Volume 14, No. 2 October 1984

ASTIN BULLETIN

A Journal of the International Actuarial Association

CONTENTS:

TIETO LTD

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 the Secretary:

F E. GUASCHI
Moorfields House, Moorfields, London EC2Y 9AL
England.

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 £15.00.

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.

PRINTED IN GREAT BRITAIN by J W ARROWSMITH LTD, BRISTOL

PREMIUM CALCULATION FOR DEDUCTIBLE POLICIES WITH AN AGGREGATE LIMIT

By Thomas Mack*

Münchener Rückversicherungs-Gesellschaft, Munich

Abstract

In Industrial Fire insurance an aggregate limit for the amount retained by the policyholder under a deductible policy has been agreed upon more frequently in recent times. This agreement is equivalent to a stop-loss cover on the retained loss amount. For the Poisson-lognormal model the corresponding stop-loss net premium is calculated using various methods (normal power, translated gamma, various discretisations) and the methods are compared. Finally, the influence of the model parameters is examined and it is demonstrated how a variety of parameter value combinations can be reduced to only a few rating curves.

Keywords

Deductibles, aggregate limit, stop-loss premium, Industrial Fire insurance, Poisson-lognormal model.

1. Introduction

On a number of markets the practice of adding an aggregate limit to an Industrial Fire insurance policy with a deductible has increased in recent times. An aggregate limit means that the maximum accumulated amount of losses to be retained by the policyholder is limited for each year; the insurer then takes over payment should this maximum be exceeded. The advantage for the policyholder is quite obvious: the risk retained under the deductible is limited, not only in terms of each loss event but also on an annual basis. A policy with a deductible but no aggregate limit, however, may lead to an unexpectedly high retained aggregate loss amount if the policyholder is confronted with an accumulation of loss events. For the insurer, the calculation of deductible rebates, difficult enough as it is, becomes even more complicated With the aggregate limit, the policyholder is granted in addition a stop-loss cover on his retained losses, which leads to a reduction in the normal deductible rebate. If the size of a loss is independent of the number of losses, the normal deductible rebate depends solely on the distribution of the loss amounts, whereas when an aggregate limit is established, the distribution of the annual number of losses has to be considered too. Moreover, the risk of fluctuation, which in connection with deductibles works against the insurer anyway (cf. Sterk (1979), Mack (1980, 1983)), is increased even further by an aggregate limit.

* The author would like to thank Dieter Arndt for carrying out the considerable work of programming

In theory, there is no problem involved in developing a formula for the reduction in the deductible rebate resulting from the aggregate limit: the formula for the stop-loss net premium (i.e., the loss expectancy of the stop-loss cover) may be applied without further ado to the distribution of the aggregate retained losses. But the computation itself is a problem, as it is a well-known fact that only in rare cases a closed analytical expression can be given for the distribution of aggregate losses. In Industrial Fire insurance there is the additional problem that the distributions of loss frequency and loss amounts are not known precisely enough (at least for the individual risk to be rated), and rough estimates for some parameters of these distributions are the best we have. It is therefore necessary to use model assumptions that are flexible and cover a broad spectrum of realistic possibilities.

This paper follows the assumption that the distribution of the annual number of losses is Poisson and that the distribution of the loss amounts is lognormal.

It is widely accepted that the Poisson distribution is realistic for the number of losses in Industrial Fire portfolios. Also the validity of the lognormal model for the loss amounts has been demonstrated on several occasions in the past (e.g., BENCKERT (1962), FERRARA (1971), STRAUSS (1975)), and in the field of Industrial Fire in particular.

Generally these distributions cannot immediately be transferred to single risks due to the influence of a big fire on the loss distributions. But a policy for which an aggregate limit is agreed is usually so large that it can be considered as a small portfolio. Therefore the application of the Poisson-lognormal model seems to be an acceptable approximation.

The information available in insurance practice on the loss distribution of the risk to be rated consists for the most part of only the net premium and no more. Therefore in order to estimate the two parameters of the lognormal distribution, additional information is necessary. In this paper it is assumed that the normal deductible rebate is also known, i.e., the reduction in the loss expectancy due to the deductible without the aggregate limit being taken into account. But the calculation of deductible rebates will not be discussed in any further detail as this is dealt with excellently in STERK (1979, p. 180ff). Should the normal deductible rebate not be known, then use can be made of the results of BENCKERT (1962), FERRARA (1971) and STRAUSS (1975), where for one of the two parameters a relatively small range of values was established that is independent of the monetary unit and thus of currency, inflation, etc.

If the mean loss amount, the net premium for full insurance cover, the deductible amount and the corresponding deductible rebate are known, the parameters of the Poisson-lognormal model are determined in full (mean number of losses = net premium/mean loss amount). And in practice these figures are on hand as a rule or they can at least be estimated with a sufficient degree of accuracy by the underwriter. With these figures, the distribution of aggregate losses is determined for the policyholder's retained amount under the deductible before accounting for the aggregate limit. Then for the calculation of the stop-loss net premium, defined by the aggregate limit on this aggregate retained loss, three different ways of approximating the stop-loss net premium are used:

- The "normal power" and "translated gamma" methods based on an analytical approximation of the distribution of aggregate losses. These procedures are extremely simple to handle and require no programming. Up to now, however, little is known of the quality of the results in such cases as the one here with a rather low mean number of losses.
- The method of approximating the loss amount distribution by means of very simple discrete distributions (one, two and three point distributions) for which the stop-loss net premium can be calculated explicitly and simply. Due to the limitation of the amount of each loss by the deductible these methods turn out to give excellent results.
- The recursive procedure for arithmetic distributions, first described by PANJER (1980); the required discretisation follows the "matching moments" method developed by GERBER (1982). This procedure produces results that may be as exact as required depending on the degree of discretisation.

The aim of these comparative calculations is not only to check the quality of these procedures, but first and foremost to find a procedure which is as simple as possible and which at the same time produces acceptable exactness. In addition, the final section investigates the influence of each of the model parameters and suggests a procedure for reducing the large number of possible combinations to a few special cases in order to derive simple rating rules for underwriters.

2. PROBLEM AND NOTATIONS

Let the following data be known for a given risk:

b = net premium (expected value of the aggregate losses) for full insurance cover

c = mean loss amount per loss event

a = deductible amount

$r(a)$ = (net) deductible rebate = reduction in the net premium resulting from the deductible, $0 \leq r(a) \leq 1$

z = annual aggregate limit for the accumulated retained losses under the deductible; z is often expressed as a multiple $z = ka$ of the deductible amount, e.g., $k = 3$.

The expected value of the aggregate retained losses under the deductible before accounting for the aggregate limit is then given by $r(a)b$. The problem is to find the expected value $r(a, z)b$ of the aggregate retained losses considering the aggregate limit z. With $1-[r(a, z)/r(a)]$ we thus obtain the proportion by which the deductible rebate $r(a)$ is to be reduced as a result of the additional aggregate limit.

The following random variables are considered.

X = loss amount per loss event

N = number of losses per year (assumed to be independent of X)

X_a = retained loss amount (per loss event) under deductible a

$$= \begin{cases} X & \text{if} \quad X \leq a \\ a & \text{if} \quad X > a \end{cases}$$

S_a = aggregate retained losses (per year) under deductible a

$$= \begin{cases} 0 & \text{if } N = 0 \\ \sum_{i=1}^{N} (X_a)_i & \text{if } N > 0 \end{cases}$$

where $(X_a)_i$ denotes the retained amount of the ith loss

$S_{a,z}$ = aggregate retained losses (per year) under deductible a and aggregate limit z

$$= \begin{cases} S_a & \text{if} \quad S_a \leq z \\ z & \text{if} \quad S_a > z. \end{cases}$$

With the given data, the following relationships exist

$$b = E(N)E(X)$$

$$c = E(X)$$

$$r(a)b = E(S_a) = E(N)E(X_a)$$

$$r(a) = \frac{E(S_a)}{b} = \frac{E(X_a)}{E(X)}.$$

Then $E(S_{a,z}) = r(a, z)b$ is to be calculated under the assumption that N is subject to a Poisson distribution and X to a lognormal distribution, i.e., (with Φ denoting the standard normal distribution function)

$$F(x) = p(X \leq x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \quad \text{for } 0 < x < \infty,$$

$$f(x) = F'(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\tfrac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right),$$

$$p(N = i) = \frac{\lambda^i}{i!} e^{-\lambda} \quad \text{for } i = 0, 1, 2, \ldots.$$

The parameter λ is given by

$$\lambda = E(N) = \frac{b}{c},$$

and the parameters μ and σ can be deduced from the values for a, $r(a)$ and c with the aid of the formulae

$$c = E(X) = \exp(\mu + \tfrac{1}{2}\sigma^2),$$

$$r(a) = \frac{1}{E(X)}\left\{\int_0^a xf(x)\,dx + a(1 - F(a))\right\}$$

$$= \Phi\left(\frac{\ln a - \mu}{\sigma} - \sigma\right) + \frac{a}{c}\left(1 - \Phi\left(\frac{\ln a - \mu}{\sigma}\right)\right)$$

(cf. STERK 1979, p. 234). For this purpose it is convenient to introduce

$$t = \frac{a}{c}.$$

Then the equation for $r(a)$ can be rewritten as

$$r(a) = \Phi\left(\frac{\ln t}{\sigma} - \frac{\sigma}{2}\right) + t\left(1 - \Phi\left(\frac{\ln t}{\sigma} + \frac{\sigma}{2}\right)\right).$$

This equation has a unique solution σ (given t and $r(a)$) because the right-hand side is a strictly increasing function of σ. Parameter μ is thus replaced with t. Besides z we now have to work with the three model parameters t, σ and λ.

Should it happen that the deductible rebate $r(a)$ is not known, it may be possible to choose the parameter value of σ from the interval [2, 2.5] in accordance with the results of BENCKERT (1962), FERRARA (1971) and STRAUSS (1975).

If $E(S_a - z)^+$ denotes the stop-loss net premium with priority (stop-loss attachment point) z on the aggregate retained losses S_a, i.e.,

$$E(S_a - z)^+ = E(S_a) - E(S_{a,z}),$$

the required reduction in the deductible rebate comes to

$$1 - \frac{r(a, z)}{r(a)} = \frac{E(S_a - z)^+}{E(S_a)}.$$

This expression, i.e., the stop-loss net premium measured as a fraction of the mean aggregate retained losses without an aggregate limit, will be called "relative stop-loss net premium" in the following discussion. Similarly, the value

$$k = \frac{z}{a}$$

i.e., the priority expressed as a multiple of the deductible amount, is referred to as "relative priority".

The curve of values of the relative stop-loss net premium as a function of the relative priority is called "stop-loss curve"; it begins at point (0; 1), is degressively and strictly decreasing (convex) and runs to point (∞; 0). The "relevant area" is that part of the curve in which the relative stop-loss net premium amounts to between 50% and 5% as in practice the majority of cases occur in this range.

3. CALCULATION METHODS

As the distribution of the aggregate retained losses S_a, which is required for an exact calculation of the stop-loss net premium, cannot be given in closed form, various approximation methods have been developed in actuarial literature. Several of these methods have been applied in the problem here. As most of these methods are well-known, no further details are given. This applies to the following methods:

1. Normal power method, see BEARD, PENTIKAINEN and PESONEN (1968/1977, p. 43ff); BERGER (1972); KAUPPI and OJANTAKANEN (1969); PESONEN (1969). More precisely, the NP2 method was used here, i.e., the changed variable was calculated from a quadratic equation.
2. Translated gamma method, see BOHMAN and ESSCHER (1963/1964); SEAL (1977); BOWERS, GERBER, HICKMAN, JONES and NESBITT (1982). In the expression for the stop-loss net premium (cf. SEAL 1977, p. 215) the incomplete gamma function occurs.
3. Recursive calculation of the stop-loss net premium by means of an arithmetic discretisation of the loss amount distribution, see GERBER (1982) who uses the recursive procedure of PANJER (1980). In the problem here the discretisation method called "matching moments" was used where the probability weights for the discretised variable are calculated in such a way that within adjacent pairs of intervals the first two moments for the discretised loss amount are equal to those of X_a according to the lognormal distribution. As an obvious extension of the recursion formula stated by PANJER and GERBER, the occurrence of losses of amount 0 for the discretised distribution was explicitly admitted as this proved to be suitable due to the skewness of the distribution of X_a in order to avoid negative probabilities.

With the normal power and the translated gamma method an estimation of the approximation error is not possible. But for the method with a discretisation of the loss amount distribution an upper bound for the approximation error can be developed using a metric introduced by GERBER (1980) (Chapter 7.3):

$$\max_{z \geq 0} |E(S_a - z)^+ - E(\tilde{S}_a - z)^+| \leq \lambda \max_{0 \leq x \leq a} |E(X_a - x)^+ - E(\tilde{X}_a - x)^+|$$

$$= \lambda \max_{0 \leq x \leq a} \left| \int_x^a (\tilde{F}(y) - F(y))\, dy \right|,$$

where the symbol $\tilde{}$ refers to the discrete approximating distribution.

If the discretised loss amount distribution only has one or two atoms, the distribution of the aggregate losses and thus the stop-loss net premium can generally be calculated very easily without recursion formula. On account of the finite range $(0, a]$ of the retained loss amount X_a it does not seem unreasonable to approximate the distribution of X_a by such a one-point or two-point distribution. Indeed it will be shown that this method in the problem here leads to astoundingly good results. For this method too, the above formula for the error bound holds true. For the choice of the atoms of the approximating loss amount distribution there are several possibilities (one-point: lower bound, upper bound, third approach; two-point: 1st, 2nd, 3rd possibility), details of which are given in the appendix.

4. COMPARISON OF CALCULATION METHODS

In principle, the matching moments discretisation is the most accurate method of calculating as the accuracy can theoretically be improved as far as desired by raising the number $n+1$ of discretisation points. It is possible that numerical problems will arise when the value of n reaches a certain size, but here it was not necessary to go so far as a better approximation was already arrived at for a relatively small n than with the other methods. Table 1 shows a typical example

TABLE 1
RELATIVE STOP-LOSS NET PREMIUM (%) USING VARIOUS METHODS
(PARAMETERS $\sigma = 2$, $t = 1$, $\lambda = 3$)

Method	$k = 1.0$	$k = 1.5$	$k = 2.0$	$k = 2.5$
Matching moments, $n = 100$	32.573	16.375	7.4675	3.2266
$n = 30$	32.571	16.373	7.4663	3.2259
$n = 10$	32.552	16.350	7.4558	3.2187
Normal power	33.4	16.9	7.97	3.56
Translated gamma	32.1	15.9	7.44	3.33
Two-point, 1st possibility	33.4	16.1	8.03	3.218
2nd possibility	32.0	16.9	7.05	3.41
3rd possibility	32.52	16.37	7.452	3.244
One-point, lower bound	21	6	1.4	0.2
upper bound	35	23	9.6	5.8
third approach	33.5	14.8	7.30	2.97

With the matching moments method for $n = 100$ the maximum error amounts to less than $\pm 0.05\%$ according to the inequality in Section 3, i.e., the exact value e.g. for $k = 1.0$ is between 32.523% and 32.623%. The normal power method generally overestimated the stop-loss net premium; in all the parameter combinations examined ($\sigma = 2$, $t = 0.1$, 0.3, 1.0, 3.0, 10.0, $\lambda = 1$, 3, 10, 30), only for $\lambda = 1$ and $t \leq 1$ was there a small area where this was not the case. Where the aggregate loss distribution was very skewed ($t = 10$, $\lambda = 1$) the normal power method overestimated the stop-loss net premium in the relevant area by more than one half of the true value in some cases. If the aggregate loss distribution is practically a normal distribution ($\lambda = 30$, $t \leq 3$; here the skewness is <0.5) the normal power method, like the translated gamma and the two-point too, produces a very good approximation to 3 decimal places. The relative error however increases with higher priorities (i.e., with a lower stop-loss net premium).

The values produced by the translated gamma method were above and below the exact values in all of the parameter combinations examined, i.e., the translated gamma stop-loss curve intersects the exact stop-loss curve, and in the cases examined more than once. A very good approximation is produced of course near to the intersections But the accuracy of the translated gamma method is not at every point better than that of the normal power method.

The stop-loss curves of each of the two-point methods also meet the exact curve usually more than once. As shown in Table 2, the third two-point method seems to provide the best approximation apart from the matching moments method. Sometimes, however, this method can produce a small range of values, where the deviation is greater than in the normal power method.

Of the three one-point methods only the third approach produces acceptable results especially if the skewness of the aggregate loss distribution is small (e.g., <0.5). This method may underestimate the true result. The other two methods should be considered as being the simplest way of providing lower and upper bounds rather than being approximations.

Table 2 is an attempt to compare the accuracy of the various methods. For this the values of the methods per priority were put in order of accuracy; the method with the value nearest to the matching moments value was given the order number 1, going down to order number 8 for the method with the value which was furthest removed. Then for each method the mean order number was calculated for a larger number of priorities, which were chosen equidistant in the relevant area.

TABLE 2

MEAN ORDER NUMBER OF THE VARIOUS METHODS IN TERMS OF ACCURACY

Parameters ($\sigma = 2$ throughout)	Method (see key below) NP	TG	TP1	TP2	TP3	OPL	OPU	OP3
$t = 0.1$, $\lambda = 3$	4.1	3.9	3.1	2.9	1.8	8.0	7.0	5.2
$\lambda = 10$	5.1	4.2	2.5	2.5	1.6	8.0	7.0	5.2
$t = 1.0$, $\lambda = 3$	4.2	3.0	3.9	3.5	1.6	8.0	7.0	4.8
$\lambda = 10$	5.0	3.6	3.3	2.5	1.5	8.0	7.0	5.0
$t = 10$, $\lambda = 3$	3.7	2.5	4.2	3.6	1.8	7.0	8.0	5.2
$\lambda = 10$	3.9	2.5	4.5	3.9	1.7	7.5	7.5	4.6

Key NP = normal power; TG = translated gamma; TPi = two-point ith possibility, OPL = one-point, lower bound, OPU = one-point, upper bound, OP3 = one-point, third approach

The results in this table cannot however be simply transferred to other parameter combinations. For $t = 1$ and $\lambda = 30$, for example, procedure TP2 has a lower mean order number than TP3; here however, all the methods are exact to three decimal places. For $\lambda = 1$ the value according to normal power in the relevant area is sometimes higher than the upper bound given by OPU.

Finally in this connection certain computing problems must be mentioned too. The possible occurrence of negative probabilities in the matching moments method has already been pointed out by GERBER (1982). But these negative probabilities do not seem to have any distorting influence on the stop-loss net premiums calculated with them. With the two-point methods it is possible that when the Poisson probabilities for high priorities are calculated, an underflow will occur, meaning that values are produced that are too small to be expressed in the computer. As these are summands and not very many either, they can be given a value of zero without having any noticeable effect on the accuracy but in general an appropriate instruction should be included in the computer program to avoid an abnormal program termination. An overflow in the translated gamma method occurred for $\lambda = 30$ and $t = 0.1$ or 0.3 in the calculation of the incomplete gamma function, i.e., values were produced which were so large they could not be expressed in the computer. This error can only be avoided by means of applying special techniques in calculating the incomplete gamma function ratio (see KHAMIS and RUDERT 1965). Difficulties in calculating were encountered in all the methods apart from the normal power method—and, of course, the one-point methods. As the normal power method produces results that are nearly always on the safe side and as the safety loading increases relative to the decrease in the stop-loss net premium as it should, this method can be generally recommended, especially if the results have to be produced quickly and without any programming.

5. DEPENDENCY OF THE RESULTS ON THE PARAMETERS

In practice it is recommended that underwriters are given simple rating tables or curves so that they do not have to consult the actuarial department each time a policy with an aggregate limit comes up. In view of the dependency of the stop-loss net premium on three parameters (σ, λ, t) it does not seem possible to provide a calculation model of the kind mentioned. Surprisingly enough however, it is possible to eliminate all three parameters to a large extent if a slight reduction in accuracy is acceptable. In view of the uncertainty of the parameter values pertaining to any one risk, this loss in accuracy can be ignored.

At first it is not automatically clear what influence the variation of one single parameter will have on the relative stop-loss net premium where the other parameters and the relative priority remain constant, as the incorporation of relative values may produce different results to those produced by absolute values. In the case where parameter λ increases, the mean number of losses increases too while the priority (both relative and absolute) remains unchanged. It is therefore obvious that the stop-loss net premium increases overproportionally and leads to an increase in the relative stop-loss net premium. In case of variation of the parameter σ or t it is best to observe the shape of the density function of the amount of retained loss X_a. As t increases, the proportion $E(X_a)/a$ decreases too, meaning that the distribution of X_a is skewed more and more to the right. If σ, λ and k are constant, therefore, the absolute priority ka will increase in relation to the mean aggregate retained losses $E(S_a) = \lambda E(X_a)$ so that the relative stop-loss net premium decreases. The same applies when parameter σ is raised:

a higher σ means a more skewed lognormal distribution with a higher expected value $E(X)$. If t is constant, the deductible $a = tE(X)$ increases too as that the distribution of X_a becomes more skewed. If λ is constant, the absolute priority therefore becomes greater in relation to $E(S_a)$ so that the relative stop-loss net premium decreases.

On account of this, with a constant λ for two pairs of parameters (σ, t), $(\tilde{\sigma}, \tilde{t})$, a similar stop-loss net premium is to be expected if the same proportion $E(X_a)/a = r(a)/t$ is produced in both cases, i e., if

$$\frac{1}{t}\Phi\left(\frac{\ln t}{\sigma}-\frac{\sigma}{2}\right)-\Phi\left(\frac{\ln t}{\sigma}+\frac{\sigma}{2}\right)=\frac{1}{\tilde{t}}\Phi\left(\frac{\ln \tilde{t}}{\tilde{\sigma}}-\frac{\tilde{\sigma}}{2}\right)-\Phi\left(\frac{\ln \tilde{t}}{\tilde{\sigma}}+\frac{\tilde{\sigma}}{2}\right).$$

Table 3 shows that this is in fact the case. Here the value of t for the various values of σ is selected in such a way that the equation above holds with $\tilde{\sigma} = 2$ and $\tilde{t} = 1$.

TABLE 3

COMPENSATING A VARIATION OF t WITH A VARIATION OF λ

Parameter Values		Relative Stop-Loss Net Premium (%) for $\lambda = 3$			
σ	t	$k = 1$	$k = 1.5$	$k = 2$	$k = 2.5$
1.6	1.70	31.4	15.2	6.75	2.79
1.8	1.33	32.0	15.8	7.14	3.02
2.0	1.00	32.6	16.4	7.47	3.23
2.2	0.72	33.0	16.8	7.72	3.39
2.4	0.50	33.3	17.2	7.94	3.55

This being so, it is possible to transpose parameter values $\sigma \neq 2$ to the case $\sigma = 2$ by an appropriate alteration of the parameter t without any essential change in the stop-loss curve. In this way parameter σ is practically eliminated.

If σ is constant a similar situation arises for the influence of variations of the parameters t and λ. Table 4 shows that an increase of t can be compensated by an appropriate increase of λ so that the stop-loss curve remains almost unchanged.

