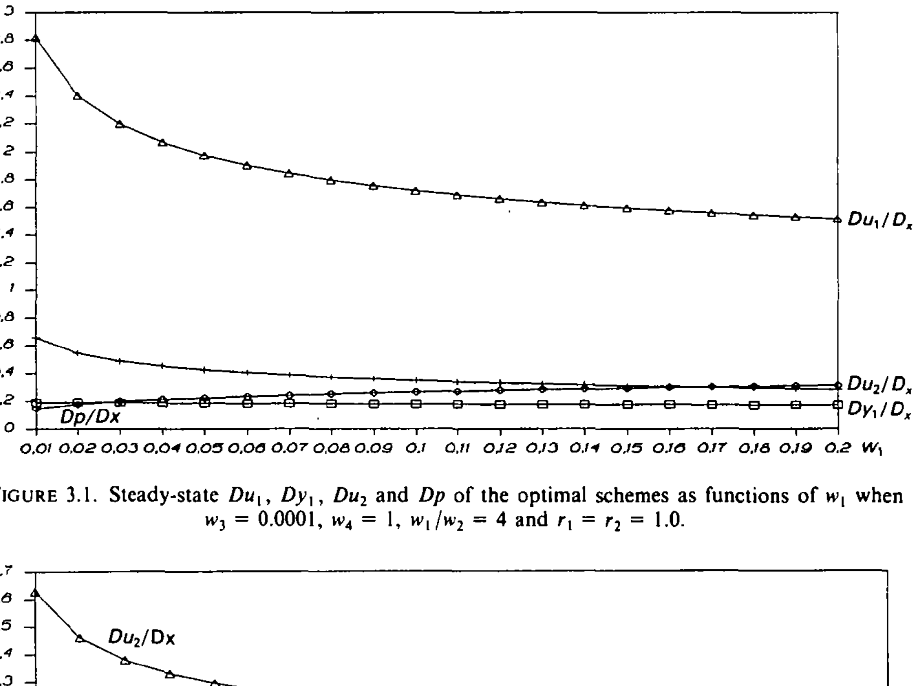

FIGURE 3.1. Steady-state Du_1, Dy_1, Du_2 and Dp of the optimal schemes as functions of w_1 when $w_3 = 0.0001$, $w_4 = 1$, $w_1/w_2 = 4$ and $r_1 = r_2 = 1.0$.

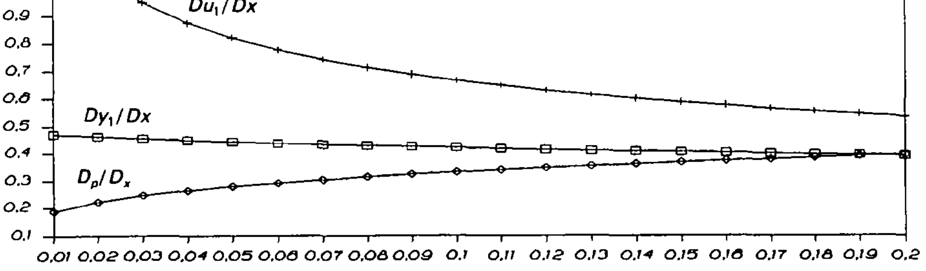

FIGURE 3.2. As Figure 3.1 but $w_1/w_2 = 1$.

In both cases Du_1, Du_2 and Dy_1 are decreasing functions of w_1, whereas Dp increases with w_1. For Du_1 and Dy_1 this is natural since the increasing w_1 means that an increase Du_1 is considered more serious and a smoother flow of u_1 is achieved by a smoother y_1. The decrease in Du_2 obviously emerges from the constancy of the ratio w_1/w_2; i.e. when w_1 increases w_2 also increases.

EXAMPLE 2. Assume that $s = q = 0$ and $d = 1$; i.e. $x(t)$ is a random walk process. As noted above, this case can be viewed as a cautious approximation which in a way constitutes an "upper limit" for actual claims processes. Now transformation (3.5) reads

$$(3.20)\quad \begin{pmatrix} Z(1,t+1) \\ Z(2,t+1) \\ Z(3,t+1) \\ Z(4,t+1) \end{pmatrix} = \begin{pmatrix} r_1+1 & 1 & 0 & 0 \\ -r_1 & 0 & 0 & 0 \\ 0 & 0 & r_2+1 & 1 \\ 0 & 0 & -r_2 & 0 \end{pmatrix} \begin{pmatrix} Z(1,t) \\ Z(2,t) \\ Z(3,t) \\ Z(4,t) \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta y_1(t) \\ \Delta p(t) \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \varepsilon(t)$$

Choose $Q_0 = 0_4$, $Q_1 = \begin{pmatrix} w_1 & 0 & 0 & 0 \\ 0 & 0.0001 & 0 & 0 \\ 0 & 0 & w_2 & 0 \\ 0 & 0 & 0 & 0.0001 \end{pmatrix}$ and Q_2 as in (3.16).

Thus, instead of Dy_1 and Dp we now consider $D(\Delta y_1)$ and $D(\Delta p)$. Note also that Dp has now to be infinite if Du_1 and Du_2 are to be finite. Take $r_1 = r_2 = 1.0$ and $w_1 = 0.01$, $w_2 = 0.05$, $w_3 = 0.5$ and $w_4 = 1.0$. The two elements on the diagonal of Q_1 other than w_1 and w_2 cannot be taken as zero, since they must be positive in order to obtain a positive definite matrix. However, they are so small that their effect on the results is insignificant. Then the steady-state solution is in the feedback form

$$(3.21)\quad \begin{cases} \Delta y_1(t) = -0.433 u_1(t-1) - 0.352 u_1(t-2) + 0.294 u_2(t-1) + 0.172 u_2(t-2) \\ \Delta p(t) = 0.374 u_1(t-1) - 0.317 u_1(t-2) - 0.521 u_2(t-1) - 0.403 u_2(t-2) \end{cases}$$

with corresponding variances

$$(3.22)\quad \begin{cases} D^2 u_1 = 6.02\sigma_\varepsilon^2 \\ D^2(\Delta y_1) = 0.14\sigma_\varepsilon^2 \\ D^2 u_2 = 4.19\sigma_\varepsilon^2 \\ D^2(\Delta p) = 0.43\sigma_\varepsilon^2 \end{cases}$$

Figures 3.3-3.4 show the steady-state standard deviations Du_1, $D(\Delta y_1)$, Du_2 and $D(\Delta p)$ of the optimal schemes as a functions of w_3 when $w_1 = 0.01$, $w_4 = 1$, $w_3/w_2 = 10$ or $= 1$.

4. Concluding remarks

The results of the paper should not be seen as suggestions for explicit solutions to be used in reinsurance treaties. In practical situations there are many factors to be taken into account, which however cannot easily be included in a mathematical model. The main emphasis of the paper is on demonstrating an

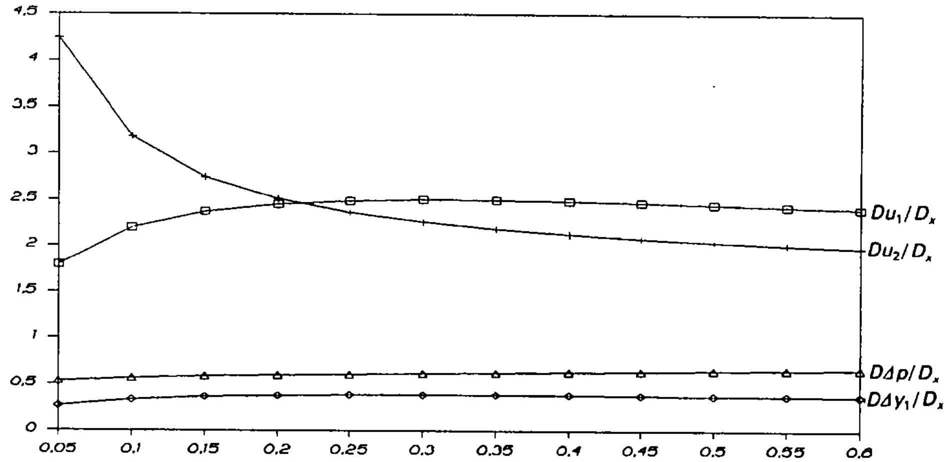

FIGURE 3.3. Steady-state Du_1, $D(\Delta y_1)$, Du_2 and $D(\Delta p)$ of the optimal schemes as functions of w_3 when $w_1 = 0.01$, $w_4 = 1$ and $w_3/w_2 = 10$ and $r_1 = r_2 = 1.0$.

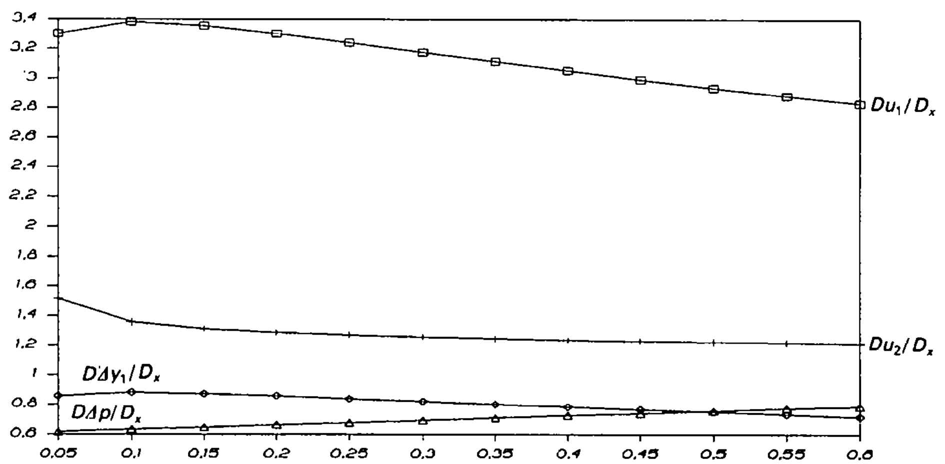

FIGURE 3.4. As figure 3.3 but $w_3/w_2 = 1$.

approach which would be considered as a rational means of tackling reinsurance problems. That is

1) cedant's and reinsurer's share of the claims are functions of the total claims amount in the reinsured part of the portfolio (i.e. they do not depend on individual risks)
2) the agreement is made on a long-term basis
3) an explicit definiton of the goals and criteria of both parties involved (such as acceptable variations in accumulated profits and in annual profits,

profitability in the long run, the rating procedure of the cedant etc.) (compare also BOHMAN (1986) and GERATHEWOHL-NIERHAUS (1986)).

In this way one may succeed in giving more weight to the most relevant factors related to a reinsurance treaty than in a heuristic approach.

This paper concentrates on point (3): how methods of stochastic control theory might be used in a search for the optimal reinsurance formulas (in Section 3 also for the rating formla), when the goals and criteria are expressed in terms of the variances of certain important variables. These rules could be applied if a sufficient consensus on the criteria and on the stochastic properties of the claims process is achieved. If there is considerable uncertainty about those properties then the formula candidates should be tested against various claims process alternatives.

APPENDIX 1

MINIMIZATION OF $Dx_1(t)$ SUBJECT TO A CONSTRAINT ON $Du_2(t)$ WITH CONSTANT PREMIUM RATES

It is assumed that the claims rate process $x(t)$ is a weakly stationary process given by equation

(A1.1) $$x(t) = \Psi(B)\,\varepsilon(t) = \varepsilon(t) + \psi_1\varepsilon(t-1) + \psi_2\varepsilon(t-2) + \ldots,$$

where $\varepsilon(t)$ is the noise process of uncorrelated random variables with mean zero and with variance σ_ε^2, and ψ_j's are the weights of past $\varepsilon(t)$'s such that $\Sigma\psi_j^2 < \infty$ and B is the backward shift operator: $B\varepsilon(t) = \varepsilon(t-1)$. However, the explicit solution is given only for the case where ψ_j's are generated by an AR(2) claims process.

It is assumed that $x(t)$, $x(t-1)$, ... are used to determine $x_1(t)$. Thus the optimal scheme can be written as the output of a linear filter $L(B)$:

(A1.2) $$x_1(t) = L(B)\,\varepsilon(t),$$

or equivalently

(A1.3) $$x_1(t) = L(B)\,\Psi^{-1}(B)\,x(t),$$

where $^{-1}$ denotes the inverse operator. If $x_1(t)$ should be a function of delayed $x(t)$'s: $x(t-d)$, $x(t-1-d)$, ... with $d < 0$ then $L(B)$ should be replaced by $B^d L(B)$ and the formulas and equations to be presented below should be correspondingly modified (see RANTALA (1984), Appendices I and II).

Let $-\mu(B)$ be the linear filter corresponding to (A1.3) and transforming $\varepsilon(t)$ into $u_2(t)$; i.e.

(A1.4) $$u_2(t) = -\mu(B)\,\varepsilon(t) = -\mu(B)\,\Psi^{-1}(B)\,x(t),$$

where we have temporarily assumed that $p = p_1 = p_2 = 0$.

Thus $\mu(B)$ and $L(B)$ are connected via equation

(A1.5) $$L(B) = -(1-r_2 B)\,\mu(B) + \Psi(B).$$

Obviously the minimum possible variance of $u_2(t)$ is zero, which results with the reinsurance scheme $L(B) = \Psi(B)$; i.e. the total business is taken over by the cedant.

The optimization problem stated in the title can be solved by finding the unrestricted minimum of

$$\frac{D^2 x_1(t)}{\sigma_\varepsilon^2} + v \cdot \left[\frac{D^2 u_2(t)}{\sigma_\varepsilon^2} - w \right], \tag{A1.6}$$

where v is the Lagrange multiplier and $w\sigma_\varepsilon^2$ the value allowed for $D^2 u_2(t)$.

The autocovariance-generating function for the autocovariances γ_k $(k = \ldots, -2-1, 0, 2, \ldots)$ is defined by (see BOX-JENKINS (1976)),

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k, \tag{A1.7}$$

where B now is a complex variable.

If $x(t) = \Psi(B)\,\varepsilon(t)$, it is easy to see that the autocovariances of $x(t)$ are generated by

$$\gamma(B) = \Psi(B)\,\Psi(F), \tag{A1.8}$$

where $F = B^{-1}$.

Applying this technique to the minimization of (A1.6) we can equivalently require an unrestricted minimum of the coefficient of $B^0 = 1$ in the expression

$$G(B) = L(B)\,L(F) + v\mu(B)\,\mu(F). \tag{A1.9}$$

Regarding (A1.5) we obtain

$$\begin{aligned} G(B) = {} & [(1-r_2 B)(1-r_2 F)+v]\,\mu(B)\,\mu(F) - \\ & - (1-r_2 B)\,\mu(B)\,\Psi(F) - (1-r_2 F)\,\mu(F)\,\Psi(B) + \Psi(B)\,\Psi(F). \end{aligned} \tag{A1.10}$$

By differentiating $G(B)$ with respect to each μ_i $(i = 0, 1, 2, \ldots)$, we obtain

$$\begin{aligned} \frac{\partial}{\partial \mu_i} G(B) = {} & [1 + r_2^2 + v - r_2 B - r_2 F]\,[B^i \mu(F) + F^i \mu(B)] - \\ & - \Psi(F)\,[B^i - r_2 B^{i+1}] - \Psi(B)\,[F^i - r_2 F^{i+1}]. \end{aligned} \tag{A1.11}$$

After selecting the coefficients of $B^0 = 1$, and equating them to zero, we obtain the following equations:

$$r_2 \mu_1 - b\mu_0 = r_2 \psi_1 - 1 \qquad (i = 0) \tag{A1.12}$$

$$r_2 \mu_{i+1} - b\mu_i + r_2 \mu_{i-1} = r_2 \psi_{i+1} - \psi_i \qquad (i \geq 1), \tag{A1.13}$$

where

(A1.14) $$b = 1 + r_2^2 + v.$$

REMARK. From (A1.12) and (A1.13) we obtain a relation for the characteristic function of μ which—if μ_0 is known—determines μ:

$$\mu(z)(r_2 + r_2 z^2 - bz) = \psi(z)(r_2 - z) - r_2 + r_2 \mu_0.$$

The solution of (A1.12)-(A1.13) is the sum of the solution of the corresponding homogeneous equation and any particular solution of the homogeneous equation.

