THE ASTIN BULLETIN

INTERNATIONAL JOURNAL FOR
ACTUARIAL STUDIES
IN NON-LIFE INSURANCE AND RISK THEORY

VOL. 10, PART 1

MAY 1978

CONTENTS

* * *

PRINTED IN THE NETHERLANDS

PROBLEMS IN THE ECONOMIC THEORY OF INSURANCE

KARL BORCH

1. INTRODUCTION

1.1 More than ten years ago I wrote a paper with the title "The Economic Theory of Insurance" [6]. I was not particularly happy about this paper, and I do not think it contributed much to the development of a sastifactory theory. The paper did however make me—and I hope some readers—acutely aware of the difficulties and problems which must be overcome before a proper theory can be constructed. These problems are still unsolved, so I have on the present occassion chosen a more modest title for a paper on substantially the same subject.

1.2. Insurance is an economic activity of some importance, and there is an obvious need for a theory to explain and analyse the activity in the insurance sector of the economy. During the last decade many economists seem to have felt the need, and to have taken it as a challenge. The results have been a fair amount of research, and a number of publications, which I shall not try to review here. It may however be useful to refer to three very recent survey articles by Farny [10], Ferry [11] and Rosa [14], which give extensive bibliographies. The three articles seem to indicate that the economics of insurance is becoming a fashionable subject of research.

2. A FEW HISTORICAL NOTES

2.1. Most economists have realised that insurance is important and interesting, even if they were unable to develop an adequate theory for this particular economic activity. The classical paper by Bernoulli [3] contains several references to insurance problems, and Adam Smith's [15] remarks about insurance are often quoted. He observed that the profit of insurance companies was modest, compared to the profits made by organizing lotteries. This observation implies that the inclination to gamble in some way must be stronger than the risk aversion in the economy as a whole.

2.2. An early attempt at a systematic analysis of the problems which are central in insurance is found in Böhm-Bawerk's first book [4], actually his thesis, or "Habilitationsschrift". In this book he considers what we today would call "conditional claims".

If your property is stolen, you have the right to recover it, if the police should catch the thief. Böhm-Bawerk studied the value one should attach to such rights. It is curious that it never seemed to occur to him that insurance companies, as a matter of routine, would have to evaluate such rights. If he had seen the connection, Böhm-Bawerk might well have become the first student of the IBNR-problem.

2.3. There were other Austrians who were intrigned by the problems in economic theory which were suggested by insurance. In a paper presented to the 6th International Congress of Actuaries in Vienna, Tauber [16] suggested that reinsurance premiums should be determined as equilibrium prices in a market where conditional claims (Ansprüche) were bought and sold. Beyond presenting this idea, he did not contribute much to the development of an economic theory of insurance, apparently because he, like many actuaries of his generation, became too fascinated by his own mathematical manipulations.

A more remarkable contribution was made by another Austrian Lindenbaum [12], who argued that the theory of insurance must be based on the "supply of security" (Sicherheitsangebot) and the "demand for risk" (Risikennachfrage). The paper was however published in 1932, and we may assume that economists in the following years were preoccupied with other problems. In any case, nobody seems to have followed up the ideas of Lindenbaum, and his paper is virtually forgotten.

2.4. In America an attempt at developing a complete theory of insurance was made by Willett [17] at the beginning of this century. His book is in many ways remarkable, but it seems somehow out of touch with the contemporary economic theory, and this may be why it has not inspired other economists to continue Willett's research. The same remarks can be applied to the book by Pfeffer [13], published 55 years later, which also seems to have had little influence on research in the two following decades.

It is probably fair to say that the present interest in the economics of insurance springs from the theory of the economics of uncertainty which has been developed during the last twenty years. The pioneering work in this field is certainly Arrow's paper from 1952 [2]. This short elegant paper does really contain an economic theory of insurance as a special case. In the following sections we shall do little more than discussing this special case in some detail.

3. Insurance and Market Equilibrium

3.1. It is convenient to begin this section with a brief restatement of the classical theory of markets of pure exchange.

We consider a market of m persons and n goods. In the initial situation person i holds an amount x_{ij} of good j. Hence the initial allocation is described by a matrix $\{x_{ij}\}$. The persons exchange goods among themselves, and arrive at a final allocation described by the matrix $\{y_{ij}\}$.

If goods are neither produced nor destroyed during the exchanges, the following "conservation" condition must be satisfied

$$\sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{m} y_{ij} \qquad \text{for } j = 1, 2, \ldots, n. \tag{1}$$

It is usually assumed that all exchanges have to take place at market prices, so that the market value of a person's holdings of goods does not change during the transactions. This assumption gives the condition

$$\sum_{j=1}^{n} p_j x_{ij} = \sum_{j=1}^{n} p_j y_{ij} \qquad \text{for } i = 1, 2, \ldots, m \tag{2}$$

where p_j is the price of good j.

The behavioral assumption leading to condition (2) is of course very restrictive. It rules out free bargaining and negotiations over the exchange of goods.

Further it is usual to assume that the preferences of person i can be represented by a utility function

$$u_i(y_{i1}, \ldots, y_{in}) = u_i(y_{i\cdot}) \qquad i = 1, 2, \ldots, m. \tag{3}$$

This assumption is not completely trivial. It implies complete selfishness, in the sense that a person will only consider "his own row" when he evaluates an allocation matrix.

3.2. With these assumptions, person i will maximize (3) subject to condition (2)—his "budget equation". This problem can be solved for any n-tuple of prices. The conditions (1) must however also be satisfied, and this will make it possible to determine the prices. Hence under reasonable assumptions about the shape of the utility functions, we obtain a solution, consisting of a final allocation $\{y_{ij}\}$, and an n-tuple of equilibrium prices. This solution is usually called "competitive equilibrium". The final allocation in this solution is *Pareto optimal,* i.e. there exists no other allocation $\{\bar{y}_{ij}\}$ such that

$$u_i(\bar{y}_{i\cdot}) \geq u_i(y_{i\cdot})$$

with at least one strict inequality.

It is easy to see that the set of Pareto optimal allocations can be found by maximizing

$$\sum_{i=1}^{m} k_i u_i(y_{i\cdot}) \tag{4}$$

subject to condition (1). Here $k_1, \ldots, k_m$ are arbitrary positive constants. Since the maximand (4) is homogeneous in the k's, it follows that the set of Pareto optimal allocations is a manifold of m-1 dimensions.

We get a single element in this set if we impose the behavioral assumptions behind the conditions (2), i.e. if we assume that all exchanges have to take place at equilibrium prices, and that each person has to satisfy his budget equation.

3.3. If we want to adapt this model to insurance, it is natural to assume that in the initial situation person i is exposed to a risk which can cause him a loss, represented by a stochastic variable x_i, with the distribution $F_i(x)$. It is natural to assume that $F_i(x_i)$ is the marginal distribution of a joint probability distribution $F(x_1, \ldots, x_m)$.

If the attitude to risk of person i is represented by the utility function $u_i(x)$, his expected utility in the initial situation will be

$$\int_0^\infty u_i(-x)\, dF_i(x).$$

In some cases it is convenient to replace this expression with

$$\int_0^\infty u_i(S_i - x)\, dF_i(x)$$

where S_i is interpreted as the "initial wealth" of person i.

In the model we have outlined, we can assume that the m persons exchange risks among themselves. There is however no natural units of risk, to which prices can be assigned, so it seems a little artificial to analyse the situation as a classical market of pure exchange.