Therefore, parameter t can be eliminated by an appropriate correction of the value of λ. Finally, with σ and t constant and a given value for the relative

TABLE 4

COMPENSATING A VARIATION OF σ WITH A VARIATION OF t

Parameter Values		Relative Stop-Loss Net Premium (%) for $\sigma = 2$				
t	λ	$k = 1$	$k = 1.5$	$k = 2$	$k = 2.5$	$k = 3$
0.1	2.4	52.0	35.4	21.9	13.6	8.03
0.3	3.4	52.6	35.4	22.3	13.5	7.68
1.0	6	53.4	35.7	22.6	13.4	7.59
3.0	12	54.2	36.1	22.6	13.2	7.32
10.0	31	56.0	37.5	23.1	13.2	7.01

stop-loss net premium, there is an almost linear connection between λ and that value of relative priority which leads to the given stop-loss net premium for this λ:

Example for $\sigma = 2$, $t = 1$:

Parameter Value of λ	1	3	10	30
Relative priority k corresponding to 10% stop-loss net premium	1.09	1.83	3.96	9.74
Straight line $0.31\lambda + 0.81$	1.12	1.74	3.91	10.11
Relative priority k corresponding to 30% stop-loss net premium	0.69	1.06	2.54	6.83
Straight line $0.21\lambda + 0.46$	0.67	1.09	2.56	6.76

This makes it possible to derive from the stop-loss curves for 2 values of the parameter λ the curves for the other values of λ approximately by means of interpolation or extrapolation.

To sum up then, we may say: the stop-loss curves resulting from the lognormal distribution by a deductible with an aggregate limit have very similar shapes for the relevant parameter values of σ, t and λ, and with the aid of appropriate parameter transformations they can be approximately interchanged. It is therefore possible to represent the effect of an aggregate limit on the expected losses in such a way that it can be determined using only a few curves or tables without any great reduction in accuracy.

APPENDIX

Calculation of the Stop-Loss Net Premium by Simple Discrete Approximations of the Distribution of Loss Amounts with One-Point or Two-Point Distributions

This appendix will deal with one-point and two-point distributions as well as a special three-point distribution for the loss amounts, i.e., distributions that allow for only one, two or three different loss amounts. For such distributions, the distribution of the aggregate losses and thus the stop-loss net premium can be calculated exactly without great difficulty. On account of the finite range $(0, a]$ of the loss amount X_a it does not seem unreasonable to approximate the distribution of X_a by such a distribution.

In the following, we shall frequently be needing the first three moments about zero of the retained loss X_a; for $i = 1, 2, 3 \ldots$ we have:

$$E(X_a)^i = \int_0^a x^i \, dF(x) + a^i(1 - F(a))$$
$$= \exp\left(i\mu + \tfrac{1}{2} i^2 \sigma^2\right) \Phi\left(\frac{\ln a - \mu - i\sigma^2}{\sigma}\right) + a^i\left(1 - \Phi\left(\frac{\ln a - \mu}{\sigma}\right)\right).$$

Here again it is convenient to replace parameter μ with t by writing

$$a = tc = tE(X) = t \exp(\mu + \tfrac{1}{2}\sigma^2)$$

for the mean loss amount. This results in

$$E(X_a)^i = c^i \left\{ \exp\left(\tfrac{1}{2} i (i-1)\sigma^2\right) \Phi\left(\frac{\ln t}{\sigma} - (i - \tfrac{1}{2})\sigma\right) + t^i \left(1 - \Phi\left(\frac{\ln t}{\sigma} + \tfrac{1}{2}\sigma\right)\right)\right\}.$$

The values

$$h_i = c^{-i} E(X_a)^i$$

now only depend on $t = a/c$ and σ. Note that $h_1 = r(a)$ applies. In the following discussion it will always be assumed that the values t, σ and λ are known and therefore h_1, h_2, h_3 too.

A1. *Approximation by Means of One-Point Distributions*

The most simple way of approximating is to work only with a constant loss amount $\theta = E(X_a)$, i.e. to approximate the aggregate retained losses S_a by means of θN. According to a theorem of BUHLMANN, GAGLIARDI, GERBER and STRAUB (1977), this results in a lower bound for the stop-loss net premium for each priority z, i.e.,

$$E(S_a - z)^+ \geq E(\theta N - z)^+ = \theta E\left(N - \frac{z}{\theta}\right)^+.$$

Another approximation stemming from BENKTANDER only uses losses of the (maximum) amount a. So that the expected value of the aggregate losses remains unchanged, the mean number of losses must be reduced mechanically to

$$\lambda^* = \frac{\theta}{a}\lambda.$$

If N^* denotes the Poisson variable belonging to λ^* then S_a will be approximated by aN^* and we have

$$E(S_a - z)^+ \leq E(aN^* - z)^+ = aE\left(N^* - \frac{z}{a}\right)^+,$$

i.e., an upper bound for the stop-loss net premium is obtained. This also results from the theorem of BUHLMANN, GAGLIARDI, GERBER and STRAUB (1977). Here the fact is employed that the distribution of aggregate losses based on the number of losses N^* and the constant loss amount a is identical with the distribution of aggregate losses which results from the number of losses N and the loss amount "a with probability θ/a or 0 with probability $1 - \theta/a$". This is a special feature of the Poisson distribution.

In these cases, the explicit calculation of the stop-loss net premium is quite simple:

$$E(N-u)^+ = \sum_{i>u} (i-u)p(N=i)$$

$$= \sum_{i>u} i\frac{\lambda^i}{i!}e^{-\lambda} - up(N>u)$$

$$= \lambda \sum_{i>u} \frac{\lambda^{i-1}}{(i-1)!}e^{-\lambda} - up(N>u)$$

$$= \lambda p(N=[u]) + (\lambda - u)p(N>u),$$

$[u]$ denoting the integer part of u.

In the special problem here, with a priority of $z = ktc$, this produces for the stop-loss net premium $E(S_a - z)^+$

- a lower bound

$$ch_1\{\lambda p(N=[v]) - (v-\lambda)(1-p(N \leqslant v))\}$$

with $v = kt/h_1$

- and an upper bound

$$ch_1\{\lambda p(N^*=[k]) - (v-\lambda)(1-p(N^* \leqslant k))\}$$

with $v = kt/h_1$, N^* being Poisson distributed with parameter $(h_1/t)\lambda$.

In a third approach, due to BENKTANDER (1974), the distribution of the aggregate losses S_a is directly approximated by the distribution of $\zeta\tilde{N}$ where $\tilde{N}$ is Poisson distributed with parameter $E(\tilde{N}) = \tilde{\lambda}$ and the values of ζ, $\tilde{\lambda}$ are determined by the equations

$$E(\zeta\tilde{N}) = \zeta\tilde{\lambda} = E(S_a),$$

$$\operatorname{Var}(\zeta\tilde{N}) = \zeta^2\tilde{\lambda} = \operatorname{Var}(S_a).$$

This yields

$$\zeta = \frac{\operatorname{Var}(S_a)}{E(S_a)} = \frac{E(X_a)^2}{E(X_a)}$$

$$\tilde{\lambda} = \frac{(E(S_a))^2}{\operatorname{Var}(S_a)} = \frac{(E(X_a))^2}{E(X_a)^2}\lambda.$$

For the stop-loss net premium with priority z we then get the approximation

$$E(S_a - z)^+ \approx E(\zeta\tilde{N} - z)^+ = \zeta E\left(\tilde{N} - \frac{z}{\zeta}\right)^+.$$

In the problem to be solved here this leads to (with priority $z = ktc$)

$$ch_1\{\lambda p(\tilde{N}=[\tilde{v}]) - (v-\lambda)(1-p(\tilde{N} \leqslant \tilde{v}))\}$$

with $v = kt/h_1$, $\tilde{v} = kth_1/h_2$ and $\tilde{N}$ being Poisson distributed with parameter $(h_1^2/h_2)\lambda$.

It is easy to see that generally $\tilde{\lambda} \leq \lambda$ and $\theta \leq \zeta \leq a$ holds and using again the theorem of BUHLMANN, GAGLIARDI, GERBER and STRAUB (1977) it can be shown that the stop-loss net premium according to this third approach is between the bounds defined above, i.e.,

$$E(\theta N - z)^+ \leq E(\zeta \tilde{N} - z)^+ \leq E(aN^* - z)^+$$

A2 *Approximation by Means of Two-Point Distributions*

Calculating the stop-loss net premium with priority z is still a quite simple matter for a distribution of loss amounts that only provides for two different loss amounts $x < y$ with probability p for x and $q = 1 - p$ for y: if W_{xy} denotes the corresponding variable of aggregate losses with λ-Poisson distributed number of losses N, then the aggregate losses, in the case of exactly $N = j$ losses, i of which have an amount y, come to

$$W_{xy} = iy + (j-i)x = i(y-x) + jx, \qquad i = 0, 1, \ldots, j,$$

with probability

$$\binom{j}{i} q^i p^{j-i} \frac{\lambda^j}{j!} e^{-\lambda}.$$

The constraint $W_{xy} \leq z$ is equivalent to $i \leq (z - jx)/(y-x)$. As $[z/x]$ losses may occur for $W_{xy} \leq z$ at the most, this leads to

$$E(z - W_{xy})^+ = \sum_{j < z/x} \frac{\lambda^j}{j!} e^{-\lambda} \sum_{i \leq w} \binom{j}{i} p^{j-i} q^i (z - jx - i(y-x))$$

with $w = \min(j, (z-jx)/(y-x))$.

This can be worked out on a programmable pocket calculator with 10 memory registers; the size of the factorials does not constitute a problem either as long as the corresponding summands are calculated recursively. Finally the stop-loss net premium is given by

$$\begin{aligned} E(W_{xy} - z)^+ &= E(W_{xy}) - z + E(z - W_{xy})^+ \\ &= \lambda(px + qy) - z + E(z - W_{xy})^+. \end{aligned}$$

There are several ways of approximating the retained loss amount X_a by means of a two-point distribution.

1st possibility:
The distribution of X_a is produced by truncating the distribution of X at the point a, i.e., the distribution of X_a always has a point mass amounting to $1 - F(a)$ at point a. Particularly where low deductibles are concerned, the obvious way is therefore to choose the two-point distribution in such a way that $y = a$. Then the other point x and its point mass p are selected so that the first two moments are

equal to those of X_a, i.e.:

$$px+(1-p)a=E(X_a)$$
$$px^2+(1-p)a^2=E(X_a)^2.$$

This leads to

$$p=\frac{(t-h_1)^2}{t^2-2th_1+h_2}$$
$$q=1-p$$
$$x=\frac{c}{p}(h_1-qt)$$
$$y=tc.$$

2nd possibility:
If a value for y other than a is admitted, the two-point distribution can be selected in such a way that the first three moments are equal to those of X_a, that is

$$px+(1-p)y=E(X_a)$$
$$px^2+(1-p)y^2=E(X_a)^2$$
$$px^3+(1-p)y^3=E(X_a)^3.$$

If ξ denotes the skewness of X_a, i.e.,

$$\xi=(h_3-h_1(3h_2-2h_1^2))(h_2-h_1^2)^{-3/2},$$

then we get

$$p=\frac{1}{2}+\frac{\xi}{2\sqrt{4+\xi^2}}$$
$$q=1-p$$
$$x=c\left(h_1-\sqrt{\frac{q}{p}(h_2-h_1^2)}\right)$$
$$y=c\left(h_1+\sqrt{\frac{p}{q}(h_2-h_1^2)}\right).$$

For the corresponding aggregate losses W_{xy}, the first three moments are equal to those of the aggregate retained losses S_a, as is the case too with the normal power and translated gamma methods.

3rd possibility (special three-point distribution):
If a procedure is desired whereby a point mass of $y=a$ is retained as in the first possibility and at the same time the first three moments of X_a are considered as in the second possibility, then this is feasible, similiarly as in the upper bound one-point distribution, if a point mass at loss amount 0 is added and the Poisson

parameter is adjusted accordingly. More precisely a three-point distribution is adjusted with points 0, x, a and point masses u, v, w so that

$$u+v+w=1$$

$$vx+wa=E(X_a)$$

$$vx^2+wa^2=E(X_a)^2$$

$$vx^3+wa^3=E(X_a)^3.$$

This yields

$$w=\frac{h_1h_3-h_2^2}{(h_1t^2-2h_2t+h_3)t}$$

$$v=\frac{(h_1-wt)^2}{h_2-wt^2}$$

$$u=1-v-w$$

$$x=\frac{c}{v}(h_1-wt).$$

In the case of a λ-Poisson distributed number of losses, the corresponding aggregate loss distribution does not change if the loss amount 0 and its point mass u are omitted and the other point masses v and w are raised accordingly and the parameter λ reduced, i.e.,

$$p=\frac{v}{v+w}$$

$$q=\frac{w}{v+w}=1-p$$

$$\lambda^*=\lambda(v+w).$$

The corresponding stop-loss net premium can therefore be calculated using the two-point distribution given by p, q, x and $y=a=tc$; in this case the reduced mean number of losses λ^* is to be used instead of λ.

Further possibilities:
If a value for y is admitted in the 3rd possibility other than $y=a$, it is possible to have the same first four moments of the special three point distribution as those of X_a. Another possibility is to break the interval $[0, a]$ into two intervals and to apply each of the first two one-point methods to each of these intervals. In this way, two-point distributions are produced which give improved upper and lower bounds for the stop-loss net premium.

REFERENCES

BEARD, R E, PENTIKAINEN, T and PESONEN, E (1969, 1977) *Risk Theory* 2nd ed Chapman and Hall London

BENCKERT, L-G (1962) The Lognormal Model for the Distribution of One Claim *ASTIN Bulletin* **2**, 9–23

BENKTANDER, G (1974) A Motor Excess Rating Problem· Flat Rate with Refund *ASTIN-Colloquium Turku*

BERGER, G (1972) Integration of the Normal Power Approximation *ASTIN Bulletin* **7**, 90–95

BOHMAN, H and ESSCHER, F (1963, 1964) Studies in Risk Theory With Numerical Illustrations Concerning Distribution Functions and Stop-Loss Premiums *Skandinavisk Aktuarietidskrift* **46**, 173–225 and **47**, 1–40

BOWERS, N L, GERBER, H U, HICKMAN, J C, JONES, D A and NESBITT, C J (1982) *Risk Theory* Society of Actuaries U S A

BUHLMANN, H, GAGLIARDI, B, GERBER, H U and STRAUB, E (1977) Some Inequalities for Stop-Loss Premiums *ASTIN Bulletin* **9**, 75–83

FERRARA, G (1971) Distributions des Sinistres Incendie Selon leur Cout *ASTIN Bulletin* **6**, 31–41

GERBER, H U (1980) *An Introduction to Mathematical Risk Theory* Irwin Homewood, Illinois

GERBER, H U (1982) On the Numerical Evaluation of the Distribution of Aggregate Claims and its Stop-Loss Premiums *Insurance Mathematics and Economics* **1**, 13–18.

KAUPPI, L and OJANTAKANEN, P (1969) Approximations of the Generalized Poisson Function *ASTIN Bulletin* **5**, 213–226

KHAMIS, S H and RUDERT, W (1965) *Tables of the Incomplete Gamma Function Ratio* Justus von Liebig-Verlag Darmstadt

MACK, T (1980) Uber die Aufteilung des Risikos bei Vereinbarung einer Franchise. *Blatter der Deutschen Gesellschaft für Versicherungsmathematik* **14**, 631-650

MACK, T (1983) The Utility of Deductibles from the Insurer's Point of View *ASTIN-Colloquium Lindau*

PANJER, H H (1980) The Aggregate Claims Distribution and Stop-Loss Reinsurance. *Transactions of the Society of Actuaries* **32**, 523–545

PESONEN, (1969) NP-Approximation of Risk Processes *Skandinavisk Aktuarietidskrift* **52**, Supplement, 63–69

SEAL, H L (1979) Approximations to Risk Theory's $F(x, t)$ by Means of the Gamma Distribution *ASTIN Bulletin* **9**, 213–218

STERK, H-P (1979) *Selbstbeteiligung unter risikotheoretischen Aspekten* Verlag Versicherungswirtschaft Karlsruhe

STRAUSS, J (1975) Deductibles in Industrial Fire Insurance. *ASTIN Bulletin* **8**, 378-393

HOMOGENEOUS PREMIUM CALCULATION PRINCIPLES

By Axel Reich

Kölnische Rückversicherungs-Gesellschaft, Cologne

Abstract

A premium calculation principle π is called positively homogeneous if $\pi(cX) = c\pi(X)$ for all $c > 0$ and all random variables X. For all known principles it is shown that this condition is fulfilled if it is satisfied for two specific values of c only, say $c = 2$ and $c = 3$, and for only all two point random variables X. In the case of the Esscher principle one value of c suffices. In short this means that local homogeneity implies global homogeneity. From this it follows that in the case of the zero utility principle or Swiss premium calculation principle, the underlying utility function is of a very specific type.

A very general theorem on premium calculation principles which satisfy a weak continuity condition, is added. Among others the proof uses Kroneckers Theorem on Diophantine Approximations.

Keywords

Premium principles, homogeneity, utility functions.

1. Introduction

In actuarial practice one generally uses only three premium calculation principles, namely the expected value principle, the variance principle and the standard deviation principle. Apart from these there are many other principles for determining a premium for a risk: all these are examined in the new textbook by Goovaerts, De Vylder and Haezendonck (1984). A central theme is the analysis of the principles which fulfil some desirable properties such as translation invariance, (sub-)additivity, iterativity, homogeneity etc. For example the expected value principle is always additive and homogeneous, but it is iterative or translation invariant only in the case of a vanishing loading.

If a premium principle is defined by a utility function, then the above mentioned, (so-called) plausible properties are in general very restrictive: The Swiss premium calculation principle e.g. is translation invariant if and only if the corresponding utility function is exponential or linear, and it is positively homogeneous if the utility function $u(x)$ is—up to linear transformations—a power of x. Therefore, e.g., the Swiss premium calculation principle is both translation invariant and homogeneous only in the case of a linear utility function. Such an analysis has been performed already for all known principles and all properties mentioned above. If π denotes a premium calculation principle, which therefore to any real random variable X assigns a real number $\pi(X)$—the premium of X—then in all cases it turns out that, e.g., in the case of translation invariance (that means

$$\pi(X+c)=\pi(X)+c \tag{1}$$

for all risks X and all real numbers c), it is sufficient to know equation (1) for all $c \in \mathbb{R}$ and only all $X \in D_2$ (D_2 the set of all 2-atomic random variables): If (1) is valid for all $c \in \mathbb{R}$, $X \in D_2$, then automatically (1) is fulfilled also for those X not in D_2. This reduction to the essence of a property has been worked out for the property of translation invariance by REICH (1984) in a definitive sense: Any principle is already translation invariant (i.e., (1) holds for all $c \in \mathbb{R}$ and all risks X), if (1) is fulfilled for all $X \in D_2$ and two specific values of c only, say $c=1$ and $c=\sqrt{2}$. In case of the Orlicz principle (HAEZENDONCK and GOOVAERTS (1982)) a single value of c, e.g., $c=1$, suffices. A further reduction is impossible as one can see from the counterexamples in REICH (1984).

In the case of the property of homogeneity (sometimes also called proportionality)

$$\pi(cX)=c\pi(X) \tag{2}$$

(more exactly we will examine positively homogeneous principles, i.e., $c \in \mathbb{R}^+$) we will now give a similar analysis of the analogous problems. Equation (2) means for $c=\frac{1}{2}$, say, that the premium of X should be homogeneously divided in two equal parts, if the risk X is split up into two parts in a homogeneous way. The aim of this paper therefore is to give an answer to the question: How little does one really need to know, to have already property (2) in full generality (i.e., for arbitrary risk X and arbitrary $c \in \mathbb{R}^+$)? Of course, this leads to other conditions than in the case of translation invariance and other principles are now of special interest. A mere corollary from the results (still to be formulated and proved) should be mentioned here: Take for example the Swiss premium calculation principle. If (2) holds for all $X \in D_2$ only and for all $c \in [\frac{1}{3}, \frac{1}{2}]$, then (2) holds automatically for all risks X and all $c \in \mathbb{R}^+$. There is therefore no difference in homogeneity as a local or global property. This fact is a trivial consequence of theorem 2.2, which is best possible in the precise sense specified there. Moreover for every known premium calculation principle the following is true ($X \in D_2$): If (2) holds in the two special cases $c=\frac{1}{2}$ and $c=\frac{1}{3}$ only, then again (2) is fulfilled for all $c>0$.

From this one can prove that even an extremely weaker assumption than the homogeneity is (with the Orlicz principle as the only exception) very restrictive for all utility principles.

2. RESULTS AND REMARKS

Among the known principles the following are in every case (i.e., independent of the choice of the corresponding parameters or utility functions) positively homogeneous: Expected value principle, maximal loss principle, percentile principle, standard deviation principle and Orlicz principle. The variance principle on the contrary is certainly not positively homogeneous in the case of a non-vanishing loading (cf. GOOVAERTS, DE VYLDER and HAEZENDONCK (1984)).

For the remaining cases of the Swiss Premium calculation principle, the zero utility principle (which is indeed a special case of the Swiss premium calculation principle, but has for technical reasons to be treated separately) and the Esscher principle it will now be proved for example: If one has for all $X \in D_2$

$$\pi(\tfrac{1}{2}X) = \tfrac{1}{2}\pi(X) \quad \text{and} \quad \pi(\tfrac{1}{3}X) = \tfrac{1}{3}\pi(X),$$

then

$$\pi(cX) = c\pi(X)$$

holds for all X and all $c \in \mathbb{R}^+$, i.e., π is positively homogeneous. More generally and more exactly:

2A. π = *zero utility principle*

This principle was introduced by BUHLMANN (1970). One starts with a utility function u with $u'(x) \geq 0$, $u''(x) \leq 0$. For a given risk X the premium $P = \pi(X)$ is determined by

$$E[u(P - X)] = u(0). \tag{3}$$

We prove

THEOREM 2.1. *For fixed, positive* $c_1, c_2 \neq 1$ *let* $\log c_1/\log c_2$ *not be rational If for every* $X \in D_2$

$$\pi(c_1 X) = c_1\pi(X) \quad \textit{and} \quad \pi(c_2 X) = c_2\pi(X)$$

hold, then u is linear. Conversely, if u is linear, then for all X and all $c \in \mathbb{R}^+$

$$\pi(cX) = c\pi(X)$$

holds, i.e. π *is positively homogeneous.*

REMARK Theorem 2.1 is best possible in the following sense: For any pair $c_1, c_2 \in \mathbb{R}^+$ ($c_1 = c_2$ is admissible), which the condition of theorem 2.1 (i.e., $\log c_1/\log c_2 \in \mathbb{Q}$) does not fulfil, there is a non-linear utility function u such that

$$\pi(c_i X) = c_i \pi(X), \qquad i = 1, 2, X \in D_2,$$

holds. In this case the zero utility principle certainly is not positively homogeneous.

