A'STIN BULLETIN

A Journal of the International Actuarial Association

EDITOR:

CONTENTS:

Volume 15, No. 1

Peter ter Berg	P. Albrecht
Netherlands	An Evolutionary Credibility Model for Claim Numbers

EDITORIAL BOARD:

Björn Ajne Sweden

Hans Bühlmann Switzerland

Klaus Flemming FR Germany

Marc Goovaerts Belgium

Jacques Janssen Belgium

William S. Jewell USA

> Jukka Rantala Finland

> > Bjørn Sundt Norway

An Evolutionary Credibility Model for Claim Numbers	1
B. ZEHNWIRTH Linear Filtering and Recursive Credibility Estimation	19
R Norberg Unbayesed Credibility Revisited	37
Corrigendum	44
P. EMBRECHTS, M. MAEJIMA and J. TEUGELS Asymptotic Behaviour of Compound Distributions	45
L. CENTENO On Combining Quota-Share and Excess of Loss	49
Obituary	64
Editorial	66
Book Reviews	67



CHANGE OF EDITORSHIP

Please see page 66 and the Instructions to Authors on the inside back cover for details of new editorial arrangements which apply with immediate effect.

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 did result, 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 theoretic 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.

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 for ASTIN by Tieto Ltd, Bank House, 8a Hill Road, Clevedon, Avon BS21 7HH, England

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

The current subscription or back issue price per volume of 2 issues including postage is £16.00.

INDEX TO VOLUMES 1-13

The Cumulative Index to Volumes 1-13 is also published for ASTIN by Tieto Ltd at the above address and is available for the price of £5.00.

Copyright © 1985 Tieto Ltd

PRINTED IN GREAT BRITAIN by J W ARROWSMITH LTD, BRISTOL

AN EVOLUTIONARY CREDIBILITY MODEL FOR CLAIM NUMBERS

By Peter Albrecht*

University of Mannheim, FR Germany

KEY WORDS

Credibility, doubly stochastic Poisson sequences, weakly stationary sequences, generalized Pólya sequence.

1. INTRODUCTION

This paper considers a particular credibility model for the claim numbers N_1 , N_2, \ldots, N_n, \ldots of a single risk within a collective in successive periods $1, 2, \ldots, n, \ldots$ In the terminology of JEWELL (1975) the model is an evolutionary credibility model, which means that the underlying risk parameter Λ is allowed to vary in successive periods (the structure function is allowed to be time dependent). Evolutionary credibility models for claim *amounts* have been studied by BUHLMANN (1969, pp. 164-165), GERBER and JONES (1975), JEWELL (1975, 1976), TAYLOR (1975), SUNDT (1979, 1981, 1983) and KREMER (1982). Again in Jewell's terminology the considered model is on the other hand stationary, in the sense that the conditional distribution of N_i given the underlying risk parameter does not vary with *i*.

The computation of the credibility estimate of N_{n+1} involves the considerable labor of inverting an $n \times n$ covariance matrix (n is the number of observations). The above mentioned papers have therefore typically looked for model structures for which this inversion is unnecessary and instead a recursive formula for the credibility forecast can be obtained. Typically nth order stationary a priori sequences (e.g., ARMA (p, q)-processes) lead to an *n*th order recursive scheme. In this paper we impose the restriction that the conditional distribution of N_i is Poisson (which by the way leads to a model identical to the so called "doubly stochastic Poisson sequences" considered in the theory of stochastic point processes). What we gain is a recursive formula for the coefficients of the credibility estimate (not for the estimate itself!) in case of an arbitrary weakly stationary a priori sequence. In addition to this central result the estimation of the structural parameters is considered in this case and some more special models are analyzed. Among them are EARMA-processes (which are positive-valued stationary sequences possessing exponentially distributed marginals and the same autocorrelation structure as ARMA-processes) as a priori sequence and models which can be considered as (discrete) generalizations of the Pólya process.

* I thank the editor and an anonymous referee for valuable suggestions

ALBRECHT

2. DEFINITION OF THE MODEL AND BASIC PROPERTIES

Let Λ_i , denote the risk parameter in period i and let $U_n(\lambda_1, \ldots, \lambda_n)$ —the structure function of the considered collective—denote the joint distribution function of $\Lambda_1, \ldots, \Lambda_n$. We make the following assumptions:

Assumption 1

(1)
$$P(N_1 = k_1, \ldots, N_n = k_n | \{\Lambda_i\}) = \prod_{i=1}^n P(N_i = k_i | \Lambda_i).$$

This means that the $\{N_i\}$ are conditionally independent given the $\{\Lambda_i\}$.

Assumption 2. The conditional distribution of N_i given $\Lambda_i = \lambda$ is a Poisson distribution

(2)
$$P(N_i = k_i | \Lambda_i = \lambda) = \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}.$$

It is Assumption 2 which creates the difference to the other above mentioned evolutionary models. The price we have to pay is the specification of the conditional distribution—which, however, is very natural for claim number models what we get on the other hand are more specific and useful results.

Combining (1) and (2) we obtain the multivariate distribution of the claim numbers

(3)
$$P(N_1 = k_1, \ldots, N_n = k_n) = \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^n \left\{ \frac{\lambda_i^{k_i}}{k_i!} e^{-\lambda_i} \right\} dU_n(\lambda_1, \ldots, \lambda_n).$$

This, however, means that the sequence $\{N_i\}_{i \in \mathbb{N}}$ is a "doubly stochastic Poisson sequence". Such sequences have been studied by GRANDELL (1971, 1972, 1976) as a special case of the doubly stochastic Poisson process, which itself can be considered as an evolutionary credibility model for claim numbers in continuous time. We will for practical purposes, however, consider only the discrete time model. A main implication of (3) is that it is possible to establish more properties of the model than just the form of the conditional linear forecast of N_{n+1} as in the usual credibility models. E.g., one can solve other statistical problems and one can give limit theorems for the process. For a lot of detailed results, cf. GRANDELL (1971, 1972, 1976) and SNYDER (1975).

If we denote

(4)
$$\begin{cases} E(\Lambda_i) = m_i, \quad \text{Cov}(\Lambda_i, \Lambda_j) = r_{ij}, \\ \text{Var}(\Lambda_i) = r_{ii} = r_{ii}, \end{cases}$$

we obtain the corresponding moments of $\{N_i\}$ as

(5)
$$\begin{cases} E(N_i) = m_i, & \text{Cov}(N_i, N_j) = r_{ij}, & i \neq j \\ \text{Var}(N_i) = r_i + m_i. \end{cases}$$

From (2) we see that the marginal distributions of the process $\{N_i\}$ are mixed Poisson distributions

(6)
$$P(N_i = k) = \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} dU_{\Lambda_i}(\lambda).$$

This implies that $P(N_i = k)$ can be calculated for various mixing distributions $U_{\Lambda_i}(\lambda)$. For some recent results see ALBRECHT (1984). The multivariate counting distribution of the process is given by (3), but can alternatively be derived as follows.

Let $L_n^{\Lambda}(s_1, \ldots, s_n)$ denote the Laplace functional of $(\Lambda_1, \ldots, \Lambda_n)$ and let $\Phi_n^N(t_1, \ldots, t_n)$ denote the probability generating functional of (N_1, \ldots, N_n) . As $e^{-\lambda(1-t)}$ is the probability generating function of a Poisson variable with

As $e^{-\lambda(1-t)}$ is the probability generating function of a Poisson variable with parameter λ , we obtain from (3)

(7)
$$\Phi_n^N(t_1,\ldots,t_n) = \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^n E[t_i^{N_i}|\Lambda_i = \lambda_i] \, dU_n(\lambda_1,\ldots,\lambda_n)$$
$$= \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^n e^{-\lambda_i(1-t_i)} \, dU_n(\lambda_1,\ldots,\lambda_n)$$
$$= L_n^\Lambda(1-t_1,\ldots,1-t_n).$$

The multivariate counting distribution then is given by the relation

(8)
$$P(N_1 = k_1, \dots, N_n = k_n) = \left[\prod_{i=1}^n \frac{1}{k_i!}\right] \frac{\partial^{\sum k_i} \Phi_n^N(t_1, \dots, t_n)}{\partial t_n^{k_n} \dots \partial t_1^{k_1}} \bigg|_{t_i = 0}$$

We now come to the central problem of credibility, the calculation of the optimal linear forecast of N_{n+1} given the N_1, \ldots, N_n . If $f_n(N_1, \ldots, N_n) = a_0 + \sum_{i=1}^n a_i N_i$ denotes the linear forecast function, the parameters which make $E\{N_{n+1} - f_n(N_1, \ldots, N_n)\}^2$ a minimum are determined in the following way (this is easily established by straightforward calculation, or as a special case from the general result of JEWELL (1971, p. 15) or GRANDELL (1976, p. 128)).

 a_0 is given by a single equation which makes the forecast unbiased

(9)
$$a_0 = E(N_{n+1}) - \sum_{i=1}^n a_i E(N_i) = m_{n+1} - \sum_{i=1}^n a_i m_i$$

The remaining coefficients are given by the $n \times n$ system of linear equations

(10)
$$\sum_{j=1}^{n} \operatorname{Cov}(N_{n}, N_{j})a_{j} = \operatorname{Cov}(N_{n}, N_{n+1}), \quad i = 1, ..., n$$

or more specifically

(11)
$$a_i m_i + \sum_{j=1}^n r_{ij} a_j = r_{in+1}, \quad i = 1, \ldots, n.$$

We note, that because of the identical expectation and covariance structure the optimal linear forecast of N_{n+1} given the N_1, \ldots, N_n equals the optimal linear

ALBRECHT

forecast of $\Lambda_{n+1} = E[N_{n+1}|\Lambda_{n+1}]$ given N_1, \ldots, N_n . In turn this means that it is also identical to the optimal linear forecast of $Var(N_{n+1}|\Lambda_{n+1}) = \Lambda_{n+1}$ given N_1, \ldots, N_n .

We now consider in detail a rather general class of doubly stochastic Poisson sequences, which turns out to have nice properties with respect to the calculation of the credibility forecast and the estimation of the structural parameters.

3. WEAKLY STATIONARY A PRIORI SEQUENCES

We require that $\{\Lambda_n\}_{n \in \mathbb{N}}$ is a weakly stationary sequence characterized by the following moment structure:

(12)
$$E(\Lambda_i) = m \text{ for all } i \in \mathbb{N}$$

(13)
$$\operatorname{Cov}(\Lambda_{i},\Lambda_{j}) = r_{|i-j|} \text{ for all } i, j \in \mathbb{N}$$

The main result in connection with this special model is that we are able to simplify the calculation of the credibility forecast. Whereas the general case only allows that the inverse of $C(n) = (\text{Cov}(N_i, N_j))_{i=1,\dots,n}$ can be calculated recursively we are able to give a recursive formula for the optimal coefficients $a_{i,n}$ however, *not* a recursive formula for the credibility forecast.

Let now

(14)
$$f_n^*(N_1,\ldots,N_n) = a_0(n) + \sum_{i=1}^n a_i(n)N_i$$

denote the optimal linear forecast of N_{n+1} given N_1, \ldots, N_n and

(15)
$$C(n) = (\text{Cov}(N_i, N_j))_{i,j=1,...,n} = (c_{ij})$$

denote the covariance matrix of (N_1, \ldots, N_n) .

We have

(16)
$$c_{ij} = \begin{cases} r_0 + m & i = j, \\ r_{|i-j|} & i \neq j. \end{cases}$$

Let

(17)
$$a(n) = (a_1(n), \ldots, a_n(n))'$$

(18)
$$\tilde{a}(n) = (a_n(n), \dots, a_1(n))'$$

(19)
$$\mathbf{r}(n) = (r_1, \ldots, r_n)^{\prime}$$

and

(20)
$$\tilde{\boldsymbol{r}}(n) = (r_n, \ldots, r_1)'.$$

From (10) we obtain that the optimal coefficients of the credibility forecast are given by

(21)
$$\boldsymbol{a}(n) = \boldsymbol{C}^{-1}(n)\tilde{\boldsymbol{r}}(n).$$

The following lemma gives the form of the inverse of a partitioned matrix.

LEMMA 1. Let the symmetric (n, n) matrix C be decomposed to

$$\boldsymbol{C} = \left(\begin{array}{c|c} \boldsymbol{c}_{11} & \boldsymbol{u}' \\ \boldsymbol{u} & \boldsymbol{D} \end{array} \right),$$

where **D** is of order (n-1, n-1). Then we have

(22)
$$\boldsymbol{C}^{-1} = \left(\begin{array}{c|c} \frac{1}{s} & -\frac{1}{s} \boldsymbol{v}' \\ \hline -\frac{1}{s} \boldsymbol{v} & \boldsymbol{D}^{-1} + \frac{1}{s} \boldsymbol{v} \boldsymbol{v}' \\ \hline -\frac{1}{s} \boldsymbol{v} & \boldsymbol{D}^{-1} + \frac{1}{s} \boldsymbol{v} \boldsymbol{v}' \end{array} \right),$$

where

$$\boldsymbol{v} = \boldsymbol{D}^{-1}\boldsymbol{u}$$
$$\boldsymbol{s} = \boldsymbol{c}_{11} - \boldsymbol{v}'\boldsymbol{u} = \boldsymbol{c}_{11} - \boldsymbol{u}'\boldsymbol{D}^{-1}\boldsymbol{u}.$$

The following lemma gives some useful elementary properties of the covariance matrix C(n).

Lемма 2.

1. C(n+1) can for $n \ge 1$ be decomposed in the following way:

(23)
$$\boldsymbol{C}(n+1) = \left(\frac{\boldsymbol{r}_0 + \boldsymbol{m} \mid \boldsymbol{r}(n)'}{\boldsymbol{r}(n) \mid \boldsymbol{C}(n)}\right).$$

2.

(24)
$$\boldsymbol{C}(n)\tilde{\boldsymbol{a}}(n) = \boldsymbol{r}(n).$$

This implies

3.

(25)
$$\boldsymbol{C}^{-1}(n)\boldsymbol{r}(n) = \tilde{\boldsymbol{a}}(n).$$

We now define (the $a_1(n)$ are the coefficients of the credibility forecast $f_n^*(N_1, \ldots, N_n)$)

(26)
$$s(n) = r_0 + m - r(n)'\hat{a}(n) = r_0 + m - \sum_{i=1}^n r_i a_{n-i+1}(n), \quad n \ge 1$$

(27)
$$k(n) = r_{n+1} - r(n)'a(n) = r_{n+1} - \sum_{i=1}^{n} r_i a_i(n), \qquad n \ge 1.$$

REMARK. $s(n) = E\{N_{n+1} - f_n^*(N_1, \ldots, N_n)\}^2$, i.e., the minimum mean square error of a linear forecast of N_{n+1} , given N_1, \ldots, N_n .

We now come to the central result.

THEOREM. For the coefficients $a_0(n+1)$, a(n+1) of the credibility forecast $f_{n+1}(N_1, \ldots, N_{n+1})$ the following relations are valid $(n \ge 1)$:

(28)
$$a_0(n+1) = \left(1 - \frac{k(n)}{s(n)}\right) a_0(n)$$

(29)
$$a_1(n+1) = \frac{k(n)}{s(n)}$$

(30)
$$a_i(n+1) = a_{i-1}(n) - \frac{k(n)}{s(n)} a_{n-i+2}(n), \quad 2 \le i \le n+1.$$

The starting values are $a_0(1) = m(1 - r_1/(r_0 + m))$ and $a_1(1) = r_1/(r_0 + m)$.

REMARK. (30) can alternatively be written as

(31)
$$(a_2(n+1),\ldots,a_{n+1}(n+1))' = a(n) - \frac{k(n)}{s(n)} \tilde{a}(n).$$

Proof.

(32)
$$a(n+1) = C^{-1}(n+1)\tilde{r}(n+1).$$

From the decomposition (23) of C(n+1), we obtain in the notation of lemma 1, using (25):

$$v = C^{-1}(n)r(n) = \tilde{a}(n)$$
$$s = (r_0 + m) - \tilde{a}(n)'r(n) = s(n).$$

The following partitioned form of $C^{-1}(n+1)$ results:

$$\boldsymbol{C}^{-1}(n+1) = \left(\begin{array}{c|c} \frac{1}{s(n)} & -\frac{1}{s(n)} \, \tilde{\boldsymbol{a}}(n)' \\ \hline -\frac{1}{s(n)} \, \tilde{\boldsymbol{a}}(n) & \boldsymbol{C}^{-1}(n) + \frac{1}{s(n)} \, \tilde{\boldsymbol{a}}(n) \tilde{\boldsymbol{a}}(n)' \end{array} \right).$$

From (32), the relations (29) and (30) easily follow. Then (28) is obtained from (9).

COROLLARY. For the mean square error s(n) of the credibility forecast the following recursive formula is valid:

(33)
$$s(n+1) = s(n) - \frac{k(n)^2}{s(n)}, \quad n \ge 1; \qquad s(1) = r_0 + m - \frac{r_1^2}{r_0 + m}.$$

PROOF. From (26)

$$s(n+1) = r_0 + m - \sum_{i=1}^n r_i a_{n+2-i}(n+1) - r_{n+1}a_1(n+1);$$

using (29), (30) this simplifies to

$$r_{0}+m-\sum_{i=1}^{n}r_{i}\left\{a_{n+1-i}(n)-\frac{k(n)}{s(n)}a_{i}(n)\right\}-r_{n+1}\frac{k(n)}{s(n)}.$$

Using (26), (27) this in turn simplifies to (33).

The theorem allows recursive calculation of the credibility forecast of N_{n+1} in case of a known risk structure. To obtain an *empirical* credibility forecast, we have to estimate the unknown parameters, which here are: m, r_0, r_1, r_2, \ldots

The estimation problem exhibits the second important property of the model considered in this section. If we assume that the *a priori* sequence $\{\Lambda_i\}$ is weakly stationary, then we obtain from (5), that the observable sequence $\{N_i\}$ is a weakly stationary one, too. We then have the possibility to apply results from the well-developed theory of the statistical analysis of weakly stationary time series, see e.g., HANNAN (1960, Chapters II-IV) or DOOB (1953, Chapter X). For example a *spectral analysis* of the sequence $\{N_i\}$ is possible. Some results in this direction can be found in GRANDELL (1976, Chapter 7.2). We will here, however, confine to the above mentioned estimation problem. Up to now we have only considered the claim number sequence of a single risk, observed for *n* years. We now assume that we observe a collective of *K* independent risks, each having the same probability law of its claim number sequence.

Let

(34)
$$N_{ji}$$
 = number of claims of risk *i* in year *j*

i = 1, ..., K; j = 1, ..., n.

From standard results of time series analysis, e.g., HANNAN (1960, pp. 30-33), we obtain the following natural estimators of the above mentioned parameters.

(35)
$$\hat{m} = \frac{1}{Kn} \sum_{i,j=1}^{n} N_{ij}$$

(36)
$$\hat{r}_{k} = \frac{1}{K(n-k)-1} \sum_{j=1}^{K} \sum_{j=1}^{n-k} (N_{j} - \hat{m}) (N_{j+k,j} - \hat{m}), \text{ for } k \ge 1$$

(37)
$$\operatorname{Var}(N_{j_i}) = \frac{1}{Kn-1} \sum_{i=1}^{K} \sum_{j=1}^{n} (N_{j_i} - \hat{m})^2.$$

A natural estimate for r_0 then is

(38)
$$\hat{r}_0 = \frac{1}{Kn-1} \sum_{i=1}^{K} \sum_{j=1}^{n} (N_{ji} - \hat{m})^2 - \hat{m}.$$

As pointed out by the referee the expected value of (37) is given by

Var
$$(N_{j_i}) - \frac{1}{Kn-1} \frac{1}{n} \sum_{j=1}^{n-1} r_j(n-j),$$

which implies a slight bias.

ALBRECHT

The theorem shows, how the coefficients of the credibility forecasts can be calculated recursively in the case of an arbitrary stationary a priori sequence. It is, however, not possible to develop a recursive formula for the credibility forecast itself for the general case. It would be interesting to examine special classes of stationary a priori sequences which give rise to recursive formulae for the credibility forecast itself. For a more general type of evolutionary models KREMER (1982) has considered ARMA (p, q) processes as a special class of stationary a priori sequences. In the model of this paper the a priori sequences have to be positivevalued to be admissible. Therefore the ARMA (p, q) processes are not admissible in general. However Lewis and a number of co-authors (see LAWRENCE and LEWIS (1980) for the most recent results) have developed models for positivevalued stationary time series $\{X_i\}_{i \in \mathbb{N}}$ which, being in general rather distinct from the ARMA-models, possess the same autocorrelation structure as the ARMAprocesses. These processes are called EARMA (p, q)-processes, the E stemming from the additional feature of all these processes: they have an exponential marginal distribution!

The results of KREMER (1982) cannot be translated into the present context for several reasons, one being that the form of the linear regressions of the EARMA-processes have not yet been established. Another drawback of the EARMA-processes is that the statistical analysis of these processes is not yet well developed in general, contrary to the ARMA-processes. In the following, we consider some examples.