3.4. It seems more natural to assume that the m persons in some way will negotiate their way to some risk-sharing arrangement. A general arrangement of this kind is defined by m functions

$$y_i(x_1 +, \dots + x_m) = y_i(x) \qquad i = 1, 2, \dots, m$$

where $y_i(x)$ is the amount to be contributed by person i, if the sum of individual losses is x. Since the model is closed so that all losses have to be born by the group of m persons, we must have

$$\sum_{i=1}^{m} y_i(x) = \sum_{i=1}^{m} x_i = x. \tag{5}$$

It can be shown [5] that the set of Pareto optimal risk-sharing arrangements is given by the m-tuple of functions $y_i(x)$ which satisfy the condition (5) and

$$u_i'(y_i(x)) = k_i u_1'(y_1(x)) \qquad i = 1, 2, \dots, m. \tag{6}$$

Here $k_1 = 1$, and $k_2, \dots, k_m$ are arbitrary positive constants. This result is valid only if all utility functions are increasing and concave, i.e. if $u_i'(.) >$ and $u_i''(.) < 0$.

3.5. The y-functions which represent Pareto optimal arrangements will usually have a complicated form. It can be shown [7] that they will be linear, i.e.

$$y_i(x) = a_i x + b_i$$

only if the utility functions of all persons belong to one of the following three classes

(i) $u_i(x) = (x - c_i)^\alpha$

(ii) $u_i(x) = \log(x - c_i)$

(iii) $u_i(x) = 1 - e^{-\alpha_i x}$.

Positive linear transformations of these functions will of course give the same results, since $u(x)$ and $w(x) = Au(x) + B$, with $A > 0$ represent the same preference ordering over any set of probability distributions.

Any of these three classes seems too narrow to give room for the different individual attitudes to risk which one would expect to find in the real world. The classes (i) and (ii) imply that all persons have the same basic attitude to risk. Differences in preferences are such that they can be explained by differences in "initial wealth". Class (iii) gives room for differences in risk aversion, but implies that preferences are independent of initial wealth.

3.6. In practice it does not often happen that a group of people negotiate a scheme for sharing risks, i.e. create their own insurance arrangement. The institutions in the real world which come closest to our model, may be the P & I clubs, which can be seen as rather exclusive mutual insurance companies, created by ship owners. The risk sharing in most P & I clubs is virtually linear, and this may for all practical purposes be a Pareto optimal arrangement. It is not unreasonable to assume that members of the club have similar preferences, and that these preferences can be represented approximately by utility functions in one of the three classes in the preceding paragraph.

Most persons who want to participate in a risk-sharing arrangement will have to go to an insurance company. Usually the company will offer a fair, but limited choice of standard insurance contracts, and people choose according to their preferences. In this way a risk-sharing arrangement is created between customers of the company, and if the company has share holders, they will also participate in the arrangement. Through exchange of reinsurance between companies, the arrangement can be extended until it becomes virtually universal. It seems however unlikely that a risk-sharing arrangement built up in this way should satisfy conditions (5) and (6) in para 3.4 and be Pareto optimal.

3.7. These considerations lead us to our main point. Economic theory gives us some information about the form of optimal risk-sharing arrangements in an idealized world represented by our model. The practical question is then if it is possible to get reasonably close to an optimum through the existing framework of insurance institutions. If the risk-sharing arrangements which we observe in the real world seem far from any optimum, we should examine if this necessarily must be so. If the answer is in the negative, we should study the possibility of reaching better arrangements through institutional changes, or changes in insurance practice.

I do not propose to answer such far-reaching questions in this paper. Instead we shall examine some of the assumptions behind the theoretical results derived in the preceding paragraphs.

4. Insurance and the Assumptions in Economic Theory

4.1. In the classical market model it is fairly safe to assume that a person has a preference ordering over collections of goods, and that this ordering can be represented by a utility function. When uncertainty is introduced, it may be slightly more risky to assume that a person has a consistent preference ordering over a set of probability distributions. If we make this assumption, the existence of a utility function follows, and the objective of the person will be to maximize expected utility. It is, however, easy to construct simple examples which throw doubt upon this assumption.

4.2. Consider a person with an initial wealth S, which includes an asset worth A, which can be lost with a probability p. Assume that he can obtain insurance against the loss of the asset in the following form: If he pays a premium kP to an insurance company, he will receive a compensation kA if the asset is lost. His problem is then to determine the optimal value of k.

For an arbitrary value of k, the expected utility is

$$U(k) = (1 - p)\, u(S - kP) + pu(S - kP - A + kA).$$

The first derivative is

$$U'(k) = - (1 - p)\, Pu'(S - kP) + p(A - P)\, u'(S - kP - A + kA)$$

and we find

$$U'(1) = \{pA - P\}\, u'(S - P).$$

If $P = pA$, i.e. if the premium is equal to the expected compensation, we have $U'(1) = 0$. Normally the premium is loaded, so that we have $P > pA$, and $U'(1) < 0$.

It is easy to show that $U''(k) < 0$, provided that $u''(x) < 0$, i.e. if the person has risk aversion. Hence, if the premium is loaded, the person will not find it optimal to take full insurance cover.

4.3. The conclusion we have reached above seems to be contradicted by observations. A person may decide not to insure some of his assets. If however he decides to take insurance, he will usually insure the asset for its full value. We would be surprised if we observed that a person deliberately insured his house, car or baggage for, say 60% of its value.

Such observations from "household" insurance may not be conclusive. The consumer does not always behave as rationally as assumed in economic theory. "Impulse buying" is a well known concept in the theory of marketing, even if it has no place in the model which was outlined in Section 3. It seems however that we can observe the same effect in corporations where we must assume that insurance decisions are made after careful considerations.

Fire insurance on industrial plant is usually written for the full value.

In ocean hull and hull interest insurance we may find arrangements which seem to imply a deliberate under-insurance, and hence may be consistent with the theoretical results we have derived. These cases are however difficult to judge, since the market value of a ship may bear little relation to the loss which the owner will suffer if the ship is lost.

4.4. In the example above we assumed proportionality between premium and compensation. This may be realistic, but it is clearly an unnecessary restriction on the choice offered to the customer. As a more general example consider a person exposed to a risk represented by the probability distribution $F(x)$, and assume that he by paying an insurance premium $P(y)$, will be entitled to a compensation $y(x)$, if the loss amounts to x.

We shall further assume that

$$P(y) = (1 + \lambda) \int_0^\infty y(x)\, dF(x).$$

This means that the premium is proportional to the expected compensation, with λ as the loading factor.

Let S stand for the initial wealth of the person considered. For a given functional $P(y)$, his problem is then to determine the function $y(x)$ which maximizes the expected utility

$$\int_0^\infty u(S - P(y) - x + y(x))\, dF(x).$$

This problem was first formulated by Arrow [1], who showed that the solution is of the following form

$$y(x) = 0 \qquad \text{for } x < D$$
$$y(x) = x - D \qquad \text{for } x > D.$$

Under this contract the insured will carry all losses smaller than the deductible D, and all excesses will be completely covered by the insurance company.

4.5. Arrow's result appears as a special case of the Pareto optimal risk-sharing arrangements presented in para 3.4, if the insurance company is risk neutral. If the customer has preferences represented by the usual concave utility function $u_1(x)$, and if the company's utility function is linear, i.e. $u_2(x) = ax + b$, the optimal risk-sharing arrangement is given by the functions

$$y_1(x) = x \quad \text{and} \quad y_2(x) = 0 \qquad \text{for } x < D$$

and

$$y_1(x) = D \quad \text{and} \quad y_2(x) = x - D \quad \text{for } x > D.$$

This result should have considerable interest. It shows that a simple and frequently used insurance contract can bring about a Pareto optimal arrangement. Before jumping to conclusions we should however scrutinize the two assumptions which led to this result.