2B. π = *Swiss premium calculation principle*

This principle was introduced by BUHLMANN, GAGLIARDI, GERBER and STRAUB (1977). If $z \in [0, 1]$ and u is a strictly monotonic, continuous function on $\mathbb{R}$ the premium $P = \pi(X)$ for a risk X is given by the equation

$$E[u(X - zP)] = u((1-z)P). \tag{4}$$

In the case $z = 1$ and by the substitution $u(x) \to -u(-x)$ one just gets the zero utility principle. By a different proof than in the case 2A one proves for $0 \leq z < 1$

THEOREM 2.2. *For fixed, positive* c_1, $c_2 \neq 1$ *let* $\log c_1 / \log c_2$ *not be rational. If for every* $X \in D_2$

$$\pi(c_1 X) = c_1 \pi(X) \quad \textit{and} \quad \pi(c_2 X) = c_2 \pi(X)$$

hold, then for suitable α, β, γ, $r \in \mathbb{R}$ (*with* $\beta\gamma > 0$, $r > 0$)

$$u(x) = \begin{cases} \alpha + \beta x^r, & x \geq 0 \\ \alpha - \gamma(-x)^r, & x < 0. \end{cases} \tag{5}$$

Conversely, if u *has the form* (5), *then for all* X *and all* $c \in \mathbb{R}^+$

$$\pi(cX) = c\pi(X)$$

holds, i.e., π *is positively homogeneous.*

REMARK. Theorem 2.2 is best possible in the following sense: For any pair c_1, $c_2 \in \mathbb{R}^+$, which does not fulfil the conditions of theorem 2.2, there is an admissible utility function u, not of the form (5), such that

$$\pi(c_i X) = c_i \pi(X), \qquad i = 1, 2, X \in D_2$$

holds. In this case the Swiss premium calculation principle certainly is not positively homogeneous.

2C. π = *Esscher principle*

This principle was introduced in BUHLMANN (1980) and so named in view of the formal similarity to the Esscher transform. Given $\alpha \geq 0$ the premium $P = \pi(X)$ is determined explicitly by the equation

$$\pi(X) = \frac{E[X \exp(\alpha X)]}{E[\exp(\alpha X)]}.$$

It is very easy to see (cf. GOOVAERTS, DE VYLDER and HAEZENDONCK (1984)), that the Esscher principle is positively homogeneous only in the case $\alpha = 0$, i.e., Esscher premium = net premium. A simple proof will give the following sharp result:

THEOREM 2.3. *If for a fixed* $c_0 \neq 1$, *a single (non-degenerated)* $X_0 \in D_2$ *the equation*

$$\pi(c_0 X_0) = c_0 \pi(X_0)$$

holds, then $\alpha = 0$.

2D. *General premium principles*

The result mentioned in the introduction, namely that the property of positive homogeneity is already fulfilled in the global sense if it is known only locally with respect to the variable c, now follows easily: If π denotes the zero utility principle, the Swiss premium calculation principle or the Esscher principle, then the theorems above yield at once

COROLLARY 2.4. *If for all $X \in D_2$ and all c in a given (arbitrarily small) bounded interval in $\mathbb{R}^+$ one has*

$$\pi(cX) = c\pi(X),$$

then $\pi(cX) = c\pi(X)$ holds for all X and all $c \in \mathbb{R}^+$.

Finally one should pay attention to a very general result, which on the one hand makes the results above more transparent, on the other hand is true for general, possibly still unknown principles: Denote by π any premium principle with the very weak and plausible continuity condition, that for every convergent sequence $(\gamma_k) \subset \mathbb{R}^+$ and every $X \in D_2$

$$\lim_{k\to\infty} \pi(\gamma_k X) = \pi(\lim_{k\to\infty} \gamma_k \cdot X).$$

For such principles one has throughout

THEOREM 2.5. *For fixed, positive c_1, $c_2 \neq 1$ let $\log c_1/\log c_2$ not be rational. If*

$$\pi(c_1 X) = c_1 \pi(X) \quad \text{and} \quad \pi(c_2 X) = c_2 \pi(X), \qquad X \in D_2,$$

then $\pi(cX) = c\pi(X)$ holds for all $X \in D_2$ and even all $c > 0$.

As a corollary (because in any interval I, however small it may be, there are of course always two numbers c_1, $c_2 \neq 1$ in I such that $\log c_1/\log c_2 \notin \mathbb{Q}$) we note: The (global) property of such premium calculation principles of being positively homogeneous is always a local property in the following sense:

If $\pi(cX) = c\pi(X)$ holds for only all $c \in I$, then automatically also for even all $c \in \mathbb{R}^+$.

REMARK. Simple and explicit examples for pairs of numbers c_1, $c_2 \neq 1$, which satisfy $\log c_1/\log c_2 \notin \mathbb{Q}$, are the following:

(i)	$c_1 = 2$,	$c_2 = 3$,
(ii)	$c_1 = \frac{1}{2}$,	$c_2 = \frac{1}{3}$,
(iii)	$c_1 = 1.1$,	$c_2 = 1.2$,
(iv)	$c_1 = 2$,	$c_2 = \pi$,
(v)	$c_1 = 2$,	$c_2 = e^{\pi}$,
(vi)	$c_1 = e$,	$c_2 = e^{\pi}$,
(vii)	$c_1 = 1.25$,	$c_2 = 1.1$.

As was pointed out by GERBER (1979), pp. 73–74, the global property of positive homogeneity is not reasonable for practical reasons. From example (vii) e.g., one can deduce by the corollary of theorem 2.5 more precisely that it is in the same way again unreasonable to accept the homogeneity property as a local property with respect to c only: If for all X

$$\pi(cX) = c\pi(X)$$

holds for, say all c between 1 and 1.25 (local property), then automatically by the results above also for all $c > 0$ (global property). In accordance with GERBER (1979) the quotient $\pi(cX)/\pi(X)$ should depend not only on c but also on X.

3. PROOFS

Ad 2A: First of all we consider the zero utility principle π with strictly monotonic utility function u, such that $u'(x) \geq 0$, $u''(x) \leq 0$. To prove theorem 2.1 we assume $u(0) = 0$ without loss of generality, because for a given risk X the premium $P = \pi(X)$ does not change if in (3) u is substituted by $u - u(0)$.

LEMMA 3.1. *If for a fixed* $c_0 > 0$

(6) $$\pi(c_0 X) = c_0 \pi(X), \qquad X \in D_2,$$

then there exists a $\beta_0 = \beta_0(c_0)$ *such that*

$$u(c_0 x) = \beta_0 u(x)$$

holds for all $x \in \mathbb{R}$.

PROOF. For the present let be $x > 0$, and for $a, b \in \mathbb{R}$, $q \in [0, 1]$ let $X = X_{a,b,q} \in D_2$ be defined by

$$\text{pr}\,(X = a) = 1 - q, \qquad \text{pr}\,(X = b) = q.$$

With the abbreviation $P = P(a, b, q) = \pi(X)$ one has by (3)

(7) $$(1-q)u(P-a) + qu(P-b) = u(0),$$

and by (6)

(8) $$(1-q)u(c_0(P-a)) + qu(c_0(P-b)) = u(0).$$

(7) and (8) yield

(9) $$\frac{1-q}{q} = \frac{u(P-b)-u(0)}{u(0)-u(P-a)},$$

(10) $$\frac{1-q}{q} = \frac{u(c_0(P-b))-u(0)}{u(0)-u(c_0(P-a))}.$$

Putting $x' = -1$, $a = 2$, $\gamma = 1$, $b = 1 - x$ one gets

(11) $$a = \gamma - x', \qquad b = \gamma - x \quad \text{and} \quad b < \gamma < a.$$

If

$$q = \frac{u(0) - u(x')}{u(x) - u(x')}, \tag{12}$$

then $0 < q < 1$ in view of the strict monotonicity of u. Because of $x = \gamma - b$, $x' = \gamma - a$ and by (12) one concludes

$$(1-q)u(\gamma - a) + qu(\gamma - b) = u(0),$$

therefore $\gamma = P(a, b, q)$ according to (7). From this, in view of (10), (11) it follows

$$\frac{1-q}{q} = \frac{u(c_0 x) - u(0)}{u(0) - u(c_0 x')} \tag{13}$$

for arbitrary $x > 0$. Together with (9) this leads to

$$\frac{u(c_0 x) - u(0)}{u(0) - u(c_0 x')} = \frac{u(x) - u(0)}{u(0) - u(x')}$$

Therefore

$$u(c_0 x) - u(0) = \frac{u(0) - u(c_0 x')}{u(0) - u(x')} [u(x) - u(0)],$$

$$u(c_0 x) = \frac{u(0) - u(c_0 x')}{u(0) - u(x')} u(x) + u(0)\left[1 - \frac{u(0) - u(c_0 x')}{u(0) - u(x')}\right]$$

$$= \frac{u(c_0 x')}{u(x')} u(x)$$

for all $x > 0$, respecting the normalization $u(0) = 0$. With $\beta_0 = u(c_0 x')/u(x')$ this is the assertion for $x > 0$.

In the case $x \leq 0$ one proves in an analogous way the existence of a real number γ_0 such that

$$u(c_0 x) = \gamma_0 u(x) \tag{14}$$

holds for all $x \leq 0$. Now certainly $\beta_0 = \gamma_0$ (this is exactly the statement of lemma 3.1), because with regard to (9) and $u(0) = 0$ one has

$$\frac{1-q}{q} = -\frac{u(P-b)}{u(P-a)} \tag{15}$$

Correspondingly by (10)

$$\frac{1-q}{q} = -\frac{u(c_0(P-b))}{u(c_0(P-a))} = -\frac{\beta_0}{\gamma_0} \cdot \frac{u(P-b)}{u(P-a)} \tag{16}$$

if $b < P < a$, which is true in view of (11) and $P = \gamma$. Comparing (15) and (16) one has $\beta_0 = \gamma_0$.

PROOF OF THEOREM 2.1. Under the assumptions of theorem 2.1 there are by lemma 3.1 real numbers β_1, β_2 such that

$$(17) \qquad u(c_1x)=\beta_1 u(x) \quad \text{and} \quad u(c_2x)=\beta_2 u(x), \qquad x\in\mathbb{R}.$$

Successive application of these relations gives for $x=1$

$$u(c_1^n c_2^m)=\beta_1^n\beta_2^m u(1), \qquad n, m\in\mathbb{Z}.$$

Together with (17) it follows for $x\in\mathbb{R}$, n, $m\in\mathbb{Z}$

$$(18) \qquad u(c_1^n c_2^m x)=u(c_1^n c_2^m)u(x)/u(1).$$

By assumption one has $\log c_1/\log c_2\notin\mathbb{Q}$, therefore according to Kronecker's approximation theorem (cf. REICH (1984), Appendix) the set

$$\{k\log c_1+l\log c_2 \;\; k, l\in\mathbb{Z}\}$$

is dense in $\mathbb{R}$. From this it follows at once that for every given number $y>0$ there are two sequences $k(n)$, $l(n)\in\mathbb{Z}$ such that

$$(19) \qquad y=\lim_{n\to\infty} c_1^{k(n)}c_2^{l(n)}.$$

By (18) and the continuity of u one concludes

$$(20) \qquad u(yx)=u(y)u(x)/u(1).$$

for arbitrary $y>0$, $x\in\mathbb{R}$. The only continuous solution of this functional equation are

$$(21) \qquad u(x)=u(1)x^r, \qquad x>0,$$

with some $r\in\mathbb{R}$, as is well known. Because u is strictly increasing, (21) holds for all $x>0$ with suitable $r>0$. Moreover, if $x<0$ then it follows from (20)

$$u(x)=u(-1)(-x)^r,$$

so indeed there are numbers $\beta>0$, $\gamma<0$ such that

$$u(x)=\begin{cases}\beta x^r, & x\geq 0\\ \gamma(-x)^r, & x<0\end{cases}.$$

In the case $r=1$ one has certainly $\beta=-\gamma$ for continuous u', therefore u is linear. The case $r\geq 2$ is impossible in view of $u''(x)\leq 0$, the case $0<r<2$, $r\neq 1$, is impossible according to the existence of $u''(0)$. Because of the assumed normalization of u theorem 2.1 is proved.

The remark after theorem 2.1 can be easily proved.

Ad 2B:

PROOF OF THEOREM 2.2 (Swiss premium calculation principle). Let be $z<1$ and without any restriction of generality let u be strictly increasing. Assume as in theorem 2.2 that two numbers c_1, $c_2 \neq 1$ are given and that $\pi(c_i X) = c_i \pi(X)$ holds for $i = 1, 2$. According to (4) one has more precisely

(22) $$E[u(c_i X - z c_i P)] = u((1-z) c_i P), \qquad X \in D_2, \qquad i = 1, 2.$$

Defining $g_i(x) = u(c_i x)$ equation (22) gives

$$E[g_i(X - zP)] = g_i((1-z)P).$$

By GOOVAERTS, DE VYLDER and HAEZENDONCK (1984), theorem 2, p. 72, there exist real numbers α_i, β_i, $i = 1, 2$, such that

(23) $$u(c_i x) = \alpha_i + \beta_i u(x), \qquad x \in \mathbb{R}.$$

Without any restriction one can assume u to be normalized, especially $u(0) = 0$. Then, of course, $\alpha_i = 0$ and

(24) $$u(c_i x) = \beta_i u(x), \qquad x \in \mathbb{R}.$$

From this it follows immediately for arbitrary n, $m \in \mathbb{Z}$, $x \in \mathbb{R}$

(25) $$u(c_1^n c_2^m x) = \beta_1^n \beta_2^m u(x) = u(c_1^n c_2^m) u(x) / u(1).$$

The condition $\log c_1 / \log c_2 \notin \mathbb{Q}$ leads via Kronecker's approximation theorem and the continuity of u to

(26) $$u(yx) = u(y) u(x) / u(1)$$

for all $y > 0$, $x \in \mathbb{R}$.

In the case $x \geq 0$ one introduces $u_1(x) = u(x)/u(1)$ and gets by (26)

$$u_1(yx) = u_1(y) u_1(x), \qquad x, y > 0.$$

As is well known for continuous u_1 it follows that u_1 is monomial, therefore u too.

$$u(x) = u(1) x^r, \qquad x \geq 0,$$

with suitable $r > 0$.

In the case $x < 0$ one defines $z = -x$ and $u_2(z) = -u(z)$. By (26)

$$u_2(yz) = u_2(y) u_2(z) / u_2(1),$$

therefore it follows in a similar way that u_2 is a monomial. This means

$$u(x) = -u(-1)(-x)^s, \qquad x < 0,$$

for suitable $s > 0$. In view of (26) $r = s$ holds, so the first part of theorem 2.2 is proved The second part is trivial.

The remark after theorem 2.2 is easily proved and the proof is omitted.

Ad 2C:

PROOF OF THEOREM 2.3 (Esscher principle). Let $X_0 \in D_2$ not be degenerated, say

$$\text{pr}(X_0 = a) = 1 - q, \qquad \text{pr}(X_0 = b) = q$$

for some a, $b \in \mathbb{R}$, $a \neq b$, $q \in (0, 1)$. If for a fixed $c_0 \neq 1$

$$\pi(c_0 X_0) = c_0 \pi(X_0)$$

is true, then

$$\frac{c_0 a(1-q)\exp(\alpha c_0 a) + c_0 b q \exp(\alpha c_0 b)}{(1-q)\exp(\alpha c_0 a) + q\exp(\alpha c_0 b)} = c_0 \frac{a(1-q)\exp(\alpha a) + bq\exp(\alpha b)}{(1-q)\exp(\alpha a) + q\exp(\alpha b)}.$$

Multiplication yields

$$(a-b)\exp[\alpha(c_0 a + b)] = (a-b)\exp[\alpha(c_0 b + a)],$$

therefore in view of $a \neq b$

$$\exp[\alpha(c_0 a + b)] = \exp[\alpha(c_0 b + a)].$$

Assume $\alpha > 0$, then $c_0 a + b = c_0 b + a$, therefore $a = b$ according to $c_0 \neq 1$. This is a contradiction, so necessarily $\alpha = 0$.

Ad 2D:

If $I = (a, b)$, $0 \leq a < b \leq \infty$, is any interval in $\mathbb{R}^+$, then by trivial arguments there are two numbers c_1, $c_2 \in I$ such that $\log c_1 / \log c_2 \notin \mathbb{Q}$. From this it is clear that Corollary 2.4 follows from the preceding Theorems.

Proof of Theorem 2.5. By induction one immediately proves for n, $m \in \mathbb{N}$

$$\pi(c_1^n X) = c_1^n \pi(X), \qquad \pi(c_2^m X) = c_2^m \pi(X). \tag{27}$$

Because of

$$\pi(X) = \pi\left(c_i \cdot \frac{1}{c_i} X\right) = c_i \pi\left(\frac{1}{c_i} X\right), \qquad i = 1, 2,$$

equation (27) even holds for n, $m \in \mathbb{Z}$. Then one has

$$\pi(c_1^n c_2^m X) = c_1^n \pi(c_2^m X) = c_1^n c_2^m \pi(X). \tag{28}$$

Given an arbitrary $c > 0$ there are according to Kronecker's approximation theorem (cf. Reich (1984), Appendix) two sequences $n(k)$, $m(k) \in \mathbb{Z}$ such that

$$\gamma_k = c_1^{n(k)} c_2^{m(k)} \to c.$$

The continuity condition for π and (28) yields

$$\pi(cX) = \pi(\lim_{k\to\infty} \gamma_k \cdot X) = \lim_{k\to\infty} \pi(\gamma_k X) = \lim_{k\to\infty} \gamma_k \cdot \pi(X) = c\pi(X),$$

therefore the assertion.

REFERENCES

BUHLMANN, H (1970) *Mathematical Methods in Risk Theory* Springer Berlin–Heidelberg–New York

BUHLMANN, H, B GAGLIARDI, H U GERBER and E. STRAUB (1977) Some Inequalities for Stop-Loss Premiums *ASTIN Bulletin* **9**, 75–83

BUHLMANN, H (1980) An Economic Premium Principle *ASTIN Bulletin* **11**, 52-60

GERBER, H U (1979) *An Introduction to Mathematical Risk Theory.* Huebner Foundation Philadelphia

GOOVAERTS, M J, F DE VYLDER and J HAEZENDONCK (1984) *Insurance Premiums* North-Holland: Amsterdam

HAEZENDONCK, J and GOOVAERTS, M J (1982) A new Premium Calculation Principle based on Orlicz Norms. *Insurance Mathematics and Economics* **1**, 41-53

REICH, A (1984). Premium Principles and Translation Invariance. *Insurance· Mathematics and Economics* **3**, 57-66

APPROXIMATING THE DISTRIBUTION OF A DYNAMIC RISK PORTFOLIO

BY WILLIAM S. JEWELL

Department of Industrial Engineering and Operations Research
University of California, Berkeley

ABSTRACT

In a previous paper, Jewell and Sundt showed how to approximate a distribution of total losses from a large, fixed, heterogeneous portfolio, using a recursive algorithm developed by Panjer for the distribution of a random sum of random variables (a single casualty contract). This paper extends the approximation procedure to large, dynamic heterogeneous portfolios, in order to model either a portfolio of correlated casualty contracts, or a future portfolio, whose composition is not known with certainty.

0. INTRODUCTION

The problem of finding the distribution of $\tilde{y} = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_2 + \cdots + \tilde{x}_N$, where the $(\tilde{x}_i)$ are a *fixed* and large set of independent, *nonidentically* distributed, integer-valued random variables was considered in JEWELL and SUNDT (1981) (hereinafter referred to as JS). Although, in theory, the discrete density of $\tilde{y}$ is just the N-fold convolution of the individual densities, this computation is very time-consuming, and various forms of approximation must be used; moreover, in many risk applications, the use of a normal approximation gives very bad results, even for large N, because of the skewness and long tails of the density. However, if the probability $p_i = \Pr\{\tilde{x}_i = 0\}$ is significant for most $i = 1, 2, \ldots, N$, it turns out that a very good approximation can be obtained using newly-developed procedures for the related problem of calculating the distribution of the sum of a *random number* of independent and *identically* distributed random variables.

In many risk applications, especially in insurance and investment management, there are an ever-changing number of risks of different types, and it is of interest to predict the distribution of a portfolio whose future composition is not known with certainty. This paper develops a general model for this situation, and shows how the approximation procedure described in JS can be extended.

1. THE DYNAMIC PORTFOLIO MODEL

Let $i = 1, 2, \ldots, N$ index a number of different risk classes (insurance policies or types of investment) in a given portfolio, and let $\tilde{n}_i \in [0, 1, \ldots]$ be the random number of independent risks of type i, giving a grand total number of risks in the portfolio

$$\tilde{n}_T = \sum_{i=1}^{N} \tilde{n}_i. \tag{1.1}$$

Risks of type i are *similar*, in the sense that, if $\tilde{x}_{ij}$ is the random monetary gamble from the jth risk of type i, then its discrete density, $f_i^0(x)$, is the same for all j, i.e.:

$$\Pr\{\tilde{x}_{ij}=x\}=f_i^0(x), \qquad (i=1,2,\ldots,N)(j=1,2,\ldots,n_i). \tag{1.2}$$

We shall only consider discrete gambles, with the common range of the $(\tilde{x}_{ij})$ as $[0,1,2,\ldots,R]$. As mentioned above, we assume for the moment that the $(\tilde{x}_{ij})$ are statistically independent of each other and the $(\tilde{n}_i)$, but we do *not* assume that the $(\tilde{n}_i)$ are independent. (But see Appendix A.)

The total monetary gamble for all risks of type i is then the sum of a random number of random variables:

$$\tilde{x}_i=\begin{cases}0, & (\tilde{n}_i=0)\\ \tilde{x}_{i1}+\tilde{x}_{i2}+\cdots+\tilde{x}_{i\tilde{n}_i}, & (\tilde{n}_i>0),\end{cases} \qquad (i=1,2,\ldots,N) \tag{1.3}$$

and the grand total monetary risk is then the fixed-term random sum:

$$\tilde{y}=\tilde{x}_1+\tilde{x}_2+\cdots+\tilde{x}_N. \tag{1.4}$$

Note that the $(\tilde{x}_i)$ are now *dependent* random variables, if the $(\tilde{n}_i)$ are.

If $g(y)$ and $\pi(n_1,n_2,\ldots,n_N)$ are the discrete densities of the total risk sum and the number of risks of each type, respectively, we have then the discrete density of y as:

$$\begin{aligned}\Pr\{\tilde{y}=y\}=g(y)=&\sum_{n_1}\sum_{n_2}\cdots\sum_{n_N}\pi(n_1,n_2,\ldots,n_N)[f_1^0(y)]^{n_1^*}\\ &*[f_2^0(y)]^{n_2^*}*\cdots*[f_N^0(y)]^{n_N^*}, \qquad (y=0,1,2,\ldots)\end{aligned} \tag{1.5}$$

which, of course, is a lengthy and laborious computation (In JS, the special case of $(\tilde{n}_i)$ deterministic was considered.)