EXAMPLE 1. EAR (1)-process as *a priori* sequence. A stationary version of the first order autoregressive model with exponential marginals with "finite past" can be obtained as follows (cf. GAVER and LEWIS (1980, p. 732):

(39)
$$\begin{cases} \Lambda_n = \rho \Lambda_{n-1} + I_n E_n, & n \ge 2\\ \Lambda_1 = \rho E_0 + I_1 E_1, & (0 \le \rho < 1) \end{cases}$$

where $\{I_n\}_{n\geq 1}$ is a sequence of i.i.d. Bernoulli-variables with $P(I_n = 0) = \rho$ and $\{E_n\}_{n\geq 0}$ is an independent sequence of i.i.d. exponentially distributed variables with parameter λ . The resulting sequence is a first order Markov process, the Λ_n are exponentially distributed with parameter λ and can alternatively be obtained in the usual first-order autoregressive form $\Lambda_n = \rho \Lambda_{n-1} + \varepsilon_n$ with a suitable $\{\varepsilon_n\}$. For the second order structure we obtain

(40)
$$\begin{cases} m = E(\Lambda_n) = 1/\lambda, & r_0 = \operatorname{Var}(\Lambda_n) = 1/\lambda^2\\ r_k = \operatorname{Cov}(\Lambda_n, \Lambda_{n+k}) = \rho^k/\lambda^2 = \rho^k r_0, & k \ge 1 \end{cases}$$

From (40) we see that the r_k fulfill the property (10) of SUNDT (1981, p. 7), which in our context reads:

(41)
$$r_{i+1-j} = \rho_i \cdot r_{i-j} \text{ for all } i \ge j, \text{ for all } j \ge 1.$$

Clearly $\rho_i = \rho$ for all *i* and from Sundt's result (11) we obtain the following recursive formula for the credibility forecast.

As (notation as in SUNDT (1981)) $\varphi_i = E\{\text{Var}(N_i|\Lambda_i)\} = E(\Lambda_i) = 1/\lambda$, we define

(42)
$$\gamma_n = \frac{1}{\lambda s(n-1)}, \quad n \ge 2; \qquad \gamma_1 = \frac{1}{1+1/\lambda}$$

(43)
$$\chi = 2\rho^2 + (1-\rho^2)\left(1+\frac{1}{\lambda}\right)$$

and obtain

(44) $\gamma_{n+1} = (\chi - \gamma_n \rho^2)^{-1}$

(45)
$$\begin{cases} f_n^*(N_1, \dots, N_n) = \rho[(1 - \gamma_n)N_n + \gamma_n f_{n-1}^*(N_1, \dots, N_{n-1})] + (1 - \rho)\lambda \\ f_0 = 1/\lambda. \end{cases}$$

This is the desired recursive formula for the credibility forecast.

It is interesting to note that the regressions of this *a priori* sequence are all linear, precisely

(46)
$$E(\Lambda_{n+1}|\Lambda_1,\ldots,\Lambda_n) = E(\Lambda_{n+1}|\Lambda_n) = \rho \Lambda_n + (1-\rho)/\lambda.$$

However, we have not been able to show that the regressions of the *a posteriori* process $\{N_i\}$ are linear too, i.e., the credibility forecast in the best forecast of N_{n+1} based on N_1, \ldots, N_n .

We now come to the estimation of the unknown parameters λ and ρ and consider again a *collective* of K independent risks each having the same law of its claim number sequence. Let N_{j_i} be defined as in (34); noticing that $E(N_i) = 1/\lambda$ and $r_1 = \text{Cov}(N_i, N_{i+1}) = \rho/\lambda^2$ we obtain from (35) and (36) the following (consistent) natural estimators of λ and ρ :

(47)
$$\hat{\lambda} = 1 \left/ \frac{1}{Kn} \sum_{i,j=1}^{n} N_{ji} \right|$$

(48)
$$\hat{\rho} = \frac{\hat{\lambda}^2}{K(n-1)-1} \sum_{i=1}^{K} \sum_{j=1}^{n-1} (N_{ji} - \hat{\lambda}^{-1}) (N_{j+1,i} - \hat{\lambda}^{-1}).$$

A drawback of the model is, that all correlations $\rho_k = \text{Corr} (\Lambda_v \Lambda_{i+k})$ are positive. Indeed, one can show that there does not exist an autoregressive sequence $\Lambda_n = \rho \Lambda_{n-1} + \varepsilon_n$ possessing exponentially distributed marginals and $\rho < 0$! However, GAVER and LEWIS (1980, p 741) present models of similar autocorrelation structure and negative correlation, which still possess the property of having an exponential marginal distribution. GAVER and LEWIS (1980, pp. 736-737) consider also an autoregressive process of first order with a gamma marginal distribution and a similar autocorrelation structure, the GAR (1)-process.

EXAMPLE 2. EMA(1)-process as a priori sequence. A first-order moving average model with exponential marginal distribution, was considered by LAWRANCE and LEWIS (1977) and can be obtained as follows (forward formulation)

(49)
$$\Lambda_n = \beta \varepsilon_n + I_n \varepsilon_{n+1}, \qquad (0 \le \beta \le 1),$$

ALBRECHT

where the $\{I_n\}_{n\geq 1}$ are i.i.d. Bernoulli variables with $P(I_n = 1) = 1 - \beta$ and $\{\varepsilon_n\}_{n\geq 1}$ is an independent sequence of i.i.d. exponentially distributed (parameter λ) random variables. The process is not Markovian and the second order structure is given by

(50)
$$\begin{cases} m = E(\Lambda_n) = 1/\lambda, & r_0 = \operatorname{Var}(\Lambda_n) = 1/\lambda^2 \\ r_1 = \operatorname{Cov}(\Lambda_n, \Lambda_{n+1}) = \beta(1-\beta)r_0 \\ r_k = 0 & \text{for } k \ge 2 \end{cases}$$

To obtain a recursive formula for the credibility forecast we can use Theorem 2 in SUNDT (1981, p. 6).

We obtain the following recursive relation for the estimation error s(n):

(51)
$$\begin{cases} s(n) = \frac{1}{\lambda} \left(1 + \frac{1}{\lambda} \right) - \frac{\beta^2 (1 - \beta)^2}{\lambda^4 s(n - 1)}, \quad n \ge 2\\ s(1) = \frac{1}{\lambda} \left(1 + \frac{1}{\lambda} \right) - \frac{\beta^2 (1 - \beta)^2}{\lambda^2 + \lambda^3} \end{cases}$$

and the following recursive formula for the credibility forecast

(52)
$$\begin{cases} f_n^*(N_1, \dots, N_n) = \frac{\beta(1-\beta)}{\lambda^2 s(n-1)} N_n - \frac{\beta(1-\beta)}{\lambda^2 s(n-1)} f_{n-1}^*(N_1, \dots, N_{n-1}) + \frac{1}{\lambda}, \\ f_1^*(N_1) = \frac{1}{\lambda} - \frac{\beta(1-\beta)}{(1+\lambda)\lambda} + \frac{\beta(1-\beta)}{1+\lambda} N_1. \end{cases}$$

A natural estimator of the unknown parameter β (λ is estimated as under (47)) is given by

(53)
$$\hat{\beta} = \frac{1}{2} + \frac{1}{2} \sqrt{1 - 4\hat{\lambda}^2} \frac{1}{K(n-1) - 1} \sum_{j=1}^{K} \sum_{j=1}^{n-1} (N_{j_j} - \hat{\lambda}^{-1}) (N_{j+1,j} - \hat{\lambda}^{-1}).$$

A drawback of the model is that the first-order autocorrelation $\rho_1 = \beta(1-\beta)$ is always nonnegative (one can show in addition, that it is always bounded from above by 1/4).

The regressions of the a priori process are given by

(54)
$$E(\Lambda_{n+1}|\Lambda_n) = \frac{1}{\lambda} \left[\beta \lambda \Lambda_n + \frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta} e^{-\lambda(1-\beta)\Lambda_n/\beta} \right]$$

and are therefore not linear.

EXAMPLE 3. EARMA (1, 1) process as a priori sequence. A first order mixed autoregressive-moving average process with an exponential marginal distribution was considered by JACOBS and LEWIS (1977) and can be obtained as follows ("backward formulation").

(55)
$$\begin{cases} \Lambda_n = \beta \varepsilon_n + U_n A_{n-1} & (0 \le \beta \le 1) \\ A_n = \rho A_{n-1} + V_n \varepsilon_n & (0 \le \rho \le 1) \\ A_0 = \varepsilon_0 \end{cases},$$

where $\{U_n\}$ and $\{V_n\}$ are independent sequences of independent Bernoulli variables with $P(U_n = 0) = \beta$, $P(V_n = 0) = \rho$ and $\{\varepsilon_n\}$ is an independent sequence of i.i.d. exponentially distributed (parameter λ) random variables. The resulting process $\{\Lambda_n\}$ is stationary and in general non Markovian. The second order structure of the process is given by

(56)
$$\begin{cases} m = E(\Lambda_n) = 1/\lambda, & r_0 \operatorname{Var}(\Lambda_n) = 1/\lambda^2 \\ r_1 = \operatorname{Cov}(\Lambda_n, \Lambda_{n+1}) = r_0(1-\beta)[\beta + \rho(1-2\beta)] \\ r_k = \rho^{k-1} r_1. \end{cases}$$

Again we can apply Theorem 2 of SUNDT (1981) to obtain a recursive formula for the credibility forecast. The result is as follows:

(57)
$$\begin{cases} s(n) = (r_0 + m) + \rho^2 (r_0 + m) - 2\rho r_1 - \frac{[\rho(r_0 + m) - r_1]^2}{s(n-1)}, & n \ge 2 \\ s(1) = (r_0 + m) - \frac{r_1^2}{r_0 + m} \end{cases}$$

(58)
$$\begin{cases} f_n^*(N_1, \dots, N_n) = \left(\rho - \frac{\rho(r_0 + m) - r_1}{s(n-1)}\right) N_n \\ + \frac{\rho(r_0 + m) - r_1}{s(n-1)} f_{n-1}^*(N_1, \dots, N_{n-1}) + m(1-\rho), & n \ge 2 \\ f_1^*(N_1) = m \left(1 - \frac{r_1}{r_0 + m}\right) + \frac{r_1}{r_0 + m} N_1. \end{cases}$$

4. SOME SPECIAL MODELS

We first treat two models which can be considered as generalizations of the Pólya-process in discrete time. The Pólya-process is a mixed Poisson process with the gamma distribution as mixing distribution.

Model A. A natural generalization, which was already considered by BATES and NEYMAN (1952), is to assume

(59)
$$\Lambda_{I} = a_{I} \cdot \Lambda,$$

where Λ follows a gamma distribution with parameters b and p. The a priori moments are given by

(60)
$$E(\Lambda_i) = a_i \frac{p}{b}, \quad \text{Var}(\Lambda_i) = a_i^2 \frac{p}{b^2}, \quad \text{Cov}(\Lambda_i, \Lambda_j) = a_i a_j \frac{p}{b^2}.$$

SNYDER (1975, p. 288) considers the continuous time analogue, which he terms "inhomogeneous Pólya process".

Model A seems to be the only known double stochastic Poisson sequence for which the multivariate counting distribution can be given explicitly. Bates and Neyman showed that

(61)
$$P(N_1 = n_1, ..., N_k = n_k) = \left(1 + \frac{a}{b}\right)^{-p} n! \binom{n+p-1}{n} \prod_{i=1}^k \frac{1}{n_i!} \left[\frac{a_i/b}{1+a/b}\right]^{n_i},$$

where $a = \sum_{i=1}^{k} a_i$ and $n = \sum_{i=1}^{k} n_i$.

Comparing (61) with JOHNSON and KOTZ (1969, p. 292, (32)) shows, that the multivariate counting distribution of the "discrete inhomogeneous Pólya process" is just a multivariate negative binomial distribution $(N = p, P_i = a_i/b)$ in their notation).

JOHNSON and KOTZ (1969, p. 295) show also, that in case of a multivariate negative binomial distribution the regressions are always linear. Especially we obtain

(62)
$$E(N_{n+1}N_1, ..., N_n) = p \frac{a_{n+1}}{b+a} + \frac{a_{n+1}}{b+a} \sum N_n$$
$$= \frac{b}{b+a} E(N_{n+1}) + \frac{a_{n+1}}{b+a} \sum N_n$$

This implies that in case of the "discrete inhomogeneous Pólya process" the optimum forecast function (with respect to the mean square error) is identical to the best *linear* forecast function (the credibility forecast).

If we want to calculate the credibility forecast with the method of chapter 1 (equations (9) and (10)), we can apply a result of JEWELL (1976, pp. 16-17), because $\text{Cov}(N_i, N_i)$ can be factored into $a_i \cdot ((p/b^2)a_i)$.

It is interesting to note that already BUHLMANN (1969, pp. 164–165) considered a similar model. He considered a sequence of conditionally Poisson distributed claim variables $\{X_n\}$ with the property

(63)
$$E(X_n|\theta) = a_n \cdot \theta,$$

where $a_n = n + c$, c is a constant independent of n and θ follows a gamma distribution.

In addition to Buhlmann's results we show in the following how the structural parameters (especially c) can be estimated.

Assume that we have given a sample of size m of observations of (N_1, \ldots, N_k) . Let

$$n_{ij} = i$$
th observation of N_{ij} , $i = 1, ..., m; j = 1, ..., k$.

Let

$$n_i = \sum_{j=1}^k n_{ij}, \quad \tilde{n}_j = \sum_{i=1}^m n_{ij}, \quad n = \sum_{i=1}^m n_i, \quad r = \frac{1}{2}k(k+1).$$

Then the log-likelihood-function of the observations is given by

(64)
$$\log L = -(n+pm) \log \left(1 + \frac{r+kc}{b}\right) + \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log (p-1+j) + \sum_{j=1}^{k} \tilde{n}_j \log \left(\frac{j+c}{p}\right).$$

The likelihood equations then are given by

(65)
$$\frac{\partial \log L}{\partial p} = -m \log \left(1 + \frac{r+kc}{b}\right) + \sum_{j=1}^{m} \sum_{j=1}^{n_j} \frac{1}{p-1+j} = 0$$

(66)
$$\frac{\partial \log L}{\partial c} = \frac{-k(n+pm)}{b+r+kc} + \sum_{j=1}^{k} \frac{\tilde{n}_j}{j+c} = 0$$

(67)
$$\frac{\partial \log L}{\partial b} = \frac{(n+pm)(r+kc)}{b^2+b(r+kc)} - \frac{n}{b} = 0.$$

If \hat{p} , \hat{c} , \hat{b} denote the maximum likelihood estimators of p, c, b, then from (67) we obtain

(68)
$$\hat{b} = -\frac{m}{n}(r+k\hat{c})\hat{p}$$

Substituting (68) in (66), we obtain that \hat{c} is the solution of

(69)
$$\sum_{j=1}^{k} \frac{\tilde{n}_j}{j+\hat{c}} = \frac{nk}{(r+k\hat{c})}$$

Substituting (69) in (65), we obtain that \hat{p} is the solution of

(70)
$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{1}{\hat{p} - 1 + j} = m \log \left(1 + \frac{n}{m\hat{p}} \right).$$

Model B Another way to obtain a generalization of the Pólya process is to replace the gamma mixing distribution by a multivariate analogue, a multivariate gamma distribution for $(\Lambda_1, \ldots, \Lambda_n)$.

A natural way to obtain a multivariate gamma distribution, more precisely a multivariate χ^2 -distribution is the following, cf. also JOHNSON and KOTZ (1972, chapter 40.3) or KRISHNAIAH and RAO (1961). The χ^2 -distribution with *n* degrees of freedom is a special gamma distribution and is the distribution of $\sum_{i=1}^{n} X_i^2$, where the X_i are independent and identically N(0, 1)-distributed (normal distribution with mean 0 and variance 1). A natural multivariate analogue is obtained by starting with *m* independent and identically multivariate normal distributed random vectors $Y_i = (Y_{i1}, \ldots, Y_{in}), i = 1, \ldots, m$. Precisely Y_i follows a $N(0, \Sigma)$ distribution, where $\Sigma = (\Sigma_y)$ is the variance covariance matrix of (Y_{i1}, \ldots, Y_{in}) and we assume that $\Sigma_{ii} = 1$.

The a priori vector

(71)
$$(\Lambda_1,\ldots,\Lambda_n) = \left(\sum_{i=1}^m Y_{i1}^2,\ldots,\sum_{i=1}^m Y_{in}^2\right)$$

then follows a distribution, which can be considered as a multivariate χ^2 -distribution. Especially each Λ_i is χ^2 -distributed with *n* degrees of freedom. The Laplace functional $L_n^{\Lambda}(s_1, \ldots, s_n) = E[e^{-\Sigma s_i \Lambda_i}]$ is given by

(72)
$$L_n^{\Lambda}(s_1,\ldots,s_n) = |\mathbf{I}+2s_{\Delta}\boldsymbol{\Sigma}|^{-m/2},$$

where s_{Δ} is a diagonal matrix with diagonal elements s_1, \ldots, s_n . From (7) it follows that the probability generating functional of N_1, \ldots, N_n is given by

(73)
$$\Phi_n^N(t_1,\ldots,t_n) = |\mathbf{I} + 2(\mathbf{I} - \mathbf{t}_\Delta)\boldsymbol{\Sigma}|$$

where $I - t_{\Delta}$ is a diagonal matrix with diagonal elements $(1 - t_1, \dots, 1 - t_n)$.

A simple special case is obtained when we assume a first-order correlation for the Y_{y} , i.e., Σ is of the form



We then obtain the following second order recursive relations for $\varphi_n(t_1, \ldots, t_n) = |I + 2(I - t_{\Delta})\Sigma|$:

(74)
$$\begin{cases} \varphi_{n+2}(t_1, \dots, t_{n+2}) = (3 - 2t_{n+2})\varphi_{n+1}(t_1, \dots, t_{n+1}) \\ -4(t_{n+2} - 1)(t_{n+1} - 1)r^2\varphi_n(t_1, \dots, t_n) & \text{for } n \ge 0 \\ \varphi_1(t_1) = (3 - 2t_1), \qquad \varphi_0(t_0) = 1. \end{cases}$$

The probability generating functional in this special case then is given by $\Phi_n^N(t_1, \ldots, t_n) = \varphi(t_1, \ldots, t_n)^{-m/2}$.

We obtain that

(75)
$$\frac{\partial^k \Phi_1(t_1)}{\partial t_1^k} = 2^k \frac{\Gamma((m/2) + k)}{\Gamma(m/2)} \varphi_1(t_1)^{-((m/2) + k)}$$

From (8) we obtain

(76)
$$P(N_1 = k) = \frac{1}{k!} \frac{\partial^k \Phi_1(t_1)}{\partial t_1^k} \bigg|_{t_1 = 0} = \frac{2^k}{k! 3^{(m/2)+1}} \frac{\Gamma((m/2) + k)}{\Gamma(m/2)}.$$

This result is identical (for t = 1) with a result of ALBRECHT (1984), who calulated P(N(t) = n) for a mixed Poisson process N(t) with a χ^2 -mixing distribution.

In addition we obtain after some calculation that

(77)
$$\frac{\partial^{k_1+k_2}\Phi_2(t_1, t_2)}{\partial t_1^{k_1}\partial t_2^{k_2}} = (-1)^{k_1} \frac{\Gamma((m/2)+k_1)}{\Gamma(m/2)} \sum_{k=0}^{k_2} \binom{k_2}{k} \{\varphi_2(t_1, t_2)^{-((m/2)+k_1)}\}^{(k)} \times \{[(4-4r^2)t_2+4r^2-6]^{k_1}\}^{(k_2-k)}\}$$

where (k) denotes the kth derivative with respect to t_2 .

We obtain after some calculation

(78)
$$P(N_{1} = k_{1}, N_{2} = k_{2}) = \frac{1}{k_{1}!} \frac{1}{k_{2}!} \frac{\partial^{k_{1}+k_{2}} \Phi_{2}(t_{1}, t_{2})}{\partial t_{1}^{k_{1}} \partial t_{2}^{k_{2}}} \bigg|_{t_{1}=t_{2}=0} = \frac{(-1)^{k}}{k_{1}!k_{2}!\Gamma(m/2)}$$
$$\times \sum_{k=0}^{k_{1}} \binom{k_{2}}{k} (-1)^{k} \frac{\Gamma((m/2)+k_{1}+k)\Gamma(k_{1}+1)}{\Gamma(k_{1}-k_{2}+k+1)}$$
$$\times (9-4r^{2})^{-((m/2)+k_{1}+k)} (4r^{2}-6)^{2k+k_{1}-k_{2}} (4-4r^{2})^{k_{2}-k_{2}}$$

Even in the simple first-order case we have not been able to develop an expression for $E(N_{n+1}|N_1, \ldots, N_n)$, the "best" estimate of N_{n+1} given N_1, \ldots, N_n .