First the solution of the homegeneous difference equation

(A1.15) $$r_2 \mu_{i+2} - b\mu_{i+1} + r_2 \mu_i = 0 \qquad (i = 0, 1, 2, \ldots)$$

is sought. The characteristic equation is

(A1.16) $$r_2 z^2 - bz + r_2 = 0;$$

i.e.

(A1.17) $$r_2 z + r_2 z^{-1} = b.$$

Thus if z_0 is a solution so is z_0^{-1} and the general solution of (A1.15) is

(A1.18) $$\mu_i = A z_0^i + A' z_0^{-i} \qquad (i = 0, 1, 2, \ldots).$$

Now, if z_0 has a modulus less than or equal to one, then z_0^{-1} has a modulus greater than or equal to one, and since $u_2(t)$ in the optimal solution must have finite variance, A' must be zero. Because of the property (A1.17) it is easy to see that z must be real. Thus the general solution of (A1.15) is $\mu(B) = A(1 - z_0 B)^{-1}$.

In deriving the particular solution of (A1.12)-(A1.13) we confine ourselves to autoregressive processes of at most order two; i.e. we assume that the weights are given by

(A1.19) $$\Psi(B) = (1 - \phi_1 B - \phi_2 B^2)^{-1}$$

and ϕ_1 and ϕ_2 are constants satisfying stationary conditions (2.1.2).

It can be shown (see RANTALA (1984), Appendix II) and is easy to check that the solution of (A1.12)-(A1.13) is then

(A1.20) $$\mu(B) = A(1 - z_0 B)^{-1} + (W_1 + W_2 B)(1 - \phi_1 B - \phi_2 B^2)^{-1},$$

where the second term on the r.h.s. is a particular solution. Coefficients A, W_1 and W_2 are given by equations

(A1.21)
$$\begin{cases} W_1 = \sqrt{-\phi_2}\,(D_1 \cos\theta + D_2 \sin\theta) \\ W_2 = -\phi_1 W_1 - \phi_2 (D_1 \cos 2\theta + D_2 \sin 2\theta) \\ \tan\theta = \sqrt{\dfrac{-\phi_1^2 - 4\phi_2}{\phi_1}} \qquad (0 \le \theta \le \pi) \\ D_1 = \dfrac{C_1 E_1 + C_2 E_2}{E_1^2 + E_2^2}\sqrt{-\phi_2} \\ D_2 = \dfrac{C_2 E_1 - C_1 E_2}{E_1^2 + E_2^2}\sqrt{-\phi_2} \\ E_1 = \dfrac{r_2 \phi_1}{2\sqrt{-\phi_2}}(1-\phi_2) - b\sqrt{-\phi_2} \\ E_2 = r_2 \sqrt{1 + \phi_1^2/4\phi_2} \cdot (1+\phi_2) \\ C_1 = r_2 \phi_1 - 1 \\ C_2 = \dfrac{(r_2\phi_1 - 1)\phi_1 + 2 r_2 \phi_2}{\sqrt{-\phi_1^2 - 4\phi_2}} \\ A = r_2^{-1} z_0 \cdot [D_1 (r\sqrt{-\phi_2}\cos\theta - b) + D_2 r \sqrt{-\phi_2}\sin\theta - r\phi_1 + 1] \end{cases}$$

when the roots of

(A1.22)
$$z^2 - \phi_1 z + \phi_2 = 0$$

are complex, and

(A1.23)
$$\begin{cases} W_1 = D_1 K_1 + D_2 K_2 \\ W_2 = -K_1 K_2 (D_1 + D_2) \\ D_1 = \dfrac{C_1 K_1}{r_2 K_1^2 - b K_1 + r_2} \\ D_2 = \dfrac{C_2 K_2}{r_2 K_2^2 - b K_2 + r_2} \\ C_1 = \dfrac{K_1 (1 - r_2 K_1)}{K_2 - K_1} \\ C_2 = -\dfrac{K_2 (1 - r_2 K_2)}{K_2 - K_1} \\ A = r_2^{-1} z_0 \cdot [D_1 (r K_1 - b) + D_2 (r K_2 - b) - r\phi_1 + 1] \end{cases}$$

when the roots K_1 and K_2 of (A1.22) are real and distinct.

When $K_1 = K_2 = K$ the following equations are obtained

$$
\text{(A1.24)} \quad \begin{cases}
C_1 = 2r_2K - 1 \\
C_2 = r_2K - 1 \\
D_2 = \dfrac{C_2K}{r_2K^2 - bK + r_2} \\
D_1 = \dfrac{C_1K + r_2D_2(1-K^2)}{r_2K^2 - bK + r_2} \\
W_1 = (D_1 + D_2)K \\
W_2 = -D_1K^2 \\
A = r_2^{-1}z_0 \cdot [(D_1 + D_2)(rK - b) - r\phi_1 + 1] .
\end{cases}
$$

Now the optimal reinsurance scheme may be found by substituting (A1.20) into (A1.5). As can be seen from equations (2.1), (A1.2)-(A1.5), the resulting difference equations for x_1, u_1 and u_2 are

$$\text{(A1.25)} \qquad x_1(t) = [-(1-r_2B)\mu(B)\Phi(B)+1]\,x(t)$$

or equivalently

$$\text{(A1.26)} \qquad x_1(t) = [-(1-r_2B)\mu(B)+\Phi^{-1}(B)]\,\varepsilon(t),$$

$$\text{(A1.27)} \qquad (1-r_1B)\,u_1(t) = -[-(1-r_2B)\mu(B)\Phi(B)+1]\,x(t)+p_1$$

and

$$\text{(A1.28)} \qquad \Phi^{-1}(B)\,u_2(t) = -\mu(B)\,x(t)+p_2/(1-\phi_1-\phi_2)(1-r_2).$$

In (A1.27) and (A1.28) the effects of non-zero premium rates are taken into account. Processes $x_1(t)$, $u_1(t)$ and $u_2(t)$ are ARMA processes whose variances are easy to compute from the presentations based on the noise process $\varepsilon(t)$.

APPENDIX 2

MINIMIZATION OF $D(\Delta x_1(t))$ SUBJECT TO A CONSTRAINT ON $Du_2(t)$ WITH CONSTANT PREMIUM RATES

Assume again that the total claims rate $x(t)$ is given by (A1.1). Moreover, in order to shorten the notations assume that $p = p_1 = p_2 = 0$.

By defining the change in the retained claims rate in the optimal linear scheme as

(A2.1) $$\Delta x_1(t) = (1-B)\,x_1(t) = L(B)\,\varepsilon(t)$$

we can proceed analogously to Appendix 1. The resulting difference equations are

(A2.2) $$(i=0): r_2\mu_2-(r_2+1)^2\mu_1+c\mu_0 = r_2\psi_2-(2r_2+1)\,\psi_1+(r_2+2),$$

(A2.3) $$(i=1): r_2\mu_3-(r_2+1)^2\mu_2+c\mu_1-(r_2+1)^2\mu_0$$
$$= r_2\psi_3-(2r_2+1)\,\psi_2+(r_2+2)\,\psi_1-1$$

(A2.4) $$(i\ge 2): r_2\mu_{i+2}-(r_2+1)^2\mu_{i+1}+c\mu_i-(r_2+1)^2\mu_{i-1}+r_2\mu_{i-2}$$
$$= r_2\psi_{i+2}-(2r_2+1)\,\psi_{i+1}+(r_2+2)\,\psi_i-\psi_{i-1},$$

where

(A2.5) $$c = 2(1+r_2+r_2^2)+v\,.$$

Thus we have to solve a difference equation of order four. The homogeneous equation is solvable by the methods presented in BOX-JENKINS (1976), Section 13.2.

The characteristic equation corresponding to difference equation (A2.4) is

(A2.6) $$r_2z^4-(r_2+1)^2z^2+cz^2-(r_2-1)^2z+r_2 = 0\,.$$

Hence, if z is a solution so is z^{-1}. Let the roots be K_1, K_1^{-1}, K_2 and K_2^{-1} with $|K_1| < 1$ and $|K_2| < 1$. If $v = 0$ then the roots of (A2.6) are 1, r_2 and r_2^{-1}. Then the modulus of only one root is less than 1. To rule out this case we assume that $v > 0$.

In subsequent applications we need only coefficients $k_0 = K_1+K_2$ and $k_1 = K_1K_2$. They can be found by the following procedure (see BOX-JENKINS (1976)):

(I) Compute $M = (1+r_2)^2/r_2$ and $N = [(1+r_2)^2+(1+r_2^2)+v]/r_2$

for a series of values of v chosen to provide a suitable range for Du_2 and $D\Delta x_1$.

(II) Compute $z_1 = 0.5(N-2) + \sqrt{0.25(N-2)^2+2N-M^2}$

and $z_2 = 0.5(N-2) - \sqrt{0.25(N-2)^2+2N-M^2}\,.$

(III) Compute $k_1 = 0.5z_1 - \sqrt{(0.5z_1)^2-1}$

and $k_0 = \sqrt{k_1(z_2+2)}\,.$

The general solution of the homogeneous equation is

(A2.7) $$\mu_i = A_1K_1^i+A_1'K_1^{-1}+A_2'K_2^i+A_2'K_2^{-i} \qquad (i=0,1,2,\ldots)\,.$$

In this solution A_1' and A_2' must be zero because in the optimal solution the solvency rate cannot have infinite variance. Hence

(A2.8) $\mu_i = A_1 K_1^i + A_2 K_2^i, \quad |K_1| < 1, \quad |K_2| < 1 \quad (i = 0, 1, 2, \ldots).$

This solution is the same, apart from coefficients A_1 and A_2, for every $x(t)$ process. The exact solution contains features which are specific to individual $x(t)$ processes; i.e. it depends on the particular solution of (A2.2)-(A2.4).

For the case $\Psi(B) = (1-\phi B)^{-1}$ with $|\phi| < 1$ a particular solution of (A2.2)-(A2.4) is easy to find. In fact, a particular solution is given by

(A2.9) $$\mu_i = D\phi^i \qquad (i = 1, 2, \ldots),$$

where

(A2.10) $$D/\phi = \frac{r_2(\phi-1)^2(\phi-r_2^{-1})}{r_2\phi^4-(r_2+1)^2\phi^3+c\phi^2-(r_2+1)^2\phi+r_2}.$$

Constants A_1 and A_2 can be determined from initial conditions (A2.2) and (A2.3), giving

(A2.11) $$\begin{cases} A_1 = \dfrac{K_1^2\left(\dfrac{r_2 DK_2}{\phi^2}+\dfrac{K_2}{\phi}-\dfrac{r_2 D}{\phi}\right)}{r_2(K_1-K_2)} \\[2ex] A_2 = \dfrac{K_2^2\left(\dfrac{r_2 DK_1}{\phi^2}+\dfrac{K_1}{\phi}-\dfrac{r_2 D}{\phi}\right)}{r_2(K_2-K_1)}. \end{cases}$$

In deriving $\mu(B)$ and $L(B)$ it is useful to observe that

(A2.12) $$A_1+A_2 = Dk_1/\phi^2+k_1/r_2\phi-Dk_0/\phi$$

and

(A2.13) $$A_1K_2+A_2K_1 = -k_1D/\phi.$$

The final solution is

(A2.14) $$\mu_i = A_1K_1^i+A_2K_2^i+D\phi^i \qquad (i = 0, 1, 2, \ldots)$$

or equivalently

(A2.15) $$\mu(B) = \frac{\mu_0+\mu_1 B}{1-k_0B+k_1B^2}+\frac{D}{1-\phi B},$$

where (see (A2.12) and (A2.13))

(A2.16) $$\mu_0 = A_1+A_2$$

and

(A2.17) $$\mu_1 = -(A_1 K_2 + A_2 K_1).$$

Thus the final formulas are:

(A2.18) $$x_1(t) = [-(1-r_2 B)(1-\phi B)\mu(B)+1]\, x(t)$$

or

(A2.19) $$x_1(t) = [-(1-r_2 B)\mu(B)+(1-\phi B)^{-1}]\,\varepsilon(t),$$

(A2.20) $$(1-r_1 B)\, u_1(t) = -x_1(t),$$

(A2.21) $$u_2(t) = -\mu(B)\,\varepsilon(t).$$

The necessary coefficients can be found from equations (A2.5), procedure I-III, (A2.10), (A2.11)-(A2.13) and (A2.15)-(A2.17).

The corresponding variances can most easily be calculated from the presentations containing $\varepsilon(t)$'s. Note that the effect of the constant premium rates p, p_1 and p_2 is not shown in equations (A2.18)-(A2.21), since we assumed the rates to be identically zero.

Next, the random walk claims process is considered. For this purpose we take a slightly more general process by assuming that

(A2.22) $$\Delta x(t) = (1-\theta B)\,\varepsilon(t)$$

with $\varepsilon(t)$'s uncorrelated; i.e. $x(t)$ is an ARIMA (0, 1, 1) process.

When looking for the solution we can proceed analogously with the considerations earlier in this Appendix. Now the following difference equations are obtained:

(A2.23) $$r_2\mu_2-(r_2+1)^2\mu_1+c\mu_0 = 1+(r_2+1)\theta \qquad (i = 0)$$

(A2.24) $$r_2\mu_3-(r_2+1)^2\mu_2+c\mu_1-(r_2^2+1)^2\mu_0 = -\theta \qquad (i=1)$$

(A2.25) $$r_2\mu_{i+2}-(r_2+1)^2\mu_{i+1}+c\mu_i-(r_2+1)^2\mu_{i-1}+r_2\mu_{i-2} = 0 \quad (i \geq 2)$$

The solution of this difference equation is exactly the same as that of the homogeneous equation above; i.e.

(A2.26) $$\mu_i = A_1 K_1^i + A_2 K_2^i, \quad |K_1| < 1, \quad |K_2| < 1 \quad (i = 0, 1, 2, \ldots)$$

and K_1 and K_2 are the solutions of equation (A2.6). Constants A_1 and A_2 can be computed from intial conditions (A2.23) and (A2.24).

For all θ $\mu(B)$ is of the form

(A2.27) $$\mu(B) = \frac{\mu_0+\mu_1 B}{1-k_0 B+k_1 B^2},$$

where

$$\mu_0 = A_1+A_2 = r_2^{-2}[r_2-r_2\theta-\theta]\, k_1 + r_2^{-1}\theta k_0$$

and

$$\mu_1 = -(A_1K_2+A_2K_1) = -r_2^{-1}k_1\theta .$$

White noise case $\theta = 1$ gives $\mu_0 = -r_2^{-2}k_1+r_2^{-1}k_0$ and $\mu_1 = -r_2^{-1}k_1$ and the random walk case $\theta = 0$ gives $\mu = r_2^{-1}k_1$ and $\mu_1 = 0$.

REFERENCES

BEARD, R.E., PENTIKÄINEN, T. and PESONEN, E. (1984) *Risk Theory*. Chapman and Hall, London.

BOHMAN, H. (1986) Reinsurance, Profitability and Insurability. *The Geneva Papers on Risk and Insurance* **11**, 126–129.

BOX, G.P.E. and JENKINS, G.M. (1976) *Time-series analysis*. Holden Day, New York.

GERATHEWOHL, K. and NIERHAUS, F. (1986) Princing Proportional Reinsurance. *Nordisk Försäkringstidsskrift* **66**, 274–288.

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

GERBER, H.U. (1984) Chains of Reinsurance. *Insurance: Mathematics and Economics* **3**, 43–48.

KREMER, E. (1982) Credibility Theory for Some Evolutionary Models. *Scandinavian Actuarial Journal*, 129–142.