2 INTERPRETATIONS

Before describing a method of approximating (1.5), we give some possible practical interpretations of the model

In insurance applications, the simplest interpretation is that i refers to different, distinguishable types of insurance policies in a given portfolio; for instance, similar policies in personal lines of insurance could refer to ordinary life insurance policies with the same face values issued to persons of the same age. For the current year, we know exactly the number of policies of type i and hence, following JS, can find an approximation to the current $g(y)$. However, an approximation to (1.5) would be necessary to predict total portfolio risk for *next* year, after some policies are withdrawn, some policies have paid out benefits, or new policies have been added, and still others have shifted type. By specifying the stochastic law governing this "drop-add" mechanism, we can get $\pi(n_1,n_2,\ldots,n_N)$ for next year. Possible reasons for leaving correlation between the $(\tilde{n}_i)$ are that we may have a precise idea of how new sales are distributed among the different types of policies, but may be uncertain about the total new business; or, the new business total may be accurately estimated but the distribution may be uncertain; or, there may be an uncertain number of policies which are shifting type (as in aging of life insurance insured), etc.

A second insurance interpretation is the so-called casualty claim model, in which multiple claims may occur on a policy during a given exposure year. Here i indexes each of a fixed number of policies, $f_i^0(x)$ is the individual claim ("severity") density, $\tilde{n}_i$ is the random number of claims ("frequency"), and $\tilde{x}_i$ is now the total monetary claim on the single policy i. Of course, if the $(\tilde{n}_i)$ were independent, then this application could be handled by making $\sum \pi(n_i)[f_i^0(y)]^{n_i^*}$ the basic density used in the procedure described in JS; but this would require prior calculation of this compound law (see also Sections 8 and 9, below). Moreover, external factors, such as weather and economics, often affect the number of claims of all types of contracts in a given portfolio in the same way, thus introducing correlation and the need for a more general model.

In most insurance portfolios, a great deal of effort is used to assure that the $(\tilde{x}_{ij})$ are statistically independent of each other. However, there remains always the possibility that risks of the same type i are influenced by the same exogenous factors. In Appendix A, we consider the case when risks of the same type are *exchangeable random variables*, which leads to a weak form of dependence on the $(\tilde{x}_{ij})$.

In investment portfolios, it is unusual to have independent risks of the same type, i.e., requiring the same investment level, and having the same outcome distribution; instead, we usually have a different amount of money invested in different risks. If we let $\tilde{n}_i$ be the level of investment in type i and $\tilde{x}_i$ the net return from this investment, then (1.3) holds only if the $(\tilde{x}_{ij})$ are perfectly correlated, or what is the same, if (1.3) is replaced by $\tilde{x}_i = \tilde{n}_i \tilde{x}_{i1}$. Another limitation on investment modelling is that it is usually possible to have negative net returns, which is discussed in Section 10. It should be remembered also that our approximation is usually successful only if the problem is modelled so that the probability of zero net return is substantial, i e., all "sure thing" return has been eliminated.

Technological risk applications are based upon the compound law interpretation; for instance, in reliability engineering, $\tilde{n}_i$ may refer to the random number of mechanical, electrical, or thermal shocks of type i which affect a given piece of equipment; in fire damage analysis, $\tilde{n}_i$ is the random number of fires of a given type (size, type of dwelling or land classification) which occur; and so forth. In technology applications, the primary modelling challenge is to express damage in appropriate, additive units for situations where there is no accepted monetary surrogate for the risk

3. NOTATION AND MOMENTS

The success of the approximation procedure to be described depends upon the assumption that most of the total risks, $(\tilde{x}_i)$, have a high probability of being zero; this can occur either because $f_i^0(0)$ is large, or because $\tilde{n}_i$ is often zero. We now change to a traditional notation (see JS) which emphasizes the distribution of risk when it is positive. Let

$$(3.1)\quad \Pr\{\tilde{x}_{ij}=0\}=f_i^0(0)=p_i=1-q_i, \qquad (i=1,2,\ldots,N)(j=1,2,\ldots,\tilde{n}_i)$$

$$(3.2)\quad f_i(x)=\Pr\{\tilde{x}_{ij}=x|\tilde{x}_{ij}>0\}=f_i^0(x)/q_i, \qquad (x=1,2,\ldots,R)$$

and define the first two moments of non-zero risk as:

$$(3.3)\quad m_i=\mathscr{E}\{\tilde{x}_{ij}|\tilde{x}_{ij}>0\}=\mathscr{E}\{\tilde{x}_{ij}\}/q_i,$$

$$(3.4)\quad v_i=\mathscr{V}\{\tilde{x}_{ij}|\tilde{x}_{ij}>0\}=[\mathscr{V}\{\tilde{x}_{ij}\}/q_i]-p_i(m_i)^2.$$

From the joint counting density, we get the marginal densities.

$$(3.5)\quad \pi_i(n)=\Pr\{\tilde{n}_i=n\}, \qquad (i=1,2,\ldots,N)(j=1,2,\ldots,\tilde{n}_i)$$

and the first two moments:

$$\begin{aligned}(3.6)\quad \lambda_i&=\mathscr{E}\{\tilde{n}_i\},\\ (3.7)\quad \gamma_{ik}&=\mathscr{C}\{\tilde{n}_i;\tilde{n}_k\},\end{aligned} \qquad (i,k=1,2,\ldots,N).$$

The approximation itself is based upon moment-matching with the first two moments of the exact density (1.5), which we now find in a straightforward manner. First, from (1.3) and the assumptions:

$$\mathscr{E}\{\tilde{x}_i|n_i\}=n_i\mathscr{E}\{\tilde{x}_{ij}\},$$

$$\mathscr{C}\{\tilde{x}_i;\tilde{x}_k|n_i,n_k\}=\begin{cases}n_i\mathscr{V}\{\tilde{x}_{ij}\}, & (i=k)\\ 0, & (i\neq k)\end{cases}$$

so that, unconditioning, we have:

$$(3.8)\quad \mathscr{E}\{\tilde{x}_i\}=\mathscr{E}\{\tilde{n}_i\}\mathscr{E}\{\tilde{x}_{ij}\},$$

$$(3.9)\quad \mathscr{C}\{\tilde{x}_i;\tilde{x}_k\}=\begin{cases}\mathscr{E}\{\tilde{n}_i\}\mathscr{V}\{\tilde{x}_{ij}\}+\mathscr{V}\{\tilde{n}_i\}[\mathscr{E}\{\tilde{x}_{ij}\}]^2, & (i=k)\\ 0+\mathscr{C}\{\tilde{n}_i;\tilde{n}_k\}\mathscr{E}\{\tilde{x}_{ij}\}\mathscr{E}\{\tilde{x}_{kl}\}, & (i\neq k).\end{cases}$$

Then, using (1 4) and notation defined above, we find the first two moments of total portfolio risk as:

$$(3.10)\quad \mathscr{E}\{\tilde{y}\}=\sum_{i=1}^{N}\lambda_i q_i m_i,$$

and

$$(3.11)\quad \mathscr{V}\{\tilde{y}\}=\sum_{i=1}^{N}\lambda_i q_i(v_i+p_i m_i^2)+\sum_{i=1}^{N}\sum_{k=1}^{N}\gamma_{ik}q_i q_k m_i m_k.$$

The (q_i), (m_i), and (v_i) are presumed known from past portfolio statistics on each type i, and the (λ_i) and (γ_{ik}) are gotten from modelling assumptions regarding the future composition of the portfolio; so, we shall assume that these moments are *given parameters*.

Of course, if the $(\tilde{n}_i)$ are statistically independent, the last term become $\sum \gamma_{ii} q_i^2 m_i^2$. In the static portfolio model in JS, the composition was fixed, with all $\tilde{n}_i = 1$; an equivalent, but slightly generalized model can be gotten from the above, with $\tilde{n}_i = n_i = \lambda_i$, and all $\gamma_{ik} = 0$.

4. THE APPROXIMATING RISK COLLECTIVE MODEL

In the approximation, we replace the original portfolio by a *homogeneous* "risk collective", that is, we assume that $\tilde{y}$ is approximately:

$$\tilde{y} = \begin{cases} 0, & (\tilde{n}_e = 0) \\ \tilde{w}_1 + \tilde{w}_2 + \cdots + \tilde{w}_{\tilde{n}_e}, & (\tilde{n}_e > 0) \end{cases} \tag{4.1}$$

where $\tilde{n}_e$ is the random number of *equivalent positive claims* $(\tilde{w}_i)$, assumed to be independent of each other and $\tilde{n}_e$, and *identically distributed*, according to *prototypical counting* and *individual risk* densities:

$$\begin{aligned} \pi(n) &= \Pr\{\tilde{n}_e = n\}, \qquad (n = 0, 1, 2, \ldots); \\ f(w) &= \Pr\{\tilde{w} = w\}, \qquad (w = 1, 2, \ldots), \end{aligned} \tag{4.2}$$

leading to the usual *compound law* of risk theory for the density of $\tilde{y}$:

$$g(y) = \sum_{n=0}^{\infty} \pi(n)[f(y)]^{n*}. \tag{4.3}$$

As mentioned earlier, the rationale behind this approximation is that, in many applications, the $(\tilde{x}_{ij})$ are zero with high probability; the $(\tilde{w}_i)$ then represent just the positive $(\tilde{x}_{ij})$. (See also JS and GERBER (1979).)

If the prototypical moments are:

$$\lambda = \mathscr{E}\{\tilde{n}_e\}; \qquad \gamma = \mathscr{V}\{\tilde{n}_e\}, \tag{4.4}$$

$$m = \mathscr{E}\{\tilde{w}\}; \qquad v = \mathscr{V}\{\tilde{w}\}, \tag{4.5}$$

then the moments of the random sum in the approximating model will be:

$$\mathscr{E}\{y\} = \lambda m, \tag{4.6}$$

$$\mathscr{V}\{\tilde{y}\} = \lambda v + \gamma m^2. \tag{4.7}$$

For a good approximation, the moments (4.6), (4.7) must be matched as closely as possible with the true values (3.10), (3.11). In addition, the forms of the $\pi(n)$ and $f(w)$ chosen may also be varied.

5. THE ADELSON–PANJER RECURSIVE ALGORITHMS

At this point, we should stop and consider whether the computation of the compound law (4.3) can be effected in any efficient manner; otherwise, it is not much improvement over (1.5). A traditional approximation (for the static portfolio problem) used in actuarial circles was to make $\pi(n)$ a Poisson law; this was because (further) approximations to the compound law had been developed in the early risk theory literature (see, e.g., GERBER (1979)).

However, the recent extension by PANJER (1981) of a recursive scheme of ADELSON (1966) now provides an efficient and direct way to compute (4.3). Essentially, if $f(w)$ is discrete over $[1, 2, \ldots]$ and the counting distribution is chosen from a certain (a, b)-*family* for which:

$$\pi(n) = \left(a + \frac{b}{n}\right)\pi(n-1), \qquad (n = 1, 2, \ldots) \tag{5.1}$$

then $g(y)$ can be calculated recursively via.

$$\begin{aligned} g(0) = \pi(0) &= \begin{cases} (1-a)^{(a+b)/a}, & (a \neq 0) \\ e^{-b}, & (a = 0) \end{cases} \\ g(y) &= \sum_{x=1}^{\min(y,\,R)} \left(a + b\frac{x}{y}\right) f(x) g(y-x), \qquad (y = 1, 2, \ldots). \end{aligned} \tag{5.2}$$

This is clearly an efficient computational procedure, provided the (a, b)-family is a useful one. As elaborated upon in SUNDT and JEWELL (1981), the only members of this family, apart from the degenerate density, are:

(5.3a) (Poisson) $$\pi(n) = \frac{\lambda^n e^{-\lambda}}{n!}, \qquad (a = 0;\, b = \lambda);$$

(5.3b) (Binomial) $$\pi(n) = \binom{M}{n} p^n (1-p)^{M-n}, \qquad (a = -p/(1-p);\; b = -a(M+1));$$

(5.3c) (Negative Binomial) $$\pi(n) = \binom{\alpha + n - 1}{n} p^n (1-p)^{\alpha}, \qquad (a = p;\, b = p(\alpha - 1)).$$

These counting distributions are useful, since they are often used in modelling compound risk laws. Furthermore, since:

$$\lambda = \mathscr{E}\{\tilde{n}_e\} = \frac{a+b}{1-a}; \qquad \gamma = \mathscr{V}\{\tilde{n}_e\} = \frac{a+b}{(1-a)^2}; \tag{5.4}$$

we get:

$$a = 1 - \frac{\lambda}{\gamma}; \qquad b = \frac{\lambda(\lambda+1)}{\gamma} - 1; \tag{5.5}$$

and

$$\frac{\mathscr{V}\{\tilde{n}_e\}}{\mathscr{E}\{\tilde{n}_e\}} = \frac{\gamma}{\lambda} = \frac{1}{1-a}. \tag{5.6}$$

The importance of the ratio (5.6) in modelling empirical counting processes is well known. From (5.3), we see that this family covers a wide range of such ratios, with the Binomial giving $(\gamma/\lambda)<1$ and the Negative Binomial (Pascal) giving $(\gamma/\lambda)>1$; the Poisson $(\gamma=\lambda)$ distribution is the dividing line.

Therefore, for computational simplicity, we propose to use the (a, b)-family to model the counting distribution $\pi(n)$ and the recursive procedure (5.2) to compute the approximate density (4.3). Note that, if $a<0$ in (5.5), we are not completely free in our choice of b, since M must be an *integer* in the Binomial law (5.3b); however, this is not usually a serious limitation (see JS).

6. THE FINAL APPROXIMATION

Having selected $\pi(n)$ on the basis of computational convenience, we must now choose the prototypical density, $f(w)$. The form which will give the best approximation in all cases is not known. However, a natural way, consistent with the interpretation given in Section 4, is to weight the individual densities (3.2) with weights proportional to the expected number of risks with positive outcome in the corresponding class, i.e., to fix.

$$f(w)=\frac{\sum \lambda_i q_i f_i(w)}{\sum \lambda_j q_j}=\frac{\sum \lambda_i f_i^0(w)}{\sum \lambda_j q_j}, \qquad (w=1,2,\ldots,R). \tag{6.1}$$

This choice is consistent with JS for the static risk portfolio model, and also provides the greatest simplification to the formulae below. Using (3.3), (3.4) in (6.1), we find first m and v in (4.5), then substitute into (4.6), (4.7) to find the first two moments of the approximating model; these moments are then equated with the exact results (3.10), (3.11), obtaining finally the *first two moments of the prototypical counting density in terms of the original parameters*:

$$\lambda=\mathscr{E}\{\tilde{n}_e\}=\sum_{i=1}^{N}\lambda_i q_i; \tag{6.2}$$

$$\gamma=\mathscr{V}\{\tilde{n}_e\}=\sum_{i=1}^{N}\lambda_i q_i\left[1-q_i\left(\frac{m_i}{m}\right)^2\right]+\sum_{i=1}^{N}\sum_{k=1}^{N}\gamma_{ik}q_i q_k\left(\frac{m_i m_k}{m}\right)^2; \tag{6.3}$$

where the mean prototypical severity is:

$$m=\mathscr{E}\{\tilde{w}\}=\frac{\sum \lambda_i q_i m_i}{\sum \lambda_j q_j} \tag{6.4}$$

and the severity variance is:

$$v=\frac{\sum \lambda_i q_i(v_i+m_i^2)}{\sum \lambda_j q_j}-m^2. \tag{6.5}$$

To summarize: In the final approximation, we would first calculate the $f_i(x)$ and the moments of Section 3 using the data, then compute $f(w)$ from (6.1) and use it in the approximating model (4.3), together with one of the $\pi(n)$ of Section 5, with (a, b) selected using (5.5), the approximate density is computed recursively via (5.2).

In the static portfolio case considered in JS, all γ_{ik} are identically zero, so that $\gamma < \lambda$, and a Binomial counting law results. This raises the integrality problem for M previously mentioned, and means that the resulting values of (a, b) do not exactly match $\mathcal{V}\{\tilde{y}\}$ in the original and approximating models; however, the resulting error is not serious in the example analyzed in that paper.

In contrast, the dynamic portfolio model of this study can give $\gamma/\lambda > 1$, and hence Negative Binomial $\pi(n)$, if the (γ_{ik}) are large enough. To see this, consider the case of *independent*, but still random, $(\tilde{n}_i)$. (6.3) then becomes:

$$\gamma = \mathcal{V}\{\tilde{n}_e\} = \sum_{i=1}^{N} \lambda_i q_i + \sum_{i=1}^{N} \left(\frac{q_i m_i}{m}\right)^2 (\gamma_{ii} - \lambda_i). \tag{6.6}$$

Thus, we see that, if a sufficient number of (marginal) counting densities (3.5) have $\gamma_{ii}/\lambda_i > 1$, then also $\gamma/\lambda > 1$, a most reasonable result.

7. THE COMPOUND MULTINOMIAL COUNTING DISTRIBUTION

One natural way in which the number of risks in the different classes, $\tilde{n} = (\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_N)$, might be generated in a predictive, dynamic model is from a Multinomial law, with given total number of risks, n_T, and a set of *selection probabilities*, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$, viz:

$$\Pr\{\tilde{\boldsymbol{n}} = \boldsymbol{n} | n_T; \boldsymbol{\pi}\} = \binom{n_T}{n_1, n_2, \ldots, n_N} \prod_{i=1}^{N} \pi_i^{n_i}, \qquad (\textstyle\sum n_i = n_T)(\sum \pi_i = 1). \tag{7.1}$$

With fixed n_T and $\boldsymbol{\pi}$, there are already correlations between the counts in different classes, as:

$$\mathcal{E}\{\tilde{n}_i | n_T; \boldsymbol{\pi}\} = \pi_i n_T; \qquad (i = 1, 2, \ldots, N) \tag{7.2}$$

$$\mathcal{C}\{\tilde{n}_i; \tilde{n}_k | n_T; \boldsymbol{\pi}\} = \begin{cases} \pi_i n_T - \pi_i^2 n_T, & (i = k) \\ -\pi_i \pi_k n_T, & (i \neq k). \end{cases} \tag{7.3}$$

However, to give more modelling flexibility, we now permit both n_T and $\boldsymbol{\pi}$ to be a random scalar and random vector, respectively, but require that they be *independent* of each other, for simplicity. This "collective" model dependency gives a more complex covariance structure.

Define:

$$\mathcal{E}\{\tilde{n}_T\} = \lambda_T = \sum_{i=1}^{N} \lambda_i; \qquad \mathcal{V}\{\tilde{n}_T\} = \gamma_T = \sum_{i=1}^{N} \sum_{k=1}^{N} \gamma_{ik}; \tag{7.4}$$

then, unconditioning (7.2), (7.3), we obtain the moments for use in (6.2), (6.3):

$$\mathcal{E}\{\tilde{n}_i\} = \lambda_i = \lambda_T \mathcal{E}\{\pi_i\}; \tag{7.5}$$

$$(7.6)\qquad \mathscr{C}\{\tilde{n}_i;\tilde{n}_k\}=\gamma_{ik}=\begin{cases}\lambda_T\mathscr{E}\{\tilde{\pi}_i\}+(\gamma_T-\lambda_T+\lambda_T^2)\mathscr{V}\{\tilde{\pi}_i\} & (i=k)\\ \quad +(\gamma_T-\lambda_T)\mathscr{E}^2\{\tilde{\pi}_i\}, & \\ (\gamma_T-\lambda_T+\lambda_T^2)\mathscr{C}\{\tilde{\pi}_i;\tilde{\pi}_j\} & \\ \quad +(\gamma_T-\lambda_T)\mathscr{E}\{\tilde{\pi}_i\}\mathscr{E}\{\tilde{\pi}_j\}, & (i\neq k).\end{cases}$$

It is easy to show that these satisfy (7.4), by using $\sum \tilde{\pi}_i = 1$.

It seems to the author that practical modelling variations might fall into one of two extremes: either (1) the (π_i) might be known rather precisely, and forecasting uncertainty might be associated with the total number of risks or, (2) there would be a relatively stable number of risks, but prediction uncertainty would remain about their distribution over the different risk classification types. (For the casualty claim model, only the first variation would probably be relevant.)

An interesting special case of the compound Multinomial coccurs when the (π_i) are fixed, and $\gamma_T=\lambda_T$. It then follows from (7.5), (7.6) that $\lambda_i=\gamma_{ii}$ ($i=1,2,\ldots,N$) and ($\tilde{n}_i$, $\tilde{n}_k$; $i\neq k$) are *uncorrelated.* This then simplifies (6.2), (6.3), (6.6) to $\lambda=\gamma$, that is, $a=0$, $b=\lambda$, and a Poisson counting distribution would be used in the approximation of Section 5! One obvious way in which this could happen is if $\tilde{n}_T$ were Poisson with parameter, say μ; it is then well known that the $(\tilde{n}_i)$ must be statistically mutually independent, with marginal densities that are Poisson with parameters $(\pi_i\mu)$.

8. AN EXACT RESULT

There is one case in which the proposed procedure gives an *exact* result. Consider a risk portfolio of fixed size N, with each contract $i=1,2,\ldots,N$ having an individual claim density $f_i^0(x)$, with parameters q_i, m_i, v_i, and an *independent* claim number density that is Poisson, with parameter μ_i. This is the basic model used in casualty insurance.

Following the procedure in Sections 5 and 6, we get the same special results described in the previous section, namely, $\gamma_{ii}=\lambda_i$, $\gamma_{ik}=0$, $(i\neq k)$ and $\lambda=\gamma=\sum \mu_i q_i$. In other words, once $f(w)$ is determined from (6.1), the recursive algorithm (5.2) is used with the Poisson density (5.3a) to find the approximate $g(y)$.

However, it is easy to show, using generating functions, that the exact form (1.5) reduces to a compound Poisson law with parameter λ, and a severity density $f(w)$. Thus, the dynamic portfolio approximation is, in fact, *exact* for independent Poisson claims. This is true even if $p_i=0$ for all i!!

Unfortunately, the same line of proof shows also that independent Binomial or Negative Binomial claim densities (with different parameters for each i) can only lead to an *approximation* of the true $g(y)$. However, it follows from Section 6 that the approximating law for $\tilde{n}_e$ would be Binomial or Negative Binomial, respectively.

9 MODELLING WITH FIXED AND RANDOM NUMBER OF COUNTS

To highlight the differences between the model and procedure of this paper and the static portfolio model in JS, it is instructive to re-examine how the independent

Poisson casualty claim model of Section 8 would be handled according to the JS procedure. We use *primes to* designate the equivalent parameters of this paper, in terms of the given model parameters μ_i, q_i, m_i, v_i.

First of all, since all $\tilde{n}_i \equiv 1$ in the JS model, we would have to estimate or calculate separately the N individual *total severity* densities for each contract risk, $\tilde{x}_i$:

$$f_i^{0\prime}(x) = g_i(x) = \sum_{n=0}^{\infty} \frac{(\mu_i)^n e^{-\mu_i}}{n!} [f_i^0(x)]^{n*}. \tag{9.1}$$

(This could be done by N applications of the Adelson algorithm, or might be approximated from real total severity data.)