(i) Firstly the arrangement will be illusory if the company should be unable to fulfill its obligations under the contract. Hence the result is valid only if the supervision is so strict that the probability of ruin is negligible

(ii) Secondly we assumed that the insurance company was risk-neutral. This cannot be correct if the company is a cedent in the reinsurance market. Hence the result can be valid only for relatively small risks, of the type that the company does not reinsure.

It seems that these two conditions often will be satisfied in the real world, and this immediately leads to a practical question. Why do not insurance companies offer a larger choice of deductibles in the insurance contracts sold to the ordinary households? For most kinds of simple property insurance there should be no serious technical difficulties involved. The rating system would however become more complicated, and this would probably make the whole risk-sharing arrangement more expensive to operate.

4.6. In most situations covered by liability insurance, there is theoretically no limit to the loss which the prospective insurance buyer can suffer. In such cases the insurance contract will however usually be drawn up so that the company's liability is limited. A similar procedure is used for many insurance contracts covering medical expenses.

This kind of insurance is not very satisfactory to the customer. It leads to the complaint that the insurance is not effective when it is most needed.

If a company is reluctant about accepting unlimited liability—against a premium with proportional loading—the company evidently has a positive risk aversion. This was explicitly assumed away in the preceding paragraph, so the argument based on Pareto optimality does no longer apply. It seems however that in many cases it should be possible to devise contracts with unlimited liability and non-proportional loading which would bring about a risk-sharing arrangement closer to an optimum than the existing methods can do.

5. Final Remarks

5.1. In economic theory the model of a pure exchange market is generalized by bringing in production. The new elements in the generalized model are:

(i) An initial endowment of input factors, described by a matrix $\{w_{ih}\}$. The interpretation is that person i owns an amount w_{ih} of input factor h. The input factors may be labour or raw materials.

(ii) An n-tuple of production facilities, described by production functions

$$x_j = f_j(w_1, w_2, \ldots) \qquad j = 1, 2, \ldots, n,$$

which define how input factors can be transformed into consumer goods.

It is usually assumed that each production facility is operated so that its profit is maximized.

Each person will then sell a part, or all of his endowment to the production facilities. He will use the proceeds, and any profits he may receive from the production facilities, to buy consumer goods.

5.2. The model we have outlined leads to a problem which can be solved. The solution will consist of: Equilibrium prices for all input factors and consumer goods, and of a matrix $\{x_{ij}\}$ describing the final allocation of consumer goods.

Elements of this model can certainly be applied to insurance, and the possibilities have been explored by a number of authors, i.a. Eisen [8] and Farny [9], and they have obtained a number of potentially useful results.

It seems however, to me at least, that insurance is essentially an exchange of risks, and that it is artificial to apply the theory of production to the design of contracts for such exchanges. Nevertheless the approach may prove fruitful. Administrative costs are high in many insurance companies, and it is important to find contract forms which are inexpensive to issue, control and fulfill. This means of course that managers of insurance companies, as managers in industry, always will have to look for ways of reducing production costs.

References

[1] Arrow, K., (1974), "Optimal Insurance and Generalized Deductibles", *Scandinavian Actuarial Journal*, pp. 1-42

[2] Arrow, K., (1953), "Le rôle de valeurs boursières pour la répartition la meilleure des risques", *Colloques Internationaux du CNRS*, XL, pp. 41-48 English translation in *Review of Economic Studies*, 1964, pp. 91-96

[3] Bernoulli, D., (1954), "Specimen Theoriae Novae de Mensura Sortis", St. Petersburg 1738. English translation in *Econometrica*, pp. 23-46

[4] Böhm-Bawerk, E.; (1881), "*Rechte und Verhältnisse vom Standpunkte der Volkswirtschaftlichen Guterlehre*", Innsbruck

[5] Borch, K., (1960), "The Safety Loading of Reinsurance Premiums", *Skandinavisk Aktuarietidskrift*, pp 163-184.

[6] Borch, K., (1967), "The Economic Theory of Insurance", *The ASTIN Bulletin*, Vol IV pp 252-264.

[7] Borch, K., (1968), "Economic Equilibrium under Uncertainty", *International Economic Review*, pp 339-347

[8] Eisen, R., (1971), "Zur Produktionsfunktion der Versicherung", *Zeitschrift fur die Gesamte Versicherungswissenschaft*, pp. 407-419

[9] Farny, D., (1965), *Produktions- und Kostentheorie der Versicherung*, Karlsruhe

[10] Farny, D., (1977), "Ansätze einer betriebswirtschaftlichen Theorie des Versicherungsunternehmens", *The Geneva Papers on Risk and Insurance*, No. 5, February, pp 9-21

[11] Ferry, C., (1977), "L'Approche Empirique de la Demande d'Assurance-Vie", *The Geneva Papers on Risk and Insurance*, No 5, February, pp. 22-34.

[12] Lindenbaum, J., (1932), "Ein Vierteljahrhundert der Bedarfstheorie der Versicherung", *Zeitschrift fur Nationalökonomie*, pp 75-99.

[13] Pfeffer, I., (1956), *Insurance and Economic Theory*, Richard Irwin

[14] Rosa, J. J., (1977), "La Demande d'Assurance Non-Vie", *The Geneva Papers on Risk and Insurance*, No. 5, February, pp. 35-42

[15] Smith, A., (1776), *The Wealth of Nations*, Edinburgh.

[16] Tauber, A., (1909), "Über Risiko und Sicherheitszuschlag", *Transactions of the Sixth International Congress of Actuaries*, Vol. 1, pp 781-842.

[17] Willett, A., (1901), *The Economic Theory of Risk and Insurance*, Columbia University Press.

RISK BEARING AND THE INSURANCE MARKET

HANS BÜHLMANN AND HANS U. GERBER

1. INTRODUCTION

Stimulated by Karl Borch's paper [3] we have tried to analyze the paper written by K. Arrow [1] in 1953. Contrary to Borch's opinion we have some doubt whether this work contains a theory of insurance as a special case. Nevertheless, it has inspired us to this note, which tries to develop a somewhat more realistic model. As a matter of fact, our development is more in the spirit of another paper by Arrow [2]. We, however, have chosen a more general setup, and we believe that our treatment is also different.

2. ARROW'S MODEL (INTERPRETED FREELY)

Arrow considers an economy of exchange with C *commodities* (labelled $c = 1, \ldots, C$) and a "world" that will be in one of S different *states* ($s = 1, \ldots, S$). The problem is to distribute the total supply of each commodity c in state s among I individuals in a Pareto-optimal fashion. According to a standard result in economic equilibrium theory every Pareto-optimal allocation can be realized by a system of perfectly competitive markets. The latter means that there are prices $\bar{p}_{sc}$ (the price for a unit of commodity c if state s occurs) and that each individual has a certain amount of money, which he then will spend to maximize his own utility. The beauty of this approach lies in its simplicity: Each individual has his own maximization problem (irrespective of the others). Thus it is enough to focus our attention on a *particular individual*. Let y denote his spendable money, let $x_{sc} \geq 0$ denote the amount of commodity c contingent to the occurrence of state s purchased, and let $V(x_{11}, \ldots, x_{SC})$ denote the "value" (or utility) of this decision.