LEMAIRE, J. and QUAIRIERE, J.-P. (1986) Chains of Reinsurance Revisited. *ASTIN Bulletin* **16**, 77–88.

PENTIKÄINEN, T. and RANTALA, J. (ed.) (1982) *Solvency of Insurers and Equalization Reserves*, Vol. I and II. The Insurance Publishing Company, Helsinki.

PESONEN, M. (1984) Optimal Reinsurance. *Scandinavian Actuarial Journal*, 65–90.

RANTALA, J. (1984) *An application of stochastic control theory to insurance business*. Ph.D. Thesis, University of Tampere.

RANTALA, J. (1986) Experience rating of ARIMA claims processes by using Kalman filter. *ASTIN Bulletin* **16**, 19–31.

RANTALA, J. (1988) Fluctuations in Insurance Business Results: Some Control-Theoretical Aspects. General lecture on Subject 1. The 23rd International Congress of Actuaries, Helsinki.

ÅSTRÖM, K.J. (1970) *Introduction to Stochastic Control Theory*. Academic Press, New York.

DR. JUKKA RANTALA
Ministry of Social Affairs and Health, P.O. Box 267, SF-00171 Helsinki, Finland.

THE PROBABILITY OF EVENTUAL RUIN IN THE COMPOUND BINOMIAL MODEL

BY ELIAS S.W. SHIU

University of Manitoba, Canada

ABSTRACT

This paper derives several formulas for the probability of eventual ruin in a discrete-time model. In this model, the number of claims process is assumed to be binomial. The claim amounts, premium rate and initial surplus are assumed to be integer-valued.

KEYWORDS

Compound binomial process; Probability of eventual ruin; Ultimate ruin probability; Infinite-time ruin probability; Risk theory; Random walk; Gambler's ruin; Lagrange series.

1. INTRODUCTION AND NOTATION

This paper is motivated by the recent paper GERBER (1988b), which discusses the probability of eventual ruin in a discrete-time model. We shall derive some of GERBER'S results by alternative methods. As we shall point out below, our formulation and notation are not exactly the same as GERBER'S.

We consider a discrete-time model, in which the number of insurance claims is governed by a binomial process $N(t)$, $t = 0, 1, 2, \ldots$. In any time period, the probability of a claim is q (denoted by p in GERBER'S paper) and the probability of no claim is $1-q$. The occurrences of a claim in different time periods are independent events. The individual claim amounts $X_1, X_2, X_3, \ldots$ are mutually independent, identically distributed, positive and integer-valued random variables; they are independent of the binomial process $N(t)$. Put $X = X_1$, and let $p(x) = \Pr(X = x)$. The value of the probability density function $p(x)$ is zero unless x is a positive integer. We also assume that the premium received in each period is one and is larger than the net premium $qE(X)$. Put $E(X) = \mu$; then the last assumption is

$$1 > q\mu \,. \tag{1.1}$$

For $k = 1, 2, 3, \ldots$, define

$$S_k = X_1 + X_2 + \ldots + X_k \,. \tag{1.2}$$

Put $S_0 = 0$. Let the initial risk reserve be a nonnegative integral amount u. The probability of eventual ruin (ultimate ruin probability, infinite-time ruin probability) $\psi(u)$ is the probability that the risk reserve

$$U(t) = u + t - S_{N(t)} \tag{1.3}$$

is ever negative. Since GERBER (1988b) defines ruin as the event that the risk reserve $U(t)$ becomes nonpositive for some t, $t > 0$, the formulas derived below will not be exactly the same as his.

2. THE PROBABILITY OF NONRUIN

It is somewhat easier to work with the nonruin function

$$\phi(u) = 1 - \psi(u).$$

For $u < 0$, $\phi(u) = 0$. Consider an initial risk reserve of amount j, $j \geq 0$. If there is no claim in the first period, the risk reserve becomes $j+1$ at the end of the period; if there is a claim of amount x in the first period, the risk reserve becomes $j+1-x$. Hence, by the law of total probability,

$$\phi(j) = (1-q)\,\phi(j+1) + qE[\phi(j+1-X)], \qquad j = 0, 1, 2, \ldots. \tag{2.1}$$

Rearranging (2.1) yields

$$\phi(j+1) - \phi(j) = q\{\phi(j+1) - E[\phi(j+1-X)]\}, \qquad j = 0, 1, 2, \ldots. \tag{2.2}$$

Summing (2.2) from $j = 0$ to $j = k-1$, we have

$$\phi(k) - \phi(0) = q\left\{\sum_{j=1}^{k} \phi(j) - E\left[\sum_{j=1}^{k} \phi(j-X)\right]\right\}, \quad k = 1, 2, 3, \ldots,$$

or

$$\phi(k) - (1-q)\,\phi(0) = q\left\{\sum_{j=0}^{k} \phi(j) - E\left[\sum_{j=1}^{k} \phi(j-X)\right]\right\}, k = 1, 2, 3, \ldots \tag{2.3}$$

Let 1_+ denote the function defined by

$$1_+(j) = 1, \qquad j = 0, 1, 2, \ldots,$$
$$1_+(j) = 0, \qquad j = -1, -2, \ldots.$$

For each pair of functions f and g, let $f*g$ denote their convolution,

$$(f*g)(j) = \sum_{i=-\infty}^{\infty} f(j-i)\,g(i). \tag{2.4}$$

Note that, if $f(i) = g(i) = 0$ for all negative integers i, then (2.4) becomes

$$(f*g)(j) = \sum_{i=0}^{j} f(j-i)\,g(i).$$

Since the convolution operation can be regarded as a multiplication operation between functions, we sometimes write $(f*g)(j)$ as $f(j)*g(j)$.

The first sum in the right-hand side of (2.3) is $(\phi * 1_+)(k)$. As X is a positive random variable,

$$\sum_{j=1}^{k} \phi(j-X) = \sum_{j=0}^{k} \phi(j-X) = (\phi * 1_+)(k-X). \tag{2.5}$$

Hence, (2.3) becomes

$$\begin{aligned} \phi(k)-(1-q)\,\phi(0) &= q\{(\phi * 1_+)(k) - E[(\phi * 1_+)(k-X)]\} \\ &= q[(\phi * 1_+)(k) - (\phi * 1_+ * p)(k)], \quad k = 1, 2, 3, \ldots. \end{aligned} \tag{2.6}$$

Since $p(0) = 0$, it is easy to check that (2.6) also holds for $k = 0$. To solve for ϕ in (2.6), we first extend it as an equation for all integers k, positive and negative:

$$\phi(k)-(1-q)\,\phi(0)\,1_+(k) = q[(\phi * 1_+)(k) - (\phi * 1_+ * p)(k)]. \tag{2.7}$$

Let δ be the function defined by $\delta(0) = 1$ and $\delta(j) = 0$ for $j \neq 0$. Then the right-hand side of (2.7) can be expressed as

$$q\{\phi(k) * 1_+(k) * [\delta(k) - p(k)]\}.$$

Rearranging (2.7) and writing

$$c = (1-q)\,\phi(0) \tag{2.8}$$

yields

$$\phi(k) * (\delta(k) - q\{1_+(k) * [\delta(k) - p(k)]\}) = c 1_+(k). \tag{2.9}$$

Equation (2.9) is a Volterra equation of the second kind. To solve for ϕ, we invert

$$\delta(k) - q\{1_+(k) * [\delta(k) - p(k)]\}$$

as the Neumann series [Brown and Page (1970, p. 226), Riesz and Sz.-Nagy (1955, p. 146)]

$$\sum_{n=0}^{\infty} q^n \{1_+(k) * [\delta(k) - p(k)]\}^{*n}. \tag{2.10}$$

(We use the notation: $f^{*0} = \delta$ and $f^{*n} = f^{*(n-1)} * f$, $n = 1, 2, 3, \ldots$.). Hence,

$$\phi(k) = c \sum_{n=0}^{\infty} q^n \{[\delta(k) - p(k)]^{*n} * 1_+^{*(n+1)}(k)\}. \tag{2.11}$$

Since

$$[\delta(k)-p(k)]^{*n} = \sum_{j=0}^{n} \binom{n}{j} (-1)^j p^{*j}(k) ,$$

(2.12) $$1_+^{*(n+1)}(k) = \binom{k+n}{n} 1_+(k) ,$$

$$\binom{n}{j}\binom{k+n}{n} = \binom{k+j}{j}\binom{k+n}{n-j} ,$$

$$\sum_{n=j}^{\infty} \binom{k+n}{n-j} q^{n-j} = \left(\frac{1}{1-q}\right)^{k+j+1}$$

and

$$p^{*j}(k) * f(k) = E[f(k-S_j)] ,$$

by an interchange of the order of summation (2.11) becomes

$$\phi(k) = c \sum_{j=0}^{\infty} (-q)^j \left\{ p^{*j}(k) * \left[\binom{k+j}{j} \left(\frac{1}{1-q}\right)^{k+j+1} 1_+(k) \right] \right\}$$

(2.13) $$= \phi(0) \sum_{j=0}^{\infty} \left(\frac{-q}{1-q}\right)^j E\left[\binom{k+j-S_j}{j} (1-q)^{S_j-k} 1_+(k-S_j)\right] .$$

As $S_j \geq j$, there are at most $k+1$ nonzero terms in the right-hand side of (2.13). This formula corresponds to (4.6) of SHIU (1988) and (3.14) of SHIU (1989a).

To derive the value of $\phi(0)$, we return to formula (2.6). Let P denote the probability distribution function of the individual claim amount random variable X. Then

$$P = 1_+ * p .$$

As k tends to positive infinity, the left-hand side of (2.6) tends to

$$1-(1-q)\,\phi(0) ,$$

while the right-hand side tends to

$$q \sum_{j=-\infty}^{\infty} [1_+(j) - P(j)] = q \sum_{j=0}^{\infty} [1-P(j)]$$

$$= q\mu$$

by the Lebesgue dominated convergence theorem. Hence.

$$\phi(0) = \frac{1-q\mu}{1-q}. \tag{2.14}$$

3. GAMBLER'S RUIN

As a verification of formulas (2.13) and (2.14), let us consider the special case that $X \equiv 2$. This is a classical problem in the theory of random walk. The probability that, with an initial reserve of u (a nonnegative integer), the company's risk reserve will ever become -1 is known to be $[q/(1-q)]^{u+1}$.

Since $S_j = 2j$, formula (2.13) becomes

$$\phi(u) = \frac{\phi(0)}{(1-q)^u} \sum_{j=0}^{\infty} [-q(1-q)]^j \binom{u-j}{j} 1_+(u-2j)$$

$$= \frac{1-2q}{(1-q)^{u+1}} \sum_{j=0}^{[\![u/2]\!]} [-q(1-q)]^j \binom{u-j}{j}. \tag{3.1}$$

For a real number r, we let $[\![r]\!]$ denote the greatest integer less than or equal to r. The polynomial

$$\sum_{n=0}^{[\![k/2]\!]} \binom{k-n}{n} x^n \tag{3.2}$$

is related to the Chebyshev polynomials of the second kind and can be expressed as [KNUTH (1973, problem 1.2.9.15), RIORDAN (1968, p. 76)]

$$\frac{(1+\sqrt{1+4x})^{k+1} - (1-\sqrt{1+4x})^{k+1}}{2^{k+1}\sqrt{1+4x}}. \tag{3.3}$$

Now,

$$\sqrt{1-4q(1-q)} = |2q-1|$$
$$= 1-2q$$

by assumption (1.1). Hence,

$$\phi(u) = 1 - \left(\frac{q}{1-q}\right)^{u+1} \tag{3.4}$$

as required.

For the case that $X \equiv m > 2$, formula (2.13) cannot be simplified. It has been given by BURMAN (1946). Also see GIRSHICK (1946, p. 290), SEAL (1962, p. 23; 1969, p. 101) and GERBER (1988b, (43)).

4. ANOTHER RUIN PROBABILITY FORMULA

GERBER (1988b) has derived another formula for the probability of eventual ruin, which is complementary to (2.13). It follows from condition (1.1) that

$$\Pr\left[\lim_{t\to\infty} U(t) = +\infty\right] = 1 .$$

If ruin occurs, there is necessarily a last upcrossing of the risk reserve $U(t)$ from level -1 to level 0. By considering the number of claims n, prior to this last upcrossing, and the time t at which it occurs, we have

$$\text{(4.1)} \quad \psi(u) = \left[\sum_{n=1}^{\infty}\sum_{t=n}^{\infty}\binom{t}{n} q^n(1-q)^{t-n}\Pr(S_n = u+t+1)\right](1-q)\,\phi(0)\,.$$

Since

$$\sum_{t=n}^{\infty}\binom{t}{n}(1-q)^t\Pr(S_n = u+t+1)$$

$$= E\left[\binom{S_n-u-1}{n}(1-q)^{S_n-u-1}\,1_+(S_n-u-n-1)\right],$$

we obtain the formula

$$\text{(4.2)} \qquad \psi(u) = (1-q\mu)\sum_{n=1}^{\infty}\left(\frac{q}{1-q}\right)^n$$

$$E\left[\binom{S_n-u-1}{n}(1-q)^{S_n-u-1}\,1_+(S_n-u-n-1)\right].$$

Continuous-time analogues of (4.2) can be found in PRABHU (1965, (5.55)), GERBER (1988a, (27)) and SHIU (1989a, (1.6)).

5. GERBER'S FANCY SERIES

Using the identity

$$(-1)^j\binom{-a}{j} = \binom{a+j-1}{j},$$

we can rewrite (2.13) as

$$\text{(5.1)} \quad \phi(u) = (1-q\mu)\sum_{j=0}^{\infty}\left(\frac{q}{1-q}\right)^j E\left[\binom{S_j-u-1}{j}(1-q)^{S_j-u-1}\,1_+(u-S_j)\right].$$

Since

$$\phi(u)+\psi(u) = 1$$

and u is an integer, adding (5.1) to (4.2) yields

$$\frac{1}{1-q\mu} = \sum_{n=0}^{\infty} \left(\frac{q}{1-q}\right)^n E\left[\binom{S_n+x}{n}(1-q)^{S_n+x}\right], \tag{5.2}$$

if we put $x = -(u+1)$. This interesting formula is Theorem 1a of GERBER (1988b). In this section we present some alternative proofs for (5.2); the assumption that x is an integer will not be used.

Assume that all the moments of the random variable X exist. Consider the linear operator G on the linear space of polynomials defined by

$$(Gf)(y) = E[f(y+X)]. \tag{5.3}$$

[Such operators have been considered by FELLER (1971, section VIII.3)]. As f is a polynomial, the random variable $f(y+X)$ in (5.3) can be expressed as

$$\sum_{j\geq 0} \frac{X^j f^{(j)}(y)}{j!}. \tag{5.4}$$

Consequently, the linear operator G can be represented as a power series in terms of the differentiation operator D:

$$G = \sum_{j\geq 0} \frac{E(X^j)}{j!} D^j. \tag{5.5}$$

Since

$$G-I = \mu D + \tfrac{1}{2} E(X^2) D^2 + \ldots,$$

we have, for each nonnegative integer n,

$$(G-I)^n x^n = n!\,\mu^n \tag{5.6}$$

and, for nonnegative integers n and m, $m < n$,

$$(G-I)^n x^m = 0. \tag{5.7}$$

It follows from (5.6) and (5.7) that

$$(G-I)^n \binom{x}{n} = \mu^n. \tag{5.8}$$

Multiplying (5.8) with q^n and summing from $n = 0$ and $n = \infty$ yields

(5.9) $$\sum_{n=0}^{\infty} q^n (G-I)^n \binom{x}{n} = \frac{1}{1-q\mu}.$$

Applying the formulas

$$(G-I)^n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} G^k,$$

$$\binom{x}{n}\binom{n}{k} = \binom{x}{k}\binom{x-k}{n-k}$$

and

$$\sum_{n=k}^{\infty} (-1)^{n-k} \binom{x-k}{n-k} q^{n-k} = (1-q)^{x-k},$$

we obtain

(5.10) $$\sum_{k=0}^{\infty} \left(\frac{q}{1-q}\right)^k G^k \left[\binom{x}{k}(1-q)^x\right] = \frac{1}{1-q\mu}.$$

Since

$$(G^k f)(x) = E[f(x+S_k)], \qquad k = 0, 1, 2, \ldots,$$

formula (5.10) is the same as (5.2).