As the second-order structure of the sequence N_n is given by

(79)
$$\begin{cases} E(N_{i}) = m, & \operatorname{Var}(N_{i}) = 3m, \\ r_{1} = \operatorname{Cov}(N_{i}, N_{i+1}) = 2mr^{2} \\ r_{k} = \operatorname{Cov}(N_{i}, N_{i+k}) = 0, \quad k \ge 2 \end{cases}$$

we can apply Theorem 2 of SUNDT (1981) to obtain a recursive formula for the credibility forecast. The result is as follows:

(80)

$$\begin{cases} s(n) = 3m - \frac{4m^2 r^4}{s(n-1)}, \quad n \ge 2 \\ s(1) = 3m - \frac{4}{3}mr^4 \end{cases}$$
(81)

$$\begin{cases} f_n^*(N_1, \dots, N_n) = \frac{2mr^2}{s(n-1)}(N_n - f_{n-1}^*(N_1, \dots, N_{n-1})) + m, \quad n \ge 2 \\ f_1^*(N_1) = m(1 - \frac{2}{3}r^2) + \frac{2}{3}r^2N_1 \end{cases}$$

Model C (a priori sequence with independent increments). If we assume that the a priori sequence { $\Lambda_0 \equiv 0, \Lambda_1, \Lambda_2, \ldots$ } possesses independent increments, this means—cf. Doob (1953, p. 96)—that for all $n \ge 3$ and $i_1 < i_2 < \cdots < i_n$ the random variables $\Lambda_{i_2} - \Lambda_{i_1}, \ldots, \Lambda_{i_n} - \Lambda_{i_{n-1}}$ are mutually independent. An additional assumption is that $E(\Lambda_i) = m$; let $V_i = \text{Var}(\Lambda_i)$, then we obtain for i < j

$$Cov (\Lambda_i, \Lambda_j) = Cov (\Lambda_i - \Lambda_0, \Lambda_j - \Lambda_i + \Lambda_i)$$

= Var (\Lambda_i) + Cov (\Lambda_i - \Lambda_0, \Lambda_j - \Lambda_i)
= Var (\Lambda_i),

i.e., in general

(82)
$$\operatorname{Cov} \left(\Lambda_{i}, \Lambda_{j} \right) = \operatorname{Var} \left(\Lambda_{\min(i,j)} \right).$$

A credibility model with the above moment structure for the *a priori* variables was already considered by GERBER and JONES (1975, pp. 98-99), they show that the credibility forecast $f_n^*(N_1, \ldots, N_n)$ of N_{n+1} is of the "updating type"

(83)
$$f_n^*(N_1,\ldots,N_n) = (1-Z_n) f_{n-1}^*(N_1,\ldots,N_{n-1}) + Z_n N_n.$$

The weights can be calculated recursively, we have

$$Z_{1} = \frac{V_{1}}{m + V_{1}}$$
$$Z_{n} = \frac{V_{n} - V_{n-1} + mZ_{n-1}}{V_{n} - V_{n-1} + mZ_{n-1} + m}.$$

Additional models for the *a priori* sequence are considered in GRANDELL (1972) (e.g., $\{\Lambda_j\}$ is in the form of a linear regression model, pp. 106-108) and GRANDELL (1976) (e.g. $\{\Lambda_j\}$ is a stationary alternating Markov chain, pp 153-157).

REFERENCES

- ALBRECHT, P (1984) Laplace Transforms, Mellin Transforms and Mixed Poisson Processes Scandinavian Actuarial Journal, 58-64
- BATES, G E and NEYMAN, J (1952) Contributions to the Theory of Accident Proneness University of California Publications in Statistics 1, 215-276
- BUHLMANN, H (1969) Experience Rating and Credibility ASTIN Bulletin 5, 157-165

DOOB, J L (1953) Stochastic Processes Wiley New York

- GAVER, D G and LEWIS, P A W (1980) First-Order Autoregressive Gamma Sequences and Point Processes Advances in Applied Probability 12, 727-745
- GERBER, H U and JONES, D A (1975) Credibility Formulas of the Updating Type In KAHN (1975), pp 89-105
- GRANDELL, J (1971) On Stochastic Processes Generated by a Stochastic Intensity Function Skandinavisk Aktuarietidskrift 54, 204-240

GRANDELL, J (1972) Statistical Inference for Doubly Stochastic Poisson Processes In Stochastic Point Processes Statistical Analysis, Theory and Applications, edited by P A W LEWIS Wiley New York, pp 90-121

- GRANDELL, J (1976) Doubly Stochastic Poisson Processes Lecture Notes in Mathematics, No 529 Springer Berlin
- HANNAN, E J (1960) Time Series Analysis Methuen London.
- JACOBS, P A and LEWIS, P. A W (1977) A Mixed Autoregressive-Moving Average Exponential Sequence and Point Process, EARMA(1,1) Advances in Applied Probability 9, 87-104

JEWELL, W S (1975) Model Variations in Credibility Theory In KAHN (1975), pp 193-244

- JEWELL, W S (1976) Two Classes of Covariance Matrices Giving Simple Linear Forecasts Scandinavian Actuarial Journal, 15-29
- JOHNSON, N L and KOTZ, S (1969) Distributions in Statistics Discrete Distributions Wiley New York
- JOHNSON, N L and KOTZ, S (1969) Distributions in Statistics Continuous Multivariate Distributions Wiley New York
- KAHN, P M (ed) (1975) Credibility New York Academic Press
- KREMER, E (1982) Credibility Theory for some Evolutionary Models Scandinavian Actuarial Journal, 129-142
- KRISHNAIAH, P R and RAO, M M (1961) Remarks on a Multivariate Gamma Distribution American Mathematical Monthly 68, 342-346

- LAWRANCE, A J and LEWIS, P A W (1977) An Exponential Moving Average Sequence and Point Process, EMA (1) Journal of Applied Probability 14, 98-113
- LAWRANCE, A J and LEWIS, P A W (1980) The Exponential Autoregressive-Moving Average EARMA (p, q) Process Journal of the Royal Statistical Society B 42, 150-161
- SNYDER, D L (1975) Random point processes Wiley, New York
- SUNDT, B (1979) An Insurance Model with Collective Seasonal Random Factors Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker, 57-64
- SUNDT, B (1981) Recursive Credibility Estimation Scandinavian Actuarial Journal, 3-21
- SUNDT, B (1983) Finite Credibility Formulae in Evolutionary Models Scandinavian Actuarial Journal, 106-116
- TAYLOR, G C (1975) Credibility for Time-Heterogeneous Loss Ratio In KAHN (1975), pp 363-389

LINEAR FILTERING AND RECURSIVE CREDIBILITY ESTIMATION

BY BEN ZEHNWIRTH

Macquarie University, North Ryde, Australia

Abstract

Recursive credibility estimation is discussed from the viewpoint of linear filtering theory. A conjunction of geometric interpretation and the innovation approach leads to general algorithms not developed before. Moreover, covariance characterizations considered by other researchers drop our elegantly as a result of geometric considerations. Examples are presented of Kalman type filters valid for non-Gaussian measurements

Keywords

Credibility, filtering theory, linear Bayesian theory, geometry, Kalman filter, prospective ratemaking, Gram-Schmidt, Fourier series.

1. INTRODUCTION AND SUMMARY

There have appeared a number of papers, fairly tightly connected, concerned with recursive credibility formulae. An early paper that occupies a somewhat central position is that of GERBER and JONES (1975), which develops credibility formulae of the updating type, valid if and only if the covariance structure (5.12) holds. The other papers notably, JEWELL (1976), SUNDT (1981, 1983) and KREMER (1982) develop recursive formulae for a variety of other evolutionary type models, the last emphasizing the relationships with modern models of time series. Last, but not least, the paper of DE JONG and ZEHNWIRTH (1983) relates some credibility models to the Kalman filter, perhaps, the most important algorithm in linear stochastic system theory.

The basic purpose of the present paper is to unify many existing results in recursive credibility theory and moreover develop more general ones. To achieve this, we adopt a geometric interpretation of recursive linear least squares estimation theory in the spirit of GERBER and JONES (1975) and DE VYLDER (1976). There is also a side benefit to be had by adopting a geometric approach—it reduces both the conceptual and algebraic burdens. The practical importance to actuaries of the present paper lies in the fact that once a model for premium rate-making is postulated, the estimators of parameters, premium forecast and associated errors may be derived quite readily using the general results contained herein. Moreover, the recursive nature of the formulae affords economy of computing space and time.

The main results here are established with the aid of KAILATH's (1974) innovation technique which has found fruitful applications in linear filtering theory. It

ASTIN BULLETIN Vol 15, No 1

ZEHNWIRTH

is intimately related to the well known Gram-Schmidt orthogonalization scheme and Fourier series.

Suppose \hat{Y} is a forecast of the random quantity Y with associated mean-square error C, based on some past measurements. Given a new measurement X we wish to update our forecast of Y and its associated mean-square error C. Let \hat{X} represent the forecast of X based on the past measurements. The innovation, $e = X - \hat{X}$, represents what is "new" in the new measurement X. The updated forecast of Y is

$$(1.1) \qquad \qquad \hat{Y} + Ke$$

where the weight K is given by

(1.2)
$$K = E[Ye] \{ E[e^2] \}^{-1}.$$

The mean-square error of the updated forecast (1.1) is

$$(1.3) C - KE[Ye].$$

The foregoing results are treated in elaborate detail in Sections 3 and 4. In Section 5 we consider a general prospective ratemaking framework and indicate how covariance structures considered by GERBER and JONES (1975), JEWELL (1976) and SUNDT (1981) drop out elegantly as a result of the geometric interpretation of the problem. Finally, in Section 6 Kalman type filters are derived for two different models using results developed earlier in the paper. The filters are related to the work of SUNDT (1981, 1983) and DE JONG and ZEHNWIRTH (1983).

2. HILBERT SPACE OF SQUARE-INTEGRABLE RANDOM VARIABLES

For the purposes of the present paper it is convenient to formulate some definitions and terminology and to state two classical projection theorems.

Consider a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The Hilbert space $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P})$ is the linear space of measurable functions from Ω into \mathbb{R} whose second moment exist. We identify with the element $X \in \mathcal{H}$, the equivalence class $\{\tilde{X} : \tilde{X} = X \text{ a.e.}\}$. The inner product $\langle X, Y \rangle$ for any two elements X and Y in \mathcal{H} is defined by

$$\langle X, Y \rangle = E[XY].$$

Accordingly, the corresponding $\|\cdot\|$ is defined by

$$||X|| = (E[X^2])^{1/2}$$

It is beneficial to extend the definition of the inner product $\langle \cdot, \cdot \rangle$ to random vectors. Suppose $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_m)$ where $X \in \mathcal{H}^n$ and $Y \in \mathcal{H}^m$. Define $\langle X, Y \rangle$ by

$$\langle X, Y \rangle = E[XY'].$$

This is not an inner product in the true sense—it is a matrix. However, if we ignore this deficiency, the projection theorem can be used as a quick mnemonic way of obtaining the approximate optimal estimators (theorem 3.2)

The following properties of the bilinear functional $\langle \cdot, \cdot \rangle$ are noted.

(2.1)
$$\langle AX, BY \rangle = A \langle X, Y \rangle B'$$

for any two matrices A and B of appropriate dimensions.

(2.2)
$$||AX||^2 = A||X||^2 A'$$

and

(2.3)
$$\langle X, Y \rangle' = \langle Y, X \rangle.$$

We state two classical projection theorems applicable to any Hilbert space (borrowed from LUENBERGER (1969)).

THEOREM 2.1. Let \mathcal{H} be a Hilbert space and \mathcal{L} a closed subspace of \mathcal{H} . Corresponding to any vector $Y \in \mathcal{H}$, \exists a unique $X^* \in \mathcal{L}$ such that

$$\|Y - X^*\|_* = \inf_{X \in \mathscr{L}} \|Y - X\|_*.$$

where $\|\cdot\|_{*}$ is the norm defined on \mathcal{H} .

Furthermore, a necessary and sufficient condition that $X^* \in \mathscr{L}$ be the unique minimization vector is that $Y - X^*$ be orthogonal (\bot) to \mathscr{L} .

In what follows denote by $\mathscr{P}(Y|\mathscr{L})$ the projection of Y onto \mathscr{L} , that is $\mathscr{P}(Y|\mathscr{L}) = X^*$.

THEOREM 2.2. Let \mathcal{L} be a closed subspace of a Hilbert space \mathcal{H} . Suppose N is a closed subspace of \mathcal{L} so that $\mathcal{L} = N \oplus N^{\perp}$ where N^{\perp} is the orthogonal complement of N in \mathcal{L} . If $Y \in \mathcal{H}$ then

$$\mathcal{P}(\mathbf{Y}|\mathcal{L}) = \mathcal{P}(\mathbf{Y}|\mathbf{N}) + \mathcal{P}(\mathbf{Y}|\mathbf{N}^{\perp}).$$

3. LINEAR ESTIMATION OF A RISK PARAMETER

One of the key problems in credibility theory is the estimation of a risk parameter. Suppose $Y \in \mathcal{H}$ is a (non-observable) risk parameter and X_0, X_1, \ldots, X_n are (observable) measurements in \mathcal{H} . A linear estimator of Y based on X_0, X_1, \ldots, X_n is any linear combination

$$Y^* = \sum_{i=1}^n a_i X_i, \qquad (a_i \in \mathbb{R})$$

with mean-square error

 $||Y - Y^*||^2$.

Denote by $\mathcal{L}_k = \mathcal{L}(X_0, X_1, \dots, X_k)$ the closed linear subspace spanned by the elements X_0, X_1, \dots, X_k . Also for notational simplification denote by $\mathcal{P}_k(\tilde{X})$ the projection $\mathcal{P}(\tilde{X}|\mathcal{L}_k)$, of \tilde{X} onto \mathcal{L}_k where $\tilde{X} \in \mathcal{H}$.

The following fundamental result is based on the projection theorem in \mathcal{H} . It is discussed in LUENBERGER (1969) and appears under various guises in NORBERG (1979) and references therein. It is included here for the sake of completeness.

THEOREM 3.1. Suppose $X = (X_0, X_1, \ldots, X_n)' \in \mathcal{H}^{n+1}$ and $\mathcal{P}_n(Y) = \hat{a}' X$ where $\hat{a} = (\hat{a}_0, \ldots, \hat{a}_n)'$.

Then,

$$\hat{a}' = \langle Y, X \rangle \|X\|^{-2}$$

and the mean-square error $||Y - \mathcal{P}_n(Y)||^2$ is

$$||Y - \mathcal{P}_n(Y)||^2 = ||Y||^2 - \langle Y, X \rangle ||X||^{-2} \langle X, Y \rangle.$$

PROOF. The projection theorem 2.1 gives

$$Y - \mathcal{P}_n(Y) \perp X_i; \qquad i = 0, 1, \ldots, n$$

whence,

$$\hat{a}'\langle X, X_i \rangle = \langle Y, X_i \rangle; \qquad i = 0, 1, \ldots, n.$$

The expression for \hat{a}' follows from the last set of equalities whereas the expression concerning the mean-square error follows by noting that $Y - \mathcal{P}_n(Y) \perp \mathcal{P}_n(Y)$. We remark that the matrix $G = ||X||^2$ is called the *Gram* matrix.

COROLLARY. If $X_0 \equiv 1$ then $\mathcal{P}_n(Y)$ is the inhomogeneous linear Bayes rule which may be written

$$\langle Y, 1 \rangle + C[Y, X^*]C^{-1}[X^*](X^* - \langle X^*, 1 \rangle)$$

with associated mean-square error (Bayes risk)

$$C[Y] - C[Y, X^*]C^{-1}[X^*]C[X^*, Y]$$

where the vector $X^* = (X_1, \ldots, X_n)$ and the covariances $C[\cdot, \cdot]$ and $C[\cdot]$ are defined as follows:

For any two vectors $U \in \mathcal{H}^n$ and $V \in \mathcal{H}^m$

$$C[U, V] = \langle U, V \rangle - \langle U, 1 \rangle \langle 1, V \rangle$$

and

$$C[U] = C[U, U].$$

We now discuss straightforward extensions of the abovementioned results to vector parameters.

Suppose $Y = (Y_1, \ldots, Y_m)' \in \mathcal{H}^m$ is a vector risk parameter to be estimated on the basis of the measurement vector $X = (X_0, X_1, \ldots, X_n) \in \mathcal{H}^{n+1}$. We restrict attention to linear estimators, namely $\sum a_{ij}X_j$, of each component Y_i of the vector

Y. Write, $A = (a_n)$, an $m \times n$ matrix. The optimal linear estimator minimizes

$$\sum_{i=1}^{m} \left\| Y_{i} - \sum_{j=1}^{n} a_{ij} X_{j} \right\|^{2}$$

over all matrices A of dimension $m \times n$.

THEOREM 3.2 (Luenberger). If $\hat{A}X$ is the optimal linear estimator of Y then

$$\hat{A} = \langle Y, X \rangle \| X \|^{-2}$$

and the error covariance matrix of $\hat{A}X$ is given by

$$\|\boldsymbol{Y} - \hat{\boldsymbol{A}}\boldsymbol{X}\|^2 = \|\boldsymbol{Y}\|^2 - \langle \boldsymbol{Y}, \boldsymbol{X} \rangle \|\boldsymbol{X}\|^{-2} \langle \boldsymbol{X}, \boldsymbol{Y} \rangle.$$

PROOF. The results follow from the observation that the optimization decomposes into a separate problem for each component Y_i of the risk parameter vector Y. The *i*th subproblem is simply that of finding $\mathcal{P}_n(Y_i)$ That is

$$AX = (\mathcal{P}_n(Y_1), \dots, \mathcal{P}_n(Y_m))'.$$
$$= \mathcal{P}_n(Y), \quad \text{say.}$$

We remark that trace $||Y - \hat{A}X||^2$ represents the mean-square error of $\mathcal{P}_n(Y)$. It is also known as the Bayes risk of $\mathcal{P}_n(Y)$ relative to squared error loss function.

COROLLARY 1. If T is a fixed $r \times m$ matrix then the optimal linear estimator of TY is $T\mathcal{P}_n(Y)$ with error covariance $T || Y - \mathcal{P}_n(Y) ||^2 T'$.

COROLLARY 2. If $X_0 \equiv 1$ then $\mathcal{P}_n(Y)$ is the inhomogeneous linear Bayes rule for Y, which may be written

$$\langle Y, I \rangle + C[Y, X^*]C^{-1}[X^*](X^* - \langle X^*, I \rangle)$$

with error covariance matrix,

$$C[Y] - C[Y, X^*]C^{-1}[X^*]C[X^*, Y].$$

All the foregoing results are well known to both linear filtering theorists and credibility theorists.

4. THE GEOMETRY OF RECURSIVE RISK PARAMETER ESTIMATION

In many practical situations the elements X_0, X_1, X_2, \ldots represent measurements taken sequentially in time. The optimal linear estimator of a risk parameter Y based on the measurements to time n, viz., X_0, X_1, \ldots, X_n is $\hat{Y}_n = \mathcal{P}_n(Y)$ with mean-square error

$$C_n = \|Y - \hat{Y}_n\|^2.$$

If X_{n+1} is the next measurement then its best linear estimator based on \mathscr{L}_n is $\mathscr{P}_n(X_{n+1})$. Accordingly, the innovation of the new information acquired at time n+1 is

$$e_{n+1} = X_{n+1} - \mathcal{P}_n(X_{n+1}).$$

Put $e_0 = X_0$ and write $e_0^* = e_0 / ||e_0||$, then by virtue of theorem 3.1

$$\mathcal{P}_0(X_1) = \langle X_1, e_0^* \rangle e_0^*$$

whence,

 $e_1 = X_1 - \langle X_1, e_0^* \rangle e_0^*.$

By virtue of the projection theorem 2.1, $e_0 \perp e_1$ and $\mathcal{L}_1 = \mathcal{L}(e_0, e_1)$. It follows that

$$e_2 = X_2 - \langle X_2, e_0^* \rangle e_0^* - \langle X_2, e_1^* \rangle e_1^*$$

where $e_1^* = e_1 / ||e_1||$.

Subsequently,

(4.1)
$$e_{n+1} = X_{n+1} - \sum_{j=0}^{n} \langle X_{n+1}, e_j^* \rangle e_j^*$$

where

$$e_j^* = e_j / ||e_j||; \qquad j = 0, 1, 2, \dots$$

We observe that the normalized innovations $\{e_j^*\}$ represent the orthonormal system obtained by the well-known Gram-Schmidt orthogonalization process. It follows, trivially, that the innovation sequence $\{e_i\}$ is orthogonal.

The closed linear subspace \mathscr{L}_{n+1} may be decomposed

(4.2)
$$\mathscr{L}_{n+1} = \mathscr{L}_n \oplus \mathscr{L}(e_{n+1}).$$

In view of the projection theorem 2.2,

(4.3)
$$\hat{Y}_{n+1} = \hat{Y}_n + \mathcal{P}(Y|e_{n+1})$$

where application of theorem 3.1 yields,

(4.4)
$$\mathscr{P}(Y|e_{n+1}) = \langle Y, e_{n+1} \rangle ||e_{n+1}||^{-2} e_{n+1}.$$

Alternatively decompose \mathscr{L}_{n+1} thus:

(4.5)
$$\mathscr{L}_{n+1} = \mathscr{L}(e_0) \oplus \cdots \oplus \mathscr{L}(e_{n+1}).$$

The Fourier series of Y based on \mathscr{L}_{n+1} is

(4.6)
$$\hat{Y}_{n+1} = \sum_{j=0}^{n+1} \langle Y, e_j^* \rangle e_j^*$$

whereas the Fourier series based on \mathscr{L}_n is

(4.7)
$$\hat{Y}_n = \sum_{j=0}^n \langle Y, e_j^* \rangle e_j^*.$$

The difference between expression (4.6) and (4.7) yields expression (4.3) We note that the key element in the foregoing analysis is the orthogonality property of the innovation sequence $\{e_i\}$.