Then, in terms of the parameters of this paper, we would get:

$$\lambda_i' = 1; \qquad q_i' = 1 - e^{-\mu_i q_i}; \qquad m_i' = \left(\frac{q_i}{q_i'}\right)\mu_i m_i \tag{9.2}$$
$$v_i' = \left(\frac{q_i}{q_i'}\right)\mu_i v_i + (m_i - p_i' m_i') m_i'.$$

Thus, the static portfolio approach of JS would use the Panjer recursive algorithm with.

$$f'(w) = (\textstyle\sum f_i^{0\prime}(w))/(\sum q_j'), \qquad (w = 1, 2, \ldots) \tag{9.3}$$

and a *Binomial* counting density with moments:

$$\lambda' = \sum q_i' < \lambda; \tag{9.4}$$
$$\gamma' = \sum q_i' \left[1 - q_i' \left(\frac{m_i'}{m'}\right)^2\right].$$

The resulting $g(y)$ would then only approximate the true density, which could be obtained exactly in this case. Thus, one might be tempted to dismiss the JS procedure in compound claims applications. However, we can imagine situations in practice where the actuary has used empirical data to estimate the densities, $g_i(x)$ and $\pi(n_i)$. Then the question of the best approximation procedure is still open.

We remind the reader that, if the $(\tilde{n}_i)$ are, in fact, deterministic, then the procedures of the two papers are equivalent; conversely, if the $(\tilde{n}_i)$ are correlated, only the procedure described here applies.

10. OTHER VARIATIONS

In JS, an improved approximation for the example considered was obtained by modifying the $\pi(0)$ of the Binomial (5.3b) to enable an exact match of $\mathcal{V}\{\tilde{y}\}$, together with an integral value of M. This modification could be used with the model of this paper whenever $(\gamma/\lambda) < 1$, and requires only a trivial change in the recursive algorithm. But this refinement is not necessary in the other cases, as $\mathcal{V}\{\tilde{y}\}$ is matched exactly. Of course, one might try matching other moments or values of the exact distribution by modifying the initial values of the prototypical counting density (see the discussion in JS).

It would also be desirable, particularly in investment applications, to extend the range of permitted $(\tilde{x}_{ij})$ to negative values The difficulty then is that the relationship (5.2) is no longer recursive, and must be solved by other means, such as iterative methods. This point is discussed in SUNDT and JEWELL (1981), where possible procedures for the Binomial and Poisson cases are suggested; exact recursion with negative values in the Negative Binomial case $(\gamma/\lambda) > 1$ does not seem to be possible.

11. COMPUTATIONAL CONCLUSIONS AND ACKNOWLEDGEMENT

The limited computations carried out thus far indicate that the same general kinds of approximation error result as in JS; in other words, the underlying severity density should not be too "lumpy" if there are only a few risk types. Errors also seem higher in strongly correlated cases, as expected. A future paper will explore computational results in more detail.

The author would like to thank the referee who found several errors in the original formulae.

APPENDIX A
DEPENDENT RISKS

In Section 1, it was assumed that the individual risk severities $(\tilde{x}_{ij})$ were statistically independent of each other and of the counts $(\tilde{n}_i)$. In this appendix, we consider the modifications necessary if the risks are *exchangeable random variables* within each type i, but still independent of the counts. As is well known, this weak dependency is equivalent to assuming that, for each type $i = 1, 2, \ldots, N$, there exists a random *parameter*, $\tilde{\theta}_i$, such that the individual risks are independent if $\tilde{\theta}_i = \theta_i$ is known, and depend in the same way upon θ_i. Thus, the basic density (1.1) is replaced by:

$$\text{(A.1)} \qquad \Pr\{\tilde{x}_{ij} = x | \theta_i\} = f_i^0(x|\theta_i), \qquad (i = 1, 2, \ldots, N), (j = 1, 2, \ldots, n_i)$$

giving a joint density within type i, given $\tilde{n}_i = n_i$ similar risks, of:

$$\text{(A.2)} \qquad \Pr\left\{\bigcap_{j=1}^{n_i} \tilde{x}_{ij} = x_{ij} | n_i\right\} = \mathscr{E} \prod_{j=1}^{n_i} f_i^0(x_{ij}|\tilde{\theta}_i),$$

and a common marginal density for any risk of type i:

$$\text{(A.3)} \qquad \Pr\{\tilde{x}_{ij} = x\} = \mathscr{E} f_i^0(x|\tilde{\theta}_i) = f_i^0(x).$$

(Expectations in the above are over the random values of $\tilde{\theta}_i$.) Exchangeable random variables thus have the property that they have the same marginal density (and self moments), their arguments may be permuted in any fashion in their joint density (A.2), and they have common cross moments.

In addition to the dependency between different types introduced by the correlation between different counts, we will also permit the different parameters in $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)$ to be statistically dependent, with arbitrary joint d.f. $U(\boldsymbol{\theta})$. In short, our new model substitutes for (1.5) the general form:

$$g(y) = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_N} dU(\boldsymbol{\theta}) \sum_{n_1} \sum_{n_2} \cdots \sum_{n_N} \pi(\boldsymbol{n}) [f_1^0(y|\theta_1)]^{n_1^*} * [f_2^0(y|\theta_2)]^{n_2^*} * \cdots * [f_N^0(y|\theta_N)]^{n_N^*}. \tag{A.4}$$

Intuitively, we can think of θ_i as representing *exogenous factors*, such as the economy, weather, political factors, etc. that influence the random outcome of all risks of type i jointly. This type of "collective behaviour" model is often used in casualty insurance, where it is recognized that all risk classification schemes are imperfect, and that residual correlations still exists among risks of a given type due to the unexplained inhomogeneity still present within the class i. Further, there might be common factors between the different classes, which would account for the dependency between $\tilde{\theta}_i$ and $\tilde{\theta}_k$ $(i \neq k)$.

Proceeding in a manner similar to Section 3, we define the positive risk densities $f_i(x|\theta_i)$, the probabilities $p_i(\theta_i)$ and $q_i(\theta_i)$, and the first two moments, $m_i(\theta_i)$ and $v_i(\theta_i)$, all dependent upon the risk parameter. (3.8), (3.9) still have the same form, except that they express only the conditional mean total risk, $\mathscr{E}\{\tilde{x}_i|\theta_i\}$, and conditional covariance of total risks between different classes, $\mathscr{C}\{\tilde{x}_i; \tilde{x}_k|\theta_i; \theta_k\}$ in terms of the conditional moments of individual risk, and the (non-$\tilde{\boldsymbol{\theta}}$-dependent) moments (3.6), (3.7) of the counts.

Now all that remains is to uncondition these moments, using the relationships:

$$\mathscr{E}\{\tilde{y}\} = \sum_{i=1}^{N} \mathscr{E}\mathscr{E}\{\tilde{x}_i|\tilde{\theta}_i\}, \tag{A.5}$$

$$\mathscr{V}\{\tilde{y}\} = \sum_{i=1}^{N} \sum_{k=1}^{N} [\mathscr{E}\mathscr{C}\{\tilde{x}_i; \tilde{x}_k|\tilde{\theta}_i; \tilde{\theta}_k\} + \mathscr{C}\{\mathscr{E}\{\tilde{x}_i|\tilde{\theta}_i\}; \mathscr{E}\{\tilde{x}_k|\tilde{\theta}_k\}\}]. \tag{A 6}$$

(Innermost operators are over the total risks $(\tilde{x}_i)$; outermost operators are over the risk parameters $(\tilde{\theta}_i)$.)

We define the unconditional versions of $q_i(\theta)$, $m_i(\theta)$, $v_i(\theta)$ as:

$$q_i = \mathscr{E}\{q_i(\tilde{\theta})\}; \qquad \bar{m}_i = \mathscr{E}\{m_i(\tilde{\theta}_i)\}; \qquad \bar{v}_i = \mathscr{E}\{v_i(\tilde{\theta}_i)\}. \tag{A.7}$$

By the theorem of conditional expectation, $q_i = \Pr\{\tilde{x}_{ij} > 0\}$ is the same as in (3.1). However, as the referee reminds us, $\bar{m}_i$ and $\bar{v}_i$ are *not* the same as m_i and v_i in (3.3), (3.4) unless the variation due to $\tilde{\theta}_i$ vanishes; hence, the different notation. In fact, in the current notation, we see that:

$$m_i = \mathscr{E}\{\tilde{x}_{ij}|\tilde{x}_{ij} > 0\} = \bar{m}_i + \mathscr{C}\{q_i(\tilde{\theta}_i); m_i(\tilde{\theta}_i)/q_i\} \tag{A.8}$$

In addition to correlations, we shall also need higher-order cross-moments, so we define:

$$Q_i(\theta_i) = q_i(\theta_i) - q_i; \qquad M_i(\theta_i) = m_i(\theta_i) - \bar{m}_i; \qquad V_i(\theta_i) = v_i(\theta_i) - \bar{v}_i; \tag{A.9}$$

and use notation like.

(A.10)
$$\overline{Q_iQ_k} = \mathscr{E}\{Q_i(\tilde{\theta}_i)Q_k(\tilde{\theta}_k)\} = \mathscr{C}\{q_i(\tilde{\theta}_i);\, q_k(\tilde{\theta}_k)\};$$
$$\overline{Q_iM_i} = \mathscr{E}\{Q_i(\tilde{\theta}_i)M_i(\tilde{\theta}_k)\};$$
$$\overline{Q_iM_iM_k} = \mathscr{E}\{Q_i(\tilde{\theta}_i)M_i(\tilde{\theta}_i)M_k(\tilde{\theta}_k)\};$$

and so forth

In place of (3.10), we have:

(A 11)
$$\mathscr{E}\{\tilde{y}\} = \sum_i \lambda_i q_i \bar{m}_i + \left\{\sum_i \lambda_i \overline{(Q_iM_i)}\right\},$$

and, in place of (3.11), we obtain:

(A.12)
$$\begin{aligned}\mathscr{V}\{\tilde{y}\} &= \sum_i \lambda_i q_i(\bar{v}_i + p_i(\bar{m}_i)^2) + \sum_i\sum_k \gamma_{ik} q_i q_k \bar{m}_i \bar{m}_k \\ &\times \sum_i \lambda_i(q_i p_i \overline{M_i^2} - (\bar{m}_i)^2\overline{Q_i^2}) \\ &+ \sum_i \lambda_i[\overline{Q_iV_i} + 2\bar{m}_i(p_i - q_i)\overline{Q_iM_i} + (p_i - q_i)\overline{Q_iM_i^2} - 2\bar{m}_i\overline{Q_i^2M_i} - \overline{Q_i^2M_i^2}] \\ &+ \sum_i\sum_k 2\gamma_{ik} q_i \bar{m}_i \overline{Q_kM_k} \\ &+ \sum_i\sum_k (\gamma_{ik} + \lambda_i\lambda_k)[q_iq_k\overline{M_iM_k} + \bar{m}_i\bar{m}_k\overline{Q_iQ_k} + 2q_i\bar{m}_k\overline{Q_iM_k} \\ &+ 2q_i\overline{Q_kM_iM_k} + 2\bar{m}_i\overline{Q_iQ_kM_k} + \overline{Q_iQ_kM_iM_k}] \\ &- \left[\sum_i \lambda_i\overline{Q_iM_i}\right]^2.\end{aligned}$$

The term in braces in (A.11) gives a correction term to the calculation of λ in (6.2) (with, of course, m_i and v_i replaced by $\bar{m}_i$ and $\bar{v}_i$); similarly, the terms in braces in (A.11) and (A.12) give two correction terms to the calculation of γ in (6.3).

In many applications, these corrections simplify because either the probability of a claim *or* the moments are independent of $\tilde{\theta}_i$. For instance, in life insurance, $m_i = \bar{m}_i$ and $v_i = \bar{v}_i$ are the moments of the face value of policies of type i, which do not usually change with exogenous conditions, while the expiration probability, $q_i(\theta)$, would probably vary with external effects, this would eliminate *all* terms in (A.11), (A.12) with M_i, M_k, or V_i! Conversely, in casualty insurance, the probability of a claim, q_i, might be relatively fixed several years in a row, but the severity moments, $m_i(\theta_i)$ and $v_i(\theta_i)$, might be relatively uncertain in view of inflation, etc.; in this case, all terms in (A.11), (A.12) involving Q_i and Q_k can be eliminated!

A more complex model can also be developed by permitting the $(\tilde{n}_i)$ to depend upon $\tilde{\boldsymbol{\theta}}$; details are left to the reader.

REFERENCES

ADELSON, R M (1966) Compound Poisson Distributions, *Operations Research Quarterly*, **17**, 73–75

GERBER, H (1979) *An Introduction to Mathematical Risk Theory*, Huebner Foundation Monograph, Irwin, R D Homewood, Ill

JEWELL, W S and SUNDT, B (1981) Improved Approximations for the Distribution of a Heterogeneous Risk Portfolio, *Bulletin of the Association of Swiss Actuaries*, 221–240

PANJER, H H (1981) Recursive Evaluation of a Family of Compound Distributions, *ASTIN Bulletin*, **12**, 22–26

SUNDT, B and JEWELL, W S (1981) Further Results on Recursive Evaluation of Compound Distributions, *ASTIN Bulletin*, **12**, 27–39

BOOK REVIEW

R. V Hogg and S. A. Klugman (1984). *Loss Distributions.* xii+235, £28.45. Chichester: John Wiley & Sons Limited

This book is written from the point of view of mathematical statistics and attacks the problem how to model the probability distribution of the financial severity of a single claim.

Two actuaries, *C. C. Hewitt* and *G. S. Patrik*, functioned as sparring partners for the authors and they were also responsible for part of the written text. As such this book represents a good blend of the pure statistical and practical aspects of the reality of loss phenomena.

The authors advocate the philosophy that a mathematical model for the empirical loss distribution, with parameters estimated from the data, provides a better tool for forecasting and pricing than the empirical distribution itself. I agree, especially when one has to account for structural change through time.

The book consists of five chapters and an appendix, containing the major distributions and their properties. The inclusion of exercises in all chapters provides further insight and makes it useful as a text.

The introductory chapter addresses itself to matters of terminology, kinds of coverages and, what I like to coin, accounting induced descriptive insurance statistics. Most of it will be well known to most of us, but it is always good to see how others put it down on paper.

Hereafter there are two chapters, titled *Models for Random Variables* and *Statistical Inference*, which form about 40% of the whole book. These two chapters are essentially an introduction to mathematical statistics, motivated by and written from the insurance point of view. Even actuaries with a strong training in mathematical statistics and nonlinear optimization procedures may find various instructive examples and models.

Various probability density functions, which should be candidates to graduate the loss distribution, are introduced. These include Gamma, lognormal, Pareto, Weibull, Burr, a generalization of Fisher's F, etc. In my opinion the inverse Gaussian distribution should have been included too.

Parameter estimation for grouped data is thoroughly discussed.

The next chapter, *Modelling Loss Distributions*, applies the procedures, described in the statistical chapters, to various real data sets, which include hurricane, homeowners physical damage, theft and fire, long-term disability, automobile bodily injury and hospital malpractice. Here also an interesting digression on the allocation of loss adjustment expenses can be found. Most of the data are in grouped format.

The final chapter, *Applications of Distributional Models* forms in my view the crown of this book. Here we find such matters as inflation, deductibles and leveraging, limits and layers, loss elimination ratios, and all this with fitted mathematical models.

This is a good book. It should be seminal for the analysis of loss data in practice Therefore it should attract to all actuaries who take an active interest in the statistical analysis of loss data.

P. TER BERG

ANNOUNCEMENT ASTIN WORKSHOP

In this Bulletin appears for the first time a new section called ASTIN Workshop. It is intended to attract papers on practical applications and related to the daily work of the actuary.

We are aware of the fact that such papers might be different in nature and/or style from those regularly published in this bulletin. Nevertheless we feel it to be important to have also such papers in our journal. Of course they should contain a valuable message which is important to the readers.

P. TER BERG

WORKSHOP

RATE MAKING AND SOCIETY'S SENSE OF FAIRNESS

BY G W. DE WIT AND J. VAN EEGHEN

Nationale-Nederlanden N.V.—Research Department
P. O Box 796, 3000 AT Rotterdam, Netherlands

1 INTRODUCTION AND SUMMARY

Actuaries have always been in search of ways to determine premiums which match the risks insured as closely as possible. They do this by differentiating between them on the basis of observable risk factors. In practice, many examples of such risk factors are being used. age and sex for life insurance; location, type of building etc. for fire insurance. Motor insurance is perhaps the most characteristic branch with respect to this phenomenon: in tariffs we find factors like weight, price or cylinder capacity of the car, age of the driver, area of residence, past claims experience (Bonus/Malus), annual mileage etc.

Outsiders may not always be very positive about such a refined premium differentiation. The basis of insurance, they say, should be solidarity among insureds; premium differentiation is basically opposed to this. Another statement heard in the field is: "Premium differentiation ultimately results in letting every individual pay his own claims, it is the end of insurance".

Much confusion arises during discussions about this subject, especially between actuaries and non-actuaries. We will therefore first give a mathematical definition of solidarity, (Section 2), followed by a brief description of certain trends in society which might bring insurers to deliberately drop certain risk factors from their tariffs in order to increase solidarity (Section 3). The consequences of doing so are examined and it is shown that increased solvency requirements will in the end prove to be ineffective. A possible solution is a voluntary transfer of premium between companies (Section 4). The situation is illustrated by an example of health insurance in the Netherlands, where proposals to arrive at such transfers are presently being discussed.

2. FORMS OF SOLIDARITY

If no insurance is purchased, the situation can be briefly summarized as follows:

	Carried by Insured	Carried by Insurer
Risk	X	0
Expected risk	$E(X)$	0
Variance of risk	$\mathrm{Var}(X)$	0

where X is the random variable representing the claims of a random insured.

By the insurance transaction, the risk is transferred from the insured to the insurer. In exchange the insured pays a premium equal to $E(X)$, if we use the expected value premium principle and ignore loadings and estimation errors. The result of the transaction is:

	Carried by Insured	Carried by Insurer
Risk	$E(X)$	$X-E(X)$
Expected risk	$E(X)$	0
Variance of risk	0	$\mathrm{Var}(X)$

It has been recognized however that risks are like leaves in a tree: similar, but never identical. We therefore say that the risk of an individual is characterized by a distribution P_θ, where θ differs from one insured to another. θ is unobservable and is in turn looked upon as a realization of a random variable Θ, whose distribution is characteristic for the market. Thus the risk process is divided into two parts: first the "Θ-lottery" (which is the realization of Θ and can be viewed as an underwriting experiment: each time a risk is accepted a θ is drawn at random from the population Θ); then the "claims lottery" ruled by the probability law P_θ. Suppose for a moment that θ is observable and that the insurer fixes the premium after observing the outcome of the Θ-lottery. If the outcome of Θ is θ, the premium charged will be the conditional expectation $E(X|\Theta=\theta)$. The premium of an insured randomly drawn from the collective then becomes a random variable itself: $E(X|\Theta)$. This situation can be represented as follows:

	Carried by Insured	Carried by Insurer
Risk	$E(X\|\Theta)$	$X-E(X\|\Theta)$
Expected risk	$E(X)$	0
Variance of risk	$\mathrm{Var}\{E(X\|\Theta)\}$	$E\{\mathrm{Var}(X\|\Theta)\}$

While in the first example the insured transferred his full risk to the insurer (X is replaced by $E(X)$), he now keeps part of the risk for himself, for his premium $E(X|\Theta)$ is a random variable.

We may now define actual solidarity as the variance of the risk transferred to the insurer (i.e., shared among insureds). Full solidarity is achieved in the first example:

$$S=\mathrm{Var}(X).$$

In the second example the actual solidarity remains restricted to the "purely probabilistic" part of it:

$$S_p = E\{\text{Var}(X|\Theta)\}.$$

The part of the variance which is caused by Θ, the "risk solidarity", remains with the insured:

$$S_r = \text{Var}\{E(X|\Theta)\}.$$

The subdivision is complete now, for it can easily be checked that

$$S = S_p + S_r$$

As we have said before, θ is unobservable; the probability distribution of an individual risk is never known. However, we do have some information on the differences in distribution of the risks in our portfolio by means of observable risk factors. These risk factors can be viewed as a (vector-valued) random variable F. Mathematically, every potential F satisfies the following:

(1) for all sets A: $\Pr\{X \in A|\Theta, F\} = \Pr\{X \in A|\Theta\}$

i.e., the conditional distribution of the risk given Θ does not depend on F.

If each insured is charged a premium $E(X|F)$* (i.e., information on risk factors is taken into account in pricing), the result of the insurance transaction is as follows:

	Carried by Insured	Carried by Insurer
Risk	$E(X\|F)$	$X - E(X\|F)$
Expected risk	$E(X)$	0
Variance of risk	$\text{Var}\{E(X\|F)\}$	$E\{\text{Var}(X\|F)\}$

Now we can write:

$$\text{Var}(X|F) = E(X^2|F) - E^2(X|F)$$

(because of (1))

$$= E\{E(X^2|\Theta)|F\} - E^2\{E(X|\Theta)|F\}$$

$$= E\{\text{Var}(X|\Theta)|F\} + \text{Var}\{E(X|\Theta)|F\}$$

* We implicitely assume that a good estimate of $E(X|F)$ is available For simplicity, we assume that $E(X|F)$ (like $E(X)$) is known In actual practice however, the choice of F is limited to those factors for which a good (small variance) estimate of $E(X|F)$ is available

Hence

$$E\{\text{Var}(X|F)\} = E\{\text{Var}(X|\Theta)\} + E\{\text{Var}[E(X|\Theta)|F]\}$$
$$= \quad S_p \quad + \quad S_{ur}$$

S_{ur} can be interpreted as the part of risk solidarity S_r that remains unknown after the information contained in F has been taken into account. It may therefore be called the "unknown-risk solidarity".

Similarly, $\text{Var}\{E(X|F)\}$ can be viewed as the part of S_r that becomes known through F. It is therefore called the "known-risk solidarity", S_{kr}.

Evidently we have:*

$$S = S_p + S_r = S_p + S_{ur} + S_{kr}$$

The result of the insurance transaction (with premiums equal to $E(X|F)$) can therefore be rewritten as:

	Carried by Insured	Carried by Insurer
Risk	$E(X\|F)$	$X - E(X\|F)$
Expected risk	$E(X)$	0
Variance of risk	S_{kr}	$S_p + S_{ur}$

The endeavours of the rate making actuary can now be represented as follows:

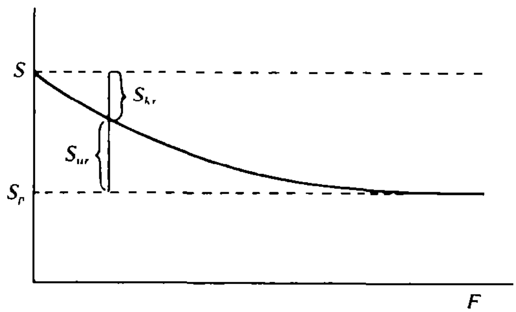


FIGURE 1 Solidarity shared among insureds

* These relations hold only if premiums are based on the expected value of X (conditionally or not to Θ or F). In social insurance however, premiums may not be related to the risk insured at all. Think e.g., of income related premiums or premiums which have to be paid in spite of the certainty one will never receive a benefit (women having to pay for a widow's pension under the Dutch social insurance). In such cases actual solidarity can be defined as

$$E[E\{(X - P(F))^2|F\}]$$

(where $P(F)$ is the premium payable) which may well exceed $S = \text{Var}(X)$.