Then the problem is to

$$\text{maximize } V(x_{11}, \ldots, x_{SC})$$

$$\text{subject to } \sum_{s=1}^{S} \sum_{c=1}^{C} x_{sc}\, \bar{p}_{sc} \leq y. \tag{1}$$

Arrow's idea is to replace this market by a two stage market. Let $q_1 > 0, \ldots, q_S > 0$ be arbitrary numbers with $q_1 + \ldots + q_S = 1$. Here q_s is the price of a security ("policy" in insurance terminology) of type s, which pays one monetary unit if state s occurs and nothing otherwise. Let p_{sc} be the price of commodity c when state s has occurred. For consistency set

$$p_{sc} = \bar{p}_{sc}/q_s. \tag{2}$$

The two decisions are now:

a) *choice of the securities*. Buy $y_s \geq 0$ securities of type $s (s = 1, \ldots, S)$ such that $\sum_{s=1}^{S} y_s q_s \leq y$.

b) *Purchase of commodities after the state s has occurred*. Let x_{sc} denote the amount of commodity c that is purchased after the state s has occurred. We must have $\sum_{c=1}^{C} x_{sc} p_{sc} \leq y_s + y - \sum_{i=1}^{S} y_i q_i$. Again, we make our decision in a) and b) to maximize the resulting utility. Obviously, this two stage problem is equivalent to the original problem (1), equivalence meaning that the same commodity bundles can be bought with the same original money amount.

From now on let us assume that the function V is of the form (according to the axioms of vonNeumann-Morgenstern)

$$V(x_{11}, \ldots, x_{SC}) = \sum_{s=1}^{S} \pi_s V_s(x_{s1}, \ldots, x_{sC}). \tag{3}$$

Here π_s is the individual's subjective probability for state s, and V_s is the utility function that applies when state s occurs. Let

$$U_s(w) = \text{maximum } V_s(x_{s1}, \ldots, x_{sC})$$
$$\text{subject to } x_{sc} \geq 0, \ \sum_{c=1}^{C} x_{sc} p_{sc} \leq w. \tag{4}$$

Thus $U_s(w)$ is the utility of w monetary units in state s. With these definitions and assumptions problem a) (optimal choice of the securities) can be isolated as follows:

$$\text{maximize } \sum_{s=1}^{S} \pi_s U_s(y + y_s - \sum_{i=1}^{S} y_i q_i)$$
$$\text{subject to } y_s \geq 0, \ \sum_{i=1}^{S} y_i q_i \leq y. \tag{5}$$

3. The Problems of Optimal Coverage

We shall study in detail the solutions of problems of the type (5). Our assumptions are as follows. a) The S utility functions $U_s(y)$ are twice differentiable, such that $U_s'(y) > 0$ and $U_s''(y) < 0$. Thus we assume that the utility functions are risk adverse. b) $q_1 + \ldots$

$+ q_S \geq 1$. If p_s is the probability that the market assigns to state s, certainly $q_s \geq p_s$. Summation over s yields the inequality above.

If $q_1 + \ldots + q_S = 1$, (as in Arrows model) we can assume that $\sum_{i=1}^{S} y_i q_i = y$ without loss of generality in (5). However, in the more interesting case where $q_1 + \ldots + q_S > 1$, this is not true anymore. This suggests that we distinguish the following two problems.

Problem A

For a fixed z, $0 \leq z \leq y$, maximize $\sum_{s=1}^{S} \pi_s U_s(y + y_s - z)$ subject to the constraints that $y_s \geq 0$ and $\sum_{s=1}^{S} y_s q_s = z$.

Problem B

Maximize $\sum_{s=1}^{S} \pi_s U_s(y + y_s - \sum_{i=1}^{S} y_i q_i)$ subject to $y_s \geq 0$, and $\sum_{s=1}^{S} y_s q_s \leq y$.

Thus in Problem A the total amount spent for premiums, z, is prescribed, while in Problem B it is variable, subject only to the upper bound y.

In either case the existence of an optimal solution is clear: The quantity to be maximized is a continuous function of the decision variables $y_1, \ldots, y_S$, which (in both cases) vary over a compact set.

4. Solution of Problem A.

Theorem 1

For any $z (0 \leq z \leq y)$ there is a unique vector $\tilde{y}_1, \ldots, \tilde{y}_S$ satisfying

(i) $\sum_{s=1}^{S} \tilde{y}_s q_s = z$, $\tilde{y}_s \geq 0$ for all s

(ii) $\frac{\pi_s}{q_s} U'_s(y + \tilde{y}_s - z) \leq K$ for all s, such that $\tilde{y}_s = 0$ whenever this inequality is strict.

This vector, and only this vector, solves problem A.

Proof

For $z = 0$, the theorem is trivially true. Hence assume $z > 0$. To show the necessity of condition (ii), consider a vector $y_1, \ldots, y_S$

for which it is violated. Then there are indices s, t such that $y_t > 0$, $y_s \geq 0$ and

$$\frac{\pi_t}{q_t} U_t'(y + y_t - z) < \frac{\pi_s}{q_s} U_s'(y + y_s - z). \tag{6}$$

Then, by increasing y_s and decreasing y_t (such that the total premium remains z) the expected utility could be increased. (Note that for this part of the proof we did not need the assumption that the utility functions are risk averse.)

The necessity (and the existence of an optimal solution) show that there is at least one vector $\tilde{y}_1, \ldots, \tilde{y}_S$ that satisfies conditions (i) and (ii) above. Let $y_1, \ldots, y_S$ be any other vector that satisfies (i). First using concavity from below of the function U_s, and then (ii), we obtain the following estimate:

$$\begin{aligned} U_s(y + y_s - z) &\leq U_s(y + \tilde{y}_s - z) + U_s'(y + \tilde{y}_s - z) \cdot (y_s - \tilde{y}_s) \\ &\leq U_s(y + \tilde{y}_s - z) + K \frac{q_s}{\pi_s} (y_s - \tilde{y}_s). \end{aligned} \tag{7}$$

Note that the first inequality is strict unless $y_s = y_s$. By summing (7) over s we see that

$$\sum_{s=1}^{S} \pi_s U_s(y + y_s - z) \leq \sum_{s=1}^{S} \pi_s U_s(y + \tilde{y}_s - z), \tag{8}$$

with a strict inequality holding unless $y_s = \tilde{y}_s$ for all s. This completes the proof of Theorem 1.

5. Solution of Problem B.

If $\sum_{s=1}^{S} q_s = 1$, solve Problem A with $z = y$. Otherwise, the following result holds.

Theorem 2

Suppose that $\sum_{s=1}^{S} q_s > 1$. Then Problem B has a unique solution, which we denote by $\tilde{y}_1, \ldots, \tilde{y}_S$. a) If $\sum_{s=1}^{S} \tilde{y}_s q_s = y$, it can be characterized by conditions (i) and (ii) in Theorem 1 with $z = y$. b) If $\sum_{s=1}^{S} \tilde{y}_s q_s < y$, it is the only vector $\tilde{y}_1, \ldots, \tilde{y}_S$ that satisfies

i) $\tilde{y}_s \geq 0$ for all s and

ii) $\frac{\pi_s}{q_s} U_s'(y + \tilde{y}_s - \sum_{t=1}^{S} \tilde{y}_t q_t) \leq \sum_{j=1}^{S} \pi_j U_j'(y + \tilde{y}_j - \sum_{t=1}^{S} \tilde{y}_t q_t)$

for all s, such that $\tilde{y}_s = 0$ whenever the inequality is strict.