FIGURE 4.1 The geometry of recursive risk parameter estimation

Figure 4.1 shows the geometry of recursive risk parameter estimation. The co-ordinate axis labelled 2 represents \mathcal{L}_n and the 1-2 plane represents \mathcal{L}_{n+1} . Observe that $Y - \hat{Y}_n \perp \mathcal{L}_n$, $Y - \hat{Y}_{n+1} \perp \mathcal{L}_{n+1}$, $e_{n+1} \perp \mathcal{L}_n$ and $e_{n+1} \perp \hat{Y}_n$.

We point out that if $X_0 \equiv 1$ then $1 \in \mathcal{L}_n$ whence we have the unbiasedness properties,

$$\langle Y - \hat{Y}_n, 1 \rangle = 0$$

(4.8) and

$$\langle e_{n+1}, 1 \rangle = 0.$$

Denote by C_n the mean-square error,

$$\|Y-\hat{Y}_n\|^2.$$

Examination of fig. 4.1 leads to

(4.9)
$$C_{n+1} = C_n - \|\mathscr{P}(Y|e_{n+1})\|^2$$

 $= C_n - \langle Y, e_{n+1} \rangle^2 ||e_{n+1}||^{-2}.$

Write,

(4.10)
$$K_{n+1} = \langle Y, e_{n+1} \rangle ||e_{n+1}||^{-2},$$

then equations (4.3) and (4.9) may be recast

(4.11) $\hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}e_{n+1}$

and

(4.12)
$$C_{n+1} = C_n - K_{n+1} \langle Y, e_{n+1} \rangle$$

respectively.

The preceding analysis also applies to the estimation of a vector risk parameter $Y \in \mathcal{H}^m$. Recall that,

$$\begin{split} \hat{\boldsymbol{Y}}_n &= \mathcal{P}_n(\boldsymbol{Y}) \\ &= (\mathcal{P}_n(\boldsymbol{Y}_1), \ldots, \mathcal{P}_n(\boldsymbol{Y}_m))' \end{split}$$

Let C_n represent the error covariance matrix of \hat{Y}_n . The following recursions are obtained.

(4.13)
$$\hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}e_{n+1}$$

and

(4.14)
$$C_{n+1} = C_n - K_{n+1} \langle e_{n+1}, Y \rangle$$

where

(4.15)
$$K_{n+1} = \langle Y, e_{n+1} \rangle ||e_{n+1}||^{-2}.$$

Finally, we remark that the preceding recursions also carry over to vector valued measurements X_0, X_1, \ldots

5. THE GEOMETRY OF RECURSIVE PROSPECTIVE RATEMAKING

In the present section we adopt the general prospective rate-making formulation of GERBER and JONES (1975).

Let X_i represent the claims cost (or loss ratio, etc.) in the *i*th period. The premium forecast for period n + 1 based on the measurements $X_0(=1), X_1, \ldots, X_n$ is denoted by P_{n+1} . This premium is the optimal affine estimator (inhomogeneous linear Bayes rule) of X_{n+1} based on the measurements X_1, X_2, \ldots, X_n .

That is,

$$P_{n+1} = \mathcal{P}_n(X_{n+1}).$$

The innovation in the measurement X_n is

$$e_n = X_n - \mathcal{P}_{n-1}(X_n).$$

Since,

$$\mathscr{L}_n = \mathscr{L}_{n-1} \oplus \mathscr{L}(e_n)$$

we have,

(5.1)
$$\mathscr{P}_n(X_{n+1}) = \mathscr{P}_{n-1}(X_{n+1}) + \mathscr{P}(X_{n+1}|e_n)$$

where

(5.2)
$$\mathscr{P}(X_{n+1}|e_n) = \langle X_{n+1}, e_n \rangle ||e_n||^{-2} e_n.$$

In keeping with GERBER and JONES (1975) write

(5.3)
$$Z_n = \langle X_{n+1}, e_n \rangle ||e_n||^{-2}$$

whence,

(5.4)
$$P_{n+1} = \mathcal{P}_{n-1}(X_{n+1}) + Z_n(X_n - P_n).$$

We emphasize that the last formula holds true in general.

We now focus on formula (4) of SUNDT (1981) which examines the situation where there exist constants b_n , c_n and d_n such that

$$P_{n+1} = b_n + c_n P_n + d_n X_n$$

Combining this with formula (5.4) above yields

$$\mathcal{P}_{n-1}(X_{n+1}) = (c_n + Z_n)P_n + (d_n - Z_n)X_n + b_n$$

As $\mathcal{P}_{n-1}(X_{n+1})$ should not depend on X_n , we must have $d_n = Z_n$, and as $P_n = \mathcal{P}_{n-1}(X_n)$ we obtain

(5.5)
$$\mathscr{P}_{n-1}(X_{n+1}) = a_n \mathscr{P}_{n-1}(X_n) + b_n$$

with $a_n = c_n + Z_n$. That is, the premium forecast for period n+1 based on \mathcal{L}_{n-1} is an affine function of the premium forecast for period n also based on \mathcal{L}_{n-1} . Since the innovations $\{e_j\}$ are orthogonal, $\mathcal{P}_{n-1}(X_{n+1})$ and $\mathcal{P}_{n-1}(X_n)(=P_n)$ have the Fourier series representations

(5.6)
$$\mathscr{P}_{n-1}(X_{n+1}) = \sum_{i=0}^{n-1} \langle X_{n+1}, e_i^* \rangle e_i^*$$

and

(5.7)
$$\mathscr{P}_{n-1}(X_n) = \sum_{i=0}^{n-1} \langle X_n, e_i^* \rangle e_i^*$$

where we recall that the sequence $\{e_i^*\}$ represents the orthonormal innovations. Substituting (5.6) and (5.7) into (5.5) gives

(5.8)
$$\langle X_{n+1}, e_i^* \rangle = a_n \langle X_n, e_i^* \rangle; \quad i = 1, ..., n-1$$

and

(5.9) $\langle X_{n+1}, e_0^* \rangle = a_n \langle X_n, e_0^* \rangle + b_n.$

Also let

(5.10)
$$b_n = E[X_{n+1}] - a_n E[X_n].$$

We are now in a position to derive the covariance characterization (5) of SUNDT (1981, p. 5).

Equation (5.8) can be written as

$$\langle X_{n+1}, e_i \rangle = a_n \langle X_n, e_i \rangle$$

and insertion of (4.1) gives

$$\left\langle X_{n+1}, X_i - \sum_{j=0}^{i-1} \langle X_j, e_j^* \rangle e_j^* \right\rangle = a_n \left\langle X_n, X_i - \sum_{j=0}^{i-1} \langle X_j, e_j^* \rangle e_j^* \right\rangle$$

that is,

$$\langle X_{n+1}, X_i \rangle - \sum_{j=0}^{i-1} \langle X_i, e_j^* \rangle \langle X_{n+1}, e_j^* \rangle = a_n \left[\langle X_n, X_i \rangle - \sum_{j=0}^{i-1} \langle X_i, e_j^* \rangle \langle X_n, e_j^* \rangle \right]$$

Combining the last equation with (5.8) we obtain

$$\langle X_{n+1}, X_i \rangle - \langle X_i, e_0^* \rangle \langle X_{n+1}, e_0^* \rangle = a_n [\langle X_n, X_i \rangle - \langle X_i, e_0^* \rangle \langle X_n, e_0^* \rangle],$$

that is,

(5.11)
$$C[X_{n+1}, X_i] = a_n C[X_n, X_i]; \quad i = 1, ..., n-1.$$

The converse is straightforward.

The case for which $a_n \equiv 1$ and $b_n \equiv 0$ in (5.5) makes (5.4) a credibility formula of the updating type in the spirit of GERBER and JONES (1975). Equation (5.11) now reduces to the covariance structure.

(5.12)
$$C[X_{i}, X_{j}] = \begin{cases} V_{i} + W_{i}; & i = j \\ V_{i}; & i < j \end{cases}$$

in agreement with GERBER and JONES (1975).



FIGURE 51 The geometry of credibility formulae of the updating type in the spirit of GERBER and JONES (1975)

Figure 51 shows the geometry of credibility formulae of the updating type. The co-ordinate axis labelled 2 represents X_n and the 1-2 plane represents \mathcal{L}_n . Let E_n represent the mean-square error of P_n , that is

$$E_n = ||e_n||^2$$
.

Figure 5.1 depicts the following orthogonality relations: $e_n \perp e_{n+1}$, $e_n \perp P_n$ and $e_{n+1} \perp \mathcal{L}_n$.

These may be used to obtain a number of expressions connecting Z_n and second-order moments of X_n , e_n , P_n etc. In particular

(5.13)
$$E_{n+1} = ||X_{n+1} - X_n||^2 - (1 - Z_n)^2 E_n$$

assuming (5.12) holds.

We can also demonstrate (5.13) mathematically thus: From expression preceding (5.1)

$$X_{n+1} - X_n = P_{n+1} - P_n + e_{n+1} - e_n.$$

Substituting (5.4) with $\mathcal{P}_{n-1}(X_{n+1}) = P_n$ into the last equation gives

$$X_{n+1} - X_n = (1 - e_n)Z_n + e_{n+1}.$$

Recognizing the fact that $e_n \perp e_{n+1}$ now yields (5.13).

GERBER and JONES (1975) also derive the relations,

(5.14)
$$Z_1 = W_1 (W_1 + V_1)^{-1}$$

(5.15)
$$Z_n = (W_n - W_{n-1} + Z_{n-1}V_{n-1})(W_n - W_{n-1} + Z_{n-1}V_{n-1} + V_n)^{-1}$$

which will be alluded to in the next section.

6. KALMAN TYPE FILTERS

In the present section we examine some applications of the algorithms developed earlier to two special models and relate them to the classical Kalman filter for which both measurement and system noises are Gaussian (JAZWINSKI (1969)).

6.1. Bühlmann Model

Consider a risk characterized by a parameter Y. Associated with this risk are measurements X_1, X_2, \ldots The following assumptions are made:

ASSUMPTION 1. Conditional on Y fixed, the measurements X_1, X_2, \ldots are independent.

ASSUMPTION 2. Conditional on Y fixed, the mean and variance of each X, can be written $E[X_i|Y] = \mu(Y)$ and $C[X_i|Y] = \sigma^2(Y)$ respectively.

Without loss of generality assume $\mu(Y) = Y$ and for notational convenience write, $y_0 = E[Y]$, $\sigma_0^2 = E[\sigma^2(Y)]$ and v = C[Y]

ZEHNWIRTH

Adopting the notation and terminology of the preceding sections, the recursion for the inhomogeneous linear Bayes rule for Y is,

(6.1.1)
$$\hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}e_{n+1}.$$

In view of assumptions 1 and 2 and the unbiasedness conditions (4.8) it follows that

$$\langle X_{n+1}, e_j \rangle = \langle Y, e_j \rangle; \qquad j = 0, 1, \ldots, n.$$

This means,

$$\mathcal{P}_n(X_{n+1}) = \mathcal{P}_n(Y)(=P_{n+1})$$

whence,

$$e_{n+1} = X_{n+1} - \hat{Y}_n$$

Consider now the inner product,

$$\langle Y, e_{n+1} \rangle = \langle Y - \hat{Y}_n + \hat{Y}_n, e_{n+1} \rangle$$
$$= \langle Y - \hat{Y}_n, e_{n+1} \rangle,$$

the latter equality following from $\hat{Y}_n \perp e_{n+1}$. Further,

$$\langle Y - \hat{Y}_n, e_{n+1} \rangle = \langle Y - \hat{Y}_n, Y - \hat{Y}_n + X_{n+1} - Y \rangle$$

= C_n

the latter equality following from the orthogonality condition

$$Y - \hat{Y}_n \perp X_{n+1}$$

as a result of (4.8) and assumptions 1 and 2. Consequently,

 $(6.1.2) \qquad \langle Y, e_{n+1} \rangle = C_n.$

Moreover,

(6.1.3)
$$\|e_{n+1}\|^2 = \|X_{n+1} - \hat{Y}_n\|^2$$
$$= \|Y - \hat{Y}_n\|^2 + \|X_{n+1} - Y\|^2$$
$$= C_n + E[C[X_{n+1}|Y]]$$
$$= C_n + \sigma_0^2$$

where the second equality follows from

 $Y - \hat{Y}_n \perp X_{n+1} - Y.$

Substituting (6.1.2) and (6.1.3) into (4.10) yields,

(6.1.4) $K_{n+1} = C_n (C_n + \sigma_0^2)^{-1}.$

For continuity write (4.12) again,

$$(6.1.5) C_{n+1} = C_n - K_{n+1}C_n$$

Substituting (6.1.4) into (6.1.5) gives

(6.1.6)
$$C_{n+1}^{-1} = C_n^{-1} + \sigma_0^{-2}.$$

In summary, we have developed the Kalman filter

(6.1.7)
$$\hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}(X_{n+1} - \hat{Y}_n)$$

(6.1.8)
$$K_{n+1} = C_n (C_n + \sigma_0^2)^{-1}$$

(6.1.9)
$$C_{n+1}^{-1} = C_n^{-1} + \sigma_0^{-2}$$

with initial conditions

$$\hat{Y}_0 = y_0$$
 and $C_0 = v$.

We point out that if Y has a Gamma distribution and $X_i|Y$ is Poisson with mean Y (implying that $\sigma^2(Y) = Y$ and $y_0 = \sigma_0^2$) then

 $\hat{Y}_n = E[Y|X_1, \dots, X_n]$

and

$$C_n = E[C[Y|X_1,\ldots,X_n]].$$

Moreover, by virtue of a fundamental result in linear Bayes theory (HARTIGAN (1969)), the same classical Kalman filter (6.1.7) to (6.1.9) is obtained if we assume instead that

 $Y \sim \text{Normal}(y_0, v)$

and

$$X_n | Y \sim \text{Normal}(Y, \sigma_0^2).$$

See DE JONG and ZEHNWIRTH (1983) for more details.

In passing we also note that since $\mathcal{P}_n(X_{n+1}) = \mathcal{P}_n(Y)$ it follows that

$$K_{n+1} = Z_{n+1}.$$

Combining (6.1.4) and (6.1.5) gives

$$Z_{n+1} = Z_n (Z_n + 1)^{-1},$$

which is also a consequence of expression (5.15). Moreover,

(6.1.10)
$$P_{n+1} = P_n + Z_n (X_n - P_n)$$

and

(6.1.11)
$$E_{n+1} = [(E_n - \sigma_0^2)^{-1} + \sigma_0^{-2}]^{-1} + \sigma_0^2.$$

The last expression also follows from (5.13).

6.2. Evolutionary Risk Parameter Model

In the present sub-section we imagine that we have a sequence of risk parameters Y_1, Y_2, \ldots and corresponding measurements X_1, X_2, \ldots . The measurement

equations are given by

Assumption 1

$$X_n | Y_n \sim \text{Poisson}(Y_n)$$

and

$$C[X_i, X_j | Y_i, Y_j] = 0, \qquad i \neq j$$

The system equations (that is, the equations indicating how the parameters evolve over time) are given by

ASSUMPTION 2

$$E[Y_n|Y_{n-1}] = Y_{n-1}$$

and

$$C[Y_n | Y_{n-1}] = \nu_n, \qquad (\nu_n \in \mathbb{R}).$$

We also assume independence between the measurement and system "noises". That is,

Assumption 3

 $C[X_n, Y_{n+1}|Y_n] = 0.$

Now, put $E[Y_1] = y_0$, a constant, and write

$$C_{n+1|n} = \|Y_{n+1} - \hat{Y}_n\|^2$$

and

$$C_{n+1} = \|Y_{n+1} - \hat{Y}_{n+1}\|^2$$

where in the present context,

$$\hat{Y}_n = \mathcal{P}_n(Y_n).$$

Applying the projection theorem to the decomposition (4.2) gives

$$\mathcal{P}_{n+1}(Y_{n+1}) = \mathcal{P}_n(Y_{n+1}) + K_{n+1}e_{n+1}$$

where now

$$K_{n+1} = \langle Y_{n+1}, e_{n+1} \rangle ||e_{n+1}||^{-2}$$

In view of assumptions 2 and 3

$$\langle Y_{n+1}, e_j \rangle = \langle Y_n, e_j \rangle; \qquad j = 0, 1, \dots, n$$

whence,

$$\mathcal{P}_n(Y_{n+1}) = \mathcal{P}_n(Y_n).$$

Similarly, in view of assumption 1

$$\mathscr{P}_n(X_{n+1}) = \mathscr{P}_n(Y_{n+1})$$

It follows that,

(6.2.1)
$$\hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}e_{n+1}$$

where

$$e_{n+1} = X_{n+1} - \hat{Y}_n.$$

Consider the inner product

(6.2 2)
$$\langle Y_{n+1}, e_{n+1} \rangle = \langle Y_{n+1} - \hat{Y}_n + \hat{Y}_n, e_{n+1} \rangle$$
$$= \langle Y_{n+1} - \hat{Y}_n, e_{n+1} \rangle$$
$$= \langle Y_{n+1} - \hat{Y}_n, Y_{n+1} - \hat{Y}_n + X_{n+1} - \hat{Y}_{n+1} \rangle$$
$$= C_{n+1|n},$$

the second equality follows by noting that

$$\hat{Y}_n \perp X_{n+1} - \hat{Y}_n$$

and the last equality follows by noting that

$$Y_{n+1} - \hat{Y}_n \perp X_{n+1} - \hat{Y}_{n+1}.$$

Now,

$$C_{n+1|n} = ||Y_{n+1} - Y_n + Y_n - \hat{Y}_n||^2$$
$$= C + ||Y_{n+1} - Y_n||^2,$$

since

$$Y_{n+1} - Y_n \perp Y_n - \hat{Y}_n.$$

Hence,

(6.2.3)
$$C_{n+1|n} = C_n + \nu_n.$$

Turning now to the computation of $||e_{n+1}||^2$ we have

(6.2.4)
$$\|e_{n+1}\|^2 = \|Y_{n+1} - \hat{Y}_n\|^2 + \|X_{n+1} - Y_{n+1}\|^2$$
$$= C_{n+1|n} + E[Y_{n+1}]$$
$$= C_{n+1|n} + y_0,$$

the second equality following from assumptions 1 and 2.

Application of (4.9) with Y_{n+1} playing the role of Y gives

(6.2 5)
$$C_{n+1} = C_{n+1|n} - K_{n+1} \langle Y_{n+1}, e_{n+1} \rangle.$$

Combining equations (6.2.1) to (6.2.5) yields the Kalman filter

(6.2.6)
$$\hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}(X_{n+1} - \hat{Y}_n)$$

(6.2.7)
$$K_{n+1} = C_{n+1|n} (C_{n+1|n} + y_0)^{-1}$$

(6.2.8)
$$C_{n+1}^{-1} = C_{n+1|n}^{-1} + y_0^{-1}$$

and

(6.2.9)
$$C_{n+1|n} = C_n + \nu_n$$

with initial conditions

$$\hat{Y}_0 = y_0$$
 and $C_{1|0} = \nu_0$.

We point out again the connection with the classical Kalman filter That is, if instead of assumption 1 we have:

Assumption 1

$$X_n | Y_n \sim \text{Normal} (Y_n, y_0).$$

In addition to assumptions 2 and 3 we also assume

Assumption 4

$$Y_n | Y_{n-1} \sim \text{Normal} (Y_{n-1}, \nu_n).$$

The same Kalman filter (6.2.6) to (6.2.9) is obtained. The prospective rating algorithm is given by

(6.2.10)
$$P_{n+1} = P_n + Z_n (X_n - P_n)$$

and

(6.2.11)
$$E_{n+1} = \{(E_n - y_0)^{-1} + y_0^{-1}\}^{-1} + y_0$$

where again

$$Z_n = K_n$$

Although the two preceding models satisfy $(5\ 12)$ we conclude by emphasizing that the general algorithms presented in Sections 4 and 5 may be applied to any model and in particular the models considered by SUNDT (1981, 1983) satisfying the more general structure (5.10) and (5.11).
REFERENCES

DE JONG, P and ZEHNWIRTH, B (1983) Credibility Theory and the Kalman Filter Insurance Mathematics and Economics 2, 281-286

DE VYLDER, F (1976) Geometrical credibility Scandinavian Actuarial Journal 121-149

- GERBER, H U and JONES, D A (1975) Credibility formulas of the Updating Type In Credibility Theory and applications, edited by P M Kahn Academic Press New York, pp 89-105
- HARTIGAN, J. A. (1969) Linear Bayesian Methods. Journal Royal Statistical Society Series B 31, 446-454
- JAZWINSKI, A H (1970) Stochastic Processes and Filtering Theory Academic Press New York
- JEWELL, W S (1976) Two Classes of Covariance Matrices Giving Simple Linear Forecasts Scandinavian Actuarial Journal, 15-29
- KAILATH, T (1974) A View of Three Decades of Linear Filtering Theory IEEE Trans Inform Thy 19, 750-760
- KREMER, E (1982) Exponential Smoothing and Credibility Theory Insurance Mathematics and Economics 1, 213-217
- LUENBERGER, D G (1969) Optimization by vector space methods John Wiley New York
- NORBERG, R (1979) The Credibility Approach to Experience Rating Scandinavian Actuarial Journal, 181-221

SUNDT, B. (1981) Recursive Credibility Estimation Scandinavian Actuarial Journal, 3-21

SUNDT, B (1983) Finite Credibility Formulae in Evolutionary Models Scandinavian Actuarial Journal, 106-116

UNBAYESED CREDIBILITY REVISITED

By RAGNAR NORBERG University of Oslo, Norway

Abstract

The unbayesed credibility procedure proposed by Gerber is revisited. Its performance is discussed, connections are drawn to earlier literature, and some possible ideas of generalizations are investigated (and found fruitless).