As F moves "towards full information about Θ" (i.e., towards the right in the above graph)—which is what the actuary tries to achieve—the unknown risk solidarity S_{ur} will tend towards 0, and the solidarity shared among insureds will remain restricted to S_p. If no F is used for rate making purposes, i.e., F is at the origin of the graph, then there will be full solidarity.

3. SOCIETY'S SENSE OF FAIRNESS

In the past, it used to be very difficult to discover risk factors both in a qualitative and in a quantitative sense. Solidarity was therefore—unavoidably—considerable. But recent developments have changed this situation:

- with the help of computers it has become possible to make thorough risk analyses, and consequently to arrive at further premium differentiation,
- the consumer's attitude towards tariffs became more critical He requested more information and, if he was a good risk, objected to pay the same premium as the bad ones.

Both developments have their own special character. The first one shows that in recent years the actuary has been successful in his travel to the right of the F-axis of fig. 1 (see for instance DE WIT (1982) or VAN EEGHEN, GREUP and NIJSSEN (1983)). With the help of large data files and the possibility to analyse such data in detail, he is on his way to reduce actual solidarity to purely probabilistic solidarity S_p. This S_p is the smallest possible value of the actual solidarity shared among insureds. It can be considered as a limit-situation in which F contains all information about Θ. BICHSEL (1983) has shown that an insurance system in which each insured is charged a premium equal to the expected value of future claims leads to the optimization of the total result of the economy. Along the same lines, one might argue that the minimization of solidarity through further refinements of tariffs, leads to the optimization of the total result of the economy as well.

The consumer's attitude is of a completely different nature. In the past, we believe he would be more inclined to simply accept the premium charged, but today things seem to be different. Premiums have increased a great deal, coverages have been extended and the risks of society have grown. Because of the relative level of premiums, the consumer has become more sensitive to price differences. Price sensitivity is probably also closely related to the general economic situation. In days of rapid economic growth and an ever increasing level of personal consumption, people will pay less attention to premium differences than in times of stagnation and budget squeezes. These developments tend to decrease total solidarity.

Nevertheless a changing attitude is starting to become apparent, caused by a critical review of today's society. It is this change that gives rise to the type of statements mentioned in the introduction. People start realizing that a certain restauration of solidarity might be desirable. For insurance, this seems to apply especially to those branches which are in the closest relation to people themselves.

Therefore: healthy persons will be paying for the less healthy ones. And: should someone who works under circumstances which endanger his physical condition have to pay a higher premium in spite of the fact his work is of vital importance to the economy? The answer to such questions is often determined by the degree of influence a person has on his own risk. Should solidarity be extended to cover those people who harm their own health by their voluntarily chosen way of life? The answer to this question would generally be affirmative, but a non-smoker's discount, for instance, denies this form of solidarity. One might object against this form of solidarity, because it reduces one's own responsibility and has an anti-prevention character. It therefore seems justified to restrict solidarity to factors for which one is not personally responsible. Alternatively, the community can impose solidarity by safety rules (helmets for motorcyclists, safety belts in cars etc.)

But solidarity is not merely related to "personal" branches of insurance, but applies to more material fields also. Should someone who, for economic reasons, lives in a certain area pay a higher motor insurance premium, because of the higher traffic risks? The higher rent he has to pay in such an area may even be compensated by special subsidies. This brings us to a totally different aspect of solidarity. Should premiums be such that everyone can afford insurance? In the past, this question used to be relevant for social insurance only The reasoning of private insurers was: if you cannot pay for insurance, don't buy it. But times have changed. Many types of insurance have become such common commodities, that they are being considered as basic needs and must therefore be affordable by everyone. If private insurers do not want to see their tasks taken over by social insurance they should keep this aspect in mind.

Where the foregoing considerations have a mainly social character, legal aspects may (or will) also be important, in the form of restrictions which preclude insurance companies from using certain risk factors, even if these factors can be proved to be statistically significant. We are thinking of:

- emancipation. It will no longer be allowed to distinguish between men and women for rate making purposes. For the European Community this rule will be laid down in a forthcoming directive;
- discrimination. Tariffs are not allowed to differentiate between racial groups.

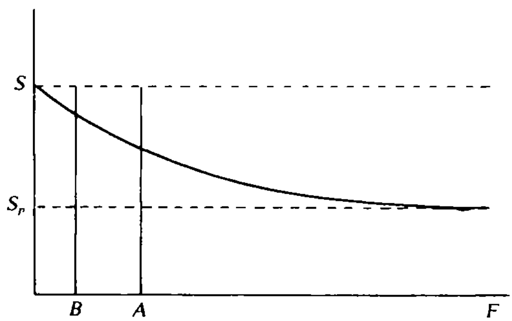


FIGURE 2

What we are saying is that there may be reasons (mostly related to society's sense of fairness) for us not to use certain parts of the risk information available in our tariffs. Although we might be able to push premium differentiation to point A in figure 2, we prefer to stick with situation B In practice, the difference between A and B can be quite significant.

4. CONSEQUENCES OF DELIBERATELY IGNORING RISK INFORMATION

If it is decided that not all information available will be used for the rating structure, this policy should be adopted by all companies operating on the same market. It will be clear that, if with one company two different risks have to pay the same premium, while with another one the good risks pay a low premium and the bad risks a higher one, the former will attract relatively more bad risks and thus will incur a loss.* This implies that in a market where every company is free to fix its own rating structure (like the Dutch or the British ones) it will be less easy to respond to social pressure for "fairness" than in markets where rating structures are imposed by the authorities (like in Germany or Belgium).

But even in a regulated market there may be problems. If certain aspects of risk information are ignored in the rating structure, they may be used for underwriting purposes. Applications by unwanted risk groups may be refused or "forgotten". By doing so, insurance companies can increase their profitability, while other companies will see their profits reduced.

This may also occur with "special character" companies: in almost every country there are companies (often mutuals) which sell insurance to agriculturers or to civil servants only or which operate in one specific area. These groups may turn out to consist of non-average risks.

To illustrate the effects of such a situation, let us assume that we are dealing with a two-company market. A risk factor F has been discovered but it is not used in the tariffs. So both companies charge a level premium $E(X)$. Suppose that, due to the market mechanism described above, the insureds of company I are characterized by $F \in \mathcal{F}$ and those of company II by $F \notin \mathcal{F}$, where $\mathcal{F}$ is a subset of the possible outcomes of F. The expected risk carried by the companies now becomes.

$$E(X|F \in \mathcal{F}) - E(X) \quad \text{for company I}$$

and

$$E(X|F \notin \mathcal{F}) - E(X) \quad \text{for company II}$$

one of which, say for company I, may be positive

This situation clearly leads to modifications in the profit and loss accounts of companies I and II In a free market (but still assuming that F remains removed from the tariffs), company I might consider adjusting its overall premium level.

* In reality, the difference has to be substantial before the effect becomes noticeable Moreover, we simplify by considering the risk process only and by ignoring expenses, marketing and client service aspects

But this would mean that F is effectively used as a rating factor, not on a company level but on the market level. Risks characterized by $F \in \mathcal{F}$ would be charged a higher premium (by company I) than risks characterized by $F \notin \mathcal{F}$ (insured with company II). Since this is what we were trying to avoid on grounds of social fairness, the situation is not very satisfactory and seems to call for another solution, especially if the premium differences are very large.

Another possibility is that the companies do not adjust their premiums to reflect the special character of the risks of their portfolio. The resulting positive value of the expected risk for company I is a risk theoretical impossibility: the insurer will soon be ruined. In practice however, premiums contain loadings for security, expenses, profits etc.

This loading will now turn out to be lower than expected, because of the special risk selection represented by $F \in \mathcal{F}$. It may very well be possible that the company can still live and survive with this smaller loading. But its existence will have become subject to more risk and a larger safety buffer may therefore be required. See also DE HULLU (1984).

Let us return to fig. 2. It illustrates that by deliberately ignoring some risk information, we find ourselves in situation B instead of A. Solidarity between insureds $S_p + S_{ur}$, which was defined as the variance carried by the insurer, was thus increased. This increase of the variance is a second indication that the solvency margin of an insurer is to be increased when not all possible risk information is used in determining premiums. Such an increase would be based on risk variance grounds and therefore its nature is different from the one which reflects premium inadequacy due to risk selection ($F \in \mathcal{F}$).

The necessary provisions in situations A and B can thus be written as:

$$R_A = RV_A + RS_A$$

$$R_B = RV_B + RS_B$$

where

$$RV = k_1 \sqrt{n}\,\sqrt{S_p + S_{ur}}$$

is the variance part of the provision (n is the portfolio size), and

$$RS = k_2 n \sqrt{S_{ur}}$$

is the risk selection part of the provision. We have

$$RV_A < RV_B;$$

$$RS_A < RS_B$$

and hence

$$R_A < R_B.$$

It should further be noted that RV expressed as a percentage of premium income, tends to zero when the size of the portfolio increases, This is not true for RS, which shows that the law of large numbers is not the whole story of insurance as is popularly believed. RS will be zero when all risk information is reflected in premium differentiation. While much literature exists on the determination of the level of the variance part of the reserve, RV, it is hard to say anything general about the level of RS. We will simply mention two of the factors that can influence it:

- the explanatory power of the deliberately omitted risk factor F. The more $E(X|F=f)$ varies with f, the greater S_{un} and the greater will be the premium inadequacy resulting from adverse risk selection. The risk pattern in fig. 3a is more dangerous than the one in fig. 3b.

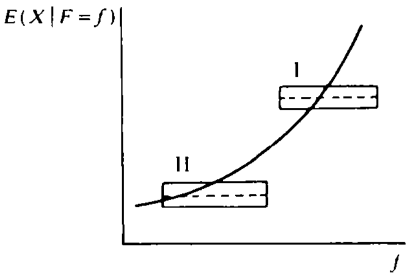


FIGURE 3a

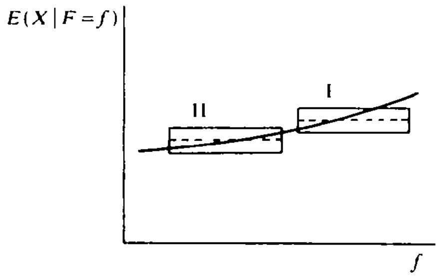


FIGURE 3b

- the possibility of the market to arrive at an effective risk selection The most dangerous risk selection would result from a choice of $\mathcal{F}$ such that $E(X|F\in\mathcal{F})-E(X|F\notin\mathcal{F})$ is maximized. In practice however, not all choices of $\mathcal{F}$ are possible. Social tolerance can be important in this respect A health insurance company for instance, cannot openly say that it accepts insureds under age 30 only, without being highly controversial. So full risk selection (through age) would be impossible, but some degree of risk selection may be possible by means of carefully planned marketing campaigns. Such aspects are reflected in the value of k_2

As we have seen, an extra security buffer may be necessary for protection against risk selection effects in a market where part of the risk information is not reflected in tariffs. This may however not be sufficient. The possibilities of fighting premium inadequacy by setting up extra provisions are limited. The difference between the net premium charged ($E(X)$) and the necessary net premium ($E(X|F\in\mathcal{F})$) may be too large to be financed from the premium loading. In such cases another solution is necessary

Since considerations of social fairness have led to the decision of non-differentiation of premiums with respect to F, this same sense of social fairness suggests a transfer of premium income from the companies characterized by $F\notin\mathcal{F}$ to those portfolios for which $F\in\mathcal{F}$ holds. Since F is an observable risk factor, with known effect on the expected losses, the level for such a "fair transfer" can be computed

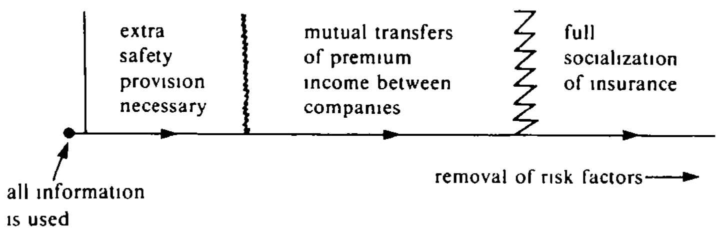
To realize such transfers in practice, a consensus between companies is necessary. This consensus can only be reached when there is full agreement on what is "socially fair". Public opinion and political pressure can prove to be major factors in the process of reaching a final agreement.

These principles have been applied in the Dutch health insurance market. "Age of the insured" is a highly significant risk factor but it is not used in tariffs. Deliberate and undeliberate risk selection by some companies have had a great impact on the profit and loss accounts of the Dutch companies. As a consequence of the free market mechanism premium levels have been adjusted, leading to highly differing premium levels from company to company. Thus age has factually returned as a rating factor. At present, proposals are being discussed to arrange transfers of premium income between companies, to cure the situation. The theoretically necessary safety provisions could then be reduced to a realistic level. Details can be found in the Appendix.

5. CONCLUSION

In this contribution we have tried to show that, depending on social circumstances, practical tariffs should not always reflect all the risk information available.

If the reduction of the relationship on a micro-level between risks insured and premiums charged is pushed to an extreme, the nature of the insurance industry will change profoundly, the end being a full socialization of insurance, with for instance income-related premiums. The different stages of such a process can be summarized as follows:



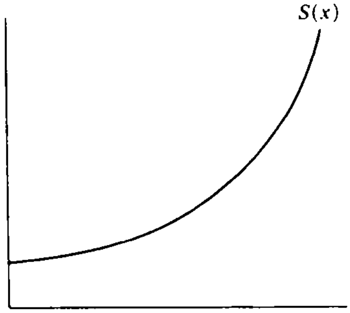
It is not our intention to plead for tariffs from which as many risk factors as possible have been removed. But in modern society, we should be aware of the fact that some aspects of the free market mechanism need to be adjusted. This also applies to insurance, and insurance companies need not feel threatened by these developments. The intention of this paper is therefore to take better notice—especially in a quantitative sense—of the consequences of solidarity transfers. A better knowledge of these transfers may help the insurance industry to react adequately upon general social developments. If such a reaction should result in a reduced premium differentiation, the insurance industry will have to reach a consensus on whether to increase solvency requirements or to neutralize the effects of risk selection by mutual transfers of premium income.

APPENDIX

Health Insurance in the Netherlands

In the Netherlands, 30% of the population obtains full health coverage from private insurance companies.

The most important risk factor (F) for this type of risk is age of the insured. The expected loss for an individual of age x is: $E(X|F=x)=S(x)$.



Clearly, age is a very significant risk factor. Premiums however, originally do not depend on age, but only on type of coverage and level of deductible. So if $N(x)$ is the number of insureds of age x in the portfolio, the net premium applicable to all insureds is:

$$P=\left(\sum_x N(x)S(x)\right)\Big/\sum_x N(x).$$

An actuarial provision is not formed; the financing system used is pay-as-you-go.

This situation is a typical example of a rating structure where an important risk factor has been deliberately ignored, and the features described in the previous paragraph are observable in the market. Premiums differ widely from company to company, as a function of the age composition of the portfolio. Companies with a "young" portfolio have low premiums and therefore attract the largest number of new (mainly young) insureds. Companies with an "old" portfolio have high premiums, they will therefore not be able to attract enough new insureds, the portfolio therefore grows older and as a result they have to increase their premiums, possibly to an unacceptably high level. This feature is reinforced by the steepness of the $S(x)$-curve: $S(85)$ is about 7 times as high as $S(20)$. The situation in the market can therefore well be represented by fig. 3a of the main text (for two companies). Due to the free market mechanism, age is de facto used as a rating factor.

An extra provision necessary to protect companies from possible premium inadequacies can be formed as follows.

Suppose there is a portfolio with a given age composition which attracts no more new insureds. The age composition in subsequent years is therefore fully determined by aging and mortality.

If p_x is the one-year survival probability of an x-year old insured, the expected number of insureds of age $x+t$ after t years will be.

$$N_t(x+t) = p_x \cdot p_{x+1} \cdots p_{x+t-1} N(x).$$

Suppose the premiums are to be kept constant at a level P, the discounted (interest $= 100i\%$) premium income for the company will be:

$$DP = P \sum_t (1+i)^{-t} \sum_x N_t(x+t)$$

while the discounted yearly claims total is:

$$DC = \sum_t (1+i)^{-t} \sum_x N_t(x+t) S(x+t).$$

The difference $DC - DP$ could be considered as a theoretically necessary extra provision.

For the average Dutch portfolio this provision would be 440% of net premium income, an amount which is simply not available and is equal to the actuarial provision in a fully capitalized health insurance system without inflation.

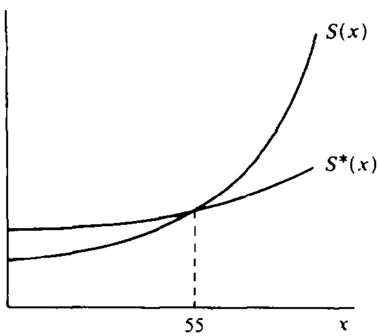
To cope with this problem and with the problem of widely differing premium levels in the market (part of which is due to differing age compositions), the Dutch health insurance companies are in the process of deciding to share the costs of older insureds. Basically the proposal is as follows:

- for each insured of age $x > 55$, the insurer receives $r[S(x) - S(55)]$ where $0 \leq r \leq 1$;
- for each insured of age $x < 55$, the insurer pays $\lambda[S(55) - S(x)]$,

where λ is determined in such a way that for the market as a whole and for fixed r, the balance of income and expenditure is zero.

The result of these transfers is a new (less steep) curve of expected losses:

$$\begin{aligned} S^*(x) &= (1-\lambda)S(x) + \lambda S(55) && x < 55 \\ &= S(55) && x = 55 \\ &= (1-r)S(x) + rS(55) && x > 55. \end{aligned}$$



Due to the reduction of the slope of the S-curve, premium differences in the market will decrease, so that the resulting situation might be represented by figure 3b of the main text.

If we compute the level of the necessary safety provision in a similar way as above, but with $S(x)$ replaced by $S^*(x)$, we find, for the values of r and λ suggested by the Dutch insurers, a provision of 20% of premium income, which is more in line with the financial position of the Dutch health insurance industry than the previous 440%.

This is a mixed solution. The S^*-curve is not completely horizontal. The factor age continues to be of importance, also because the premium transfers relate only to a specific part of the total health insurance coverage. An extra provision (20%) thus remains necessary. Premium differences due to differences in age composition are strongly reduced. Solidarity between younger and older insureds is thus secured through an agreement between insurance companies which does not interfere with normal, healthy competition.

REFERENCES

BICHSEL, F (1983) Versicherungspramien und optimales Verhalten von Konsumenten und Produzenten gegenuber Risiken *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, 1983, 105–113

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

HULLU, A DE (1984) A Management Oriented Approach to Solvency Transactions 22nd International Congress of Actuaries, Sydney, Volume 3, 21–29

WIT, G W DE (1982) Premium Calculation in Theory and Practice *General Insurance Bulletin (Australia)*, nr 34 (June), also published in *Bulletin de l'Association Royale des Actuaires Belges*, nr 76

THE INFLUENCE OF EXPENSE LOADINGS ON THE FAIRNESS OF A TARIFF

By Jean Lemaire

Université Libre de Bruxelles, Brussels, Belgium

Abstract

In non-life insurance, it is nearly always assumed that the expense loading is a fraction of the risk premium. This may deeply affect the fairness of a tariff, as illustrated in the case of the Belgian bonus-malus system.

Keywords

Rate making, Expense loadings.

1. Introduction—Summary

The exponential growth of the number of papers dealing with the theory of premium calculation principles is one of the significant trends of the actuarial science during the last decade. Also noteworthy is the fact that all of those papers concentrate on the risk premium (pure premium and safety loading) and deliberately leave aside the determination of the loading for expenses, commissions, taxes, profit, ... We shall attempt to show in this paper that this neglect has some severe consequences, that it is futile to try to assess the risk premium with great precision if the expense loading is only grossly estimated, that risk premiums with desirable characteristics in terms of the principles of risk classification are distorted through the loading process (this should be obvious since in many cases the expense loading is greater than the risk premium). Note that the same remark was made by Jewell (1980): "The next step in premium setting is to determine the additional 50–200% increase which determines the commercial premium by adding expense and profit loadings. Except in life insurance where there are specific cost models for sales commissions (in many cases of regulated form), there seems to be no further modelling principles used, except [multiplying the risk premium by a factor $1+\alpha$]. This lacuna in the literature is all the more surprising, as it is in sharp contrast to the fields of engineering and business management, where extensive and sophisticated cost allocation and modelling are the order of the day. Are these activities outside the realm of the actuary, ... ?"

2. Apparent and Real Risk Premiums

In all lines of insurance the policy-holders are partitioned according to some criteria that significantly affect the risk (like use and power of the car, age and occupation of the driver in motorcar insurance). Let s be the number of cells, and $\{b_i; i=1,\dots,s\}$ the commercial tariff: b_i is the premium to be paid by a policy-holder that belongs to cell i. b_i is the sum of two components: the risk premium r_i and the expense loading e_i, that contains the company's general expenses g_i, the commissions c_i, the taxes t_i, and, in some cases, a profit loading p_i:

$$b_i = r_i + e_i, \qquad i=1,\dots,s$$

where $e_i = g_i + c_i + t_i + p_i$.

In non-life insurance, it is nearly always assumed* that the expense loading is a fraction of the risk premium:

$$b_i = r_i(1+\alpha), \qquad \alpha > 0, \qquad i=1,\dots,s.$$

The loading coefficient $\alpha = \alpha_g + \alpha_c + \alpha_t + \alpha_p$, where

α_g = loading coefficient for general expenses

α_c = loading coefficient for commissions

α_t = loading coefficient for taxes

α_p = loading coefficient for profits.

This proportional approach is certainly open to some criticisms. Why should the salesmen of the company (brokers, agents, ...) be paid more for bad risks than for good risks (on the contrary we feel that they should be rewarded for bringing good risks to the company)? Is it fair that young drivers pay more taxes than older policy-holders? Is there any reason for the fact that drivers living in big cities contribute more to the profit of the company than inhabitants of small communities? If a proportional loading is applied, the high risk cells certainly pay a disproportionate share of the expenses. This means that the "real" risk premium they pay is not r_i, but $r_i' = r_i + (EX)_i$, where $(EX)_i$ is the excess charge of expenses (considered here as the "hidden" part of the risk premium)

(a) *A Special Case: Flat Expense Loading*

Suppose that there is no reason whatsoever that the high risk cells contribute more towards the expenses than the low risks, and denote by n_i the population of cell i. Instead of paying $b_i = r_i(1+\alpha)$, a risk that belongs to cell i should pay

* Among the few exceptions we found in the literature were

- the proposed new motor rating structure in the Netherlands, the authors' recommended rates are applicable for 90% of the premium income (including the part of the component for expenses contained therein), while the remaining 10% is considered to relate to expenses which should be appropriated towards each policy as a fixed amount (GREGORIUS (1982))
- a recommendation of the Massachusetts Insurance Service Office (ROY (1980)) The proposition is to allocate 75% of the operating costs as a fixed amount, and the remaining 25% as a proportional loading

$b_i' = r_i + \beta$, where

$$\beta = \frac{1}{n}\left(\alpha \sum_{i=1}^{s} n_i r_i\right)$$

$$= \frac{\alpha}{1+\alpha} \frac{\sum_{i=1}^{s} n_i b_i}{n}$$

is computed in such a way that the total income $\sum_i n_i b_i$ of the company is not modified. As this risk effectively pays b_i, he is charged a (positive or negative) excess premium of

$$(EX)_i = \alpha r_i - \beta$$

$$= \frac{\alpha}{1+\alpha}\left[b_i - \frac{\sum_i b_i n_i}{n}\right].$$

The real risk premium paid is thus

$$r_i' = r_i + (EX)_i = b_i - \beta.$$

(b) *A More General Case: Linear Loading*

Suppose now that the expense loading should be partly proportional to the risk premium, partly uniform. Instead of being charged $b_i = r_i(1+\alpha)$, a risk of cell i should contribute

$$b_i' = r_i(1+\gamma) + \beta,$$

where $\gamma = \gamma_g + \gamma_c + \gamma_t + \gamma_p$, and $\beta = \beta_g + \beta_c + \beta_t + \beta_p$.