An operational calculus proof is (5.10) can be found in SHIU (1989b).

If the random variable X in formula (5.2) is degenerate, i.e., $X \equiv \mu$, then we have

(5.11) $$\sum_{n=0}^{\infty} \binom{x+\mu n}{n} [q(1-q)^{\mu-1}]^n = \frac{1}{(1-\mu q)(1-q)^x}.$$

This result is quite well known; it and its variants can be found in PÓLYA (1922, (7)), WHITTAKER and WATSON (1927, p. 133, example 3), RIORDAN (1968, p. 147), PÓLYA and SZEGÖ (1970, p. 126, problem 216), KNUTH (1973, problem 1.2.6.26), MELZAK (1973, p. 117, example 4), COMTET (1974, p. 153), HENRICI (1974, p. 121, problem 12), ROTA (1975, p. 56), ROMAN and ROTA (1978, p. 115) and HOFRI (1987, p. 34). The standard proof of formula (5.11) is by an application of the Lagrange series formula. The proof can readily be generalized to one for (5.2), as we shall show below. (Also see section 5 of SHIU (1989a)).

Let h be an analytic function and let

(5.12) $$z = b + wh(z).$$

By the implicit function theorem, there is a unique root $z = z(w)$ which reduces to b at $w = 0$. If f is an analytic function, then $f(z) = f(z(w))$ may be expressed as follows [RIORDAN (1968, p. 146), PÓLYA and SZEGÖ (1970, p. 125), GOULDEN and JACKSON (1983, p. 17)]:

$$\frac{f(z)}{1-wh'(z)} = \sum_{j=0}^{\infty} \frac{w^j}{j!} \left[\frac{d^j}{dy^j} [f(y)[h(y)]^j] \right]_{y=b}. \tag{5.13}$$

Now, consider $b = 1-q$,

$$f(y) = y^x$$

and

$$h(y) = E(y^X).$$

Then

$$[h(y)]^j = E(y^{S_j})$$

and

$$\frac{1}{j!}\frac{d^j}{dy^j}[f(y)[h(y)]^j] = E\left[\binom{S_j+x}{j} y^{S_j+x-j}\right]. \tag{5.14}$$

With $w = q$, the right-hand side of (5.13) is the same as the right-hand side of (5.2) and equation (5.12) becomes

$$z = (1-q)+qE(z^X).$$

Thus $z = 1$ and the left-hand side of (5.13) is identical to the left-hand side of (5.2).

6. REMARKS

(i) Consider formula (2.14). Since $X \geq 1$ by hypothesis, the number $\phi(0)$ is always bounded above by one as it should be. If $1 \leq q\mu$, then ruin is guaranteed; but this is ruled out by condition (1.1). It follows from (2.14) that

$$\psi(0) = \frac{q(\mu-1)}{1-q}. \tag{6.1}$$

However, GERBER'S (1988b) result is that

$$\psi(0) = q\mu.$$

This discrepancy exists because GERBER defines ruin to occur when the risk reserve $U(t)$ becomes nonpositive, while we consider the insurance company to be solvent even if its risk reserve is zero. An anonymous referee has kindly pointed out that our definition of ruin is equivalent to DUFRESNE'S (1988, section 3) and (2.14) is DUFRESNE'S formula (37).

(ii) GERBER (1988b) first obtained formula (5.2) and then derived a formula corresponding to (4.1). With these two formulas, he derived formulas corresponding to (2.14) and (2.13).

(iii) Formula (2.12) is a special case of the combinatorial identity

$$\sum_{k=0}^{r} \binom{r-k}{m} \binom{s+k}{n} = \binom{r+s+1}{m+n+1},$$

where m, n, r and s are nonnegative integers and $n \geq s$ [RIORDAN (1968, p. 35, problem 13), KNUTH (1973, p. 58), HOFRI (1987, p. 39, problem 2b)].

(iv) Formula (2.1) can written as

(6.2) $\phi(j+1)-\phi(j) = [q/(1-q)]\{\phi(j)-E[\phi(j+1-X)]\}, \quad j = 0, 1, 2, \ldots.$

Hence, for each positive integer k,

$$\begin{aligned} \phi(k)-\phi(0) &= [q/(1-q)]\{\phi(k)-E[\phi(k+1-X)]\} * 1_+(k-1) \\ &= [q/(1-q)]\{\phi(k) * [1_+(k-1)-P(k)]\}, \end{aligned} \tag{6.3}$$

which is reminiscent of a renewal equation in the compound Poisson model [(FELLER, 1971, (XI.7.2)), (SHIU, 1989a, (2.4))]. Let h denote the function

$$h(k) = [1_+(k-1)-P(k)]/(\mu-1), \qquad k = 0, \pm 1, \pm 2, \ldots.$$

It follows from (6.3) and (6.1) that, for all integers k,

$$\phi(k)-\phi(0)\, 1_+(k) = \psi(0)\,[\phi(k) * h(k)].$$

Define $H^{*n} = h^{*n} * 1_+$. Then

$$\phi(u) = \phi(0) \sum_{n=0}^{\infty} [\psi(0)]^n H^{*n}(u). \tag{6.4}$$

Formula (6.4) is analogous to a convolution series formula in the compound Poisson model; see SHIU (1988, (2.1); 1989a, (2.14)). Since $h(i) = 0$ for all $i \leq 0$, there are at most $u+1$ nonzero terms in the right-hand side of (6.4), i.e.,

$$\phi(u) = \phi(0) \sum_{n=0}^{u} [\psi(0)]^n H^{*n}(u). \tag{6.5}$$

As

$$\sum_{n=0}^{\infty} [\psi(0)]^n = 1/[1-\psi(0)] = 1/\phi(0),$$

we have, for each nonnegative integer u,

$$\psi(u) = [1-\psi(0)] \sum_{n=1}^{\infty} [\psi(0)]^n [1-H^{*n}(u)]. \tag{6.6}$$

Formula (6.6) has been derived by R. MICHEL and can be found in a forthcoming risk theory book by C. HIPP and R. MICHEL. Observe that, when $X \equiv 2$, $h(j) = \delta(j-1)$ and formula (3.4) immediately follows from (6.5). I thank C. HIPP for the information above.

ACKNOWLEDGMENT

Support from the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

REFERENCES

BURMAN, J. P. (1946) Sequential sampling formulae for a binomial population. *Journal of the Royal Statistical Society B* **8**, 98–103.

BROWN, A. L. and PAGE, A. (1970) *Elements of Functional Analysis.* Van Nostrand Reinhold, London.

COMTET, L. (1974) *Advanced Combinatorics: The Art of Finite and Infinite Expansions.* Reidel, Dordrecht.

DUFRESNE, F. (1988) Distributions stationnaires d'un système bonus-malus et probabilité et ruine. *ASTIN Bulletin* **18**, 31–46.

FELLER, W. (1971) *An Introduction to Probability Theory and Its Applications,* Vol. 2 (2nd edn). Wiley, New York.

GERBER, H. U. (1988a) Mathematical fun with ruin theory. *Insurance: Mathematics and Economics* **7**, 15–23.

GERBER, H. U. (1988b) Mathematical fun with the compound binomial process. *ASTIN Bulletin* **18**, 161–168.

GIRSHICK, M. A. (1946) Contributions to the theory of sequential analysis, II, III. *Annals of Mathematical Statistics* **17**, 282–298.

HENRICI, P. (1974) *Applied and Computational Complex Analysis,* Vol. 1: *Power Series – Integration – Conformal Mapping – Location of Zeros.* Wiley, New York.

HOFRI, M. (1987) *Probabilistic Analysis of Algorithms: On Computing Methodologies for Computer Algorithms Performance Evaluation.* Springer-Verlag, New York.

KNUTH, D. E. (1973) *The Art of Computer Programming,* Vol. 1: *Fundamental Algorithms* (2nd edn). Addison-Wesley, Reading, Massachusetts.

MELZAK, Z. A. (1973) *Companion to Concrete Mathematics: Mathematical Techniques and Various Applications.* Wiley, New York.

PÓLYA, G. (1922) Sur les séries entières dont la somme est une fonction algébrique. *L'Enseignement Mathématique* **22**, 38–47. Reprinted in *George Pólya: Collected Papers,* Vol. 1: *Singularities of Analytic Functions,* MIT Press, Cambridge, Massachussetts (1974), pp. 165–174.

PÓLYA, G. and SZEGÖ, G. (1970) *Aufgabe und Lehrsätze aus der Analysis,* Vol. I (4th edn). Springer-Verlag, Berlin.

PRABHU, N. U. (1965) *Queues and Inventories: A Study of Their Basic Stochastic Processes.* Wiley, New York.

RIESZ, F. and SZ.-NAGY, B. (1955) *Functional Analysis.* Ungar, New York. An English translation of *Leçons d'analyse fonctionelle* (deuxième édition), Akadémiai Kiadó, Budapest (1953).

RIORDAN, J. (1968) *Combinatorial Identities.* Wiley, New York. Reprinted with correction by Krieger, Huntington, New York (1979).

ROMAN, S. M. and ROTA, G.-C. (1978) The umbral calculus. *Advances in Mathematics* **27**, 95–188.

ROTA, G.-C. (1975) *Finite Operator Calculus*. Academic Press, New York.
SEAL, H. L. (1962) The random walk of a simple risk business. *ASTIN Bulletin* **4**, 19–28.
SEAL, H. L. (1969) *Stochastic Theory of a Risk Business*. Wiley, New York.
SHIU, E. S. W. (1988) Calculation of the probability of eventual ruin by Beekman's convolution series. *Insurance: Mathematics and Economics* **7**, 41–47.
SHIU, E. S. W. (1989a) Ruin probability by operational calculus. *Insurance: Mathematics and Economics*, forthcoming.
SHIU, E. S. W. (1989b) On Gerber's fun. *Scandinavian Actuarial Journal*, forthcoming.
WHITTAKER, E. T. and WATSON, G. N. (1927) *A Course of Modern Analysis* (4th edn). Cambridge University Press, London.

ELIAS S. W. SHIU
Department of Actuarial and Management Sciences, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

ON AN INTEGRAL EQUATION FOR DISCOUNTED COMPOUND – ANNUITY DISTRIBUTIONS

BY COLIN M. RAMSAY

Actuarial Science, University of Nebraska, Lincoln NE, USA 68588-0307, (402) 472-5823

ABSTRACT

We consider a risk generating claims for a period of N consecutive years (after which it expires), N being an integer valued random variable. Let X_k denote the total claims generated in the k^{th} year, $k \geq 1$. The X_k's are assumed to be independent and identically distributed random variables, and are paid at the end of the year. The aggregate discounted claims generated by the risk until it expires is defined as $S_N(v) = \Sigma_{k=1}^{N} v^k X_k$, where v is the discount factor. An integral equation similar to that given by PANJER (1981) is developed for the *pdf* of $S_N(v)$. This is accomplished by assuming that N belongs to a new class of discrete distributions called annuity distributions. The probabilities in annuity distributions satisfy the following recursion:

$$p_n = p_{n-1}\left(a + \frac{b}{a_n}\right), \qquad \text{for} \qquad n = 1, 2, \ldots,$$

where a_n is the present value of an n-year immediate annuity.

KEYWORDS

Annuity distributions; integral equation; aggregate discounted claims.

1. INTRODUCTION

A major problem in mathematical risk theory is the evaluation of the distribution of the aggregate claims occuring in a fixed time period. This is because the aggregate claims is usually the sum of a random number of claims. If Y_k is the size of the k^{th} claim and N is the number of claims in this time period, then the aggregate claims S is given by

$$S = \sum_{k=1}^{N} Y_k. \tag{1}$$

The Y_k's are usually assumed to be independent and identically distributed (*iid*) with common cummulative distribution function (*cdf*) $F(y)$. If the n-fold convolution of $F(y)$ with itself is given by

$$F_n(y) = \int_0^y F_{n-1}(y-z)\,dF(z), \qquad n = 1, 2, \ldots,$$

with $F_0(y) = 1$, for $y \geq 0$, and the non-defective claim number distribution is

$$p_n = \Pr[N = n],$$

for $n = 0, 1, \ldots$, then the *cdf* of S is

$$(2) \qquad G(y) = \sum_{n=0}^{\infty} p_n F_n(y).$$

Unfortunately, explicit expressions for $F_n(y)$ are usually not available, so the equation (2) is generally not very useful. Approximations for $G(y)$ are thus needed.

In order to facilitate the easy evaluation of $G(y)$ in equation (2), PANJER (1981), and SUNDT and JEWELL (1981) provided a family of claim number distributions which yielded an integral equation for the *pdf* of S when the Y_k's are absolutely continuous random variables. The random variable N must have probabilities satisfying the recursion

$$(3) \qquad p_n = p_{n-1}\left(a + \frac{b}{n}\right)$$

where a and b are constants depending on the length of the time period. This family includes the geometric, Poisson, binomial, negative binomial, logarithmic series, and the so-called extended truncated negative binomial distribution. See WILLMOT (1988) for details. PANJER (1981) proved that if p_n satisfies equation (3), then $g(y)$, the *pdf* of S, satisfies the following integral equation for $y > 0$:

$$(4) \qquad g(y) = p_1 f(y) + \int_0^y \left(a + \frac{bz}{y}\right) f(z)\, g(y-z)\, dz.$$

This integral equation can be solved numerically; see STRÖTER (1985).

Recall that S is defined as the aggregate claims over a fixed time period. If this time period T is large, i.e., extending over several years, then it many be prudent to include an interest discount factor to obtain the present value of these claims. Let T_k be the random time at which the claim Y_k occurs, and $N(T)$ be the number of claims over T years, T a positive integer. The aggregate discounted claims, denoted by $S_T^*(v)$, will be given by

$$(5) \qquad S_T^*(v) = \sum_{k=1}^{N(T)} v^{T_k} Y_k$$

where $v = 1/(1+i)$ and i is the constant annual rate of interest. Comparing equations (1) and (5), it is clear that $S_T^*(v)$ is a more complicated random variable than S, and hence will have a more complicated *cdf*. $S_T^*(v)$ can be simplified by making the traditional actuarial assumption that claims are paid at the end of the year in which they occur. This means that equation (5) reduces to

$$S_T(v) = \sum_{k=1}^{T} v^k X_k \tag{6}$$

where X_k is the aggregate claims generated in year k. We assume that the number of claims occuring during each year is an *iid* sequence, implying that the X_k's are also *iid*.

The important observation to note here is that $S_T(v)$ is now the sum of T (a fixed number) of random variables X_k. Thus we have seen that the traditional model studied by PANJER and SUNDT and JEWELL can be adapted to include an interest factor. However an expression for the *pdf* of $S_T^*(v)$ will not be similar to equation (4) when the probabilities of $N(T)$ satisfy equation (3). We will see that by making T random, it is possible that $S_T(v)$ can be extended to yield a *pdf* which satisfies an integral equation similar to (4).

2. THE MAIN RESULTS

The inclusion of interest and/or inflation factors in risk theoretic models have appeared in the literature mainly in the context of the calculation of ruin probabilities; see, for example, WATERS (1983), BOOGAERTS and CRIJNS (1987), and GARRIDO (1988) and references therein. The limiting distributions of discounted processes have been studied by GERBER (1971), and BOOGAERT, HAEZENDONCK and DELBAEN (1988). However, there has been no work in the literature on integral equations similar to that of PANJER (1981) for aggregate discounted claims.