Keywords

Unbayesed credibility, principles of statistical decisions.

1. INTRODUCTION

More than two years have elapsed since GERBER (1982) proposed a procedure for construction of estimators highlighted as Unbayesed Credibility. During this time there has been published no further work on the topic. It is, therefore, worth while having another look at the unbayesed estimations to throw some light on their properties and to inquire if further ideas ought to be pursued along the same lines.

In Section 2 of the present paper two estimation problems considered by GERBER (1982) and Gerber's unbayesed approach to their solution are briefly recapitulated. The properties of the two unbayesed estimators are discussed in Section 3; it is shown that one of them will usually have an infinite expected squared loss. Section 4 stresses the need to build adequate mathematical models and to work strictly within these in search for methods. In particular the properties of any proposed method has to be examined in terms of the performance criterion adopted. Section 5 presents a couple of variations of the unbayesed approach which show that it can lead to many different estimators; the particular form of any unbayesed estimator is due to arbitrary restrictions imposed on the estimating functions rather than being due to the structure of the model itself.

2. REVIEW OF THE UNBAYESED ESTIMATION PROCEDURE

In order to make our presentation fairly selfcontained and to state points clearly, let us recall the unbayesed set-up in neutral mathematical terms. The model framework in Sections 4 and 6 of Gerber's paper is the following.

MODEL. Let X_y , i = 1, ..., m(>1), j = 1, ..., n, be a collection of real random variables For each *i* the X_y , j = 1, ..., n, have the same distribution, which we denote by F_i . All X_y are mutually independent, and $F = (F_1, ..., F_m) \in$

ASTIN BULLETIN Vol 15, No 1

NORBERG

 $\mathscr{F} \times \cdots \times \mathscr{F}^m$, where \mathscr{F} is the nonparametric family of all those distributions on the real line which possess a finite second order moment.

Let μ and σ^2 denote the mean and the variance, which are well defined functionals on \mathcal{F} , and put $\mu_i = \mu(F_i)$ and $\sigma_i^2 = \sigma^2(F_i)$, i = 1, ..., m.

The vector of means, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$, is to be estimated. More precisely, let \mathscr{P} denote the class of all measurable *m*-vectorvalued functions of (only) the X_y 's; we seek a $\boldsymbol{P} = (P_1, \ldots, P_m) \in \mathscr{P}$ that is in some sense close to $\boldsymbol{\mu}$. Gerber considers two different measures of closeness, hence two problems, the first of which is the following (numbers in square brackets refer to formulas in Gerber's paper):

PROBLEM 1 Pick $P \in \mathcal{P}$ such that

(1) [20]
$$\sum_{i=1}^{m} E_{F}(P_{i} - \mu_{i})^{2}$$

is "small" (not "minimum" as stated by Gerber, see second remark below).

(Here E_F denotes the integral with respect to $F_1 \times \cdots \times F_1 \times \cdots \times F_m \times \cdots \times F_m$, the joint distribution of the X_{ij} 's.)

A couple of remarks are in order at this stage. First a formal one: Gerber phrases his problem as that of predicting, for each *i*, a future independent selection $X_{i,n+1}$ from F_i , the performance of a set of predictors P_i being measured by (1) with μ_i replaced by $X_{i,n+1}$. That problem is, however, equivalent to the one stated here because

$$E_F(P_i - X_{i,n+1})^2 = \sigma_i^2 + E_F(P_i - \mu_i)^2.$$

The second remark concerns realities As it stands, problem 1 is not properly stated since for each choice of P the expression in (1) is a functional depending on F. One cannot find a P minimizing (1) for all F (the choice $P_i = \mu(G_i)$, with $G_i \in \mathcal{F}, i = 1, ..., m$, is optimal in $F = (G_1, ..., G_m)$, but poor in other points F where $\sum_{i=1}^{m} {\{\mu(G_i) - \mu(F_i)\}}^2$ is large). Thus, still loosely speaking, we can only require of P that (1) should not be "too large" in "too many" points F. We leave these considerations for the time being and continue our recapitulation of the unbayesed approach.

Gerber constructs his unbayesed credibility estimator in the following manner:

METHOD 1. (i) As a first step, solve the simple problem of minimizing (1) as $P = (P_1, \ldots, P_m)$ ranges in the class \mathcal{P}' of functions with P_i of the form

where

$$\bar{X}_{i} = \frac{1}{n} \sum_{j=1}^{n} X_{ij} ,$$
$$\bar{X} = \frac{1}{m} \sum_{h=1}^{m} \bar{X}_{h} ,$$

38

and Z is a function of F (only). Minimum is readily found to be attained at

(3) [21]
$$Z = \frac{n \sum_{h=1}^{m} (\mu_h - \bar{\mu})^2}{n \sum_{h=1}^{m} (\mu_h - \bar{\mu})^2 + (m-1)\sigma^2},$$

where

$$\bar{\mu} = \frac{1}{m} \sum_{h=1}^{m} \mu_h$$
 and $\sigma^2 = \frac{1}{m} \sum_{h=1}^{m} \sigma_h^2$.

(ii) As a second step, replace numerator and denominator in (3) by their "natural" unbiased estimators [13] and [22] (Gerber's Section 4). Then the right-hand side expression in (3) turns into

(4) [23]
$$\hat{Z} = 1 - \frac{(m-1)\hat{\sigma}^2}{n\sum_{h=1}^m (\bar{X}_h - \tilde{X})^2}$$

with $\hat{\sigma}^2 = \sum_{h=1}^m \hat{\sigma}_h^2 / m$ and $\hat{\sigma}_h^2 = \sum_{j=1}^n (X_{hj} - \bar{X}_h)^2 / (n-1)$. Upon replacing Z in (2) by \hat{Z} , we obtain a function $\hat{P} \in \mathcal{P}$, which is the unbayesed credibility estimator.

The second problem is the following.

PROBLEM 2. The same as problem 1, but with (1) replaced by the componentwise expected squared error

(5) [27]
$$E_F(P_i - \mu_i)^2, \quad i = 1, ..., m,$$

(vector-valued)

The above remarks to problem 1 apply also to problem 2. The unbayesed procedure follows the same outline as in method 1:

METHOD 2. First minimize (5) as $P = (P_1, \ldots, P_m)$ ranges in the class \mathcal{P}'' of functions with P_i of the form

$$Z_{t}\bar{X}_{t}+(1-Z_{t})\bar{X},$$

where each Z_i is a function of F. Proceeding in analogy to step (ii) of method 1, Gerber arrives at the estimator \hat{P} given by

(7)
$$\hat{\vec{P}}_{i} = \bar{X}_{i} - \frac{m-1}{mn} \frac{\hat{\sigma}_{i}^{2}}{(\bar{X}_{i} - \bar{X})}, \quad i = 1, ..., m.$$

Having summarized the present state of unbayesed credibility, we now set forth to study its merits in terms of concepts from estimation theory.

3 PROPERTIES OF THE UNBAYESED ESTIMATORS

We first consider problem 1 and the unbayesed estimator given by (2) and (4).

It ought, perhaps, to be said that it is unfortunate to speak of \hat{P} as the "real solution" to the problem of minimizing (1) (Gerber's Section 4), confer the second

remark to problem 1 above. Clearly, method 1 is only a preparatory piece of motivating heuristics, and the resulting \hat{P} is so far only a candidate estimator, whose performance has to be examined in terms of the criterion (1). This task has not been undertaken—in fact, not even mentioned—in the previous literature on unbayesed credibility, and no references are made to the closely related literature on compound estimation problems. Therefore, a few remarks are added here on these matters:

Very little is known about the possibility of solving (reasonably precise versions of) problem 1 under the present model with nonparametric \mathcal{F} . Some results on restricted inadmissibility have been established: For certain simple parametric subfamilities $\mathcal{F}_0 \subset \mathcal{F}$ one can construct estimators that in all of \mathcal{F}_0^m dominate old established estimators known to be uniformly optimal on \mathcal{F}_0^m with respect to the traditional performance criterion (5) when one restricts to the class of unbiased estimators. The first results of this kind appeared in fundamental papers by STEIN (1956) and JAMES and STEIN (1961). They considered the subfamily \mathcal{F}_0 of all normal distributions with variance 1 (say) and proved that the estimator

$$\boldsymbol{P^*} = (\bar{X}_1, \ldots, \bar{X}_m),$$

which is admissible on \mathscr{F}_0^m with respect to (5) and furthermore is uniformly minimum variance unbiased, does not even remain admissible when criteron (1) is adopted. If $m \ge 3$, it is dominated by the so-called James-Stein estimator P^{**} defined by

$$P_{i}^{**} = \left(1 - \frac{m-2}{n \sum_{h=1}^{m} \bar{X}_{h}^{2}}\right) \bar{X}_{i}, \qquad i = 1, \dots, m.$$

To most statisticians this result came as a surprise, to some even as an unpleasant one, and there were signs of controversies between defenders of the traditional P^* on the one side and advocates of the new P^{**} on the other Now there is no reason to discuss which is the better of P^* and P^{**} (on \mathcal{F}_0^m), because that question is settled by emotionless mathematics once the performance criterion is chosen. What can be discussed, is only the choice of criterion. That discussion is, however, not of a purely mathematical nature, but depends on the goals and attitudes of the decision maker.

In closing our comments on problem 1, we note that admissibility on \mathscr{F}_0^m of estimators of the James-Stein type has been extensively treated in the literature, see e.g., BERGER (1976). A survey of James-Stein estimations is given by EFRON and MORRIS (1973)

Let is now examine the unbayesed estimator designed for problem 2. Again by the second remark to problem 1, it is clear that \hat{P} given by (7) does not represent the solution to the problem of minimizing (5) (Gerber's Section 6). In fact, by inspection of (7), it is readily seen that \hat{P} , assigns the value $+\infty$ to (5) on wide subsets of \mathscr{F}^m : If, for instance, the F_i are normal distributions, then $\hat{\sigma}_i^2$ is independent of $\bar{X}_i - \bar{X}$, and $(\bar{X}_i - \bar{X})^{-1}$ has no expected value. More generally, if the marginal distribution of $\bar{X}_i - \bar{X}$ has a point of increase in 0, then \hat{P}_i and hence $(\hat{P}_i - \mu_i)^2$ are usually not integrable. Problem 2 in the present nonparametric model is one of the classics of statistics, and to the knowledge of the present author no alternatives to the natural unbiased estimator (8) have been proposed in the pre-unbayesed literature. Thus, in this case it would really be surprising if \hat{P} could be shown to have any good properties. And unpleasant as the interpretation of the model and problem 2 is that the estimation problems are unrelated in every respect; the samples are drawn in an independent manner from populations that have nothing in common, and the losses incurred by error of estimation are measured separately for each problem. A reasonable task for the theory would be to put a firm basis to the intuitive feeling that the estimator of μ_i should depend only on X_{i1}, \ldots, X_{in} . If we are not able to justify the deletion of the X_{hy} , $h \neq i$, from the estimation of μ_i , then we would be in serious trouble: How could we then in a rational way choose the statistical basis for a given estimation problem? Which irrelevant data were not to be included? Which advice should we give to the practitioners?

As we have seen, the unbayesed approach gives rise to no such concerns. The traditional and intuitively sound estimator (8) remains an uncontested answer to problem 2.

4. MODEL AND METHOD

After the discussions in section 3 the question arises: What brought Gerber to enter the X_{hp} , $h \neq i$, into the estimation of μ_i ? Why didn't he use the "natural unbiased estimator" (8)? The reason seems to be that he had a particular interpretation in mind; X_{i1}, \ldots, X_{in} are spoken of as being the claim amounts in *n* different years for risk no. *i* in a portfolio of *m* insured risks. A few remarks on basic principles of statistical decisions are called for:

The first step in a statistical analysis is to separate out of the situation those features that are believed to have some bearing on the problem and work them into a mathematical model. The model should give a surveyable and, as far as possible, true picture of the phenomena. If, for instance, the data stem from similar automobile insurance risks, the model ought to give precise content to the notion of similarity between these risks. The model in Section 2 fails to reflect the essential circumstance that automobile insurance claims have something in common that distinguishes them from data on e.g. soldiers' heights and turnover of cheese. One reasonable mathematical means of expressing this similarity between the risks is to regard them as selections from one and the same structure distribution (population). Thus the structure distribution is not "essentially superfluous" (Section 4 of Gerber's paper) to those who think they can learn something about a given risk by looking at other risks of the same kind.

Having decided on a model, the purpose of the decision has to be expressed in terms of a performance criterion. When this is done, one is left with the purely mathematical problem of finding decision functions with good performance.

In the traditional credibility analysis based on models with structure distributions, credibility estimators are obtained as logical consequences of the mathematical set-up. They are justified by their (restricted) optimality properties.

NORBERG

This is not the case with unbayesed approach in method 2 above. There the particular credibility appearance of the estimator could only be obtained by the analyst's intervening into the mathematics by prescribing a certain procedure and exempting it from the requirements expressed by the performance criterion. The following quotation from Neyman (1954) seems pertinent: "... the efforts of the representatives of modern statistical theory are directed towards solving problems that depend only on the stochastical model studied and on nothing else".

Considered as a statistical framework for the analysis of related risks, problem 1 together with the model in Section 2 plays an intermediate role. As explained above, the model may be judged as inadequate, but it still represents a reasonable partial description of the situation. The connection between the different rating problems is now established through the choice of the performance criterion (1). The unbayesed method 1, however, is until further without any support whatsoever in studies of its performance.

5. Some variations of the unbayesed procedure

Until the estimator resulting from method 1 has been investigated with respect to performance, it can, of course, not be excluded. But the unbayesed device as such can be put on test in other ways. One angle of attack arises from noting that the requirement that Z in (2) be independent of i is quite arbitrary.

Looking for good estimators, we could possibly gain something by allowing Z to depend on i, that is, let P be of the form (6). But then the unbayesed procedure reduces to that of method 2 and delivers (7), which maximizes the expected loss instead of minimizing it as pointed out already in Section 3.

Let us allow for further flexibility and admit nonhomogeneous estimators of μ_i of the form

$$P_{i} = Z_{i0} + \sum_{h=1}^{m} Z_{ih} \bar{X}_{h}.$$

Then we find by the first step of the unbayesed prescription that the optimal approximation to μ_i is μ_i . And following the recipe further, we now only have to estimate μ_i by a natural unbiased estimator. Then we obtain finally (8), which would have resulted immediately if we already at the outset looked for a natural unbiased estimator.

Which ends the present discussion. Some further considerations around the topic of unbayesed estimations can be found in an unpublished report, NORBERG (1983), which can be received upon request.

REFERENCES

- BERGER, J O (1976) Admissible Minimax Estimation of a Multivariate Normal Mean with Arbitrary Quadratic Loss Annals of Statistics 4, 223-226
- EFRON, B and MORRIS, C. (1973) Stein's Estimation Rule and its Competitors—An Empirical Bayes Approach Journal of the American Statistical Association 68, 117-130
- GERBER, H U (1982) An Unbayesed Approach to Credibility Insurance. Mathematics & Economics 1, 271-276
- JAMES, W and STEIN, C (1961) Estimation with Quadratic Loss Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, 361-379 Univ of California Press
- NEYMAN, J (1954) Current Problems of Mathematical Statistics Proceedings of the International Congress of Mathematicians, Amsterdam 1954, Vol 1
- NORBERG, R. (1983) An Unbased Approach to Credibility, Statistical research report No 1, 1983 Institute of Mathematics, University of Oslo
- STEIN, C (1956) Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability 1, 361-379. Univ of California Press

CORRIGENDUM

B. Sundt (1982). Asymptotic behaviour of compound distributions and stop-loss premiums. *ASTIN Bulletin* 13, 89-98.

In the proof of Theorem 3, one uses that $g_s = H(s) - H(s+1)$ instead of the correct expression $g_s = H(s-1) - H(s)$. This means that the displayed asymptotic expression for g_s really concerns g_{s+1} , and that the correct result should be

$$g_s \sim \frac{1-p}{\nu} e^{-\kappa s}.$$

The author became aware of this error when reading Milidiu and Jewell (1984), whose Theorem 2 gives a generalization of this result.

REFERENCE

MILIDIU, R L and W S JEWELL (1984) Asymptotic approximation to compound negative-binomial distributions Submitted for publication

ASYMPTOTIC BEHAVIOUR OF COMPOUND DISTRIBUTIONS

By PAUL EMBRECHTS* Imperial College, London, U.K. Makoto Maejima Keio University, Yokohama, Japan

AND

JOZEF L. TEUGELS

Departement Wiskunde, KUL, Leuven, Belgium

Abstract

We improve on some results of SUNDT (1982) on the asymptotic behaviour of compound negative binomial distributions

Key Words

Compound negative binomial distributions, renewal theory, asymptotic estimates.

Consider the aggregate claims of an insurance company in a given period,

$$X = \sum_{i=1}^{N} Y_i$$

where the claim sizes $\{Y_i: i \in \mathbb{N}_0\}$ are i.i.d. non-negative random variables with $F(x) = P\{Y_1 \le x\}$ non-lattice (i.e., we assume the claim size distribution F to be non-discrete; take for instance F continuous), independent of the negative binomial claim arrival variable N. Then

$$p_n \equiv P\{N=n\} = {\alpha-1+n \choose n} p^n q^{\alpha}, \qquad n \in \mathbb{N}$$

where 0 , <math>p + q = 1 and $\alpha > 0$. Denote by f the Laplace-Stieltjes transform of F and assume that there exists a constant $\kappa > 0$ satisfying

(1)
$$p^{-1} = \int_0^\infty e^{\kappa x} dF(x)$$

and that

(2)
$$\nu = p \int_0^\infty x \, e^{\kappa x} \, dF(x) < \infty$$

i.e, $p^{-1} = f(-\kappa)$ and $|f'(-\kappa)| < \infty$. We now want to estimate $P\{X > x\}$ as $x \to \infty$.

* This research was supported by a grant from The Nuffield Foundation

ASTIN BULLETIN Vol 15, No 1

NOTATION. If f(n) and g(n) are two functions, in this paper we always abbreviate the statement $\lim_{n\to\infty} f(n)/g(n) = 1$ to $f(n) \sim g(n)$ as $n \to \infty$.

In SUNDT (1982), the following theorem was proved.

PROPOSITION 1 (SUNDT (1982), Theorem 5). If $e^{xx}P\{X > x\}$ is ultimately monotone, then

(3)
$$P\{X > x\} \sim (\kappa \Gamma(\alpha))^{-1} (q/\nu)^{\alpha} x^{\alpha-1} e^{-\kappa x}, \quad as \ x \to \infty.$$

The condition of ultimate monotonicity was needed because the proof in SUNDT (1982) used a Tauberian argument. In this note we want to prove that under (1) and (2), (3) always holds, as indeed was conjectured by Sundt.

THEOREM. Assume the negative binomial model above. If the claim size distribution F satisfies (1) and (2), then (3) holds.

Of course condition (2) is only needed to get a non-trivial statement in (3). The proof of the theorem differs entirely from the one given in SUNDT (1982) and essentially hinges on the following recent Blackwell type theorem for generalised renewal measures.

PROPOSITION 2 (EMBRECHTS, MAEJIMA AND OMEY (1984), Theorem 1a). Let a be a positive function such that $a(x) = x^{\beta}L(x)$, $\beta > -1$ and L slowly varying (that is, for all t > 0, $L(tx) \sim L(x)$ as $x \to \infty$). Let F be non-lattice. Then for all h > 0,

(4)
$$\sum_{n=1}^{\infty} a(n) P\{x < S_n \le x+h\} \sim h \mu^{-\beta-1} a(x), \quad as \ x \to \infty,$$

where $S_n = X_1 + \cdots + X_n$ is the random walk defined by F and μ the mean of F. Moreover, the convergence in (4) is uniform in h on compact sets.

A more general statement including $\beta \leq -1$ is given in EMBRECHTS, MAEJIMA and OMEY (1984).