Proof

a) If there is an optimal $\tilde{y}_1, \ldots, \tilde{y}_s$ with $\sum_{s=1}^{S} \tilde{y}_s q_s < y$, it has to satisfy condition (ii) above. For, if it did not, there would either be an index s such that

$$\frac{\pi_s}{q_s} U_s'(y + \tilde{y}_s - \sum_{i=1}^{S} y_i q_i) > \sum_{j=1}^{S} \pi_j U_j'(y + \tilde{y}_j - \sum_{i=1}^{S} \tilde{y}_i q_i), \tag{9}$$

in which case the expected utility could be increased by increasing $\tilde{y}_s$, or there would be an index s such that $\tilde{y}_s > 0$ and the inequality in (ii) is strict, in which case the expected utility could be increased by a reduction of $\tilde{y}_s$. (For the necessity of (ii) we again did not need the assumption that the utility functions are risk averse).

b) Suppose now that $\tilde{y}_1, \ldots, \tilde{y}_S$ is a vector that satisfies conditions (i) and (ii) of part b) in Theorem 2. Any other decision, say $y_1, \ldots, y_S$ (where $\sum_s q_s y_s = y$ is also permissible), can be compared with it as follows: For any s,

$$U_s(y + y_s - z) \leq U_s(y + \tilde{y}_s - \tilde{z}) + U_s'(y + \tilde{y}_s - \tilde{z}) \cdot (y_s - \tilde{y}_s + \tilde{z} - z)$$

$$\leq U_s(y + \tilde{y}_s - \tilde{z}) + \frac{q_s}{\pi_s} (y_s - \tilde{y}_s) \sum \pi_j U_j'(y + \tilde{y}_j - \tilde{z}) +$$

$$+ U_s'(y + \tilde{y}_s - \tilde{z}) \cdot (\tilde{z} - z), \tag{10}$$

with the convenient notation $\tilde{z} = \sum \tilde{y}_i q_i$, $z = \sum y_i q_i$. Multiplying both sides by π_s, and summing over s, we get

$$\sum_{s=1}^{S} \pi_s U_s(y + y_s - z) \leq \sum_{s=1}^{S} \pi_s U_s(y + \tilde{y}_s - z). \tag{11}$$

Furthermore, this inequality is strict unless $y_s = \tilde{y}_s$ for all s, which shows the uniqueness of any optimal solution satisfying (ii).

6. How to Find the Solutions.

To find the solution of Problem A, first relabel the states such that

$$\frac{\pi_1}{q_1} U_1'(y - z) \geq \frac{\pi_2}{q_2} U_2'(y - z) \geq \ldots \geq \frac{\pi_S}{q_S} U_S'(y - z). \tag{12}$$

Now we choose y_1 such that

$$\frac{\pi_1}{q_1} U_1'(y + y_1 - z) = \frac{\pi_2}{q_2} U_2'(y - z). \tag{13}$$

Then we increase y_1 and choose y_2 such that

$$\frac{\pi_1}{q_1} U_1'(y + y_1 - z) = \frac{\pi_2}{q_2} U_2'(y + y_2 - z) = \frac{\pi_3}{q_3} U_3'(y - z) \quad (14)$$

etc. Thus, gradually we increase the coverage, from left to right, until the total premium reaches the level z. Clearly, the resulting coverage will satisfy properties (i) and (ii) of Theorem 1.

For the further discussion, let $\tilde{y}_1, \ldots, \tilde{y}_S$ denote the optimal coverage if the premium equals z, hence

$$U(z) = \sum_{s=1}^{S} \pi_s U_s(y + \tilde{y}_s - z) \quad (15)$$

is the maximal utility at premium level z, and let $K = K(z)$ denote the upper bound in (ii) of Theorem 1. Finally, set

$$K_v(z) = \sum_{s=1}^{S} \pi_s U_s'(y + \tilde{y}_s - z). \quad (16)$$

Theorem 3

$U'(z)$ equals $K(z) - K_v(z)$ and is a non-increasing function.

Proof

Let z_1, z_2 be any two numbers, and let $\tilde{y}_s^{(i)}$ denote the optimal coverage for state s if the total premium should be z_i ($i = 1, 2$). Using the concavity from below of U_s and property (ii) in Theorem 1, we find that

$$\begin{aligned} & U_s(y + \tilde{y}_s^{(2)} - z_2) - U_s(y + \tilde{y}_s^{(1)} - z_1) \\ & \leq U_s'(y + \tilde{y}_s^{(1)} - z_1) \cdot (\tilde{y}_s^{(2)} - \tilde{y}_s^{(1)} + z_1 - z_2) \quad (17) \\ & \leq \frac{q_s}{\pi_s} K(z_1) \cdot (\tilde{y}_s^{(2)} - \tilde{y}_s^{(1)}) - U_s'(y + \tilde{y}_s^{(1)} - z_1) \cdot (z_2 - z_1). \end{aligned}$$

Multiply both sides by π_s, and summing over s, we obtain the inequality

$$U(z_2) - U(z_1) \leq (K(z_1) - K_v(z_1)) \cdot (z_2 - z_1). \quad (18)$$

By interchanging the roles of z_1 and z_2, and inverting the sign, a lower bound is obtained for $U(z_2) - U(z_1)$. Finally, assume $z_1 < z_2$.

Then these two inequalities can be written as follows.

$$K(z_2) - K_v(z_2) \leq \frac{U(z_2) - U(z_1)}{z_2 - z_1} \leq K(z_1) - K_v(z_1). \quad (19)$$

Monotonicity of $K(z) - K_v(z)$ is seen immediately from (19), and the rest of theorem 3 follows by taking the limit for $z_2 \to z_1$. Now observe the following: Let $0 \leq \tilde{z} \leq y$ be the premium spent in the optimal solution $\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_S$ of problem B (i.e. $\tilde{z} = \sum_{s=1}^{S} q_s \tilde{y}_s$). For this $\tilde{z}$ problem A must have the same solution as problem B and we conclude, that the two bounds appearing in the characterization of the solutions must be the same, hence

$$K(\tilde{z}) = K_v(\tilde{z}).$$

On the other hand theorem 3 leads to the following

Corollary

If	$K(0) \leq K_v(0)$	then	$\tilde{z} = 0$
	$K(y) \geq K_v(y)$	then	$\tilde{z} = y$

otherwise let z satisfy

$$K(z) = K_v(z) \qquad \text{then} \qquad \tilde{z} = z$$

Based on this corollary and the monotonicity of $K(z) - K_v(z)$, $0 \leq z \leq y$ one may find $\tilde{z} \neq 0$ by gradually increasing the level z of premium spent until $K(z) - K_v(z) = 0$, or if this does not happen for $z \leq y$, by putting $\tilde{z} = y$.

Note

It is sometimes more convenient, to follow the above procedure until the quotient $\frac{K(z)}{K_v(z)}$ reaches 1. To justify this alternative, we also prove that $\frac{K(z)}{K_v(z)}$ is nonincreasing $\left(\frac{K_v(z)}{K(z)} \text{ nondecreasing}\right)$ for $0 \leq z \leq y$.