Consider a risk that can produce either no claim or it produces a sequence of *iid* positive claims that are paid at the end of the year in which they occured. Such risks are pertinent to health insurance, dental insurance, etc. The sequence of claims will run for N years, starting from year 1 until year N, after which no further claims are produced. N is an integer valued non-negative random variable. The total claims produced in the k^{th} year is $X_k > 0$, $k = 1, 2, \ldots$. If interest is at rate i annually, the aggregate discounted claims will be given by $S_N(v)$ where

$$S_N(v) = \sum_{k=1}^{N} v^k X_k \tag{7}$$

Notice the difference between equations (6) and (7), the constant T is now replaced by the random variable N. These equations clearly have different interpretations.

In order to develop an integral equation for the *pdf* of $S_N(v)$, we will introduce a new family of claim number distributions for N, called annuity distributions, with probabilities p_n satisfying the following difference equation:

$$(8) \qquad p_n = p_{n-1}\left(a + \frac{b}{a_n}\right), \qquad \text{for} \qquad n = 1, 2, \dots,$$

where a_n is the present value of an n-year immediate annuity at interest rate i, i.e.,

$$(9) \qquad a_n = \frac{(1-v^n)}{i}.$$

As before, $p_n = \Pr[N = n]$.

Let $P(z)$ be the probability generating function of N, i.e.,

$$P(z) = \sum_{n=0}^{\infty} p_n z^n, \qquad \text{for} \qquad -1 \leq z \leq 1 .$$

It can easily be proven that

$$E[S_N(v)] = \frac{\mu(1-P(v))}{i}$$

and

$$\begin{aligned} \operatorname{Var}[S_N(v)] &= E[\operatorname{Var}[S_N(v) \mid N]] + \operatorname{Var}[E[S_N(v) \mid N]] \\ &= \frac{\sigma^2 v^2}{1-v^2}\left[1 - P(v^2) + \left(\frac{\mu}{i}\right)^2 [P(v^2)] - [P(v)]^2\right] \end{aligned}$$

where $\mu = E[X_k]$ and $\sigma^2 = \operatorname{Var}[X_k]$.

From equation (7) we condition on $\{N = n\}$ and define $S_n(v)$ as

$$S_n(v) = \sum_{k=1}^{n} v^k X_k, \qquad n = 1, 2, \dots .$$

Note that, because the X_k's are *iid*, $S_n(v)$ has, for each non-negative integer m, the same distribution as

$$S_n(v) = \sum_{k=1}^{n} v^k X_{m+k} .$$

Therefore, since

$$S_n(v) = vX_1 + v \sum_{k=1}^{n-1} v^k X_{k+1} ,$$

$S_n(v)$ is seen to have the same distribution as $vX_1 + vS_{n-1}(v)$. Thus if $f_n(x)$ is the probability distribution function of $S_n(v)$, then the following convolution relationships will exist:

$$f_1(x) = f\left(\frac{x}{v}\right),$$

$$f_n(x) = \int_0^x f_{n-1}\left(\frac{x-y}{v}\right) f\left(\frac{y}{v}\right) dy \tag{10}$$

for $n = 2, 3, \ldots$ and $f(x)$ is the *pdf* of the X_k's.

Before deriving the integral equation for the *pdf* of $S_N(v)$, the following lemma is needed:

LEMMA 1. If X_k, $k = 1, 2, \ldots, n$ are *iid* random variables with finite mean, and the constants w_k are positive weights, let

$$Z_n = \sum_{k=1}^{n} w_k X_k \quad \text{and} \quad W_n = \sum_{k=1}^{n} w_k,$$

then for $k \in \{1, 2, \ldots, n\}$ and $n = 1, 2, \ldots$

$$E[X_k \mid Z_n = x] = \frac{x}{W_n}. \tag{11}$$

PROOF: By the symmetry of *iid* random variables and the fact that the weights are positive constants,

$$E[w_k X_k \mid Z_n = x] \propto w_k x.$$

Let π be the constant of proportionality. Summing both sides of the above expression yields

$$x = \pi W_n x,$$

i.e.,

$$\pi = \frac{1}{W_n}.$$

So

$$E[w_k X_k \mid Z_n = x] = \frac{w_k x}{W_n}$$

and equation (11) follows.

Q.E.D.

Consider the case where $w_k = v^k$ and $W_n = a_n$, then

$$\begin{aligned} E[X_1 \mid S_{n+1}(v) = x] &= \frac{x}{a_{n+1}} \\ &= \frac{1}{f_{n+1}(x)} \int_0^x \frac{y}{v} f_n\left(\frac{x-y}{v}\right) f\left(\frac{y}{v}\right) dy. \end{aligned} \tag{12}$$

We are now able to establish the main result of this paper.

THEOREM 1. Let $S_n(v)$ be defined as in equation (7) with $pdf\, g(x)$ for $x > 0$. If N has its probabilities satisfying the recursion in equation (8) and $\Sigma_{n=0}^{\infty} p_n = 1$, then for $x > 0$,

$$g(x) = p_1 f(x/v) + \int_0^x \left(a + \frac{by}{vx}\right) g\left(\frac{x-y}{v}\right) f(y/v)\, dy \tag{13}$$

with $\Pr[S_N(v) = 0] = p_0$.

PROOF: Since the X_k's are positive, $S_N(v) = 0$ if and only if $N = 0$. So $\Pr[S_N(v) = 0] = p_0$. For $x > 0$,

$$\begin{aligned}
g(x) &= \sum_{n=1}^{\infty} p_n f_n(x) \\
&= p_1 f_1(x) + \sum_{n=1}^{\infty} p_{n+1} f_{n+1}(x) \\
&= p_1 f(x/v) + \sum_{n=1}^{\infty} p_n \left(a + \frac{b}{a_{n+1}}\right) f_{n+1}(x) \\
&= p_1 f(x/v) + \sum_{n=1}^{\infty} a p_n \int_0^x f_n\left(\frac{x-y}{v}\right) f(y/v)\, dy + \\
&\quad + \sum_{n=1}^{\infty} p_n \frac{b}{a_{n+1}} f_{n+1}(x) \\
&= p_1 f(x/v) + \int_0^x a g\left(\frac{x-y}{v}\right) f(y/v)\, dy + \\
&\quad + \sum_{n=1}^{\infty} p_n \int_0^x \frac{by}{vx} f_n\left(\frac{x-y}{v}\right) f(y/v)\, dy \\
&= p_1 f(x/v) + \int_0^x \left(a + \frac{by}{vx}\right) g\left(\frac{x-y}{v}\right) f(y/v)\, dy
\end{aligned}$$

Q.E.D.

A similar result can be established if we assume that claims are subject to inflation at rate r and there is no interest. This can be accomplished by defining $w_k = (1+r)^k$, and using a new family of discrete claim number distributions with

$$p_n = p_{n-1}\left(a + \frac{b}{\ddot{s}_n}\right), \qquad \text{for} \qquad n = 1, 2, \ldots, \tag{14}$$

where

$$\ddot{s}_n = \sum_{k=1}^{n} (1+r)^k . \tag{15}$$

In this case

$$E[X_k \mid S_n(1+r) = x] = \frac{x}{\ddot{s}_n} . \tag{16}$$

The resulting integral equation is

$$g(x) = p_1 f(x/(1+r)) + \int_0^x \left(a + \frac{by}{(1+r)x}\right) g\left(\frac{x-y}{(1+r)}\right) f(y/(1+r))\, dy. \tag{17}$$

Note that in equation (13), for $0 < v < 1$, the argument of $g(.)$ in the integrand will exceed x, so $g(x)$ will depend on values of its argument between x and x/v. This will pose problems for obtaining numerical solutions. This problem does not arise in equation (17).

3. ANNUITY DISTRIBUTIONS

Equations (8) and (14) represent two new types of claim number distributions. However, they can be viewed as belonging to the same family of discrete annuity distributions because both equations can be written in the form:

$$p_n = p_{n-1}\left(a + \frac{b}{a(n, \delta)}\right), \qquad \text{for} \qquad n = 1, 2, \dots, \tag{18}$$

where

$$a(n, \delta) = \sum_{k=1}^{n} e^{k\delta}, \qquad -\infty < \delta < \infty .$$

Here $\delta < 0$ can be viewed as the force of interest while $\delta > 0$ can be viewed as the force of inflation. This implies that from equation (9) and (15)

$$a(n, \delta) = \begin{cases} a_n & \text{if} \quad \delta < 0, \\ n & \text{if} \quad \delta = 0, \\ \ddot{s}_n & \text{if} \quad \delta > 0. \end{cases} \tag{19}$$

Thus the family of discrete distributions as described in equation (3) is a special case of the annuity distribution with $\delta = 0$.

For a non-defective annuity distribution to exist, its probabilties must sum to one, implying that

$$R(a, b, \delta) = 1 + \sum_{n=1}^{\infty} \prod_{k=1}^{n} \left(a + \frac{b}{a(k, \delta)}\right) \tag{20}$$

must coverge. There are several tests that can be used to check the convergence of $R(a, b, \delta)$, see MALIK (1984) or WILLMOT (1988). For example, the ratio-test ensures convergence if

$$\lim_{n\to\infty}\left(a+\frac{b}{a(n,\delta)}\right)=L<1\,.$$

Once $R(a, b, \delta)$ exists, the p_n's will be given by

$$(21)\qquad p_n=\begin{cases}\dfrac{1}{R(a,b,\delta)} & \text{if}\quad n=0;\\[2ex] p_0\displaystyle\prod_{k=1}^{n}\left(a+\frac{b}{a(k,\delta)}\right) & \text{if}\quad n=1,2,3,\ldots\end{cases}$$

For given a and b that ensures the convergence of $R(a, b, \delta)$, one can easily evaluate the p_n's and the moments of the distribution. Unfortunately, closed form expressions are not easily obtainable these distributions, except of course when $\delta = 0$.

Further research is needed in the distributional properties of annuity distributions, the tail thickness, and the estimation of the parameters a and b. It will also be instructive to compare the various members of the family when $\delta = 0$ to those with the same parameters a and b but with $\delta \neq 0$. One would expect that the tails of these comparable distributions to become thicker as δ decreases.

REFERENCES

BOOGAERTS, P. and CRIJNS, V. (1987) Upper bounds on ruin probabilities in case of negative loadings and positive interest rates. *Insurance: Mathematics and Economics* **6**, 221–232.

BOOGAERT, P., HAEZENDONCK, J. and DELBAEN, F. (1988). Limit theorems for the present value of the surplus of an insurance portfolio. *Insurance: Mathematics and Economics* **7**, 131–138.

GARRIDO, J. (1988) Diffusion premiums for claim severities subject to inflation. *Insurance: Mathematics and Economics* **7**, 123–129.

GERBER, H. (1971) The discounted central limit theorem and its Berry-Esseen analogue. *Annals of Mathematical Statistics*, Vol. 42, **1**, 389–392.

MALIK, S. (1984) *Introduction to convergence*. Halstead Press, New York.

PANJER, H. (1981) Recursive evaluation of a family of compound distributions. *ASTIN-Bulletin* **12**, 22–26.

STRÖTER, B. (1985) The numerical evaluation of the aggregate claim density function via integral equations. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **17**, 1–14.

SUNDT, B. and JEWELL, W. (1981) Further results on recursive evaluation of compound distributions. *ASTIN Bulletin* **12**, 27–39.

WATERS, H. (1983) Probability of ruin for a risk process with claim cost inflation. *Scandinavian Actuarial Journal* **66**, 148–164.

WILLMOT, G. (1988) Sundt and Jewell's family of discrete distributions. *ASTIN Bulletin* **18**, 17–29.

WORKSHOP

A GENERALIZATION OF AUTOMOBILE INSURANCE RATING MODELS: THE NEGATIVE BINOMIAL DISTRIBUTION WITH A REGRESSION COMPONENT

BY GEORGES DIONNE and CHARLES VANASSE

Université de Montréal, Canada *

ABSTRACT

The objective of this paper is to provide an extension of well-known models of tarification in automobile insurance. The analysis begins by introducing a regression component in the Poisson model in order to use all available information in the estimation of the distribution. In a second step, a random variable is included in the regression component of the Poisson model and a negative binomial model with a regression component is derived. We then present our main contribution by proposing a bonus-malus system which integrates a priori and a posteriori information on an *individual basis*. We show how net premium tables can be derived from the model. Examples of tables are presented.

KEYWORDS

Multivariate automobile insurance rating; Poisson model; negative binomial model; regression component; net premium tables; Bayes analysis; maximum likelihood method.

INTRODUCTION

The objective of this paper is to provide an extension of well known models of tarification in automobile insurance. Two types of tarification are presented in the literature:

1) a priori models that select tariff variables, determine tariff classes and estimate premiums (see VAN EEGHEN et al. (1983) for a good survey of these models);
2) a posteriori models or bonus-malus systems that adjust individual premiums according to accident history of the insured (see FERREIRA (1974), LEMAIRE (1985, 1988) and VAN EEGHEN et al. (1983) for detailed discussions of these models).

* Centre de recherche sur les transports and Department of Economics. This research was financed by SSHRC Canada. Comments on a previous version of the paper by ALOIS GISLER, CHRISTIAN GOURIEROUX, STUART KLUGMAN, JEAN LEMAIRE, DAVID SCOTT and two anonymous referees were very useful.

This study focuses on the selection of tariff variables using multivariate regression models and on the construction of insurance tables that integrates a priori and a posteriori information on an *individual basis*. Our contribution differs from the recent articles in credibility theory where geometric weights were introduced (NEUHAUS (1988), SUNDT (1987, 1988)). In particular, SUNDT (1987) uses an additive regression model in a multiplicative tariff whereas our nonlinear regression model reflects the multiplicative tariff structure.

The analysis begins by introducing a regression component in both the Poisson and the negative binomial models in order to use all available information in the estimation of accident distribution. We first show how the univariate Poisson model can be extended in order to estimate different individual risks (or expected number of accidents) as a function of a vector of individual characteristics. At this stage of the analysis, there is no random variable in the regression component of the model. As for the univariate Poisson model, the randomness of the extended model comes from the distribution of accidents.

In a second step, a random variable is introduced in the regression component of the Poisson model and a negative binomial model with a regression component is derived. We then present our main contribution by proposing a bonus-malus system which integrates explicitly a priori and a posteriori information on an individual basis. Net premium tables are derived and examples of tables are presented. The parameters in the regression component of both the Poisson and the negative binomial models were estimated by the maximum likelihood method.

1. The Basic Model

1.a. *Statistical Analysis*

The Poisson distribution is often used for the description of random and independent events such as automobile accidents. Indeed, under well known assumptions, the distribution of the number of accidents during a given period can be written as

$$\text{pr}\,(Y_i = y) = \frac{e^{-\lambda}\lambda^y}{y!} \tag{1}$$

where y is the realization of the random variable Y_i for agent i in a given period and λ is the Poisson parameter which can be estimated by the maximum likelihood method or the method of moments. Empirical analyses usually reject the univariate Poisson model.

Implicitly, (1) assumes that all the agents have the same claim frequency. A more general model allows parameter λ to vary among individuals. If we assume that this parameter is a random variable and follows a gamma distribution with parameters a and $1/b$ (GREENWOOD and YULE (1920), BICHSEL (1964), SEAL (1969)), the distribution of the number of accidents during a given period becomes

$$\text{pr}\,(Y_i = y) = \frac{\Gamma(y+a)}{y!\,\Gamma(a)} \frac{(1/b)^a}{(1+1/b)^{y+a}} \tag{2}$$

which corresponds to a negative binomial distribution with $E(Y_i) = \bar{\lambda}$ and $\text{Var}\,(Y_i) = \bar{\lambda}\left[1 + \frac{\bar{\lambda}}{a}\right]$, where $\bar{\lambda} = ab$.

Again, the parameters a and $(1/b)$ can be estimated by the method of moments or by the maximum likelihood method.