PROOF OF THEOREM. Define the associated distribution or Esscher transform

$$F_{\kappa}(x)=(f(-\kappa))^{-1}\int_0^x e^{\kappa y}\,dF(y),\qquad x\ge 0.$$

One easily verifies that for all $n \ge 2$, integer, because of (1)

$$F_{\kappa}^{(n)}(x)=p^n\int_0^x e^{\kappa y}\,dF^{(n)}(y),\qquad x\ge 0,$$

(here (n) denotes the nth convolution, i.e., $G^{(n)} \equiv G^{*n}$).

Now

(5)
$$P\{X > x\} = \int_{x}^{\infty} e^{-\kappa y} d\left\{\sum_{n=0}^{\infty} p^{-n} p_n F_{\kappa}^{(n)}(y)\right\}.$$

By Stirling's formula, $a(n) = p^{-n}p_n \sim (q^{\alpha}/\Gamma(\alpha))n^{\alpha-1}$ as $n \to \infty$, satisfying the condition in proposition 2 with $\beta = \alpha - 1$. Hence it follows that, with

$$H(y) = \sum_{n=0}^{\infty} p^{-n} p_n F_{\kappa}^{(n)}(y), \qquad y \ge 0,$$
$$\forall h > 0: \ H(y+h) - H(y) \sim h\nu^{-\alpha} a(y), \qquad y \to \infty.$$

In this last expression, we use $a(\cdot)$ defined on the positive real numbers, this can be achieved most easily by a(x) = a([x]) where [] denotes integer part. Therefore by uniform convergence:

$$\forall \varepsilon > 0 \exists y^* \colon \forall y \ge y^* \text{ and } \forall h, 0 \le h \le 1, \text{ say:}$$

(6)
$$\frac{1-\varepsilon}{\Gamma(\alpha)}(q/\nu)^{\alpha}hy^{\alpha-1} \leq H(y+h) - H(y) \leq \frac{1+\varepsilon}{\Gamma(\alpha)}(q/\nu)^{\alpha}hy^{\alpha-1}.$$

Take now $x \ge y = y^*(\varepsilon)$, $\Delta > 0$ fixed then it follows from (5) and (6) that

$$P\{X > x\} = \int_{x}^{\infty} e^{-\kappa y} dH(y)$$

= $\sum_{k=0}^{\infty} \int_{x+k\Delta}^{x+(k+1)\Delta} e^{-\kappa y} dH(y)$
 $\lesssim \sum_{k=0}^{\infty} e^{-\kappa (x+k\Delta)} (H(x+(k+1)\Delta) - H(x+k\Delta)))$
 $\lesssim \frac{1+\varepsilon}{\Gamma(\alpha)} (q/\nu)^{\alpha} \sum_{k=0}^{\infty} e^{-\kappa (x+k\Delta)} (x+k\Delta)^{\alpha-1}\Delta$
 $\rightarrow \frac{1+\varepsilon}{\Gamma(\alpha)} (q/\nu)^{\alpha} \int_{x}^{\infty} e^{-\kappa y} y^{\alpha-1} dy, \text{ as } \Delta\downarrow 0.$

A similar argument, replacing $(1 + \varepsilon)$ by $(1 - \varepsilon)$, proves the converse inequality (\ge). Letting $\varepsilon \downarrow 0$ we get as $x \rightarrow \infty$

$$P\{X > x\} \sim \frac{1}{\Gamma(\alpha)} (q/\nu)^{\alpha} \int_{x}^{\infty} e^{-\kappa y} y^{\alpha-1} dy.$$

The theorem follows since

$$\int_{x}^{\infty} e^{-\kappa y} y^{\alpha-1} dy \sim \kappa^{-1} e^{-\kappa x} x^{\alpha-1}, \quad \text{as } x \to \infty.$$

Obviously, there is no mistery in assuming p_n to be negative binomial. The proof easily extends to more general situations. To give the reader some idea of the generality of our approach, below we present a fairly straightforward extension of our Theorem. For further details, the reader is referred to TEUGELS (1985) in which these and related questions in insurance will be discussed.

For instance, suppose $p_n = P\{N = n\}$ satisfies the following property: there exist $\mu > 1$, L slowly varying and $\gamma \in \mathbb{R}$ such that $\mu^n p_n \sim n^{\gamma} L(n)$ as $n \to \infty$ (in the

negative binomial case, $\mu = p^{-1}$, $L(n) \equiv (q^{\alpha}/\Gamma(\alpha))$ and $\gamma = \alpha - 1$). Again, define the Esscher transform (assumed to exist!) by

$$F_{\kappa}(x) = (f(-\kappa))^{-1} \int_0^{\infty} e^{\kappa v} dF(y)$$

where $\kappa = \kappa(\mu) > 0$ is the solution of

$$\mu=\int_0^\infty e^{\kappa x}\,dF(x).$$

If now the conditions (1) and (2) hold (with p^{-1} replaced by μ) then

$$P\{X > x\} \sim \frac{(-f'(-\kappa)/f(-\kappa))^{-\gamma-1}}{\kappa} e^{-\kappa x} x^{\gamma} L(x), \quad \text{as } x \to \infty.$$

In general, the behaviour of $P\{X > x\}$ will depend on the relationship between the asymptotic behaviour of $P\{N > n\}$ and 1 - F(x) as $n, x \to \infty$. A multitude of results exist, these are all reviewed in TEUGELS (1985). In a forthcoming paper, we also plan to return to the lattice case (i.e., when the claim size distribution is discrete).

REFERENCES

- EMBRECHTS, P, MAEJIMA, M and OMEY, E (1984) A Renewal Theorem of Blackwell Type Annals of Probabiluy 12, 561-570
- SUNDT, B (1982) Asymptotic Behaviour of Compound Distributions and Stop-Loss Premiums Astin Bulletin 13, 89-98
- TEUGELS, J L (1985) Selected Topics in Insurance Mathematics Lecture Notes Catholic University of Leuven (Belgium), Department of Mathematics, to appear

ON COMBINING QUOTA-SHARE AND EXCESS OF LOSS

By Lourdes Centeno*

Heriot-Watt University, Edinburgh, Scotland

Abstract

This paper considers reinsurance retention limits in cases where the cedent has a choice between a pure quota-share treaty, a pure excess of loss treaty or a combination of the two. Our primary aim is to find the combination of retention limits which minimizes the skewness coefficient of the insurer's retained risk subject to constraints on the variance and the expected value of his retained risk. The results are given without specifying precisely how the excess of loss reinsurance premium is calculated. It is also shown that, depending to some extent on the constraint on the variance, the solution to the problem is a pure excess of loss treaty if the excess of loss premium is calculated using the expected value or standard deviation principle but that this need not be true if the variance principle is used.

Keywords

Reinsurance, Quota-share, Excess of loss, skewness, coefficient of variation, constrained optimization.

1. INTRODUCTION

This paper considers reinsurance retention limits in cases where the cedent has a choice between a pure quota-share treaty, a pure excess of loss treaty or a combination of the two Such combinations occur in practice; see, for example, GERATHEWOHL (1980, Vol. 2, p. 371).

We assess the effects on the insurer of a particular combination of reinsurance treaties by considering three moment functions of the insurer's retained risk. These functions are the skewness coefficient and the variance of the insurer's net claims and the insurer's expected net profit. Our primary aim is to find the combination of retention limits which minimizes the skewness coefficient of the insurer's net claims, subject to a maximum value for the variance of the insurer's net claims and a minimum value for the insurer's expected net profit.

In Section 3 we show that the solution to this problem is unchanged if we replace the skewness coefficient by the coefficient of variation of the insurer's net claims.

Constrained optimization as a criterion for determining optimal retention limits has been used before, see BUHLMANN (1970, pp. 114-119), but not in relation to a combination of types of reinsurance. Combinations of types of reinsurance

* This research was supported by Fundação Calouste Gulbenkian and by Instituto Superior de Economia of Lisbon I wish to thank Howard Waters for his helpful advice and criticism

ASTIN BULLETIN Vol 15, No 1

CENTENO

have not often been discussed in the mathematical insurance literature; one notable exception is LEMAIRE, REINHARD and VINCKE (1981). There are some similarities between our paper and theirs, but also some important differences. For example, in our paper we allow the claim number distribution to be more general than the Poisson (for example negative binomial). There is also a difference in the way in which we assume the reinsurance premiums are calculated. We assume the quota-share premium is calculated on a proportional basis with a commission payment to the insurer; we do not specify how the excess of loss reinsurance premium is calculated but make some assumptions about this premium which are shown to be satisfied if it is calculated using the expected value, standard deviation or variance principles.

In Section 2 we describe in detail the two reinsurance treaties and discuss our assumptions relating to the excess of loss reinsurance premium.

In Section 3 we state our problems and give the solution in general form.

In Section 4 we give the solution to our problems assuming the excess of loss reinsurance premium is calculated using the expected value, the standard deviation or the variance principle. It is shown that, provided the constraint involving the variance of the insurer's retained risk is not too restrictive, the optimal solution is a pure excess of loss treaty in the first two cases but this need not to be true in the last case.

In Section 5 we discuss briefly the necessity of the assumptions made concerning the claim number distribution.

In Section 6 we give a numerical example to illustrate our results.

2. THE REINSURANCE ARRANGEMENT AND THE COST OF THE EXCESS OF LOSS REINSURANCE

2.1. The Reinsurance Arrangement

Consider a risk for which the aggregate gross (of reinsurance) claims in some fixed time interval are denoted by a random variable Y. We assume Y has a compound distribution, so that

$$Y = \sum_{i=1}^{N} X_i$$

where $\{X_i\}$, with $0 \le x_0 < X_i < x_1 \le +\infty$, is a sequence of i.i.d. random variables, with common distribution function F, representing the amounts of the individual claims and N is a random variable, independent of the X_i 's, representing the number of claims in the time interval. We shall assume that F is continuous and that the third moments of X_i and N are finite (although this will not always be necessary). Let λ_1 , λ_2 and λ_3 denote the mean, variance and third central moment of N. Throughout this and the following two sections, which contain our main results, we shall make the following two assumptions:

$$(2.1.1) \qquad \qquad \lambda_2 - \lambda_1 \ge 0$$

In Section 5 we shall comment on the necessity of these assumptions for our results but for the present we remark that both assumptions will hold if N has either a Poisson or a negative binomial distribution.

We assume the insurer of the risk arranges a combination of quota-share and excess of loss reinsurance in the following way:

Firstly, the insurer chooses a quote-share retention level which we denote a so that the insurer's aggregate claims, net of quota-share reinsurance, are aY. We assume the cost of the quota-share reinsurance is calculated on a proportional basis with a commission payment. (See CARTER (1979, p. 87).) More precisely, let P denote the insurer's gross (of expenses and reinsurance) premium income in respect of this risk. We assume an amount eP is used to cover the insurer's expenses, irrespective of the level of reinsurance. The premium of the quote-share reinsurance is (1-a)P less a commission payment of c(1-a)P.

Secondly, the insurer chooses an excess of loss retention level which we denote M so that the insurer's aggregate claims, net of quota-share and excess of loss reinsurance, can be represented by a random variable Y(a, M), where

$$Y(a, M) = \sum_{i=1}^{N} \min(aX_i, M).$$

We denote by P(a, M) the premium paid to the reinsurer in respect of the excess of loss arrangement and we assume the premiums for the two arrangements are calculated independently of each other. (It could be argued that there should be a connection between the two reinsurance premium calculations since 100% reinsurance should cost the same for the two types of treaty but we do not make this extra assumption). Hence the insurer's net (of expenses and reinsurance costs) premium income is

$$P(c-e) + aP(1-c) - P(a, M).$$

2.2. The Cost of the Excess of Loss Reinsurance

Let C(a, M) denote the cost to the insurer of the excess of loss reinsurance arrangement, so that

(2.2.1)
$$C(a, M) = P(a, M) - E[aY - Y(a, M)]$$

Throughout this paper we make the following assumptions concerning C(a, M):

$$(2.2.2) C(a, M) \in \mathscr{C}_1 for a, M > 0,$$

where \mathscr{C}_i is the class of functions with continuous derivatives of order *i*.

(2.2.3) If
$$x_1 = +\infty$$
, $\lim_{M \to \infty} C(a, M) = 0$ for any $a \in (0, 1]$
If $x_1 < +\infty$, $C(a, M) = 0$ for any $M \ge ax_1$ and any $a \in (0, 1]$
(2.2.4) $\partial C/\partial M < 0$ for $M \in (0, ax_1)$ and any $a \in (0, 1]$
(2.2.5) $C(a, M)$ is a convex function of a and M .

Assumptions (2.2.3) are natural. (2.2.4) implies only that the cost of the excess of loss arrangement should decrease as more of the risk is retained by the insurer. (2.2.5) is a little more difficult to interpret but it holds in all our examples in Section 2.3. Roughly speaking, if we regard *a* as fixed, (2.2.4) and (2.2.5) together imply that as *M* decreases, the cost of reinsurance increases and the rate of increase of this cost should also increase.

From assumption (2.2.2), (2.2.3) and (2.2.4) we can see that

(2.2.6)
$$C(a, M) \ge 0$$
 for any $a \in [0, 1]$ and any $M \ge 0$

with

(2.2.7) C(a, M) = 0 if and only if $M \ge ax_1$.

2.3. Some Examples

In this section we discuss very briefly the assumptions of Section 2.2 when P(a, M) is calculated according to some well known principles.

When P(a, M) is calculated according to the expected value principle, standard deviation principle or variance principle (see, for example, GERBER (1979, p. 67) it is not difficult to prove that C(a, M) satisfies (2.2.2), (2.2.3) and (2.2.4). If P(a, M) is calculated according to the expected value principle it can be shown that C(a, M) satisfies (2.2.5). Now suppose that P(a, M) is calculated according to the standard deviation principle so that, after a little calculation,

 $C(a, M) = f\{\lambda_1 G^2(a, M) + (\lambda_2 - \lambda_1) H^2(a, M)\}^{1/2}.$

Where f is a positive loading factor and

$$G(a, M) = \left[\int_{M/a}^{\infty} (ax - M)^2 \, dF(x) \right]^{1/2},$$
$$H(a, M) = \int_{M/a}^{\infty} (ax - M) \, dF(x).$$

It can be shown that given any two non-negative convex functions $g_1(x)$ and $g_2(x)$ the function $g(x) = [g_1^2(x) + g_2^2(x)]^{1/2}$ is still convex. It can also be shown that H(a, M) is a convex function of (a, M) (this is equivalent to say that C(a, M) is convex if P(a, M) is calculated according to the expected value principle). Since $\lambda_2 - \lambda_1 \ge 0$ by assumption, in order to prove that C(a, M) is convex we only have to prove that G(a, M) is convex. The convexity of G(a, M) follows easily since

$$G(a, M) \in \mathscr{C}_2$$
 for $a > 0, M > 0$

 $\partial^2 G(a, M) / \partial M^2 \ge 0$ using the Cauchy-Schwarz inequality

and

$$\{\partial^2 G(a, M)/\partial M^2\} \cdot \{\partial^2 G(a, M)/\partial a^2\} - \{\partial^2 G(a, M)/\partial a \partial M\}^2 = 0.$$

The fact that C(a, M) satisfies (2.2.5) when P(a, M) is calculated according to the variance principle follows directly from the corresponding result for the

standard deviation principle (Since the square of a non-negative convex function is also convex)

3. THE PROBLEM AND ITS SOLUTION

3.1. The Problem

The problem, in broad terms, is to choose retention levels a and M which are, in some sense, optimal for the insurer. We shall assess the effects of reinsurance by considering moment functions of the distribution of the insurer's retained risk. More precisely, let W(a, M) be a random variable denoting the insurer's net (of expense and reinsurance) profit and let E[W(a, M)], V[Y(a, M)], CV[Y(a, M)] and $\gamma(Y(a, M))$ be the expected net profit and the variance, coefficient of variation and skewness coefficient of the insurer's net claims respectively. Our main problem is

PROBLEM 1. Minimize $\gamma[Y(a, M)]$ over the set Γ .

Where $\Gamma = \{(a, M): 0 \le a \le 1 \text{ and } M \ge 0 \text{ and } E[W(a, M)] \ge B \text{ and } V(Y(a, M)) \le D\}$

for some constants B and D. It is assumed that B and D are such that $\Gamma \neq \emptyset$. (Note that we assume CV[Y(a, M)] and $\gamma[Y(a, M)]$ are zero if either a = 0 or M = 0, as well as C[Y(a, M)] when a = 0.)

We shall show that any solution to problem 1 is a solution to problem 2 and vice versa, where problem 2 is

PROBLEM 2 Minimize CV[Y(a, M)] over the set Γ .

Note that V[Y(a, M)] = V[W(a, M)] and $\gamma[Y(a, M)] = -\gamma[W(a, M)]$ so that problem 1 can be expressed entirely in terms of the insurer's net profit. This is not the case for problem 2 since here is no simple relationship between CV[Y(a, M)] and CV[W(a, M)].

In order to solve the above problems it will be helpful to consider the following simpler problem:

PROBLEM 3 Minimize $\gamma[Y(a, M)]$ over the set Γ_1 .

Where $\Gamma_1 = \{(a, M): 0 \le a \le 1, M \ge 0, E[W(a, M)] \ge B\}$

or equivalently (as we will see),

PROBLEM 4. Minimize CV[Y(a, M)] over the set Γ_1 , i.e., we drop the constraint concerning the variance.

3.2. The Skewness Coefficient and the Coefficient of Variation of the Total Net Claims

The statement and proof of the following result assume, for convenience, that $x_0 > 0$.

CENTENO

RESULT 1. (i) $\gamma[Y(a, M)]$ and CV[Y(a, M)] are functions of class \mathscr{C}_1 for a, M > 0.

(ii) Both of them are strictly increasing functions of the single variable M/a for $x_0 < M/a < x_1$ and points such that $0 < M/a \le x_0$ and $M/a \ge x_1$ give minimum and maximum values respectively of the two functions over the set $\{(a, M): a, M > 0\}$.

(iii) $\gamma[Y(a_1, M_1)] > \gamma[Y(a_2, M_2)]$ if and only if $CV[Y(a_1, M_1)] > CV[Y(a_2, M_2)]$.

PROOF. (i) A little calculation gives the following formulae:

- (3.2.1) $V[Y(a, M)] = \lambda_1 (\beta_2 \beta_1^2) + \lambda_2 \beta_1^2$
- (3.2.2) $CV[Y(a, M)] = \{V[Y(a, M)]\}^{1/2}/(\lambda_1\beta_1)$
- (3.2.3) $\gamma[Y(a, M)] = \{\lambda_3 \beta_1^3 + \lambda_1(\beta_3 3\beta_1\beta_2 + 2\beta_1^3)$

$$+ 3\lambda_2\beta_1(\beta_2 - \beta_1^2) / \{V[Y(a, M)]\}^{3/2}$$

where

$$\beta_{k} = \int_{0}^{M/a} a^{k} x^{k} dF(x) + M^{k} (1 - F(M/a)).$$

Using integration by parts and the assumptions that F(0) = 0 and that F is continuous, we have

(3.2.4)
$$\beta_{k} = M^{k} - ka^{k} \int_{0}^{M/a} x^{k-1} F(x) dx$$

from which the proof of (i) follows immediately.

(ii) Let z = M/a. Then we can see that

$$(3.2.5) \qquad \qquad \beta_k = M^k \alpha_k$$

where

$$\alpha_k = \int_0^z (x/z)^k dF(x) + 1 - F(z).$$

Substituting (3.2.5) into (3.2.1), (3.2.2) and (3.2.3) we see that CV[Y(a, M)] and $\gamma[Y(a, M)]$ can be expressed as functions of the single variable z. We shall show that $d\gamma[Y(z)]/dz > 0$, for $x_0 < z < x_1$

$$\frac{d\gamma}{dz}[Y(z)] = 3(1 - F(z))z^{-5}[\lambda_1(\alpha_2 - \alpha_1^2) + \lambda_2\alpha_1^2]^{-5/2} \\ \times \left\{\lambda_1^2 \int_0^z (x^2 z^2 - x^3 z) dF(x) + \lambda_1(\lambda_2 - \lambda_1)h(z) \\ + (2\lambda_2^2 - \lambda_1\lambda_2 - \lambda_1\lambda_3)\alpha_1^2 \int_0^z (xz^3 - x^2z^2) dF(x)\right\}$$

where

$$h(z) = 2 \int_0^z x(1 - F(x)) \, dx \cdot \int_0^z (x^2 - xz) \, dF(x) \\ + \int_0^z (1 - F(x)) \, dx \cdot \int_0^z (xz^2 - x^3) \, dF(x).$$

It is easily checked that h(0) = 0 and that $dh(z)/dz \ge 0$ so that $h(z) \ge 0$ for $z \ge 0$.

That $d\gamma/dz$ is strictly positive for $x_0 < z < x_1$ then follows from assumptions (2.1.1) and (2.1.2). The proof that dCV/dz > 0 is similar to that given above but is somewhat simpler and does not require assumptions (2.1.1) and (2.1.2). The remaining part of (ii) now follows immediately and (iii) follows from (ii).