Proof

Let $N = N(z)$ denote the set of indices for which $\tilde{y}_s = 0$. Then

$$K_v(z) = K(z) \left(\sum_{s \notin N} q_s \right) + \sum_{s \in N} \pi_s U_s'(y - z), \tag{20}$$

and therefore

$$\frac{K_v(z)}{K(z)} = \sum_{s \notin N} q_s + \frac{\sum_{s \in N} \pi_s U_s'(y - z)}{K(z)}. \tag{21}$$

Since $\frac{\pi_s U_s'(y - z)}{K(z)} \leq q_s$ for $s \in N$, this shows that $K_v(z)/K(z)$ is a nondecreasing function (the numerator in the last expression is a nondecreasing function, while $K(z)$ is nonincreasing).

In the following the procedure for finding the optimum in problem B is explicitly carried out for

exponential utility (Section 7)
quadratic utility (Section 8).

7. Exponential Utility

Let $U_s(x) = 1 - e^{-\alpha(x-y_s^*)}$, $U_s'(x) = \alpha e^{\alpha y_s^*} e^{-\alpha x}$. You may interpret y_s^* as the "need for money" in state s. Suppose then y sufficiently large, such that the following property holds for the optimum $\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_S$ of problem B (according to theorem 2).

$$\frac{\pi_s}{q_s} e^{\alpha \tilde{y}_s^*} e^{-\alpha \tilde{y}_s} \leq \sum_j \pi_j e^{\alpha y_j^*} e^{-\alpha \tilde{y}_j} \quad \begin{array}{l} \text{for all } s \text{, with strict in-} \\ \text{equality only allowed if} \\ \tilde{y}_s = 0. \end{array} \tag{22}$$

With the notation

$$\pi_s^* = \pi_s e^{\alpha y_s^*} \tag{23}$$

and

$$C_s(y_1, y_2, \ldots y_S) = \frac{\frac{\pi_s^*}{q_s} e^{-\alpha y_s}}{\sum_{j=1}^{S} \pi_j^* e^{-\alpha y_j}} \tag{24}$$

(22) becomes

$$C_s(\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_S) \leq 1 \quad \begin{array}{l} \text{for all } s \text{, with strict in-} \\ \text{equality only allowed if} \\ \tilde{y}_s = 0. \end{array} \tag{25}$$

Abbreviate $\hat{z}$ for $\sum_{j=1}^{S} q_j \tilde{y}_j$. (25) may hold for $z = 0$ and then $\tilde{z} = 0$. Otherwise, increasing gradually the premium level z and adjusting $y_1, y_2, \ldots y_s$ at each level z according to the solution of problem A, $\max_{s \in S} C_s$ will monotonically decrease until it reaches 1 at $z = \hat{z}$. (See *note* after theorem 3.) Observe that in the exponential case the ordering

$$C_1(y_1, y_2, \ldots y_S) \geq C_2(y_1, y_2, \ldots y_S) \geq \ldots \geq C_S(y_1, y_2, \ldots y_s)$$

never changes during this process.

Let then m be the number of states, which are insured in the optimal solution of B (number of variables $\tilde{y}_s$ different from 0 in (25)).

From (21) we have

$$K_v(\tilde{z}) = \sum_{j=1}^{s} \pi_j^* e^{-\alpha\tilde{y}_j} = K(\tilde{z}) \sum_{j=1}^{m} q_j + \sum_{j=m+1}^{s} \pi_j^*$$

and hence from the corollary of theorem 3

$$1 = \sum_{j=1}^{m} q_j + \frac{1}{K(\tilde{z})} \sum_{j=m+1}^{s} \pi_j^*$$

$$\frac{1}{K(\tilde{z})} = \frac{1 - \sum_{j=1}^{m} q_j}{\sum_{j=m+1}^{s} \pi_j^*}$$

therefore (recall $K(\tilde{z}) = \frac{\pi_s^*}{q_s} e^{-\alpha\tilde{y}_s}$ for $s = 1, 2, \ldots m$)

$$\alpha\tilde{y}_s = \log \frac{\pi_s^*}{q_s} + \log \frac{1}{K(\tilde{z})} = \tag{26}$$

$$\log \frac{\pi_s^*}{q_s} + \underbrace{\log(1 - \sum_{j=1}^{m} q_j) - \log \sum_{j=m+1}^{s} \pi_j^*}_{\Delta m}$$

for $s \leq m$.

The optimal m is found as the first index for which

$$\frac{\pi_{m+1}^*}{q_{m+1}} \frac{1 - \sum_{j=1}^{m} q_j}{\sum_{j=m+1}^{s} \pi_j^*} \leq 1 \text{ or equivalently } \log \frac{\pi_{m+1}^*}{q_{m+1}} + \Delta_m \leq 0 \tag{27}$$

It is easily checked, that this condition also applies if $m = 0$.

Numerical Examples (In all examples the exponent $\alpha = 10^{-2}$)

First example

s	1	2	3	4	5
y_s^*	1000	100	50	10	5
π_s	0.1	0.2	0.3	0.2	0.2
q_s	0.3	0.3	0.3	0.3	0.3
π_s^*	2202.65	.544	0.495	0.221	0.210
$\frac{\pi_s^*}{q_s}$	7342.16	1.813	1.65	0.737	0.7

$1 - \sum_{j=1}^{s-1} q_j$	1	0.7
$\sum_{j=s}^{S} \pi_j^*$	2204.12	1.470
check (27)	3.33	*0.863*

Hence only state 1 is insured and from (26) $\tilde{y}_1 = 815.95$
$q_1\tilde{y}_1 = 244.78.$

Second example

Insurance becomes "horribly expensive" for $s = 1$, otherwise same as in first example.

s	*1*	*2*	*3*	*4*	*5*
y_s^*	1000	100	50	10	5
π_s	0.1	0.2	0.3	0.2	0.2
q_s	*1*	0.3	0.3	0.3	0.3
π_s^*	2202.65	0.544	0.495	0.221	0.210
$\frac{\pi_s^*}{q_s}$	2202.65	1.813	1.65	0.737	0.7
$\sum_{j=s}^{S} \pi_j^*$	2204.12				
check (27)	$<$ *1*				

Hence now *no* insurance is bought at all!

Third example

The "insurance need" is eliminated in state $s = 1$, otherwise still the same as before.

s	*2*	*3*	*4*	*5*	*1*
y_s^*	100	50	10	5	0
π_s	0.2	0.3	0.2	0.2	0.1
q_s	0.3	0.3	0.3	0.3	0.3
π_s^*	0.544	0.495	0.221	0.210	0.1
$\frac{\pi_s^*}{q_s}$	1.813	1.65	0.737	0.7	0.333
check (27)	1.155	1.126	*0.555*		

Hence insurance on $s = 2$ and 3 $\tilde{y}_2 = 31.17$ $q_2\tilde{y}_2 = 9.35$
$\tilde{y}_3 = 21.75$ $q_3\tilde{y}_3 = 6.52$