1.b. *Optimal Bonus Malus Rule*

An optimal bonus malus rule will give the best estimator of an individual's expected number of accidents at time $(t+1)$ given the available information for the first t periods $(Y_i^1, \ldots, Y_i^t)$. Let us denote this estimator as $\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t)$.

One can show that the value of the Bayes' estimator (i.e. a posteriori mathematical expectation of λ) of the true expected number of accidents for individual i is given by

$$\hat{\lambda}_i^{t+1}(Y_i^1 \ldots Y_i^t) = \int_0^\infty \lambda f(\lambda/Y_i^1 \ldots Y_i^t)\, d\lambda\,. \tag{3}$$

Applying the negative binomial distribution, the a posteriori distribution of λ is a gamma distribution with probability density function

$$f(\lambda/Y_i^1 \ldots Y_i^t) = \frac{(1/b+t)^{a+\bar{Y}_i}\, e^{-\lambda(1/b+t)}\, \lambda^{a+\bar{Y}_i-1}}{\Gamma(a+\bar{Y}_i)}, \tag{4}$$

where $\bar{Y}_i = \sum_{j=1}^{t} Y_i^j$.

Therefore, the Bayes' estimator of an individual's expected number of accidents at time $(t+1)$ is the mean of the a posteriori gamma distribution with parameters $(a+\bar{Y}_i)$ and $((1/b)+t)$:

$$\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t) = \frac{a+\bar{Y}_i}{(1/b)+t} = \bar{\lambda}\left[\frac{a+\bar{Y}_i}{a+t\bar{\lambda}}\right]. \tag{5}$$

Actuarial net premium tables can then be calculated by using (5).

2. The Generalized Model

Since past experience cannot, in a short length of time, generate all the statistical information that permits fair insurance tarification, many insurers use both *a priori* and *a posteriori* tarification systems. A priori classification is based on significant variables that are easy to observe, namely, age, sex, type of driver's license, place of residence, type of car, etc. A posteriori information is then used to complete a priori classification. However, when both steps of the analysis are not adequately integrated into a single model, inconsistencies may be produced.

In practice, often linear regression models by applying a standard method out of a statistical package are used for the a priori classification of risks. These standard models often assume a normal distribution. But any model based on a continuous distribution is not a natural approach for count data characterized by many "zero accident" observations and by the absence of negative observations. Moreover, the resulting estimators obtained from these standard models often allow for negative predicted numbers of accidents. Regression results from count data models are more appropriate for a priori classification of risks.

A second criticism is linked to the fact that *univariate* (without regression component) statistical models are used in the Bayesian determination of the individual insurance premiums. Consequently, insurance premiums are function merely of time and of the past number of accidents. The premiums do not vary *simultaneously* with other variables that affect accident distribution. The most interesting example is the age variable. Let us suppose, for a moment, that age has a significant negative effect on the expected number of accidents. This implies that insurance premiums should decrease with age. Premium tables derived from univariate models do not allow for a variation of age, even if they are a function of time. However, a general model with a regression component would be able to determine the specific effect of age when the variable is statistically significant.

Finally, the third criticism concerns the coherency of the two-stage procedure using different models in order to estimate the same distribution of accidents.

In the following section we will introduce a methodology which responds adequately to the three criticisms. First, count data models will be proposed to estimate the individual's accident distribution. The main advantage of the count data models over the standard linear regression models lies in the fact that the dependent variable is a count variable restricted to non-negative values. Both the Poisson and the negative binomial models with a regression component will be discussed. Although the univariate Poisson model is usually rejected in empirical studies, it is still a good candidate when a regression component is introduced. Indeed, because the regression component contains many individual variables, the estimation of the individual expected number of accidents by the Poisson regression model can be statistically acceptable since it allows for heterogeneity among individuals. However, when the available information is not sufficient, using a Poisson model introduces an error of specification and a more general model should be considered. Second, we will generalize the optimal bonus-malus system by introducing all information from the regression into the calculation of premium tables. These tables will take account of time, accident record and the individual characteristics.

2.a. *Statistical Analysis*

Let us begin with the Poisson model. As in the preceding section, the random variables Y_i are independent. In the extended model, however, λ may vary between individuals. Let us denote by λ_i the expected number of accidents corresponding to individuals of type i. This expected number is determined by k exogenous variables or characteristics $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ which represent different a priori classification variables. We can write

$$\lambda_i = \exp(x_i\beta) \tag{6}$$

where β is a vector of coefficients $(k \times 1)$. (6) implies the non-negativity of λ_i.

The probability specification becomes

$$\Pr(Y_i = y) = \frac{e^{-\exp(x_i\beta)}(\exp(x_i\beta))^y}{y!}. \tag{7}$$

It is important to note that λ_i is not a random variable. The model assumes implicitly that the k exogenous variables provide enough information to obtain the appropriate values of the individual's probabilities. The β parameters can be estimated by the maximum likelihood method (see HAUSMAN, HALL and GRILICHES (1984) for an application to the patents — R & D relationship). Since the model is assumed to contain all the necessary information required to estimate the values of the λ_i, there is no room for a posteriori tarification in the extended Poisson model. Finally, it is easy to verify that (1) is a particular case of (7).

However, when the vector of explanatory variables does not contain all the significant information, a random variable has to be introduced into the regression component. Following GOURIEROUX MONFORT and TROGNON (1984), we can write

$$\lambda_i = \exp(x_i\beta + \varepsilon_i) \tag{8}$$

yielding a random λ_i. Equivalently, (8) can be rewritten as

$$\lambda_i = \exp(x_i\beta)\, u_i \tag{9}$$

where $u_i \equiv \exp(\varepsilon_i)$.

As for the univariate negative binomial model presented above, if we assume that u_i follows a gamma distribution with $E(u_i) = 1$ and $\operatorname{Var}(u_i) = 1/a$, the probability specification becomes

$$\operatorname{pr}(Y_i = y) = \frac{\Gamma(y+a)}{y!\,\Gamma(a)} \left[\frac{\exp(x_i\beta)}{a}\right]^{y} \left[1 + \frac{\exp(x_i\beta)}{a}\right]^{-(y+a)} \tag{10}$$

which is also a negative binomial distribution with parameters a and $\exp(x_i\beta)$. We will show later that the above parameterization does not affect the results if there is a constant term in the regression component.

$$\text{Then } E(Y_i) = \exp(x_i\beta) \text{ and } \operatorname{Var}(Y_i) = \exp(x_i\beta)\left[1 + \frac{\exp(x_i\beta)}{a}\right].$$

We observe that $\operatorname{Var}(Y_i)$ is a nonlinear increasing function of $E(Y_i)$. When the regression component is a constant c, $E(Y_i) = \exp(c) = \bar{\lambda}$ and

$$\operatorname{Var}(Y_i) = \bar{\lambda}\left[1 + \frac{\bar{\lambda}}{a}\right]$$

which correspond, respectively, to the mean and variance of the univariate negative binomial distribution.

DIONNE and VANASSE (1988) estimated the parameters of both the Poisson and negative binomial distributions with a regression component. A priori information was measured by variables such as age, sex, number of years with a driver's license, place of residence, driving restrictions, class of driver's license and number of days the driver's license was valid. The Poisson distribution with a regression component was rejected and the negative binomial distribution with a regression component yielded better results than the univariate negative binomial distribution (see Section 3 for more details).

An extension of the Bayesian analysis was then undertaken in order to integrate a priori and a posteriori tarifications on an individual basis.

2.b. *A Generalization of the Optimal Bonus Malus Rule*

Consider again an insured driver i with an experience over t periods; let Y_i^j represent the number of accidents in period j and x_i^j, the vector of the k characteristics observed at period j, that is $x_i^j = (x_{i1}^j, \ldots, x_{ik}^j)$. Let us further suppose that the true expected number of accidents of individual i at period j, $\lambda_i^j(u_i, x_i^j)$, is a function of both individual characteristics x_i^j and a random variable u_i. The insurer needs to calculate the best estimator of the true expected number of accidents at period $t+1$. Let $\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1})$ designate this estimator which is a function of past experience over the t periods and of known characteristics over the $t+1$ periods.

If we assume that the u_i are independent and identically distributed over time and that the insurer minimizes a quadratic loss function, one can show that the optimal estimator is equal to:

$$\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1})$$

$$= \int_0^{\infty} \lambda_i^{t+1}(u_i, x_i^{t+1}) f(\lambda_i^{t+1}/Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^t)\, d\lambda_i^{t+1}. \tag{11}$$

Applying the negative binomial distribution to the model, the Bayes' optimal estimator of the true expected number of accidents for individual i is:

$$\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1}) = \dot{\lambda}_i^{t+1} \left[\frac{a + \overline{Y}_i}{a + \overline{\lambda}_i} \right] \tag{12}$$

where $\lambda_i^j = \exp(x_i^j \beta) u_i \equiv (\dot{\lambda}_i^j) u_i$, $\overline{\lambda}_i = \sum_{j=1}^{t} \dot{\lambda}_i^j$ and $\overline{Y}_i = \sum_{i=1}^{t} Y_i^j$.

When $t = 0$, $\hat{\lambda}_i^1 = \dot{\lambda}_i^1 \equiv \exp(x_i^1 \beta)$ which implies that only a priori tarification is used in th first period. Moreover, when the regression component is limited to a constant c, one obtains:

$$\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t) = \overline{\lambda} \left[\frac{a + \overline{Y}_i}{a + t\overline{\lambda}} \right] \tag{13}$$

which is (5). This result is not affected by the parametrization of the gamma distribution.

It is important to emphasize here some characteristics of the model. In (13) only individual past accidents ($Y_i^1, \ldots, Y_i^t$) are taken into account in order to calculate the individual expected numbers of accidents over time. All the other parameters are population parameters. In (12), individual past accidents and characteristics are used simultaneously in the calculation of individual expected numbers of accidents over time. As we will show in the next section, premium tables that take into account the variations of both individual characteristics and accidents can now be obtained.

Two criteria define an optimal bonus-malus system which has to be fair for the policyholders and be financially balanced for the insurer. It is clear that the estimator proposed in (12) is fair since it allows the estimation of the individual risk as a function of both his characteristics and past experience. From the fact, that $E(E(A/B)) = E(A)$, it follows that the extended model is financially balanced:

$$E(\hat{\lambda}_i^{t+1}(Y_i^1, \dots, Y_i^t;\ x_i^1, \dots, x_i^{t+1})) = \lambda_i^{t+1} \text{ since } E(u_i) = 1 .$$

3. Examples of Premium Tables

As mentioned above, Dionne and Vanasse (1988) estimated the parameters of the Poisson regression model (β vector) and of the negative binomial regression model (β vector and the dispersion parameter a) by the maximum likelihood method. They used a sample of 19 013 individuals from the province of Québec. Many *a priori* variables were found significant. For example, the age and sex interaction variables were significant as well as classes of driver's licences for bus, truck, and taxi drivers. Even if the Poisson model gave similar results to those of the negative binomial model, it was shown (standard likelihood ratio test) that there was a gain in efficiency by using a model allowing for overdispersion of the data (where the variance is greater than the mean): the estimate of the dispersion parameter of the negative binomial regression $\hat{a}$ was statistically significant (asymptotic t-ratio of 3.91). The usual χ^2 test generated a similar conclusion. The latter results are summarized in Table 1:

TABLE 1

ESTIMATES OF POISSON AND NEGATIVE BINOMIAL DISTRIBUTIONS WITH A REGRESSION COMPONENT

Individual number of accidents in a given period	Observed numbers of individuals during 1982-1983	Predicted numbers of individuals for 1982-1983	
		Poisson *	Negative binomial *
0	17,784	17,747.81	17,786.39
1	1,139	1,201.59	1,131.05
2	79	60.56	86.21
3	9	2.88	8.18
4	2	.15	.98
5+	0	0	0
	19,013	$\chi^2 = 29.91$ $\chi^2_{2,.95} = 5.99$	$\chi^2 = 1.028$ $\chi^2_{1,.95} = 3.84$
		Log Likelihood = −4,661.57	Log Likelihood = −4,648.58

* The estimated β parameters are published in DIONNE-VANASSE (1988) and are available upon request. $\hat{a} = 1.47$ in the negative binomial model.

The univariate models were also estimated for the purpose of comparison. Table 2 presents the results. The estimated parameters of the univariate negative binomial model are $\hat{a} = .696080$ and $(1/\hat{b}) = 9.93580$ yielding $\hat{\lambda} = .0701$. One observes that $\hat{a} = 1.47$ in the multivariate model is larger than $\hat{a} = .6961$ in the univariate model. This result indicates that part of the variance is explained by the a priori variables in the multivariate model.

Using the estimated parameters of the univariate negative binomial distribution presented above, table 3 was formed by applying (14) where \$ 100 is the first period premium ($t = 0$):

$$\hat{P}_i^{t+1}(Y_i^1, \ldots, Y_i^t) = 100 \frac{(\hat{a} + \overline{Y}_i)}{(\hat{a} + t\hat{\lambda})} \tag{14}$$

In Table 3, we observe that only two variables may change the level of insurance premiums, i.e. time and the number of accumulated accidents. For example, an insured who had three accidents in the first period will pay a premium of \$ 462.43 in the next period, but if he had no accidents, he would have paid only \$ 90.86.

From (14) it is clear that no additional information can be obtained in order to differentiate an individual's risk. However, from (12), a more general pricing formula can be derived:

$$\hat{P}_i^{t+1}(Y_i^1 \ldots Y_i^t;\ x_i^1 \ldots x_i^{t+1}) = M\hat{\lambda}_i^{t+1} \left[\frac{\hat{a} + \overline{Y}_i}{\hat{a} + \hat{\overline{\lambda}}_i} \right] \tag{15}$$

TABLE 2

ESTIMATES OF UNIVARIATE POISSON AND NEGATIVE BINOMIAL DISTRIBUTIONS

Individual number of accidents in a given period	Observed numbers of individuals during 1982-1983	Predicted numbers of individuals for 1982-1983	
		Poisson (exp $\hat{c}$ = 0.0701)	Negative binomial ($\hat{a}$ = 0.6960; $1/\hat{b}$ = 9.9359)
0	17,784	17,726.60	17,785.28
1	1,139	1,241.86	1,132.05
2	79	43.50	88.79
3	9	1.02	7.21
4	2	0.02	.61
5+	0	0	0
	19,013	$\chi^2 = 133.06$ $\chi^2_{2,.95} = 5.99$	$\chi^2 = 2.21$ $\chi^2_{1,.95} = 3.84$
		Log Likelihood = −4950.28	Log Likelihood = −4916.78

TABLE 3
UNIVARIATE NEGATIVE BINOMIAL MODEL
$\hat{a} = .696080 \quad \hat{\lambda} = .0701$

t \ $\bar{Y}_i$	0	1	2	3	4
0	100.00				
1	90.86	221.38	351.91	462.43	612.96
2	83.24	202.83	322.42	442.01	561.60
3	76.81	187.15	297.50	407.84	518.19
4	71.30	173.72	276.15	378.58	481.00
5	66.52	162.09	257.66	353.23	448.80
6	62.35	151.92	241.49	331.06	420.63
7	58.67	142.95	227.23	311.52	395.80
8	55.40	134.98	214.56	294.15	373.73
9	52.47	127.85	203.23	278.61	353.99

where $\hat{\lambda}_i^{t+1} \equiv \exp(x_i^{t+1}\hat{\beta})$, $\hat{\lambda}_i \equiv \sum_{j=1}^{t} \exp(x_i^j\hat{\beta})$,

and M is such that

$$1/I \sum_{i=1}^{I} \hat{\lambda}_i^{t+1} M = \$\,100$$

when the total number of insureds is I.