The total proportional part of the company's income is

$$(1+\gamma)\sum_i n_i r_i = \frac{1+\gamma}{1+\alpha}\sum_i n_i b_i.$$

In order to keep the same total income, β should then be equal to

$$\beta = \frac{1}{n}\left(\sum_i n_i b_i - \frac{1+\gamma}{1+\alpha}\sum_i n_i b_i\right)$$

$$= \frac{\alpha-\gamma}{1+\alpha}\frac{\sum_i n_i b_i}{n}.$$

The excess premium for cell i

$$(EX)_i = \alpha r_i - (\gamma r_i + \beta)$$

$$= \frac{\alpha-\gamma}{1+\alpha}\left[b_i - \frac{\sum_i n_i b_i}{n}\right].$$

Then the real risk premium is

$$r_i' = r_i + (EX)_i = \frac{1}{1+\alpha}\left[b_i(1+\alpha-\gamma) - (\alpha-\gamma)\frac{\sum_i n_i b_i}{n} \right].$$

Of course other expense allocation models are conceivable (commissions designed in such a way that the broker has an incentive to sign up good risks, for instance), but the model considered in (b) is more likely to be selected in practice, due to its simplicity.

3. AN 'APPLICATION

Since 1971, all Belgian companies are compelled to use an 18-class merit-rating system in motorcar third-party insurance. The premium levels $\{b_i; i = 1, \ldots, 18\}$ are presented in Table 1 along with the populations n_i observed in a company (columns 1 and 2). The expense loading is purely proportional, with the following coefficients

Company expenses			$\alpha_g = 0.5901$
Commissions			$\alpha_c = 0.3257$
Taxes	for the social security system	$\alpha = 0.1916$	$\alpha_t = 0.4885$
	for the fund of the handicapped	$\alpha = 0.1149$	
	for the Red Cross	$\alpha = 0.0048$	
	tax	$\alpha = 0.1722$	
		Total loading	$\alpha = 1.4043$

The expense loading thus multiplies the risk premium by 2.4!

(a) *Flat Expense Loading*

Let us assume that the fair way to allocate expenses is that each policy-holder pays a fixed amount. In our example we obtain

$$\beta = \frac{\alpha}{1+\alpha}\frac{\sum_i n_i b_i}{n} = 39.9308.$$

We then compute the excess premium, and express it as a percentage of the commercial premium b_i (see Table 1, columns 3 and 4). For instance a policy-holder of class 18 can claim that he is being overcharged by 76.88, or 38.44%! Then, we substract β from b_i in order to obtain the real risk premium (column 5). By multiplying the figures of this column by 1.6647 (in order to bring back the premium of the initial class 10 to 100), we obtain the "real" merit-rating system applied by the Belgian companies. It differs markedly from the "alleged" one: for instance the ratio between the largest and smallest premiums is 8, instead of the apparent 3.33!

TABLE 1

FLAT EXPENSE LOADING

b_i	n_i	$(EX)_i$	$\frac{(EX)_i \times 100}{b_i}$	r'_i	"Real" System
200	27	76 88	38 44	160 0692	266 47
160	28	53 52	33 45	120 0692	199 88
140	53	41 83	29 88	100 0692	166 59
130	81	36	27 69	90 0692	149 94
120	115	30 16	25 13	80 0692	133.29
115	201	27 24	23 69	75 0692	124 97
110	322	24 32	22 11	70 0692	116.65
105	507	21 40	20 38	65 0692	108 32
100	1 141	18 48	18.48	60.0692	100
100	1 429	18 48	18 48	60 0692	100
95	2 318	15 56	16 37	55 0692	91 68
90	3 385	12 64	14 04	50.0692	83 35
85	9 190	9 72	11 43	45 0692	75 03
80	9 791	6 79	8 49	40 0692	66 71
75	9 887	3 87	5 17	35 0692	58 38
70	12 231	0 95	1 36	30.0692	50 06
65	11 025	−1 95	−3 02	25.0692	41.73
60	70 962	−4.89	−8 14	20 0692	33 41
	132 693				

(b) *Linear Loading*

To be more realistic, let us compute the real "hidden" merit-rating system under the following assumptions.

(i) Commissions should be the same for every risk.

Indeed in Belgium a broker is nothing more than a salesman, and does not participate in the settlement of claims. He should not have any incentive to sign up customers that belong to the worst risk classes. So $\gamma_c = 0$ and

$$\beta_c = \alpha_c \frac{\sum_i n_i r_i}{n} = \frac{\alpha_c}{1+\alpha} \frac{\sum_i n_i b_i}{n} = 9.2608.$$

(ii) The contributions to the social security system, the fund of the handicapped and the Red Cross should be proportional to the risk premium.

Bad risks have a higher propensity to cause claims with bodily injury, thereby adding their share to the deficits of the social security system and the fund of the handicapped. It is only fair that they should pay for it. So $\gamma_t = 0.3113$.

(iii) The tax should be the same for all policy-holders. So

$$\beta_t = \frac{\alpha_t - \gamma_t}{1+\alpha} \frac{\sum_i n_i b_i}{n} = 5.0390.$$

(iv) The part of the general expenses related to the production and the administration of the policies should be uniformly distributed among the policyholders. The part related to the claims settlement should be proportional to the risk premium. In a large Belgian company, the former part accounts for 72.54% of the general expenses, the latter part for the remaining 27.46%. This leads to $\gamma_g = 0.1620$ and

$$\beta_g = \frac{\alpha_g - \gamma_g}{1+\alpha} \frac{\sum_i n_i b_i}{n} = 12.1714.$$

Assembling the three components, we have

$$\gamma = \gamma_c + \gamma_t + \gamma_g = 0.4733$$

$$\beta = \beta_c + \beta_t + \beta_g = 26.4712$$

Altogether, around one third of the total expense loading is allocated proportionally, the remaining two thirds evenly.

The computations described in Section 2 enable us to compute the "real" merit-rating system applied by the Belgian companies; it is more severe than the "official" one, since for instance the ratio between the extreme premiums is 6.18, instead of the apparent 3.33.

It was stated over and over again [see for instance LEMAIRE (1982)] that the Belgian bonus-malus system is unefficient and unfair to the best drivers, since the penalizations for claims are much too small.

The preceding considerations show that the effect of a purely proportional loading is to attenuate this unfairness.

TABLE 2

LINEAR LOADING

b_i	$(EX)_i$	$\frac{100(EX)_i}{b_i}$	r_i'	"Real" System
200	50 97	25 48	134 15	249 16
160	35 48	22 18	102 03	189 50
140	27 74	19 81	85 97	159 67
130	23 86	18 36	77 94	144 75
120	19 99	16 66	69 90	129 83
115	18 06	15 70	65 89	122 37
110	16 12	14 65	61 87	114 92
105	14 18	13 51	57 86	107 46
100	12 25	12 25	53 84	100
100	12 25	12 25	53 84	100
95	10 31	10 86	49 83	92 54
90	8.38	9.31	45 81	85 08
85	6 44	7 58	41 79	77 63
80	4 50	5 63	37 78	70 17
75	2 57	3 42	33 76	62 71
70	0 63	0 90	29 75	55 26
65	−1 30	−2 01	25.73	47 79
60	−3 24	−5 40	21 72	40 33

REFERENCES

GREGORIUS, F (1982) Development of the Study In *New Motor Rating Structure in the Netherlands* ASTIN-Groep Nederland, pp 15–37

JEWELL, W (1980) Models in Insurance Paradigms, Puzzles, Communications and Revolutions *Proc of the 21st I C A Zurich-Lausanne*

LEMAIRE, J (1982) *L'assurance Automobile Modèles Mathématiques et Statistiques* Fernand Nathan-Editions Labor

ROY, J (1980) La Tarification en Assurance Automobile au Québec Rapport du Surintendant des Assurances

L'ÉCRÊTEMENT DES SINISTRES "AUTOMOBILE"

PAR LIONEL MOREAU

Assurances Générales de France, Paris

SUMMARY

French insurance companies usually classify their agents according to their results by branch and to this effect they calculate their respective claim ratios.

As regards motor insurance many agents have argued against this practice as they believed it was unfair to be adversely classified when they had the misfortune to incur a large claim.

In this article, the merits of various statistical systems for attenuating the effects of large claims on the portfolio are considered in order to come to the most equitable solution possible, i.e., to ascertain the best method and system for attenuating the effects of large claims. The conclusion is that the excess of loss technique is best suited to this effect.

To arrive at this conclusion the author of this article relates in detail the various stages of the study, the various systems envisaged and the tests which have allowed to select the most appropriate system. The main results obtained are also given together with their respective drawbacks.

At the end of the article the author mentions the still fairly limited bibliography which deals with this question.

Dans le but d'améliorer les résultats, les compagnies d'assurance ont l'habitude en France de publier un classement de leurs Agents Généraux basé sur le ratio S/P (sinistres/primes) et de tenir compte de ce classement pour l'attribution de certains avantages commerciaux.

Or, en assurance automobile, un malaise profond s'était emparé d'un certain nombre d'entre eux, et ils avaient mis l'accent sur un point qu'ils estimaient injuste:

"Nous ne pouvons pas être pénalisés" disaient-ils, "si un très gros sinistre, atteignant 500 ou 1000 fois la prime annuelle, frappe un de nos clients, car il s'agit alors d'un phénomène aléatoire dont nous ne sommes pas responsables".

Telle est l'origine de la politique de l'*écrêtement.*

1. LA THEORIE DE BASE

Nous désignerons pour une agence i:

- les primes émises par P_i
- le nombre de véhicules par v_i
- le nombre de sinistres par n_i
- le montant des sinistres par S_i
- le montant du sinistre j par s_i^j

$$S_i = \sum_{j=1}^{n_i} s_i^j.$$

Nous voulons substituer à l'indicateur S_i/P_i brut un indicateur plus nuancé, obtenu par la répartition sur *l'ensemble des agences* de la masse des montants des sinistres excédant un seuil nommé *priorité*. Cette technique est celle bien connue en réassurance sous le nom d'excess-loss. Pour utiliser le langage habituel, nous parlerons des *sous-crêtes*, pour désigner les montants inférieurs à cette priorité et *surcrêtes* les montants supérieurs.

On peut donc se poser deux problèmes distincts:

(a) Comment calculer les surcrêtes?

(b) Comment les répartir?

(a) pour les calculer, trois méthodes sont proposées par Mlle Hess (cf. bibliographie [1]):

1. *La priorité fixe*

Doit alors être répartie, pour l'ensemble des agences, la masse R des sinistres supérieure à la priorité fixe F:

$$R=\sum_i \sum_{j\in J_i} (S_i^j - F)$$

avec $J_i = \{j: s_i^j > F\}$.

2. *La priorité dite au Kème percentile*

On veut par exemple écrêter 1% ou 2% des sinistres. La priorité F est alors le résultat d'un calcul implicite:

$$\sum_i \sum_{j\subset J_i} = K\cdot \sum_i n_i$$

et on détermine aisément F par le tracé de l'histogramme cumulatif.

3. *La priorité dite au percentile variable*

C'est la même technique mais segmentée, les agences ayant été préalablement ventilées dans des classes différentes en fonction de critères de taille, le pourcentage pouvant devenir plus important si la taille est plus petite, afin de minimiser la dispersion.

(b) Pour répartir cette masse, on peut appliquer plusieurs techniques:

1. *Répartition sur les sous-crêtes*:

Si nous posons

$$T_i = \sum_{j\notin J_i} s_i^j + F\sum_{j\in J_i} 1$$

$$S_i' = T_i \times\left(1+\frac{R}{\sum_i T_i}\right).$$

2. *Répartition sur les nombres de sinistres*:

$$S_i' = T_i + \frac{R\times n_i}{\sum_i n_i}.$$

3. *Répartition sur les nombres de véhicules*:

$$S'_i = T_i + \frac{R \times v_i}{\sum_i v_i}.$$

4. Nous ajouterons bien sûr une quatrième répartition, la répartition sur les primes

$$S'_i = T_i + \frac{R \times P_i}{\sum P_i}.$$

On doit donc étudier quel indicateur rendra le mieux compte de la situation.

Les travaux mentionnés ci-dessus nous proposent de déterminer S'_i / P_i ayant l'une des formes 1, 2 ou 3 ci-dessus.

Nous proposons d'y ajouter la quatrième et même d'étudier une forme plus classique, celle appliquée en réassurance d'excess-loss, à savoir:

$$T_i / P'_i$$

$$P'_i = P_i \times \left(1 - \frac{R}{\sum P_i}\right)$$

qui consiste à ne prendre en compte que les sous-crêtes rapportées à des primes nettes du prix de l'excess.

II. CRITERE D'OPTIMISATION

Entre toutes ces méthodes d'écrêtement et ces choix de répartition et d'indicateur, quelle est la meilleure façon de procéder? Celle bien sûr qui donne le meilleur ajustement avec ce que nous déciderons d'appeler le meilleur critère (ce qui résulte là d'un choix et non d'une vérité qui s'imposerait).

Nous avons pensé que les agents étaient "responsables" des éléments qualitatifs suivants.

- les sinistres *matériels* en fréquence et en coût moyen
- les sinistres *corporels* en fréquence.

Autrement dit, nous voulions juger les agents sur le critère $\hat{S}_i / P_i$

$$\hat{S}_i = \sum_{j \in J_i^m} s_i^j + m \sum_{j \in J_i^c} 1$$

avec·

$$J_i^m = j : s_i^j = \text{montant d'un sinistre matériel}$$

et

$$J_i^c = j : S_i^j = \text{montant d'un sinistre corporel}$$

m étant le coût moyen national:

$$m = \left(\sum_i \sum_{j \in J_i^c} s_i^j\right) \Big/ \sum_i \sum_{j \in J_i^c} 1.$$

Nous ne prétendrons pas ici que ce critère était le meilleur, mais il était celui qui correspondait le mieux à la sensibilité de notre Compagnie et de l'Assurance Automobile française qui, par la pratique du bonus, a éliminé le coût moyen des gros sinistres comme critère de jugement, polarisant son action contre les excès de fréquence. Eliminer le coût réel des matériels ne nous a pas par contre semblé judicieux pour deux raisons:

(a) cela aurait introduit un trop gros biais avec une formule d'écrêtement

(b) les agents, par l'exercice rapide des recours, et une bonne connaissance du marché local peuvent avoir une influence, légère mais réelle, sur les coûts moyens des petits matériels.

III. LE TRAVAIL ENTREPRIS

Les pages qui vont suivre vont décrire l'étude à laquelle nous nous sommes livrés en 1979 pour le compte de notre Direction Automobile.

Nous cherchions au départ, en partant du support réel de nos agences, à déterminer quelle formule et quel niveau d'écrêtement se rapprochaient le plus du critère optimal que nous avons décrit plus haut.

Ce critère en effet, ne pouvait rester que théorique: Les esprits n'auraient pas été prêts à l'accueillir, et sur le plan pratique, il était très difficile à réaliser à l'état permanent. Nous étions tenus par deux contraintes:

1. les données dont nous disposions,
2. notre marge de négociation.

Le point 2 nous imposait déjà une solution d'écrêtement parce qu'elle avait déjà fait son chemin dans les esprits et nous recommandait la voie la plus simple, si possible une priorité fixe et une répartition en primes. Le point 1 nous obligeait à faire l'inventaire de nos données statistiques.

A. Inventaire Statistique

La branche Automobile dispose d'un grand nombre de ventilations statistiques: elle est d'ailleurs tenue par la réglementation de fournir des états faisant ressortir, pour de très nombreuses catégories, les ratios S/P (Sinistres/Primes acquises) recalculés à chaque clôture d'exercice pour tous les exercices antérieurs (le taux n'étant absolument définitif que vers la 6ème année mais déjà très stabilisé après la 3ème). On peut donc distinguer les résultats des différentes garanties (responsabilité civile, dommages) dans les différents types de véhicules (2 roues, 4 roues, camions, transports publics, etc. ...) et même dans les différents types de garantie (contrats n'ayant que la Responsabilité Civile, ayant les dommages, etc. ...).

Malheureusement, et cela se comprend aisément, étant donné le nombre de divisions que cela créerait, ces statistiques sont tenues pour l'ensemble du portefeuille et non pas agence par agence. Les seules statistiques tenues par agence portent sur l'intégralité du portefeuille, tous risques et tous véhicules compris.

Nous avons donc fait une étude plus précise de l'existant et avons demandé à l'informatique de nous constituer un double fichier d'étude.

(1) un fichier par agence nous indiquant.

• le portefeuille en nombre de véhicules:	4 roues V_i^1
	2 roues V_i^2
• le portefeuille en primes:	P_i
• le nombre de sinistres·	matériels n_i^m
	corporels n_i^c
• le montant des sinistres:	matériels S_i^m
	corporels S_i^c.

(2) un fichier trié par agence composé de tous les sinistres dépassant 10 000 F fournissant leur nature (matériel ou corporel) et leur montant.

Le travail ayant été effectué en 1979 a porté sur l'exercice 1976 qui, à défaut d'être totalement clos, reflétait déjà des résultats quasiment définitifs.

Afin de fixer les idées, nous publions ci-dessous les premiers chiffres, avant toute vérification, qui sont sortis du listage de ces fichiers:

- nombre d'agences traitées: 1548
- nombre de véhicules:

$$V = \sum v_i = 1\,051\,188$$

dont

$$V^2 = 316\,567 \quad \text{(2 roues)}$$

- primes totales:

$$P = 799{,}3 \text{ mns F} \quad \text{(mns F = millions de francs)}$$

- nombre de sinistres:

$$n = 289\,625$$

dont

$$n^c = 16\,846 \quad \text{(corporels)}$$

- montant des sinistres:

$$S = 603{,}1 \text{ mns F}$$

dont

$$S^c = 336{,}5 \text{ mns F} \quad \text{(corporels)}$$

- écrêtements:

à 10 000 F:	249,8 mns F
20 000 F:	212,1 mns F
30 000 F:	188,2 mns F
40 000 F:	170,5 mns F
50 000 F:	156,0 mns F

Sinistre maximal: 4,4 mns F.

B. Raffinage des Données

Cette étape du travail, qui est la plus ingrate, ne peut être passée sous silence lorsqu'il s'agit de décrire un cas réel.

Un premier examen a fait de suite apparaître trois séries de difficultés:

• De mauvaises codifications faisaient ressortir certains sinistres dans des agences-fantômes qui n'avaient jamais existé, ceci prouvant d'ailleurs qu'il devait exister des erreurs d'imputation, phénomène moins visible.

• Les écrêtements choisis n'étaient pas suffisamment élevés puisque, à 10 000 F, on écrêtait 75% des sinistres corporels et à 50 000 F 45%.

• Enfin, on s'est aperçu que le nombre de sinistres *matériels* dépassant les divers points d'excess n'était pas négligeable et qu'il convenait de les rapatrier en "pseudo-corporels" dans les statistiques en nombre de sinistres du premier fichier.

Afin de rendre les données exploitables, nous avons donc procédé à un travail préparatoire effectué sur un terminal individuel avec des programmes souples, ce qui était possible grâce à la petite taille du fichier décrit ci-dessus.

Cet ensemble de tâches préparatoires au calcul proprement dit s'est enchaîné selon le plan suivant:

1. lecture du fichier d'agences, vérification de séquence et éclatement par compagnies (notre statistique auto portant également sur certaines filiales)
2. préparation du fichier de calcul comportant tous les renseignements documentés et laissant la place pour recevoir les zones d'écrêtement
3. lecture du fichier de sinistres, vérification de séquence et éclatement par compagnie.
4. totalisation par agence des sinistres dépassant les 5 seuils d'écrêtement retenus (20 000, 40 000, 60 000, 80 000, 100 000) et comptage, en nombre et en montant, des "gros matériels" (dépassant 10 000 F).
5. assortiment des deux fichiers—rectification des codes erronés (agences-fantômes)—copie des totaux écrêtés dans les zones prévues—rapatriement en nombre et montant des "gros matériels" en "pseudo-corporel".
6. tri par chiffre d'affaires—répartition en 7 groupes d'agences par tranche de 240 (la 7ème étant incomplète) afin de tester les meilleurs ajustements par niveau de taille.

Ces étapes ont dû être répétées 3 ou 4 fois avant d'obtenir une totale fiabilité. Le tableau ci-dessous résume la situation avant et après raffinage

	avant raffinage voir ci-dessus	après raffinage
Nombre d'agences	1 548	1 554
Nombre de véhicules	1 051 188	1 057 018
Nombre de 2 roues	316 567	318 403
Primes totales en mns F	799,3	803,7
Nombre de sinistres	289 625	290 706
Montant des sinistres	603,1	606,5
Nombre de corporels	16 846	18 532*
Montant des corporels	336,5	363,9*

* Dont 1 601 gros matériels pour 26,2 mns F rapatriés en corporels

• surcrêtes:

20 000 F:	206,7 mns F
40 000 F:	164,7 mns F
60 000 F:	138,0 mns F
80 000 F:	122,0 mns F
100 000 F:	107,6 mns F

C. Le Module de Calcul

Ce module avait pour mission de tester quelle était, pour les 5 priorités fixes ci-contre définies, et dans chacun des 7 groupes de taille, la meilleure des méthodes envisagées à savoir

1 la répartition sur les primes
2. la répartition sur les surcrêtes
3. la répartition sur le nombre de véhicules, et
0 aucune répartition sinistres nets sur primes amputées de la prime d'excess.

Pour chaque agence on a donc calculé, à côté du brut, le S/P théorique $\hat{S}_i/P_i$ défini au chapitre II, à l'aide de la moyenne nationale m du coût moyen corporel

$$m = \frac{363{,}9 \text{ mns F}}{18\,532} = 19\,636.$$

Puis au cours de 4 calculs successifs (type 1, 2, 3, 0) on a calculé, pour chacun des 7 groupes, h le biais moyen

$$\beta^h_{t,r} = \frac{1}{\sum_{i \in h} 1} \times \sqrt{\sum_{i \in h} (J_i - J'_i)^2}$$

en posant $J_i = \hat{S}_i/P_i$ et J'_i comme le ratio précédemment exposé selon le type choisi—(S'_i/P_i ou T_i/P'_i ayant un mode de calcul différent selon le type t et le seuil d'écrêtement r, voir § I).