8. Quadratic Utility

In this section

$$U_s(x) = \alpha(x - y_s^{**}) - \frac{(x - y_s^{**})^2}{2}\ ; \ x - y_s^{**} \leq \alpha$$

$$U_s'(x) = \alpha + y_s^{**} - x$$

The condition corresponding to (22) in Section 7 is then

$$\frac{\pi_s}{q_s}(\alpha + y_s^{**} - y - \tilde{y}_s + \sum_j q_j\tilde{y}_j) \leq \alpha + \bar{y}_s^{**} - y - \bar{\tilde{y}} + \sum_j q_j\tilde{y}_j \quad (28)$$

for all s, with strict inequality only allowed if $\tilde{y}_s = 0$

Abbreviations

$$\bar{y}^{**} = \sum_j \pi_j y_j^{**}$$

$$\bar{\tilde{y}} = \sum_j \pi_j \tilde{y}_j$$

Redefine

$$\alpha + y_s^{**} - y = y_s^* \qquad \text{and you obtain}$$

$$\frac{\frac{\pi_s}{q_s}(y_s^* - \tilde{y}_s + \sum_j q_j\tilde{y}_j)}{\bar{y}^* - \bar{\tilde{y}} + \sum_j q_j\tilde{y}_j} \leq 1 \quad (29)$$

for all s, with strict inequality only allowed if $\tilde{y}_s = 0$

Observe that as long as the numerator of the left side in (29) is positive, we are in the region where U_s' is positive. The numbering of the sides is defined in decreasing order of

$$C_s = \frac{\pi_s}{q_s}\frac{y_s^*}{\bar{y}_s^*}, \text{ hence } C_1 \geqslant C_2 \geqslant \dots \geqslant C_S \quad (30)$$

These quantities are the inital values at $y_1 = y_2 = \dots = y_S = 0$ of the functions

$$C_s(y_1, y_2, \dots y_S) = \frac{\frac{\pi_s}{q_s}(y_s^* - y_s + \sum q_j y_j)}{\bar{y}^* - \bar{y} + \sum q_j y_j} \quad (31)$$

We again gradually increase $z = \sum_j q_j y_j$ and for each z adapt $y_1, y_2, \dots y_S$ according to the solution of problem A; $\max_{s \in S} C_s$ will then again monotonically decrease to 1, but unfortunately the ordering of the $C_s(y_1, y_2, \dots y_S)$ (for those s which are not yet insured) may change! So while it is clear that insurance, if any, must always be bought on $s = 1$, we must if necessary try several combinations of other states to find out the optimum.

Numerical Examples

First example

s	*1*	*2*	*3*	*4*	*5*	
y_s^*	1000	100	50	10	5	
π_s	0.1	0.2	0.3	0.2	0.2	$\bar{y}^* = 138$
q_s	0.3	0.3	0.3	0.3	0.3	
C_s	2.415	0.483	0.362	0.048	0.024	

We try to insure state number 1 only. If this does achieve an optimum we must have

$$C_1(y_1, 0, 0, \ldots 0) = \frac{1}{3}\,\frac{1000 - y_1 + 0.3\,y_1}{138 - 0.1\,y_1 + 0.3\,y_1} = 1$$

from which we find

$$y_1 = 450.77$$

$$q_1 y_1 = 135.23$$

It remains to be checked whether $C_s(y_1, 0, 0, \ldots 0) \leq 1$ for $s \geq 2$

$$C_2(y_1, 0, 0, \ldots 0) = \frac{2}{3}\,\frac{100 + 135.23}{228.15} = 0.69$$

$$C_3(y_1, 0, 0, \ldots 0) = \frac{3}{3}\,\frac{50 + 135.23}{228.15} = 0.81 \text{ (has surpassed } C_2\text{!)}$$

As states 4 and 5 have the same probabilities and premiums as state 2 their C-values must be lower than that of state 2 also. This shows that just insuring state 1 with the above amounts is optimal.

Second example

If we change in the first example only q_1 from 0.3 to 1 (insurance on the state insured in the first example becomes "horribly expensive"), then all initial C-values drop below 1 which means that no insurance should be bought.

Third example

"Insurance need" in state 1 is eliminated (i.e. $y_1^* = 0$). Otherwise same as first example.

s	2	3	4	5	1	
y_s^*	100	50	10	5	0	
π_s	0.2	0.3	0.2	0.2	0.1	$\bar{y}^* = 38$
q_s	0.3	0.3	0.3	0.3	0.3	
C_s	1.75	1.32	0.18	0.09	0	

It is obvious that some insurance must be bought, certainly on $s = 2$ and probably also on some other states, $s = 3$ being a very likely candidate.

We try to find an optimal solution, where y_2 and y_3 are different from zero

$$C_2(y_2, y_3, 0, \ldots 0) = \frac{2}{3} \frac{100 - y_2 + 0.3(y_2 + y_3)}{38 - 0.2y_2 - 0.3y_3 + 0.3(y_2 + y_3)} = 1$$

$$\text{or } 860 - 14 y_2 + 3 y_3 = 0$$

$$C_3(y_2, y_3, 0, \ldots 0) = \frac{50 - y_3 + 0.3(y_2 + y_3)}{38 - 0.2y_2 - 0.3y_3 + 0.3(y_2 + y_3)} = 1$$

$$\text{or } 120 - 7 y_3 + 2y_2 = 0$$

$$y_2 = 69.35 \qquad y_3 = 36.96$$

$$q_2 y_2 = 20.80 \qquad q_3 y_3 = 11.09 \qquad \text{total premium} \quad 31.89$$

We must check that $C_4(y_2, y_3, 0, 0, 0) \leq 1$. This check suffices since $\frac{\pi_s}{q_s} \leq \frac{\pi_4}{q_4}$ for $s = 5, 1$ (C_5 and C_1 will then automatically be below 1).

$$\text{Check: } C_4(y_2, y_3, 0, 0, 0) = \frac{2}{3} \frac{10 + 31.89}{38 - 24.96 + 31.89} = 0.62,$$

which proves optimality.

9. Literature

[1] Arrow, K. J., (1964), The Role of Securities in the Optimal Allocation of Risk Bearing. Review of Economic Studies (translation of original paper in French, 1953).

[2] Arrow, J. K., (1974), Optimal Insurance and Generalized Deductibles Scandinavian Actuarial Journal.

[3] Borch, K., (1978), Problems in the Economic Theory of Insurance. The ASTIN Bulletin, 10, 1-11

PARETO-OPTIMAL RISK EXCHANGES AND RELATED DECISION PROBLEMS

HANS U. GERBER

1. SUMMARY

In various branches of applied mathematics the problem arises of making decisions to reconcile conflicting criteria. One example is the classical statistical problem, where a type 1 error cannot be arbitrarily reduced without increasing the probability for a type 2 error. Another example, quite familiar to actuaries, is graduation, where a compromise between smoothness and fit has to be reached This motivates the concept of Pareto-optimal decisions, which is discussed in section 2 There is a simple method, maximizing a weighted average of the scores, to obtain certain Pareto-optimal decisions. In section 3 a condition is given, which is satisfied in most applications, that guarantees that all the Pareto-optimal decisions can be found by this method This is applied in section 4, where the problem of risk exchange between n insurance companies is considered. The original model of Borch is generalized: it is assumed that some of the companies are not willing to contribute more than a certain fixed amount towards the aggregate loss of the other companies The theorem in section 4 gives a characterization of all the Pareto-optimal risk exchanges Because of the restrictions, these risk exchanges do not just depend on the combined surplus (which would amount to pooling) in general, and can be found by an algorithm. One benefit of this generalization of Borch's Theorem is that two seemingly unrelated results (optimality of a stop loss contract, and optimality of certain dividend formulas in group insurance) follow from it as special cases.

2. EVALUATION OF DECISIONS UNDER CONFLICTING VIEW POINTS

Often one is faced with the situation where a decision has to be made in the presence of several criteria. Mathematically, the problem can be formulated as follows.