This general pricing formula is function of time, the number of accumulated accidents and the individual's significant characteristics in the regression component. In consequence, tables can now be constructed more generally by using (15). First, it is easy to verify that each agent does not start with a premium of \$ 100. In Table 4, for example, a young driver begins with

TABLE 4
NEGATIVE BINOMIAL MODEL WITH A REGRESSION COMPONENT
Male, 18 years old in period 0, region 9, class 42

t \ $\bar{Y}_i$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	180.06	302.78	425.50	548.23	670.95
5	165.81	278.81	391.82	504.82	617.83
6	153.64	258.36	363.07	467.79	572.50
7	79.85	134.28	188.70	243.12	297.55
8	76.92	129.35	181.77	234.19	286.62
9	74.20	124.76	175.33	225.90	276.46

$ 280.89. Second, since the age variable is negatively significant in the estimated model, two factors, rather than one, have a negative effect on the individual's premiums (i.e. time and age). In Table 4, the premium is largely reduced when the driver reaches period seven at 25 years old (a very significant result in the empirical model).

For the purpose of comparison, Table 4 was normalized such that the agent starts with a premium of $ 100. The results are presented in table 5a. The effect of using a regression component is directly observed. Again the difference between the corresponding premiums in Table 3 and Table 5a come from two

TABLE 5a

TABLE 4 DIVIDED BY 2.8089

t \ $\bar{Y}_t$	0	1	2	3	4
0	100.00				
1	88.17	148.27	208.36	268.46	328.55
2	77.42	130.18	182.94	235.70	288.46
3	70.13	117.93	165.73	213.53	261.33
4	64.10	107.79	151.48	195.18	238.87
5	59.03	99.26	139.49	179.72	219.95
6	54.70	91.98	129.26	166.54	203.82
7	28.43	47.81	67.18	86.55	105.93
8	27.38	46.05	64.71	83.37	102.04
9	26.42	44.42	62.42	80.42	98.42

TABLE 5b

COMPARISON OF BASE PREMIUM AND BONUS-MALUS FACTOR COMPONENTS

	Univariate Model			Individual of Table 4		
	Base Premium	Bonus Malus Factor		Base Premium *	Bonus Malus Factor	
t \ $\bar{Y}_t$		0	1		0	1
0	100.00	1.0000		280.89	1.0000	
1	100.00	0.9086	2.2138	280.89	0.8817	1.4827
2	100.00	0.8324	2.0283	280.89	0.7742	1.3018
3	100.00	0.7681	1.8715	280.89	0.7013	1.1793
4	100.00	0.7130	1.7372	280.89	0.6410	1.0779
5	100.00	0.6652	1.6209	280.89	0.5903	0.9926
6	100.00	0.6235	1.5192	280.89	0.5470	0.9198
7	100.00	0.5867	1.4295	154.67	0.5163	0.8682
8	100.00	0.5540	1.3498	154.67	0.4973	0.8363
9	100.00	0.5247	1.2785	154.67	0.4797	0.8066

* To be compared with Table 5a, this column should be divided by 2.8089.

sources: the individual in Table 5a has particular a priori characteristics while all individuals are implicitly assumed identical in Table 3 and age is significant when the individual reaches period seven (25 years old). Finally, the above comparison shows that the Bonus-Malus factor is now a function of the individual's characteristics as suggested by (12). Table 5b separates the corresponding base premium and Bonus-Malus factor components of the total premiums in the first two columns of Table 3 and Table 4.

Moreover, when the insured modifies significant variables, new tables may be formed. In Table 4 the driver was in region # 9 (a risky region in Quebec) and had a standard driving license.

TABLE 6

NEGATIVE BINOMIAL MODEL WITH A REGRESSION COMPONENT
SAME INDIVUDUAL AS IN TABLE 4, MOVED TO MONTREAL IN PERIOD 4

t \ $\bar{Y}_i$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	119.65	201.19	282.73	364.28	445.82
5	113.18	190.32	267.45	344.59	421.73
6	107.38	180.56	253.74	326.92	400.11
7	56.98	95.81	134.65	173.48	212.32
8	55.47	93.28	131.08	168.89	206.69
9	54.04	90.87	127.70	164.53	201.36

TABLE 7

NEGATIVE BINOMIAL MODEL WITH A REGRESSION COMPONENT
SAME INVIDUAL AS IN TABLE 4, MOVED TO MONTREAL IN PERIOD 4,
CHANGED FOR CLASS 31 (TAXI) IN PERIOD 5

t \ $\bar{Y}_i$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	119.65	201.19	282.73	364.28	445.82
5	291.65	490.42	689.19	887.96	1086.73
6	256.00	430.48	604.95	779.42	953.90
7	127.26	213.99	300.72	387.45	474.18
8	119.97	201.73	283.49	365.25	447.02
9	113.47	190.80	268.13	345.47	422.80

Now if the individual moves from region # 9 to a less risky region (Montreal, for example) in period 4, the premiums then change (see Table 6).

Having two accidents, he now pays $ 282.73 in period 4 instead of $ 425.50. Finally, if the driver decides to become a Montreal taxi driver in period 5, the following results can be seen in Table 7.

Again, having two accidents, he now pays $ 689.19 in period 5 instead of $ 267.45.

CONCLUDING REMARKS

In this paper, we have proposed an extension of well-known models of tarification in automobile insurance. We have shown how a bonus-malus system, based only on a posteriori information, can be modified in order to take into account simultaneously a priori and a posteriori information on an individual basis. Consequently, we have integrated two well-known systems of tarification into a unified model and reduced some problems of consistencies. We have limited our analysis to the optimality of the model.

One line of research is the integration of accident severity into the general model even if the statistical results may be difficult to use for tarification (particularly in a fault system). Recent contributions have analyzed different types of distribution functions to be applied to the severity of losses (LEMAIRE (1985) for automobile accidents, CUMMINS et al. (1988) for fire losses, and HOGG and KLUGMAN (1984) for many other applications). Others have estimated the parameters of the total loss amount distribution (see SUNDT (1987) for example) or have included individuals' past experience in the regression component (see BOYER and DIONNE (1986) for example). However, to our knowledge, no study has ever considered the possibility of introducing the individual's characteristics and actions in a model that isolates the relationship between the occurence and the severity of accidents *on an individual basis.*

REFERENCES

BICHSEL, F. (1964) Erfahrungs-Tarifierung in der Motorfahrzeughaftpflichtversicherung. *Mitt. Verein. Schweiz. Versicherungs-Mathematiker* **64**, 119–130.

BOYER, M. and DIONNE, G. (1986) La tarification de l'assurance automobile et les incitations à la sécurité routière: une étude empirique. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* **122**, 293–322.

CUMMINS, J. D., DIONNE, G., MCDONALD, J. B. and PRITCHETT, B. M. (1988) Application of the GB2 Distribution in Modeling Insurance Loss Processes. Mimeo, Department of Insurance, University of Pennsylvania.

DIONNE, G and VANASSE, C. (1988) Automobile Insurance Ratemaking in the Presence of Asymmetrical Information. Mimeo, CRT and Economics Department, Université de Montréal.

FERREIRA, J. (1974) The Long-Term Effects of Merit Rating Plans on Individual Motorists. *Operations Research* **22**, 954–978.

GOURIEROUX, C., MONFORT, A. and TROGNON, A. (1984) Pseudo Maximum Likelihood Methods: Application to Poisson Models. *Econometrica* **52**, 701–720.

GREENWOOD, M. and TYLE, G. U. (1920) An Inquiry into the Nature of Frequency Distribution of Multiple Happenings. *Journal of Royal Statistical Society* **A 83**, 255–279.

HAUSMAN, J., HALL, B. H. and GRILICHES Z. (1984) Econometric Models for Count Data with an Application to the Patents — R & D Relationship. *Econometrica* **52**, 910–938.

HOGG, R. V. and KLUGMAN S. A. (1984) *Loss Distributions*. Wiley, New-York.

LEMAIRE, J. (1985) *Automobile Insurance: Actuarial Models*. Kluwer-Nighoff, Boston, 248 pages.

LEMAIRE, J. (1988) A Comparative Analysis of Most European and Japanese Bonus-Malus Systems. *Journal of Risk and Insurance* **LV**, 660–681.

NEUHAUS, W. (1988) A Bonus-malus System in Automobile Insurance. *Insurance: Mathematics & Economics* 7, 103–112.

SEAL, H. L. (1969) *Stochastic Theory of a Risk Business*. Wiley, New-York.

SUNDT, B. (1987) Two Credibility Regression Approaches for the Classification of Passenger Cars in a Multiplicative Tariff". *ASTIN Bulletin* **17**, 41-70.

SUNDT, B. (1988) Credibility Estimators With Geometric Weights. *Insurance: Mathematics & Economics* 7, 113–122.

VAN EEGHEN, J., GREUP, E. K. and NIJSSEN, J. A. (1983) *Surveys of Acturial Studies: Rate Making*. Research Department, Nationale — Nederlanden N.V., 138 pages.

GEORGES DIONNE

CRT, Université de Montréal, C.P. 6128, Succ. A, Montreal, H3C 3J7, Canada.

MUTUAL REINSURANCE AND HOMOGENEOUS LINEAR ESTIMATION

By WALTHER NEUHAUS

University of Oslo
and
Storebrand Insurance Company, Oslo, Norway

ABSTRACT

The technique of risk invariant linear estimation from NEUHAUS (1988) has been applied in the construction of a mutual quota share reinsurance pool between the subsidiary companies of the Storebrand Insurance Company, Oslo. The paper describes the construction of the reinsurance scheme.

1. INTRODUCTION

The Storebrand Insurance Company is the largest non-life insurer in Norway. Non-life business is written by four wholly owned stock companies, each covering a certain geographic area. The regional companies enjoy a large degree of autonomy, while certain areas, like tariffication and reinsurance, are managed centrally.

All but one of the regional companies are small, measured even by Norwegian standards. This makes their profitability subject to large fluctuations, even after deduction of external reinsurance. In 1987, the company top management issued a request to devise a way of stabilising the regional companies' profitability. The idea of additional reinsurance was launched at an early stage, and all the traditional forms of reinsurance were discussed. During the discussions a number of guidelines were formulated.

1. The reinsurance should give protection against large claims, as well as large claim numbers (typically caused by spells of bad weather).
2. No additional external reinsurance was to be bought.
3. The reinsurance should be fair, it should not take the accountability off the regional companies (other than correcting for "random" fluctuations).
4. The reinsurance should be very easy to administer.
5. Compulsory participation for the 4 regional companies.

Guidelines 1 and 4 quickly disqualified excess of loss reinsurance and surplus reinsurance. Guideline 3 disqualified stop-loss reinsurance. Left over was quota share reinsurance. The solution arrived at was a mutual quota-share pool, described briefly as follows.

a. Each regional company cedes a certain share of its business (premium and losses) to the pool. Business to be ceded is own account business, i.e. after deduction of external reinsurance.
b. The total losses ceded to the pool are redistributed amongst the participating companies in the same proportion as premium was ceded to the pool.
c. The premium ceded to the pool is returned in its entirety, thus leaving the regional companies' premium unaltered.

The arrangement described is essentially a loss pool, since only losses (not premium) are affected. A desirable side effect of this property is that the regional companies' expense ratio is left unchanged; thus eliminating the need for reinsurance commission.

The mutual quota share pool is a very traditional way of reinsurance, which does not necessarily make it a poor way of reinsurance. In the following chapter a mathematical model is given, within which the mutual quota share is optimal.

2. OPTIMAL REINSURANCE IN THE BÜHLMANN-STRAUB MODEL

Let us number the regional companies by $i = 1, \dots, I$. For company i, define P_i = premium for own account, S_i = losses for own account, $X_i = S_i/P_i$ = loss ratio for own account. Note that "for own account" in this context means business net of external reinsurance, but before application of the mutual quota-share treaty.

We make the assumptions of the BÜHLMANN-STRAUB model (BÜHLMANN & STRAUB, 1970). These assumptions are that there exists a latent parameter Θ_i so that

$$E(X_i \mid \Theta_i) = b(\Theta_i), \tag{2.1}$$

$$\operatorname{Var}(X_i \mid \Theta_i) = v(\Theta_i)/P_i, \tag{2.2}$$

where b and v are real-valued functions of Θ_i. It is then assumed that the parameters Θ_i are i.i.d. random variables, and that

$$E(b(\Theta_i)) = \beta, \tag{2.3}$$

$$E(v(\Theta_i)) = \phi, \tag{2.4}$$

$$\operatorname{Var}(b(\Theta_i)) = \lambda. \tag{2.5}$$

These assumptions obviously fit the problem to be solved very well. The function $b(\Theta_i)$ is interpreted as the underlying (long-run) loss ratio of company i, and the aim of the exercise is to estimate this quantity.

For fixed values of the parameters β, ϕ, λ, the best linear estimator of $b(\Theta_i)$ (with respect to mean squared error) is the credibility estimator

$$\bar{b}_i = z_i X_i + (1 - z_i)\beta, \tag{2.6}$$

where

$$z_i = P_i/(P_i + \kappa), \tag{2.7}$$

$$\kappa = \phi/\lambda. \tag{2.8}$$

To simplify notation, define

$$c_i = 1 - z_i, \tag{2.9}$$

and note the relation

$$P_i c_i = \kappa z_i. \tag{2.10}$$

For fixed values of ϕ, λ, and unknown β, the best linear unbiased estimator of $b(\Theta_i)$, based on $X_1, \ldots, X_I$, is

$$\bar{\bar{b}}_i = z_i X_i + (1 - z_i)\hat{\beta}, \tag{2.11}$$

where

$$\hat{\beta} = [\sum_j z_j]^{-1} \sum z_j X_j. \tag{2.12}$$

Proofs of the optimality of (2.6), (2.11) may be found in BÜHLMANN (1970).

A risk exchange between the I companies is given by the transformation

$$(S_1, \ldots, S_I) \to (\tilde{S}_1, \ldots, \tilde{S}_I) = (P_1 \bar{\bar{b}}_1, \ldots, P_I \bar{\bar{b}}_I). \tag{2.13}$$

This risk exchange is defined by replacing each company's loss ratio X_i with the estimate $\bar{\bar{b}}_i$. It is optimal in the sense of minimum mean squared error estimation of the "underlying loss ratio" $b(\Theta_i)$. That the risk exchange coincides with a mutual quota share treaty may be seen by

$$\begin{aligned} \tilde{S}_i = P_i \bar{\bar{b}}_i = P_i \{ z_i X_i + c_i \hat{\beta} \} &= z_i S_i + P_i c_i z^{-1} \sum_j z_j (S_j / P_j) \\ &= z_i S_i + \kappa z_i z^{-1} \sum_j c_j \kappa^{-1} S_j = z_i S_i + (z_i / z) \sum_j c_j S_j = z_i S_i + (z_i / z) S, \end{aligned} \tag{2.14}$$

where we have defined $z = \sum_j z_j$, $S = \sum_j c_j S_j$. The variable S is just the total losses ceded to the pool. The risk exchange (2.13) replaces the losses of company i with the sum of the retained share and a share of the pool, the share of the pool being z_i/z. To see that this share is equal to the proportion of premium ceded to the pool, note that $z_i/z = P_i c_i / \sum_j P_j c_j$.

A direct consequence of (2.14) is the identity

$$\sum_i \tilde{S}_i = \sum_i S_i, \tag{2.15}$$

which makes (2.13) a proper risk exchange in the sense of BÜHLMANN & JEWELL (1979). GISLER (1987) mentions the property (2.15); it ensures that no claims are "lost" when homogeneous credibility estimation is applied.

3. CHOICE OF MODEL

Let us consider one line of business. The risk exchange (2.13) is characterised by the value of "action parameter" κ, entering into the credibility factors z_i, see (2.7). Ideally one should use $\kappa = \phi/\lambda$, where ϕ, λ are the true variances. Since it is preposterous to try to separate empirically the variance components ϕ and λ from just 4 replications (companies), and since the author does not subscribe to subjectivism, we applied the minimax approach of NEUHAUS (1988), which is sketched in the sequel.

For a fixed $k > 0$, define the risk exchange $S \to \tilde{S}(k)$ by

$$\tilde{S}_i(k) = z_i(k) S_i + (z_i(k)/z(k)) S(k), \tag{3.1}$$

where $z_i(k) = P_i/(P_i + k)$, $z(k) = \sum_j z_j(k)$, $c_j(k) = 1 - z_j(k)$, $S(k) = \sum_j c_j(k) S_j$.