REMARKS The equivalence of problem 1 and 2, and of problems 3 and 4, follows from part (iii) of the above result.

A further implication of the result is that the locus of points (a, M) satisfying the relation $\gamma[Y(a, M)] = \text{constant}$, or CV[Y(a, M)] = constant, is a straight line passing through the origin in the (a, M)-plane, a higher value of the constant giving a line with steeper slope.

3.3. Isocost Curves

In this section we consider the locus of points (a, M) satisfying the relation E[W(a, M)] = B which is equivalent to

(3.3.1)
$$P(c-e) + a(P(1-c) - \lambda_1 E(X)) - C(a, M) = B.$$

Where B is the constant appearing in the definition of the sets Γ and Γ_1 . (See Section 3.1). It can be regarded as the set of points with a fixed reinsurance price, since (3.3.1) is equivalent to

$$(3.3.2) \quad (1-a)[P(1-c)-\lambda_1 E(X)] + C(a, M) = P(1-e)-\lambda_1 E(X) - B.$$

Where the left-hand side represents the total reinsurance cost of the arrangement (a, M).

We make the following assumptions about the parameters involved in our problems:

(3.3.3) $P(1-c) - \lambda_1 E(X) > 0$

$$(3.3.4) \qquad B < P(1-e) - \lambda_1 E(X)$$

 $(3.3.5) \qquad B > P(c-e)$

$$(3.3.6) B > \max \{ E[W(a, M)]: 0 < a \le 1, 0 \le M \le ax_0 \}.$$

Assumption (3.3.3) implies that the cost of the quota-share arrangement $[(1-a)(P(1-c)-\lambda_1 E(X))]$ is positive for $0 \le a \le 1$ and also that the cost of this arrangement decreases with the retention *a*. Then (3.3.3) together with (2.2.6)

and (2.2.7) implies that the total reinsurance cost of the arrangement (a, M) is non-negative, and is zero if and only if a = 1 and $M \ge x_1$. Assumption (3.3.4) is then natural since the right hand side represents the insurer's expected profit after expenses but without any reinsurance. Assumptions (3.3.5) and (3.3.6) imply that points such that a = 0 or $M \le ax_0$ respectively are not feasible solutions to our problems, i.e., we do not consider solutions where the whole risk is passed to the reinsurer through the quota-share arrangement or where the excess of loss retention is less than the smallest possible claim (net of quota-share reinsurance).

The following result discusses the shape of the isocost curves.

RESULT 2. Let

$$\rho(a, M) = E[W(a, M)] - B \text{ for any } a, M > 0$$
$$a_0 = [P(e-c) + B] / [P(1-c) - \lambda_1 E(X)]$$

 $A = \{a \cdot 0 < a \le 1 \text{ and there exist at least one } M, M < ax_1, \text{ such that } \rho(a, M) = 0\}.$

Then (i) $A = (a_0, 1]$.

(ii) For each $a \in A$ there is a unique M such that $\rho(a, M) = 0$ i.e., there is a function Φ mapping A into $(0, \infty)$ such that $M = \Phi(a)$ is equivalent to $\rho(a, M) = 0$.

- (iii) $\Phi(a) \in \mathscr{C}_1$.
- (iv) $\lim_{a \to a_0^+} \Phi(a) = a_0 x_1$. (v) $\lim_{a \to a_0^+} \Phi'(a) = -\infty$.

(vi) $\Phi(a)$ is convex and is strictly convex if C(a, M) is strictly convex.

PROOF First note that $\rho(a, M) = 0$ is equivalent to

$$\rho(a, M) = P(c - e) + a(P(1 - c) - \lambda_1 E(X)) - C(a, M) - B = 0.$$

(i) Let $\hat{a} \leq a_0$. It follows from the definition of a_0 and from (2.2.6) and (2.2.7) that $\rho(\hat{a}, M) < 0$ for any $M < \hat{a}x_1$. Hence $\hat{a} \notin A$. Now let $\hat{a} \in (a_0, 1]$. $\rho(\hat{a}, M)$, considered as a function of M, is continuous since C(a, M) is assumed continuous. Also

$$\lim_{M \to \hat{a}x_1^-} \rho(\hat{a}, M) = P(c-e) + \hat{a}(P(1-c) - \lambda_1 E(X)) - B > 0$$
$$\lim_{M \to \hat{a}x_0^+} \rho(\hat{a}, M) < 0 \quad \text{by (3.3.6)}.$$

Hence there is at least one M, $M < \hat{a}x_1$, such that $\rho(\hat{a}, M) = 0$.

(ii) Suppose $\rho(a, M_1) = 0 = \rho(a, M_2)$ for some a, M_1 and M_2 . Then $C(a, M_1) =$ $C(a, M_2)$ and hence, using (2.2.4), $M_1 = M_2$.

(iii) This follows from the Implicit Function theorem. See, for example, Apostol (1963).

(iv) Let $\{a_n\}$ be a sequence such that $a_n > a_0$, $\lim_{n \to \infty} a_n = a_0$ and $\lim_{n\to\infty} \Phi(a_n) = k \leq +\infty$. By continuity we have $\rho(a_0, k) = 0$, which implies that $k \ge a_0 x_1$ using the definitions of ρ and a and (2.2.6) and (2.2.7). But using part (i) above, $\Phi(a_n) < a_n x_1$ and (iv) follows

(v) If $x_1 = +\infty$ this is obvious. If $x_1 < +\infty$ we have only to notice that, using (2.2.3) and (2.2.4), both $\partial C/\partial a$ and $\partial C/\partial M$ are zero at the point (a_0, a_0x_1) . Hence $\partial \rho/\partial M$ is zero and $\partial \rho/\partial a$ is strictly positive at (a_0, a_0x_1) .

(vi) Let $a_1, a_2 \in A$ and $0 \le \lambda \le 1$. $\rho(a, M)$ is concave since C(a, M) is convex, so we have

$$\rho(\lambda a_1 + (1 - \lambda)a_2, \Phi(\lambda a_1 + (1 - \lambda)a_2))$$

= 0 = $\lambda \rho(a_1, \Phi(a_1)) + (1 - \lambda)\rho(a_2, \Phi(a_2))$
 $\leq \rho(\lambda a_1 + (1 - \lambda)a_2, \lambda \Phi(a_1) + (1 - \lambda)\Phi(a_2)).$

Using the proof of part (i) above we have

$$\lambda \Phi(a_1) + (1+\lambda) \Phi(a_2) \ge \Phi(\lambda a_1 + (1-\lambda) a_2)$$

It is clear that Φ is strictly convex if the same is true for C(a, M).

3.4. The Variance as a Function of (a, M).

We shall find the following result useful when proving our main results in the next section.

RESULT 3 (i) $\partial V[Y(a, M)]/\partial a > 0$ for $x_0 < M/a < x_1$, (ii) $\partial V[Y(a, M)]/\partial M > 0$ for $x_0 < M/a < x_1$.

PROOF. We have already seen that $V[Y(a, M)] \in \mathcal{C}_1$ for a, M > 0 (see proof of result 1(i)). Differentiating (3.2.1) we have:

$$\frac{\partial V}{\partial a} = 2\lambda_1 a \int_0^{M/a} x^2 dF(x) \\ + 2(\lambda_2 - \lambda_1) \int_0^{M/a} x dF(x) \left[\int_0^{M/a} ax dF(x) + M(1 - F(M/a)) \right] \\ \frac{\partial V}{\partial M} = 2\lambda_1 (1 - F(M/a)) \left[MF(M/a) - \int_0^{M/a} ax dF(x) \right] \\ + 2\lambda_2 (1 - F(M/a)) \left[\int_0^{M/a} ax dF(x) + M(1 - F(M/a)) \right]$$

(ii) follows directly and (i) follows from (2.1.1).

3.5. The Solution

In this section we solve our problems in general terms.

RESULT 4. (i) The non-negative constraints are redundant in our problems. (ii) The constraint $E[W(a, M)] \ge B$ is active in the optimum of our problems, i.e., in the optimum of our problems this constraint holds as an equality.

CENTENO

PROOF. (1) Follows directly from assumptions (3.3.5) and (3.3.6).

(ii) We shall prove the result for problem 1, and hence problem 2. The proof for problems 3 and 4 is similar but simpler. Let $(a_1, M_1) \in \Gamma$ be such that $E[W(a_1, M_1)] > B$. From the proof of result 2 we know that there exists $M^* < M_1$ where $a_1x_0 < M^* < a_1x_1$ and $E[W(a_1, M^*)] = B$. Using result 3(ii) and result 1(ii) we see that

$$V[Y(a_1, M^*)] \leq V[Y(a_1, M_1)] \leq D$$

and

 $\gamma[Y(a_1, M^*)] < \gamma[Y(a_1, M_1)].$

Let us now consider the solution to problem 3 (and hence to problem 4). We know that the solution lies on the isocost curve $M = \Phi(a)$ and information about the shape of this curve is contained in result 2. Figure 1 gives three examples of



FIGURE 1 Isocost curves in the (a, M)-plane

isocost curves, labelled I_1 , I_2 and I_3 . We know that each curve has slope $-\infty$ at the point (a_0, a_0x_1) and we have assumed x_1 is finite for convenience. We also know that each curve is convex although not necessarily strictly convex. From result 1 we know that straight lines through the origin in fig. 1 represent points of constant skewness, the larger the slope the higher the skewness. Hence it is clear that the solution to problem 3 is the point, or set of points, where the straight line through the origin with the smallest slope intersects the isocost curve. If the isocost curve is decreasing, as in I_1 , this point will be $(1, \Phi(1))$, i.e., pure excess of loss reinsurance will be optimal. (Note that LEMAIRE, REINHARD and

58

VINCKE (1981), by making assumptions about the calculation of the reinsurance premiums different to ours, were able to assume that the isocost curves were decreasing and hence that, in terms of our problem, excess of loss reinsurance was optimal.) Even if the isocost curve is not decreasing, as in I_2 , the point $(1, \Phi(1))$ may still be the solution to problem 3. Isocost curve I_3 shows a case where the solution is not $(1, \Phi(1))$.

It is clear that in general the solution to problem 3 will be $(1, \Phi(1))$ unless we can find a point on the isocost curve such that the gradient of the isocost curve at that point equals the slope of the line joining that point to the origin. Such a point may not be unique since the isocost curve may not be strictly convex

Summarizing we have the following result:

RESULT 5. Let

$$H = \{(a, \Phi(a)): a_0 < a \le 1, d\Phi(a)/da = \Phi(a)/a\}.$$

Then

(i) if H is empty the solution to problems 3 and 4 is the point $(1, \Phi(1))$.

(ii) if H is not empty, all the points in H are solutions to problems 3 and 4.

REMARKS. (i) We have given a geometrical proof of result 5 but it is possible to give a more formal proof using the Kuhn-Tucker conditions and the facts that E[W(a, M)] is a concave function and $\gamma[Y(a, M)]$, or CV[Y(a, M)], is a quasi-convex function of (a, M). See ARROW and ENTHOVEN (1961).

(ii) Using the definition of Φ the set H can be defined as

(3.5.1)
$$H = \{(a, M) : a \leq 1 \text{ and } E[W(a, M)] = B$$

and $B + P(e-c) + C(a, M) - a \partial C / \partial a - M \partial C / \partial M = 0\}.$

We are now in a position to solve problems 1 and 2.

RESULT 6. Let

 $a_1 = \inf \{a: (a, \Phi(a)) \text{ is a solution to problem 3} \}.$

Then

(i) if $V[Y(a_1, \Phi(a_1))] \leq D$, $(a_1, \Phi(a_1))$ is a solution of problem 1 and every solution of problem 1 is a solution of problem 3.

(ii) if $V[Y(a_1, \Phi(a_1))] > D$ the solution to problem 1 is $(a^0, \Phi(a^0))$ where

$$a^{0} = \sup \{a : a \leq 1, E[W(a, M)] = B \text{ and } V[Y(a, M)] = D\}.$$

In this case $a^0 < a_1$.

PROOF. (i) If $(a_1, \Phi(a_1))$ is a solution of problem 3 and $V[Y(a_1, \Phi(a_1))] \leq D$ then clearly $(a_1, \Phi(a_1))$ is a solution of problem 1. If $(a, \Phi(a))$ is another solution to problem 1 we must have $\gamma[Y(a, \Phi(a))] = \gamma[Y(a_1, \Phi(a_1))]$ and so $(a, \Phi(a))$ solves problem 3.

CENTENO

(ii) Using geometrical arguments and result 3, it is clear that for any a such that $a_1 \le a \le 1$ we have, assuming $V[Y(a_1, \Phi(a_1))] > D$, $\Phi(a_1) \le \Phi(a)$ and $D < V[Y(a_1, \Phi(a_1))] \le V[Y(a, \Phi(a))]$. This shows that $a^0 < a_1$.

On the other hand, $\gamma[Y(a, \Phi(a))]$ is a strictly decreasing function of a for $a_0 < a < a_1$, as is clear when we consider the geometrical proof of result 5.

So for any point $(a, \Phi(a))$ such that $a_0 < a < a^0$ we will have $\gamma[Y(a, \Phi(a))] > \gamma[Y(a^0, \Phi(a^0))]$ and for any point such that $a^0 < a \le 1$ we will have $V[Y(a, \Phi(a))] > D$, otherwise these would be a contradiction to the definition of a^0 or to the mean value theorem.

4. THE SOLUTION IN SOME SPECIAL CASES

In this section we give, briefly, the solution to problems 3 and 4 when the excess of loss reinsurance premium is calculated according to the expected value principle, the standard deviation principle or the variance principle.

RESULT 7. (i) If the excess of loss reinsurance premium is calculated using the expected value principle or standard deviation principle, the solution to Problems 3 and 4 is $(1, \Phi(1))$, i.e., a pure excess of loss arrangement.

(ii) If the excess of loss reinsurance premium is calculated using the variance principle, the solution to problem 3 and 4 is $(\hat{a}, \Phi(\hat{a}))$ where

$$\hat{a} = \min \{2[P(e-c)+B]/[P(1-c)-\lambda_1 E(X)], 1\}.$$

PROOF. (i) The proof is immediate since for both cases

(4.1)
$$B + P(e-c) + C(a, M) - a\frac{\partial C}{\partial a} - M\frac{\partial C}{\partial M} = B + P(e-c)$$

which is positive by assumption (3.3.5) and so the set H is always empty (although the relevant isocost curve is not necessarily decreasing). The result follows from result 5(i).

(ii) In this case the left-hand side of (4.1) is equal to

$$B+P(e-c)-C(a, M).$$

Hence the set H is

 $\{(a, M): E[W(a, M)] = B \text{ and } a = 2[B + P(e-c)]/[P(1-c) - \lambda_1 E(X)] \text{ and } a \le 1\}$ and the result follows from result 5.

5. DISCUSSION

We have assumed throughout Sections 2, 3 and 4 that assumptions (2.1.1) and (2.1.2) hold for the claim number distribution N. It is clear that all our results relating to the coefficient of variation, in particular the solutions to problem 2

and 4, are valid without making assumption (2.1.2), since this assumption was used only in the proof of result 1(ii), and then only in relation to the skewness coefficient.

Assumption (2.1.1) was used in relation to the coefficient of variation to show that C(a, M), and hence the isocost curve $M = \Phi(a)$, is convex when the excess of loss reinsurance premium is calculated according to the standard deviation principle or the variance principle (see Section 2.3). If (2.1.1) does not hold it is not hard to find examples where P(a, M) is calculated according to either the standard deviation principle or the variance principle and where the isocost curve is no longer convex. (One particular example assumes N to be a degenerate random variable always equal to 1, which is equivalent to assuming a combination of quota-share and stop-loss reinsurance). However if (2.1.1) does not hold we can still state result 7(i), relating to the coefficient of variation, since this result is an immediate consequence of result 5(i), and it is easy to see that this result is independent of the convexity of the isocost curves. Assumption (2.1.1) was also used for the proof of result 3(i), which was later applied in the proof of result 6(ii). It is not difficult to see that if (2.1.1) does not hold, but $a_1 = 1$ in result 6, this result is still true. So we can conclude that when P(a, M) is calculated using the expected value or standard deviation principle, the main results relating to the coefficient of variation [i.e., result 7(i) and result 6], hold without the assumption (2.1.1) being fulfilled.

When P(a, M) is calculated according to the variance principle the proofs of Results 5(ii), 6(ii) and 7(ii), relating to the coefficient of variation, are no longer valid without (2.1.1), although it may be possible to prove some of these results without this assumption.

Furthermore (2.1.1) is not a necessary condition for the proof of result 1 relating the skewness coefficient, since this result still holds when N is a degenerate random variable and when the distribution function of the individual claim amounts is absolutely continuous [see LEMAIRE, REINHARD and VINCKE (1981)]. In this particular case all the comments relating to the coefficient of variation apply to the skewness coefficient.

We have already mentioned in Section 2.1 that both (2.1.1) and (2.1.2) hold if N has a Poisson or a Negative Binomial distribution. It is also worth mentioning that (2.1.1), but not (2.1.2), holds for any mixed Poisson distribution and that (2 1 2), but not (2.1.1), holds if N has a binomial distribution or is a degenerate random variable.

6. AN EXAMPLE

In this section we discuss a numerical example that illustrates the results in the previous sections.

We assume the gross aggregate claims are generated by a compound negative binomial distribution with

$$\lambda_1 = 10; \qquad \lambda_2 = 20; \qquad \lambda_3 = 60$$

and

$$F(x) = \begin{cases} 0 & \text{if } x \le 1 \\ 1 - x^{-4} & \text{if } x > 1 \end{cases}$$

so that individual claims have a Pareto distribution. We assume

P = 24; e = 0.35; B = 1.7

and the premium loading factor, f, used in the calculation of the excess of loss reinsurance premium is 0.8, 0.45 and 0.4 when the premium calculation used is the expected value principle, the standard deviation principle and the variance principle respectively. Table 1 gives the point (a, M) which is the solution to problems 1 and 2 for various values of c and D. Note that from table 1, we can see that when c = 0.4 the isocost curves for the three premium calculation principles are not decreasing functions with a.

Excess of loss premium calculation principle	с	D	(a, M)	V[W(a, M)]	$\gamma[Y(a, M)]$	CV[(a, M)]
Expected value principle	04	33	(1, 1 676)	32 38	0 6763	0.4507
		27	(0 908, 1 57)	27	0 677	0 4511
	03	33	(1, 1.676)	32 38	0 6763	0 4507
		27	(0 863, 2 53)	27	0 6886	0 4562
Standard deviaton principle	0.4	33	(1, 1497)	30 77	0 6743	0 4495
		27	(0 921, 1 48)	27	0 6755	0 4502
	03	33	(1, 1 497)	30 77	0 6743	0 4495
		27	(0 846, 18 6)	27	07153	0 4609
Variance principle	04	33	(0 9375, 1 47)	27 7	0 6751	0 4500
		27	(0 926, 1.46)	27	0 6751	0 4500
	03	33	(1, 1575)	31 54	0 6752	0 4500
		33	(0 854, 3 42)	27	0 6952	0 4580

TABLE I

Let us now consider in more detail the case where c = .3 and the standard deviation principle is used for the excess of loss reinsurance premium. Figure 2 shows the variance and the skewness coefficient for seven different isocost curves, starting with B = 32/15 and decreasing B in steps of 4/30 until we get B = 4/3. Points with the same subscript correspond to points on the same isocost curve and the smaller the subscript, the higher the value of B. The points I_1, I_2, \ldots, I_7 correspond to pure excess of loss treaties and these points together with the points on the solid lines represent solutions of Problem 1 for some value of D. The dotted lines correspond to points that are never solutions to problem 1, although they are on the isocost curved considered. This is because, for example, points between I_5 and I''_5 have both greater skewness and greater variance than I_5 .

62



FIGURE 2 Isocost curves in the (V, γ) -plane

REFERENCES

- APOSTOL, T M (1963) Mathematical Analysis Addinson-Wesley London
- ARROW, K J, and ENTHOVEN, A C (1961) Quasi-Concave Programming Econometrica 29, 779-800
 BUHLMANN, H (1970) Mathematical Methods in Risk Theory Springer-Verlag Berlin, Heidelberg, New York
- CARTER, R L (1979) Reinsurance Kluwer Publishing Great Britain
- GERATHEWOHL, K (1980) Reinsurance Principles and Practice Verlag Versicherungswirtschaft e V Karlsruhe
- GERBER, H U (1979) An Introduction to Mathematical Risk Theory Richard D Irwin, Inc Homewood
- HADLEY, G (1967) Nonlinear and Dynamic Programming Addison-Wesley London
- LEMAIRE, J., REINHARD, J. M. and VINCKE, PH. (1981) A New approach to Reinsurance Multicriteria Analysis. *Net Retentions*, 7-42. Nederlandse Reassurantie Groep NV. Amsterdam

OBITUARY

HILARY L. SEAL

11th January 1911-25th July 1984

On July 25, 1984 actuarial science lost one of its greatest personalities: Hilary L. Seal died from the side effects of a thrombosis This was a sudden end to his activity, which was always at the 200% level On an operational scale, Hilary stayed young. It remains to our satisfaction that the fruits of his activity will be here for future generations

Hilary received his formal education in his native England, first at Birmingham University, then at University College, London, where he graduated in Statistics with first class honors. He had an actuarial position in Brazil, but returned to England to serve his country during and shortly after World War II as a statistician in the Admiralty. In 1948 he received his Ph.D. for his thesis on "Discrete Random Processes in Relation to Mortality Data". At this time it was clear that the old island was too small for Hilary; he immigrated to North America. He worked briefly for an insurance company in Toronto, then moved to New York and New Haven, where he became a successful consultant. He was able to combine his professional activity with an impressive scientific career. For twenty years he taught statistics at Yale University. When he and his family moved to Apples (Switzerland) in 1972, the Swiss Federal Institute of Technology of Lausanne and the University of Lausanne secured his services; in 1980 Hilary occupied the Chair of Honor of the Institute of Actuarial Science of the University of Lausanne.