Pour être complet, nous ajouterons qu'en ce qui concerne le type 3 (répartition par véhicule), nous n'avons compté les 2 roues que pour une fraction de véhicules, selon une équivalence de prime qui s'est avérée être un 2 roues = 0,126 véhicule 4 roues

C'est donc la minimisation du biais de $\beta^h_{t,r}$ qui donne la traduction concrète de "meilleur type de calcul" et "meilleur niveau d'écrêtement", expressions qui, à défaut d'être définies, n'avaient pas jusque là de signification réelle.

D. Les Résultats Obtenus

Une première étude sur les 240 premières agences sur les types 1, 2 et 3, a permis d'éliminer de suite le dernier comme le fait ressortir le tableau ci-dessous (si bien que l'on n'a pas examiné les autres groupes).

Entre les deux premiers types qui semblaient à très peu de chose près équivalents, nous avons bien sûr choisi le premier pour les raisons politiques exposées ci-dessus

GROUPE No 1 BIAIS EN %

Ecrêtement	Type No 1	Type No 2	Type No 3
0	38,16	38,16	38,16
20 000	12,48	13,59	35,96
40 000	11,88	12,70	36,61
60 000	11,61	12,42	30,60
80 000	11,77	12,76	26,76
100 000	11,98	13,08	23,74

L'étude de ce type 1 dans les différents groupes d'agence a fait ressortir les biais suivants:

BIAIS EN % PRIMES EN mns F

Groupe	1	2	3	4	5	6	7
Par moyenne	127,02	70,07	51,42	39 04	27,59	17,16	5,37
Sans écrêtement	38,16	34,29	36,56	53,44	65,19	52,11	215,12
20 000	12,48	13,08	13,23	14,70	17,41	18,85	23,55
40 000	11,88	11,52	12,13	13,53	15,51	16,67	27,92
60 000	11,61	10,93	11,99	13,70	15,43	16,63	37,46
80 000	11,77	11,20	12,84	14,69	16,41	18,03	48,98
100 000	11,98	11,55	13,78	15,77	17,64	19,64	59,61

Il est de suite apparu que le groupe 7 des petites agences n'était pas ajustable mais que, ce cas mis à part, on obtenait l'optimum dans une tranche d'écrêtement comprise entre 40 et 60 000 F.

On a donc refait le calcul de type 0 (sous-crêtes rapportées aux primes nettes de l'excess) sur une "bande élargie" allant de 45 000 à 65 000 par pas de 5000. On a obtenu le résultat ci-après.

Groupe	1	2	3	4	5	6	7
Par moyenne	127,02	70,07	51,42	39,04	27,59	17,16	5,37
Sans écrêtement	38,16	34,29	36,59	53,44	55,19	52,11	215,12
45 000	11,47	8,30	8,52	9,70	11,69	11,58	49,90
50 000	11,38	8,28	8,74	10,16	12,06	12,21	50,58
55 000	11,33	8,31	9,07	10,61	12,46	12,84	51,89
60 000	11,30	8,41	9,44	11,08	12,94	13,57	53,57
65 000	11,30	8,55	9,87	11,53	13,47	14,30	55,38

On était ainsi arrivé à cette conclusion fort safisfaisante que le meilleur ajustement était obtenu en utilisant le processus le plus facile à faire adopter: prendre sur les primes un "impôt" répartissant la charge des gros sinistres et tirer le ratio d'après les sous-crêtes.

IV. CONCLUSION

Cette étude a permis de prendre une décision. La priorité de 60 000 F pour 1976 a été choisie et on l'a indexée pour les exercices suivants. Aujourd'hui, le calcul est toujours en vigueur.

Lors de l'exécution de ce travail, il n'avait été effectué sur ce sujet que peu d'écrits publiés sauf, à notre connaissance, la thèse de Mlle Hess patronnée par le Groupement Technique Accidents et déjà citée [1].

Depuis lors une étude fort complète utilisant la théorie de la crédibilité a traité à fond ce sujet [2] ainsi que d'autres papiers du même auteur [3].

Bien que ce travail n'ait pas été conçu lors de son exécution dans le cadre strict d'une application de la théorie de la crédibilité, il semble clairement que les résultats concrets obtenus en soient une illustration naturelle et qu'il pourra être repris sous cette optique sans que ses conclusions en soient modifiées.

V BIBLIOGRAPHIE

[1] HESS, MARIE-PASCALE *Analyse des techniques d'écrêtements des sinistres et application à la branche Automobile* Thèse présentée devant la faculté des Sciences et l'Institut de Statistique des Universités de Paris et devant l'Institut des Actuaires Français

[2] GISLER, A (1980) Optimum Trimming of Data in the Credibility Model *Bulletin des Actuaires Suisses* 313–325

[3] BUHLMANN, H, GISLER, A, ET JEWELL, W (1982) Excess Claims and Data Trimming in the Context of Credibility Rating Procedures *Bulletin des Actuaires Suisses* 117–147

ACTUARIAL REMARKS ON PLANNING AND CONTROLLING IN REINSURANCE

By Erwin Straub

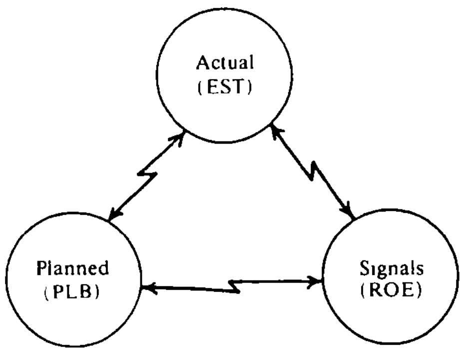
Zurich

1. SKETCH OF THE PLANNING AND CONTROLLING PROCESS IN A (RE)INSURANCE COMPANY

Planning is, or ought to be, an ever-developing process in which virtually each member of the company has to be involved. Planning without controlling, i.e., without feedback, planning on its own, is useless.

In the first part of the present note, general aspects of planning are briefly described inasfar as they are relevant to possible treatment by actuarial methods.

1.1. *"Hardware" and "Software" of Planning*



The circles in the above figure represent what could be called the hardware of planning consisting of three sets of figures, namely

- Actual figures describing the most recent reality. Most of these figures are part of what is usually called the Earnings Statement (EST).
- Forecasts for the near future, say the next three years. Most of these figures are contained in the Planning Budget (PLB).
- Control figures or signals in the sense of a "bread-line" expressing, for example, how much the company should earn as a minimum in order to remain self-financing. Such figures are calculated on the basis of so-called Return on Equity considerations and they are found in a corresponding ROE-document.

By software we mean everything done with the hardware described above, i.e., primarily the comparison of figures from different domains of the hardware on different levels within the organisation of the company, the analysis, for example, of deviations between actual and planned, the conclusions to be taken from such analyses and the planning of new actions as a consequence.

In the above graph, circles symbolize the hardware and arrows indicate the software.

Also non-numeric planning instruments (e.g., project planning, action plans, assessment and decisions) are considered to be part of the planning software.

1.2. *Earnings Statement and Planning Budget*

The Earnings Statement and the Planning Budget are the two most important numerical management accounting tools for planning and controlling a reinsurance group and/or company and/or its various profit centers. The structure of both is the same and can be sketched as follows:

Operating Result Reinsurance (Non Life) consisting of
Premiums
Underwriting Result gross
Retrocessions
Change in IBNR
Management Expenses
Standard Investment Income on technical reserves

Adding these components together—each of them to be taken with its correct positive or negative sign—yields the operating result (before tax) of the reinsurance production unit in question (e.g., a marketing department, a geographic area, a specific product of the whole company).

Looking at this operating result over a number of years we observe that it is affected by two kinds of fluctuation, namely

- cyclical variations due to pulsation of the markets and
- random variations due to the occurrence (or non-occurrence) of large claims.

While it is of vital importance to judge (past and future) cyclical variations as realistically as possible in order to be able to react both adequately and in time, random variations are of quite a different nature and therefore require a completely different statistical treatment. Such a treatment is described in Section 1.3 below.

Clearly it is not at all easy to distinguish clearly between cyclical and random fluctuations in practice because the total fluctuation of the operating result is a mixture of both. Random fluctuations appear as a kind of noise or disturbance which makes it difficult to quantify the underlying cyclical changes and trends.

1.3. *The Cat Fund Concept*

Basically random fluctuations of operating (or underwriting) results can be smoothed either by external reinsurance (or retrocession) or internally by some kind of catastrophe and large claims fund. In what follows we shall mostly speak in terms of such internal arrangements, although the inherent ideas are the same both for an actual external cover and an internal arrangement which exists only on paper.

Looking at a reinsurance company from an organizational point of view, it could be conceived as a sort of profit center hierarchy.

Example:

[PC 1] top level = level 1 = whole company

[PC 2] [PC 2] [PC 2] level 2 = e.g., geographic area

[PC 3] [PC 3] level 3 = e.g., line of business within area

We assume for the following that systematic planning and controlling is institutionalised in the company; in the present context this would mean among other things that

- an Earnings Statement (every year) and
- a Planning Budget (say every three years)

are produced for each profit center in the above figure. Clearly the Earnings Statement of a PC 2 is the sum of the Earnings Statements of all its PC 3s (and similarly for PC 1 of all its PC 2s). This sounds trivial but is nonetheless relevant in practice when there are different currencies, for example.

Whether a claim is considered as large or small depends on the size and structure of the portfolio under consideration. As a consequence of this, any mechanism designed to eliminate the "noise" must be much more efficient on the lower level profit centers (with the smaller volumes) than on top level.

This goal can be achieved in practice by working with two instruments, namely a Catastrophe Protection or Cat Cover acting on level 1 and a so-called Large Claim Compensator acting on the profit centers of level 3 defined as follows:

(i) The Cat Cover is an excess of loss arrangement where the retrocessionaire pays the excess of each claim which for the account of PC 1 exceeds a priority P (say £1 million), however not more than a certain cover amount C (say £19 million, depending on the top catastrophe exposure). This Cat Cover is either placed with external retrocessionaires or consists of a company-internal catastrophe excess of loss arrangement, a Cat Fund. In practice the whole Catastrophe Protection is often a combination of both.

(ii) The Large Claim Compensator, an internal excess of loss mechanism, is usually a lower layer to the Cat Cover, paying the excess of every claim which for the account of any PC 3 is larger than p (say £100 000) up to where the Cat Fund comes in.

Note that the point "where the Cat Cover comes in" can vary from case to case: If a large claim hits one single PC 3 only with a gross amount of £3 million, then the Cat Cover pays £2 million to PC 1 as well as to the said PC 3, and thus the Large Claim Compensator is left with an obligation of £900 000 to that same PC 3. If, however, a £3 million claim hits two PCs of level 3, say *A* with £600 000 and *B* with £2.4 million then again the Cat Fund's payment to the top PC 1 is two million, of which £400 000 to the *A*-PC 3 and £1.6 million to *B* so that the underlying Large Claim layer is left with payments of £100 000 and £700 000 to *A* and *B* respectively.

So the structure of the Earnings Statement and the Planning Budget (of any profit center from level 1 down to 3) sketched in Section 1.7 is incomplete. The full structure is rather:

Premiums	
Gross underwriting result	
Retrocessions	
Change in IBNR	
Contributions to Recuperations from	Cat Cover
Contributions to Recuperations from	Large Claim Compensator
Management Expenses	
Standard Investment Income on technical reserves	

1.4. *Return on Equity Considerations*

The purpose of calculating ROE minimum control figures is to give a quantitative answer to the following two basic questions:

(i) How much equity does a company need in order to run its business? (ER = equity required)

(ii) How much should the company earn as a minimum on its ER? (ROE minimum = minimal return on equity)

Both ER and ROE are control figures which immediately lead to further questions, such as:

(i) How does the actual (or planned) equity of the company compare with its ER?

(ii) How does the actual (or planned) overall operating result of the company compare with its ROE minimum?

The basic idea underlying the ROE calculations is the criterion that the Group company should be self-financing. This may in some cases be a very severe criterion since after all a company can only do as well as the marketplace will allow. Nevertheless, if the minimum ROE is higher than the actual result, this is, to say the least, an important piece of information to the General Management.

2 ACTUARIAL CONTRIBUTIONS TO PLANNING

Now the question is "What can an actuary do in the corporate planning department of an insurance or reinsurance company?" Below are indicated some possible answers to a few specific planning problems.

2.1. *How to Fix Cover Amounts and Priorities in the Cat Fund Arrangement*

The amount of coverage under the Cat Protection is dictated by the needs, i e., the exposures written by the company. As a rule of thumb the upper point of the catastrophe cover is in the region of 10% of the company's GNPI (underlying gross net premium income) or less.

What is a reasonable priority P?

As a guideline for fixing P under the Cat Fund we can argue that if the average operating result is x% of the GNPI then P must be much below that since, if it were of the same order of size, then one single large claim after deduction of recuperation from the Cat Cover would already destroy the entire operating result. As the latter is perhaps in the region of 1% to 5%, the priority P should be some ten times less, say 0.3% of GNPI. The same reasoning is used for fixing the priorities p under the Large Claim Compensator.

2.2. *Assessing the Cat Fund's Size and the Yearly Contribution to the Fund*

A practicable rule of thumb is to say that the fund should be able to pay a secular catastrophe claim, i.e., a catastrophe which happens in all likelihood only once in a century (such as the 1923 Tokyo earthquake or the Betsy windstorm in 1965, but also an imaginable secular man-made catastrophe which could hit the portfolio).

Another pragmatic approach is reflected in a rule of thumb of the type

$$\text{reserve} = \frac{\text{fluctuation}}{\text{premium loading} \times \text{risk willingness}},$$

the basic idea of which can be formulated as follows:

On the one hand the Cat Fund size (=reserve) must vary directly with the potential fluctuations of its claim load, while on the other hand it can be lower for higher loadings in the contribution to the fund and the more one is prepared to accept that the fund may be exhausted (=risk willingness). Intuitively this sort of connection between the above four items is pretty obvious; no actuarial model is needed to see this.

When it comes to quantifying things like "fluctuation" or "risk willingness" we cannot, however, do without a risk theoretical model. Considering the most simple actuarial model of insurance being a reservoir with steadily inflowing premiums and stochastically outflowing claims and putting equality in Cramér's inequality one obtains the following general formula of which the above mentioned rule of thumb is but a special case, namely

$$\frac{E[Z]}{U} \cdot v[Z] \cdot \left(-\frac{\ln \varepsilon}{2}\right) = \frac{E[\tilde{Y}]/E[Z]}{\operatorname{Var}[\tilde{Z}]/\operatorname{Var}[Z]}$$

of which a sketch of a proof is given under 2.6 below and where U is the reserve, Z stands for the yearly gross claims load of any insurance company or portfolio under consideration, $E[Z]$ its expected value i.e., the pure gross risk premium, $v[Z] = \operatorname{Var}[Z]/E^2[Z]$ which is the square of the coefficient of variation of Z, i.e., a measure for fluctuation, ε is the tolerated probability of ruin, $\tilde{Y}$ and $\tilde{Z}$ are the net yearly result and claims load respectively under some arbitrary reinsurance cover.

Putting $\tilde{Y} = Y$ and $\tilde{Z} = Z$ in the above formula yields

$$\frac{U}{E[Z]} = \frac{v[Z]}{(E[Y]/E[Z])(-2/\ln \varepsilon)}$$

or in other words

$$\text{reserve} = \frac{\text{fluctuation}}{\text{premium loading} \times \text{risk willingness}}$$

with

$$\text{“reserve”} = \frac{U}{E[Z]}$$

i.e., the initial reserve U is to be expressed as a multiple of $E[Z]$

$$\text{“premium loading”} = \frac{E[Y]}{E[Z]} = \text{profit margin}$$

since if $Y = P - Z$ with premiums $P = (1+\delta)E[Z]$ then $E[Y]/E[Z] = \delta$

$$\text{“risk willingness”} = \frac{-2}{\ln \varepsilon}$$

where ε denotes the probability with which we allow the fund to be exhausted at some future time.

2.3. *Breakdown of Overall Risk Capital on Subportfolios*

Risk capital (sometimes also called contingency or fluctuation reserves), catastrophe reserves and solvency requirements—though fitting different purposes and/or looked upon from different standpoints—always pose the same two problems for the actuary: the assessment of an appropriate overall reserve and the question of how to find the "right" distribution of the latter over a number of subportfolios or profit centers. For a solution of the second of these two problems put again

$$\tilde{Y} = Y \quad \text{and} \quad \tilde{Z} = Z$$

in the above mentioned general formula, thus

$$\frac{E[Z]}{U} \cdot v[Z] \cdot \left(-\frac{\ln \varepsilon}{2} \right) = \delta$$

or

$$U = v[Z] \cdot \left(-\frac{\ln \varepsilon}{2} \right) \cdot \frac{1}{\delta} \cdot E[Z].$$

If now the entire portfolio is subdivided into a number of subportfolios $j = 1, 2, \ldots, N$ with totals of claims Z_j, we calculate U_j according to

$$U_j = v[Z_j] \cdot \left(-\frac{\ln \varepsilon'}{2} \right) \cdot \frac{1}{\delta_j} E[Z_j]$$

where $\varepsilon' =$ constant, i.e., independent of j, which means that we consider a distribution of the total reserve over the subportfolios as fair if each subportfolio has the same ruin probability ε'.

Of course $\sum_{j=1}^{N} U_j = U$, i.e.,

$$\ln \varepsilon' \cdot \sum_{j=1}^{N} v[Z_j] \frac{E[Z_j]}{\delta_j} = \ln \varepsilon \cdot v[Z] \frac{E[Z]}{\delta}$$

which determines the common ruin probability ε'.

$$\ln \varepsilon' = \ln \varepsilon \cdot \frac{\dfrac{\text{Var}[Z]}{\delta E[Z]}}{\sum_{j=1}^{N} \dfrac{\text{Var}[Z_j]}{\delta_j E[Z_j]}}.$$

Assuming non-correlated Z_j thus yields

$$\frac{-\ln \varepsilon'}{-\ln \varepsilon} = \frac{\dfrac{\sum \text{Var}[Z_j]}{\sum \delta_j E[Z_j]}}{\sum \dfrac{\text{Var}[Z_j]}{\delta_j E[Z_j]}}$$

where the right-hand side is always less than one because of

$$\sum a_j b_j < \sum a_j \sum b_j$$

since for nonnegative a_j and b_j one has $a_j b_j \dot{<} a_j \sum b_k$ and by summing over j one gets $\sum a_j b_j < \sum a_j \sum b_k$.

2.4. "*Extending*" *Scarce Statistical Materials*

A main difficulty to be overcome when assessing Cat Cover premiums is, for example, the fact that we only possess scarce statistical data as a rule. Instead of—or even better, parallel to—a parametric model approach it is sometimes useful to proceed pragmatically by combining underwriting judgement with scarce statistics as follows.

1st step: take the statistics available, say of the last five years, of claims exceeding the priority of the catastrophe fund.

2nd step: add the same five years statistics to it but with a built-in artificial windstorm claim which, we believe, is likely to occur every ten years.

3rd step: repeat the ten years statistics obtained in this way by building in an additional big fire and a catastrophe air crash.

4th step: repeat the above twenty years' statistics and add a severe windstorm with a return period of forty years.

5th step: the 40 years are again doubled and reinforced by a secular earthquake catastrophe.

2.5. *Quantifying the Change in IBNR*

Clearly if the organization is such that the component "change in IBNR" of the Earnings Statement is considered to be assessed by the actuary working in the Planning section to some extent, then there are a number of different methods at his disposition. Instead of describing them here even only sketchwise we refer to the excellent monograph "Loss Reserving Methods", Issue No 1 of Surveys of Actuarial Studies prepared by the Nationale-Nederlanden N.V., The Netherlands.

2.6. *Derivation of the Above Used Formula*

We merely indicate here the main steps of a proof of the general formula

$$\frac{E[Z]}{U}\cdot v[Z]\cdot\left(-\frac{\ln\varepsilon}{2}\right)=\frac{E[\tilde{Y}]/E[Z]}{\mathrm{Var}[\tilde{Z}]/\mathrm{Var}[Z]}$$

used before.

Cramér's inequality says that if ε denotes the ruin probability and $\tilde{Y}$ the net result of the portfolio under consideration then

$$\varepsilon\leqslant e^{-RU}$$

with U = initial reserve and R = solution of $1=E[e^{-R\tilde{Y}}]$.

Putting equality in Cramér's inequality and taking logarithms on both sides yields

$$-\frac{\ln\varepsilon}{U}=R$$

where R is the positive solution of $\varphi(R)=\ln E[e^{-R\tilde{Y}}]=0$.

Taking only the first two terms of the expansion of $\varphi(R)$ we obtain

$$-RE[\tilde{Y}]+\frac{R^2}{2}\operatorname{Var}[\tilde{Y}]=0 \quad \text{i.e. } R=\frac{2E[\tilde{Y}]}{\operatorname{Var}[\tilde{Y}]}$$

and therefore

$$-\frac{\ln \varepsilon}{2U}=\frac{E[\tilde{Y}]}{\operatorname{Var}[\tilde{Y}]}, \quad \text{where } \tilde{Y}=\text{net result}=\tilde{P}-\tilde{Z}$$

Multiplying both sides by $\operatorname{Var}[Z]$ and dividing them by $E[Z]$ yields (realizing that $\operatorname{Var}[\tilde{Y}]=\operatorname{Var}[\tilde{Z}]$)

$$q=\frac{E[Z]}{U}\cdot v[Z]\cdot\left(-\frac{\ln \varepsilon}{2}\right)=\frac{E[\tilde{Y}]/E[Z]}{\operatorname{Var}[\tilde{Z}]/\operatorname{Var}[Z]}$$

where the left-hand side (which we denote by q = security factor) does not depend on the type of reinsurance (because no "˜" occurs), contrary to the right-hand side.

Interpretation of individual terms:

$\frac{U}{E[Z]}$ = initial reserve in "natural" money units $E[Z]$

$v[Z]=\frac{\operatorname{Var}[Z]}{E^2[Z]}$ = square of the coefficient of variation of Z

$\frac{E[\tilde{Y}]}{E[Z]}$ = expected net result in natural money units

$\frac{\operatorname{Var}[\tilde{Z}]}{\operatorname{Var}[Z]}$ = some sort of reciprocal measure of the efficiency of the applied reinsurance programme.

In line with the intuition that the security factor decreases with increasing initial reserves, decreasing fluctuations of the gross result and increasing tolerated ruin probability (the latter being a measure of the risk aversion).

ASTIN BULLETIN

VOLUME 14—1984

ASTIN BULLETIN

INDEX TO VOLUME 14

INSTRUCTIONS TO AUTHORS

1. Papers for publication should be sent in triplicate to:

   Peter TER BERG, Hildebrandhove 38,
   2726 AW Zoetermeer, Netherlands.

   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!