This risk exchange has obviously the same structure as (2.13), only κ is replaced by k. Let $\tilde{X}_i(k) = \tilde{S}_i(k)/P_i$ be the loss ratio after reinsurance.

The loss incurred by using k as action parameter is measured by the loss function

$$L(k, \phi, \lambda) = I^{-1} \sum_i E(\tilde{X}_i(k) - b(\Theta_i))^2, \tag{3.2}$$

the objective being to minimize (3.2). It can be shown that

$$L(k, \phi, \lambda) = I^{-1} [\phi \sum_{i,j} g_{ij}^2(k)/P_j + \lambda \sum_{i,j} (\delta_{ij} - g_{ij}(k))^2], \tag{3.3}$$

where we have defined for $1 \le i, j \le I$,

$$g_{ij}(k) = \delta_{ij} z_i(k) + c_i(k) z_j(k)/z(k). \tag{3.4}$$

Assume that data available are $P'_1, \ldots, P'_I$ and $X'_1, \ldots, X'_I$, representing premiums and loss ratios for (one or more) previous periods. Then one may estimate β by

$$\beta^* = \sum_i w'_i X'_i, \tag{3.5}$$

where $w'_i = P'_i / \sum_j P'_j$. The estimator β^* is the overall loss ratio for the period observed. The statistic

$$V^* = \sum_i w'_i (X'_i - \beta^*)^2 \tag{3.6}$$

has expectation

$$E(V^*) = \lambda \sum_i w'_i (1 - w'_i) + \phi \sum_i w'_i (1 - w'_i)/P'_i. \tag{3.7}$$

The total variance in the loss ratio is estimated by V^*.

As in NEUHAUS (1988), the parameter $k > 0$ may be chosen so that the risk exchange $S \to \tilde{S}(k)$ becomes an equaliser rule with respect to the parameter set

$$(3.8) \qquad \mathcal{N} = \{(\phi, \lambda) \mid V^* = \lambda \sum_i w_i'(1-w_i') + \phi \sum_i w_i'(1-w_i')/P_i'\},$$

i.e. $L(k, \phi, \lambda) =$ constant for $(\phi, \lambda) \in \mathcal{N}$.

The calculations needed to find k are similar to those given in NEUHAUS (1988). The reason for choosing an equaliser rule is that it will be a (restricted) minimax rule with respect to the parameter set $\mathcal{N}$. Note that $\{\kappa \mid (\phi, \lambda) \in \mathcal{N}\} = \langle 0, \infty \rangle$.

4. EXAMPLE

Consider the line "Small to medium commercial risk". Table 1 gives the relevant statistics for the year 1987. It is found that $\beta^* = 1.211$, $V^* = 0.102$,

$$\sum_i w_i'(1-w_i') = 0.595, \quad \sum_i w_i'(1-w_i')/P_i' = 0.021.$$

The value $k = 42$ makes the risk exchange an equaliser rule across the parameter set

$$(4.1) \qquad \mathcal{N} = \{(\phi, \lambda) \mid 0.102 = \lambda \cdot 0.595 + \phi \cdot 0.021\}.$$

The factors c_i (42) are displayed in the rightmost column of table 1. One sees that the large company should cede about one-third of its business to the pool, while the 3 small companies should cede about two-thirds of their business to the pool.

TABLE 1

STATISTICS FOR "SMALL TO MEDIUM COMMERCIAL RISK"

Company	P_i'	X_i'	w_i'	$w_i'(X_i'-\beta^*)^2$	$c_i(42)$
East	81.366	1.425	0.590	0.026	0.34
South	19.816	1.163	0.144	0.000	0.68
West	18.149	0.475	0.132	0.071	0.70
North	18.596	1.047	0.135	0.003	0.69
Total	137.927	$1.211 = \beta^*$		$0.102 = V^*$	

Figure 1 shows the square root of the different loss functions (3.2) dependent on the true κ, where it is assumed that $(\phi, \lambda) \in \mathcal{N}$, $\mathcal{N}$ given by (4.1). The square root is displayed because it is measured in the same scale as the estimand. The loss functions of three risk exchanges are displayed,

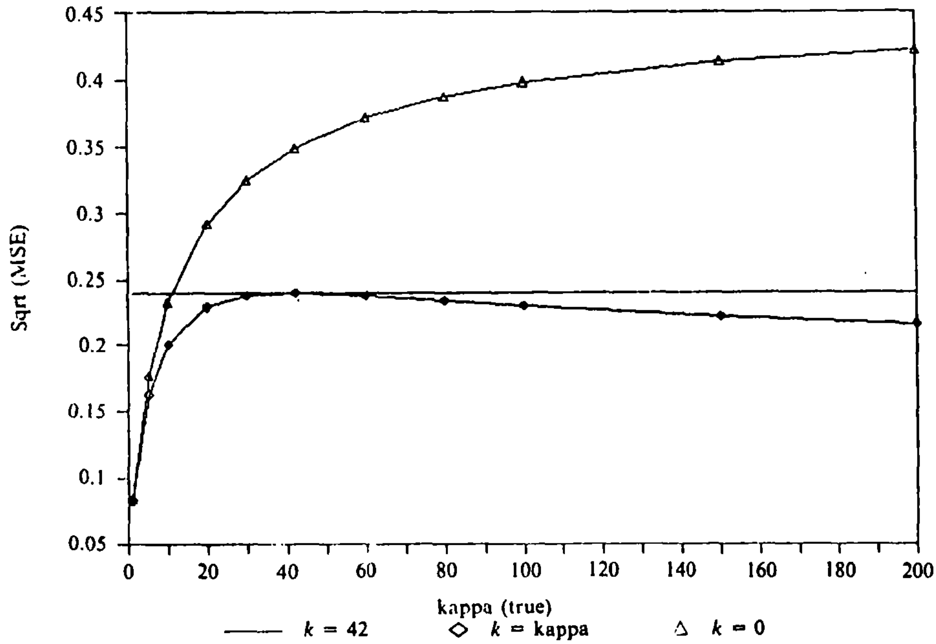

FIGURE 1

1. $k = 42$, giving the equaliser rule with respect to $\mathcal{K}$,
2. $k = \kappa$, giving the optimal risk exchange (2.13),
3. $k = 0$, meaning no reinsurance at all.

It is seen that the choice $k = 42$ gives a constant loss function across $\mathcal{K}$ and a considerable improvement over $k = 0$ (using $k = 0$ means judging each regional company only by its own loss ratio). The choice $k = \kappa$ is optimal, but the improvement it gives over $k = 42$ is very moderate over most of the parameter space displayed.

5. CONCLUDING COMMENTS

The aim of the paper has been to show that even a very traditional quota share pool reinsurance exhibits optimality properties when the shares are appropriately chosen.

Conceding that it is preposterous to separate empirically the variance components ϕ and λ, one may ask whether estimating $E(V^*)$ by V^* is any better; it is probably not, but the equaliser value of k does not depend on V^*, see NEUHAUS (1988).

The aim being to estimate the companies' loss ratio, should one include β^* in the estimator? Two arguments may be used against using β^*. The first argument is that a linear estimator using β^*, being the empirical counterpart of (2.6), would not have the desirable property (2.15), thus it does not give a proper risk exchange. The second argument goes as follows: The parameter β should not be fixed but random, $\beta = \beta(\psi)$, and (2.1)-(2.5) should be conditional relations, given ψ. This is a hierarchical credibility model; let $\xi = \text{Var}(\beta(\psi))$. The optimal inhomogeneous estimator of $b(\Theta_i)$ is then

$$\bar{b}_i = z_i(\kappa)X_i + (1 - z_i(\kappa)) \frac{\sum_j z_j(\kappa) X_j + \lambda \xi^{-1} E\,(\beta(\psi))}{\sum_j z_j(\kappa) + \lambda \xi^{-1}}, \tag{5.1}$$

see SUNDT (1979). The estimator (2.11) is obtained by letting $\xi \to \infty$, which in the Bayesian context means using a vague prior distribution for $\beta(\psi)$.

One may contend that it is unnecessary to establish a reinsurance treaty in order to assess the 4 companies' underlying loss ratio, when simple calculation of the homogeneous unbiased linear estimator would do the job. But, as experience has shown, the bottom line after mutual reinsurance is accepted by everyone as true expression of a company's profitability. On the other hand, an actuary telling company management that "*well, the loss ratio is 120, but my model says it should have been 105*" is doomed to fail. The reinsurance treaty makes the same statement more credible.

The loss function (3.2) is an unweighted average of the 4 companies' loss functions. This loss function reflects the objective of estimating the companies' underlying loss ratio, regardless of their premium volume. In an economic environment, the loss function should be weighted to reflect the fact that a unit of error in assessing the loss ratio is most serious for the large companies. It is possible to find an equaliser rule for weighted loss function, and probably the optimal k would not be changed much, see NEUHAUS (1988).

A separate risk exchange was set up for each line of business. The obvious reason was to spare the accounting staff for troublesome allocation problems. Stabilising each line of business also had the positive side effect of reducing regional demands for immediate remedial action (premium increases or discounts) in the wake of fluctuating loss ratios.

A more complicated model is needed if one wants to design a risk exchange for the I companies, which spans all lines of business. Probably the simplest model would be of the form

$$b_{ij} = \mu + \alpha_i + \beta_j, \tag{5.2}$$

where

b_{ij} is the underlying loss ratio for company i, line j,
μ is a fixed mean,
α_i is a random parameter characterising company i,
β_j is a random parameter characterising line j.

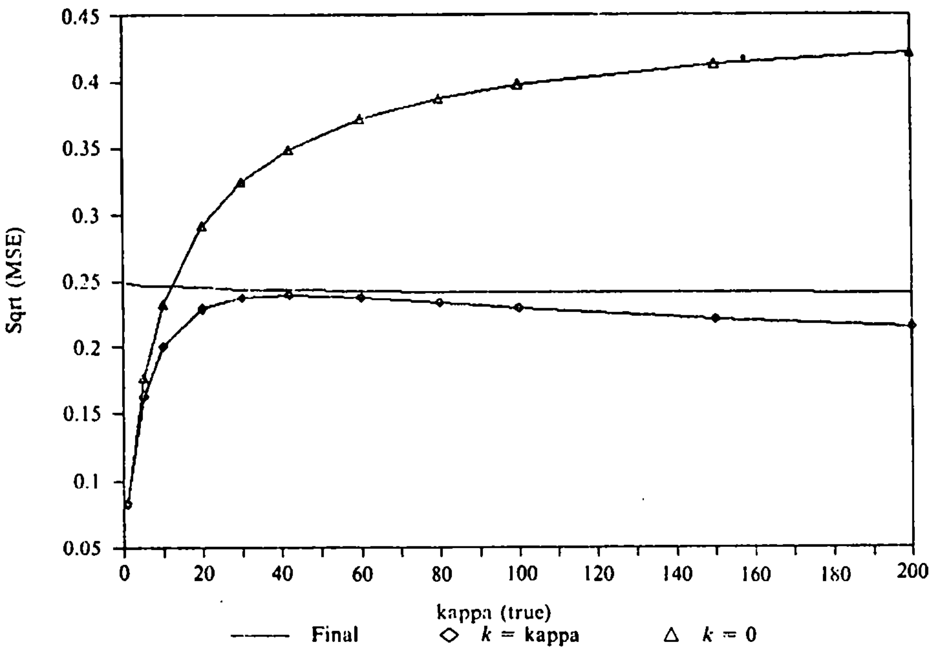

FIGURE 2

An estimator of the underlying loss ratio of company i is

$$\bar{b}_i = \bar{\mu} + \bar{\alpha}_i + [\sum_j p_{ij}]^{-1} \sum_j p_{ij}\bar{\beta}_j, \tag{5.3}$$

where $\bar{\mu}$, $\bar{\alpha}_i$, $\bar{\beta}_j$ are calculated by the credibility method described in BUCHANAN et al. (1989). Unfortunately, this method lacks the transparency which makes the estimators (2.6) and (2.11) so attractive.

A point of lengthy discussions was the choice of reinsured shares, although all but one company finally accepted the recommended shares. Figure 2 shows the loss function of the final scheme, compared with the optimal loss function ($k = \kappa$) and the loss function without reinsurance. We did not analyse whether the final scheme, being (very slightly) sub-optimal in the sense of minimaxing (3.2) over (4.1), has any other nice properties, such as Pareto-optimality. Here is a field for further analysis. Incidentally, if there is anything like empirical Pareto-optimality, the author has experienced it: Whatever modification of the scheme was suggested during the discussions, someone was certain to object.

REFERENCES

BUCHANAN, R. A., HEPPELL, I., NEUHAUS, W. (1989) A hierarchical credibility model. Paper presented to the XXI ASTIN Colloquium.

BÜHLMANN, H. (1970) Credibility procedures. *Proceedings of the sixth Berkeley Symposium on mathematical statistics and probability.*

BÜHLMANN, H. and JEWELL, W.S. (1979) Optimal risk exchanges. *ASTIN Bulletin* **10**, 243–262.
BÜHLMANN, H. and STRAUB, E. (1970) Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 111–133.
GISLER, A. (1987) Einige Bemerkungen zum hierarchischen Credibility-Modell. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 91–98.
NEUHAUS, W. (1988) Risk invariant linear estimation. *Scandinavian Actuarial Journal*, 157–190,
SUNDT, B. (1979) A hierarchical credibility regression model. *Scandinavian Actuarial Journal*, 107–114.

WALTHER NEUHAUS
Mathematical Institute, University of Oslo, P.O.Box 1053 Blindern,
N-0316 Oslo 3, Norway.

ACTUARIAL VACANCY

At the University of IOWA, IOWA CITY, there is an opening for an

Actuarial Professorship

Highly qualified candidates are sought for The Principal Financial Group Professor of Actuarial Science, beginning August 1990. The Principal Financial Group is supplementing university salary and expense allowance by at least $ 7,000/year. A reputation for excellence in teaching and actuarial research required. Selection process begins December 1st and continues until position filled.

Applications also sought for a less senior, tenure-track or tenured position, and from individuals desiring visiting appointments for one or two years. Specialties in either life or casualty acceptable, Ph. D. (or equivalent) is required for tenure-track appointment.

The University of Iowa has a history of excellence in both statistics and actuarial science. Iowa City is a clean, attractive community that offers many advantages. Send C.V. and three letters of evaluation to: Jim Broffitt, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA. Minorities and women are especially encouraged to apply. EO/AAE

GUIDELINES TO AUTHORS

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

   Hans Bühlmann,
   Mathematik, ETH-Zentrum,
   CH-8092 Zürich, Switzerland.

   D. Harry Reid,
   Eagle Star Insurance Company Ltd,
   Eagle Star House, 9 Aldgate High Street,
   London EC3N 1LD, United Kingdom.

   or to one of the Co-Editors:

   Alois Gisler,
   "Winterthur" Swiss Insurance Company,
   P.O. Box 357, CH-8401 Winterthur, Switzerland.

   David Wilkie
   Messrs. R. Watson & Sons
   Watson House, London Rd., Reigate, Surrey RH2 9PQ, United Kingdom.

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

2. Manuscripts should be typewritten on one side of the paper, double-spaced with wide margins. The basic elements of the journal's style have been agreed by the Editors and Publishers and should be clear from checking a recent issue of *ASTIN BULLETIN*. If variations are felt necessary they should be clearly indicated on the manuscript.

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

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

5. Footnotes should be avoided as far as possible.

6. Upon acceptance of a paper, any figures should be drawn in black ink on white paper in a form suitable for photographic reproduction with lettering of uniform size and sufficiently large to be legible when reduced to the final size.

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

   Examples:

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

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

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

Continued overleaf.

COMMITTEE OF ASTIN

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. 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.

11. 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.