Hilary's publications are manifold and cover the broad range of statistics and actuarial science. A bibliography is being prepared and will be published soon. The topics of his papers include estimation of decrement rates, multivariate statistics, pension mathematics, risk theory, queuing theory, numerical methods such as simulation and inversion of Laplace transforms; Hilary made good use of the computer at a time when other actuaries still relied on the abacus. His monographs are classics: Multivariate Statistical Analysis for Biologists (1964), Stochastic Theory of a Risk Business (1969), Survival Probabilities: The Goal of Risk Theory (1978).

Hilary was a brilliant speaker. When he talked, one could expect fireworks But no matter how spectacular his lectures were, they were always based on extensive research. One of Hilary's loves was reading what others had been writing. His famous library is the testimony to this passion.

For the readers of the Astin Bulletin who had neither the opportunity to listen to Hilary nor to know him, we reprint the beginning of a letter that he wrote in 1950 to the Editors of the Journal of the Institute of Actuaries Students Society. The topic is "Spot the Prior Reference", and the letter begins as follows:

A game which is fast become a favourite relaxation of the more priggish type of mathematician is one which might be called: Spot the prior reference.



Photograph shows the presentation of Corresponding Membership in the Association of Swiss Actuaries in June 1980.

The equipment is elementary—a good memory or an extensive system of card records with appropriate cross-references. The object of the game is simple—the infliction of a blow to the self-esteem of a colleague while retaining an appearance of scientific detachment.

The first move is made by an author who inadvertently omits that thorough search through the numerous volumes of *Mathematical Reviews* and the *Zentralblatt für Mathematik* which nowadays occupies as much of a mathematician's time as the preparation of a supposedly original article. The second move falls to the editor whose referees fail to notice that the work submitted has already appeared in print in a substantially similar form ten, twenty or even a hundred years earlier—and the game is on. The reviewer now appears on the scene and scores one or more points according to the number of years he can span and the amount of scorn he can convey in a politely worded account of the author's limitations. The game continues as a third and fourth writer show that even the reviewer himself has not found the site of original publication of the material presented. Final honours go to the player who has revealed the greatest number of missing references in the previous writers' articles.

Following this introduction, Hilary showed that the convolution of uniform distributions (a favorite topic of some writers of the 20th century) could be traced down to the 18th century

OBITUARY

Hilary was a Fellow of the Faculty of Actuaries in Scotland, an Associate of the Institute of Actuaries, a Fellow of the Royal Statistical Society, an Associate of the Society of Actuaries, a Fellow of the Canadian Institute of Actuaries and a member of more than a dozen other professional societies. He was one of the very few actuaries who have been elected Fellows of the American Statistical Society. At the International Congress of Actuaries in Switzerland he was made a Corresponding Member of the Association of Swiss Actuaries.

We shall miss Hilary Seal for his professional contributions. His family and his friends will miss him for much more.

H. U. Gerber

EDITORIAL

This is the last issue of *Astin Bulletin* to be published under my editorship. Hans Bühlmann and D. Harry Reid will take on the editorial responsibilities from the next issue onwards. Their addresses can be found in the *Instructions to Authors* on the inside back cover.

I feel priviledged to have serived as Editor of the Astin Bulletin, a job which I have done for almost eight years now. It brought me into contact with authors and papers and also enabled me to read referees' reports. I am confident that these inputs have influenced my own thinking on insurance matters in a positive way. It is an experience which I am glad to have had and I take this opportunity to thank all the people involved; authors, referees, members of the Editorial Board as well as the membership of Astin for having confidence in me.

BOOK REVIEWS

J. VAN EEGHEN, E. K. GREUP and J. A. NIJSSEN (1983). Rate Making. Surveys of Actuarial Studies, No. 2. Nationale-Nederlanden N.V., Rotterdam. 138 pages.

The series Surveys of Actuarial Studies, published by the Research Department of the Nationale-Nederlanden with G. W. de Wit as editor, covers, in its first two volumes, two important R's of actuarial nonlife activity: Reserving and Rate making. The first volume, *Loss Reserving Methods*, was reviewed in *Astin Bulletin* 14, No. 1.

As the name of the series implies, the present volume contains, in a condensed and logically ordered form, material from a large number of actuarial books and papers, as well as some general statistical methods. The valuable bibliography at the end of the booklet contains some eighty references.

The plan of the book is as follows. In the Introduction the rate making process in general is discussed. After that, the treatment is entirely devoted to the risk premium part of the premium. Chapter 1 treats the selection of tariff variables (rating factors). Determination of tariff classes, defined via the tariff variables chosen, is considered in Chapter 2. Chapter 3, Parameter estimation in modelled tariff structures, treats the problem of estimating the risk premium, or the claims frequency or the average claims size, as a function of the tariff class. In Chapter 4 an example of the credibility approach is given. Finally, Chapter 5 gives a brief outline of the problem of large claims.

In general, each method of analysis is given a brief but sufficiently detailed presentation. Then there are some hints on numerical computation and, in most cases, also numerical examples. Finally, the authors give their own comments on the method.

In the following I will give some of my own comments on the contents of the book.

In the Introduction, "the (known) solidarity part" of the premium is introduced as a separate premium component. This refers to an intentional omission of some premium differentiation, for social reasons. The issue has obviously attracted a great deal of interest in the Netherlands recently. It is not unknown in other parts of the world. Still I think it belongs to the larger context of the difference between risk factors, i.e., factors influencing the risk, and rating factors, which are the factors actually used for premium calculation. Personally, I would have liked this difference, which is not exclusively caused by feelings of social fairness, to be more clearly set out in the discussion.

Chapter 1 presents Lemaire's linear regression selection method, Hallin-Ingenbleek's unmodelled selection procedure and, as a nice contribution from the statistical tool-box, a method based on discriminant analysis. In the comments it is pointed out that exact significance levels are difficult to establish for the procedures. In particular, assumptions of normality and homoscedasticity will mostly be violated in practice. Especially the latter (equality of variances) I find questionable as it is generally inconsistent with the compound Poisson model. With this in mind, as the authors point out, the methods may however be efficient tools for exploratory data analysis In Chapter 2 methods of cluster analysis are applied to the problem of reducing a maximal number of basic classes, defined by the tariff variables, to a smaller number of tariff classes One method is Dickmann's application of general cluster analysis to insurance problems, based on variation within and between clusters. The other is the method of Loimaranta, Jacobsson and Lonka, which is based on likelihood estimation and tests of mixtures of distributions. The concept of the chapter is very elegant. The two methods of cluster analysis may be used to produce a set of admissible subdivisions of the portfolio. Finally, the credibility related method of Schmitter and Straub, which is also described, may be used to judge between them.

In the first part of Chapter 3 non-parametric methods of estimation in modelled tariff structures are considered: Simon-Bailey (minimum chi-square), graduation by marginal totals and least squares. Applications are made to multiplicative and additive models. Comments on methods and models are generally well elaborated. Reference is given to a number of papers according to which the additive model should produce a better fit for insurance data, even if the majority of existing rating systems rather seem to be multiplicative. So for balance I should mention that the Swedish motor insurance data from 1977, analyzed by Hallin-Ingenbleek in *SAJ* 1983, No 1, showed a somewhat better fit for the multiplicative model than for the additive one, as did data from 1979. The second part of the chapter is devoted to a careful presentation of maximum likelthood methods, containing *inter alia* Ter Berg's treatment of loglinear models for Poisson, gamma and inverse Gaussian distributions.

The last two chapters are rather short. Chapter 4 presents the elegant Bühlmann-Straub model. For tariff construction this credibility approach has to be applied with some care, according to the reviewer's experience. This is because it starts out from the assumption that the risk groups under study are similar, in the sense that their risk characteristics are assumed to be chosen at random from one and the same collective. The method therefore has a tendency to give too small differences between risk groups. The case for experience rating of individual contracts may be different.

The fifth chapter on large claims outlines methods of Schäffer-Willeke and Gisler. The problem of large claims is a nuisance in tariff construction work, at least as soon as personal injury claims or fire claims are present. So, as a practitioner one could have hoped for a fuller treatment, perhaps including the division of claims into more than two size groups (e.g., normal claims, excess claims, superexcess claims) and/or some help from the theory of outlying observations. Maybe one could hope for another volume in the series on this subject?

In summary, the authors have collected in a limited space an astonishingly rich material on general rate making methods. They have deliberately refrained from discussing the loading for commissions and expenses, and problems pertaining to special lines of business. There is no mentioning of investment income. These limitations are most natural considering the size of the book. It should be of great value to every non-life actuary. R. E. BEARD, T. PENTIKAINEN and E. PESONEN (1984). *Risk Theory* (3rd edition) Chapman & Hall Ltd., London. xvii+408 pages, £11.95 paperback/£24.50 hard-bound.

[A review on the first edition by H. Bühlmann appeared in AB 6, 178-179.]

Those readers who are familiar with the first two editions of this pioneering book on risk theory will be surprised to see that the new edition is a complete revision of the earlier editions. This renewed third edition gives an introduction to risk theory with main emphasis on the practical aspects of theoretical results. Therefore this book bridges the gap between practical problems and pure risk theory.

The first chapter provides us with some thoughts on general modelling and more specifically on insurance models. Also the notations which will be used in the subsequent sections are introduced.

In the second chapter the authors examine the Poisson process. Classical properties as well as approximations are considered. In addition they discuss the economical influences on the claim number. They distinguish four kinds of fluctuations: trends, long-period cycles, short-period oscillations and pure random fluctuations. The structural distribution is introduced to incorporate short-period oscillations. The classical characteristics of mixed Poisson distributions are subsequently examined.

In Chapter 3 the compound Poisson process is extensively studied. The distribution of the claim size, those of the aggregate claims as well as basic characteristics of the distribution are largely taken into consideration.

Possible estimation techniques for the aggregate claim size distribution are given Some problems arising from large claims are given. Analytical results are discussed as well as different types of claim distributions. The effect of a reinsurance treaty on total claim size is examined. The by now classical approximations for the compound Poisson distribution are given namely the Edgeworth expansion, the normal power approximation, the gamma-approximation etc. Also some more recent techniques, such as the inversion of the characteristic function, the recursion algorithm are also dealt with.

Mostly based on the normal power approximation of the compound Poisson distribution the authors discuss in Chapter 4 some practical problems related to a one-year time span such as: evaluation of the fluctuation range of the annual underwriting profits and losses, the reserve-funds, the problem of greatest retention, the influence of several retention limits, excess of loss reinsurance, stop-loss reinsurance, experience rating.

In Chapter 5 the variance is used as a measure of stability to design an optimal form of reinsurance to discuss reciprocity of two companies and the equitability of safety loadings.

In Chapter 6 a completely new chapter (not appearing in the previous editions) is included considering the risk processes with a time span of several years. In this case the basic parameters of the risk process are continually subject to alterations which are partially revealed as trends and partially as cyclical The effect of these phenomena is modelled for carrying out long-term considerations,

BOOK REVIEWS

e.g., the Poisson parameter is adapted to take into account the trends as well as cycles (by means of an autoregressive process). In addition the problem of forecasting the future flow of business is studied. The method is also adapted for coping with inflation. Investment is a new topic developed. Ruin probabilities for a finite time period come into the picture for discussing the problem of solvency. This chapter ends with the description of the Monte Carlo simulation of risk business.

In Chapter 7 several applications of the risk processes with a time span of several years developed in the previous chapter are given: the evaluation of net retentions, the effect of cycles, the effect of the time span, the effect of inflation, dynamic control rules and a solvency profile.

By means of cohort analysis the results of risk theory are then adapted to the life insurance branch in the following chapter.

In Chapter 9 infinite time ruin probability is studied essentially by means of the adjustment-coefficient. Some practical consequences are deduced.

The final chapter describes the application of risk theory to business planning. In the previous chapters many applications of a risk theory, such as the estimation of a suitable level for the maximum net retention, the evaluation of stability, the safety loading and the magnitude of the funds have been treated as isolated aspects of an insurance business. In this chapter a picture of the management process in its entirety is built up. An integration of the risk theoretical aspects in the context of other management aspects, not of actuarial nature, is carried out.

The book ends with some appendices containing derivations and proofs of some of the mathematical results obtained in the book: derivation of the Poisson and mixed Poisson processes, Edgeworth expansion, Infinite time ruin probability, Computation of limits of finite time ruin probabilities, Random numbers.

In addition the book contains quite a lot of interesting exercises (and their solution), an author index, a bibliography, a subject index as well as a necessary list of symbols.

In conclusion this book on risk theory where formulae are approached from the practical point of view shows to practical actuaries that some of the theoretical results lead to a better understanding of what is going on. To theoretical actuaries (at universities) the book gives a motivation for going on with theoretical research. Although this book has just appeared it is clear from discussions with students that it provides us with insurance models and material which is highly appreciated by people preparing for the actuarial profession.

M. GOOVAERTS

M. GOOVAERTS, F. DE VYLDER and J. HAEZENDONCK (1984). Insurance Premiums. Theory and Applications. North-Holland, Amsterdam. xi+406 pages, US \$63.75/Dfl.150.00

This book introduces the reader to areas of insurance mathematics which have so far not been published on this scale in the form of a textbook. The individual
BOOK REVIEWS

themes which the authors have already contributed to in numerous publications are for the first time discussed systematically, starting from the fundamentals.

A brief general introduction is followed by the central chapters on premium calculation principles (Chapter 2) and their properties (Chapter 3). The idea of introducing premium calculation principles goes back to Hans Bühlmann and it is astonishing how many such principles have meanwhile been developed. Based mainly on work by Bühlmann, Gerber and the authors, the following principles are introduced: expected value, maximal loss, variance, standard deviation, semi-variance, mean value, zero utility, Swiss premium, Orlicz, Esscher and mixtures of Esscher principles. Apart from the definition some principles are motivated by e.g., statistical reasoning or utility considerations. In addition properties and characterisations are supplied which at present are not to be found in printed form anywhere else. Because of the mathematically stringent form every expert in this field and every reader interested in mathematics too, who wishes to familiarize himself with the theoretical foundations of premium calculation, will definitely appreciate this book if only for its systematic presentation. On the other hand, it has to be said to every practitioner that it is not the purpose of this book to evaluate the various principles in contrast to each other or to examine the practical feasibility of the principles which use utility concepts. In Chapter 3 Properties of Premium Calculation Principles the properties of additivity, translation invariance, iterativity, homogeneity, multiplicativity and some generalisations are investigated as to which principles fulfill them. For example the property of additivity holds that in the case of independent risks X, Y the premium for X and Y should be just the sum of the individual premiums for Xand Y. Not all properties are plausible for insurance premiums and not all generalisations important. The question of important and reasonable premium calculation principles and properties, which has not been resolved in the field of theory, should not be dealt with separately for principles and properties, and should not be answered without involving practitioners either. On this point these chapters provide a very successful presentation which will be the basis of every scientific discussion in this very lively field in future.

In the next chapter, entitled Ordering among Risks, various possibilities of how to introduce a partial order in the set of all (e.g., bounded) risks are considered. Of course, for a fixed premium principle π a natural order is induced by $\pi(X_1) \leq \pi(X_2)$, but one is also interested in conditions independent of π . The most important term here, the net stop-loss ordering, is introduced and its behaviour as to mixing and convolution investigated. These results allow statements to be made about the influence of these orders of number and size of losses on the orders of the corresponding total losses in the usual risk theoretical model. In this connection the dangerousness of distributions is discussed; after that generalisations of the stop-loss ordering are treated with stochastic dominance, familiar from the theory of finance, and stop-loss dominance.

The chapter Bounds on Stop-Loss Premiums takes up the problem which occurs in reinsurance practice of trying to calculate stop-loss premiums with incomplete risk information. Whereas to make an exact calculation of a net stop-loss premium

BOOK REVIEWS

one requires full knowledge of the distribution function, the situation where one often only has an estimate accepted on trust for the expected value and perhaps the variance too, is looked at very realistically here. Depending on the available information e.g., about the first n moments or the symmetry of the density function, one looks for an upper bound for all stop-loss premiums of those risks which the given information applies to. The step from the noncalculable "true" stop-loss premium to the upper bound is therefore a cautious one. In some cases, e.g., when expected value and variance are regarded as known, lower bounds can also be worked out analogously, i.e., an error estimate. The mathematical tool necessary for this stems from finite dimensional analysis and is presented comprehensively, so that this chapter is self-contained. The most important aid is a duality theorem from convex analysis allowing the original maximisation problem to be transformed into a simple minimisation problem so that analytical or numerical results are gained. In the last chapter on applications the method for estimating stop-loss premiums described above is applied to the case of bounded exposure. Further applications of the ordering of risks and the procedure in case of incomplete information are indicated for questions of determining the optimal critical claim size in a bonus-malus system, for bounds of the ruin probability and bounds for stop-loss premiums for weighted compound distributions.

This book, which only requires little prior mathematical knowledge, is the most comprehensive presentation of the themes dealt with here. Because of the mathematical style the practitioner will miss explanations and a mutual comparison of the results at some points. The book reflects the current level of knowledge in some fundamental partial fields of insurance mathematics in a precise form, thereby providing a base from which theoreticians and practitioners can communicate with each other.

A. Reich

INSTRUCTIONS TO AUTHORS

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

Hans Buhlmann Mathematik, ETH-Zentrum 8092 Zurich Switzerland D. Harry Reid Eagle Star Insurance Company Ltd 1 Threadneedle Street, London EC2R 8BE England

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 referee process, which will take about three months

2. The first page of each paper should start with the title, the name(s) of the author(s), 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.

Acknowledgements and grants received should be placed as a footnote, which is not included in the count of the other footnotes. Footnotes should be kept to a minimum.

- 3. Manuscripts should be typewritten on one side of the paper, double-spaced with wide margins
- 4. Tables should only be included if really essential. Tables should be numbered consecutively. Do not use vertical lines!
- 5. Upon acceptance of a paper, any figures must be drawn in black ink on white paper in a form suitable for photographic reproduction with a lettering of uniform size and sufficiently large to be legible when reduced to final size
- 6. Important formulae should be displayed and numbered on the right hand side of the page In mathematical expressions, especially in the text, authors are requested to minimize unusual or expensive typographical requirements. This may be achieved by using the solidus instead of built-up fractions and to write complicated exponentials in the form exp()

Matrices (uppercase) and vectors (lowercase) will be printed boldface. Boldface should be indicated by wavy underlining

7. References in the text are given by the author's name in capitals, followed by the year of publication between parentheses.

Examples ROTHENBERG (1973) has done , ... should be referred to FOLKS and CHHIKARA (1978); . as shown by DE FINETTI (1937)

At the end of the paper the references should be grouped alphabetically and chronologically For journal references give author(s), year, title, journal, volume and pages. For book references give author(s), year, title, city and publisher. Illustrated with the abovementioned references, this works out as:

- FINETTI, B DE (1937). La prévision, ses lois logiques, ses sources subjectives, Annales de l'Institut Henry Poincaré 7, 1-68 Reprinted as Foresight. its Logical Laws, its Subjective Sources, in H. E Kyburg and H. G. Smokler (Eds.) (1980). Studies in Subjective Probability. Huntington. Krieger Publishing Company, Inc
- FOLKS, J L., and R S CHHIKARA (1978). The Inverse Gaussian Distribution and its Statistical Application—A Review (with Discussion), Journal of the Royal Statistical Society B 40, 263–289.
- ROTHENBERG, T J (1973). Efficient Estimation with A Priori Information. New Haven Yale University Press.

Observe that abbreviations should not be used!

COMMITTEE OF ASTIN

Jean LEMAIRE	Belgium	Chairman/Treasurer
Ragnar Norberg	Norway	Vice-Chairman
Jean-Marie REINHARD	Belgium	Deputy Secretary
Peter ter Berg	Netherlands	Editor
Hans BUHLMANN	Switzerland	IAA-Delegate
LeRoy J. Simon	USA	IAA-Delegate
Harald Bohman	Sweden	Member
Giovanna Ferrara	Italy	Member
Charles A. HACHEMEISTER	USA	Member
Jean Lamson	France	Member
Erkki Pesonen	Finland	Member
Jürgen Strauss	FR Germany	Member

Neither the COMMITTEE OF ASTIN nor TIETO LTD are responsible for statements made or opinions expressed in the articles, criticisms and discussions published in ASTIN BULLETIN.