Let D be the set of all possible decisions. We are given n real-valued functions $s_1(d), \ldots, s_n(d)$, $d \in D$. If $d_1, d_2 \in D$ and $s_i(d_1) \geq s_i(d_2)$, this means that decision d_1 is *better* than (or at least as good as) decision d_2 *with respect to criterion i*. Let

$$s(d) = (s_1(d), \ldots, s_n(d)), \; d \in D \tag{1}$$

and

$$S = \{x / x = s(d) \text{ for some } d \in D\} \tag{2}$$

denote the range of the "score function" $s(\cdot): D \to R^n$. A decision d_1 is said to be *strictly better* than a decision d_2, if $s_i(d_1) \geq s_i(d_2)$ for $i = 1, \ldots, n$, and if at least one of these inequalities is strict. A decision d is called *Pareto-optimal*, if there is not a decision that is strictly better than d. If R is any subset of R^n, a point $x \in R$ is called a Pareto-optimal point of R if the intersection of R with $Q_x = \{y/y_i \geq x_i, i = 1, \ldots, n\}$ consists only of the point x. Thus a decision d is Pareto-optimal, if and only if $s(d)$ is a Pareto-optimal point of S.

Under fairly general conditions (for example if S is finite, or if S is a closed region that is bounded by a plane whose normal vector points to the positive 2^n-tant) one should obviously chose a Pareto-optimal decision. However, we shall not discuss the question, *which* of the Pareto-optimal decisions should be chosen.

Example 1

In a class of k students n quizzes were given during the term. Let $s_i(d)$ denote the score of student d in quiz i ($i = 1, \ldots, n, d = 1, \ldots, k$). Who is the top student of the class? Thus $D = \{1, \ldots, k\}$, and clearly the Pareto-optimal students (and only these) are candidates for this honor.

Example 2

Consider the following statistical decision problem: Population i has a *pdf* $f(x; i)$, $i = 1, \ldots, n$. Given an observation, say X, the statistician tries to name the underlying population. Thus D consists of all "tests" (see [5] for example). It is convenient to allow randomized tests. Then a test δ is defined by n non-negative functions $p_1(x), \ldots, p_n(x)$ with $p_1(x) + \ldots + p_n(x) = 1$ for all x. This means that the statistician, having observed X, will name population i with probability $p_i(X)$. Let

$$s_i(\delta) = \int p_i(x) f(x; i)\, dx \tag{3}$$

be the probability for a correct guess if the observation originates from population i, $i = 1, \ldots, n$. Clearly, the statistician wants to select a test that is Pareto-optimal.

Example 3

Consider the Whittaker-Henderson Problem. Given are m ungraduated values, say $v_1, \ldots, v_m$. A decision is the choice of m graduated values, $u = (u_1, \ldots, u_m)$. Thus $D = R^m$ in this example.

Let

$$F(u) = \sum_{i=1}^{m} w_i (u_i - v_i)^2 \tag{4}$$

be a measure for "fit", where $w_1, \ldots, w_m$ are certain weights, and let

$$S(u) = \sum_{i=1}^{m-z} (\Delta^z u_i)^2 \tag{5}$$

be our measure for "smoothness", where $z < m$ is some integer, see [6]. Here $n = 2$, $s_1(u) = -F(u)$, $s_2(u) = -S(u)$, and we want to find graduated values that are Pareto-optimal in this sense.

The most important example (at least as far as this paper is concerned) will be discussed in section 4.

3. How to Find Pareto-Optimal Decisions

Certain Pareto-optimal decisions can be found by the following method: chose n positive numbers $k_1, \ldots, k_n$ and try to maximize the linear combination

$$\sum_{i=1}^{n} k_i s_i(d),\ d \in D. \tag{6}$$

For, if a decision $\tilde{d}$ has the property that there are positive constants $k_1, \ldots, k_n$ such that

$$\sum_{i=1}^{n} k_i s_i(d) \leq \sum_{i=1}^{n} k_i s_i(\tilde{d}) \tag{7}$$

for all $d \in D$, it is obviously Pareto-optimal.

In *Example 1* above this method amounts to assigning certain weights to the n quizzes, and (based on this) to determine the student(s) with the highest (weighted) average score.

In *Example 2* let

$$M(x) = \max\{k_i f(k;\, i)/i = 1, \ldots, n\}, \tag{8}$$

and let $\tilde{\delta}$ be a test, described by $\tilde{p}_1(x), \ldots, \tilde{p}_n(x)$, such that

$$\tilde{p}_i(x) = 0 \text{ whenever } k_i f(x;\, i) < M(x), \tag{9}$$

$i = 1, \ldots, n$. Thus $\tilde{\delta}$ consists of naming the population (or one of the populations), for which the maximum is attained in formula (8).

Then if δ is another test, given by $p_1(x), \ldots, p_n(x)$,

$$\begin{aligned} \sum_{i=1}^{n} k_i s_i(\delta) &= \sum_{i=1}^{n} \int k_i p_i(x)\, f(x;\, i)\, dx \\ &\leq \sum_{i=1}^{n} \int M(x) p_i(x)\, dx \\ &= \int M(x)\, dx = \sum_{i=1}^{n} k_i s_i(\tilde{\delta}). \end{aligned} \tag{10}$$

Hence a test $\tilde{\delta}$ of this form is Pareto-optimal. Note that the inequality is strict unless δ satisfies condition (9) too.

In *Example 3* the vector $\tilde{u}$ which minimizes $k_1F(u) + k_2S(u))$ is found as the solution of a certain matrix equation, see [6].

The question arises whether all the Pareto-optimal decisions can be obtained by this method. In general, the answer is no. Consider Example 1 with a class of just three students. Suppose the scores in 2 quizzes were (6, 1) for student A, (3, 3) for student B, and (1, 6) for student C. Obviously, all 3 students are Pareto-optimal. But only students A and C can be obtained by the above method.

However, if S is a closed convex region, all the Pareto-optimal points and decisions can be obtained by this method: if $\tilde{d}$ is a Pareto-optimal decision, inequality (7) holds for all $d \in D$, where $(k_1, \ldots, k_n)$ is a vector that is perpendicular to the (or a) plane that is tangent to S at $x = s(d)$. A convenient way to verify convexity of S is to show that for any two points x_0, $x_1 \in S$, the line segment $\{x/x = rx_1 + (1 - r)x_0, 0 < r < 1\}$ is contained in S. The validity of this condition can be easily seen in *Example 2*: if δ_0, δ_1 are any two tests, define a test δ_r $(0 < r < 1)$, which consists of using δ_1 with probability r and δ_0 with probability $1 - r$. Then, by the law of total probability,

$$s_i(\delta_r) = rs_i(\delta_1) + (1 - r)s_i(\delta_0) \tag{11}$$

$(i = 1, \ldots, n)$. Hence all the Pareto-optimal tests are of the form (9), which is essentially the content of the *lemma of Neyman-Pearson*, see [5] for example.

Often it is possible to show the validity of the following condition (which may hold even if S is not convex).

Condition C. For any two decisions d_0, $d_1 \in D$ there is a family of decisions $d_r \in D$, $0 < r < 1$, such that

$$s_i(d_r) \geq rs_i(d_1) + (1 - r)s_i(d_0) \tag{12}$$

for $i = 1, \ldots, n$.

If S is closed and if Condition C is satisfied, all the Pareto-optimal points and decisions can be obtained by the method described at the beginning of this section: Condition C implies that the set of Pareto-optimal points on S coincides with the set of Pareto-optimal points on the convex hull of S. In *Example 3* the validity of Condition C can be verified as follows. If $u^{(0)}$, $u^{(1)}$ are two vectors of graduated values

$$u^{(j)} = (u_1^{(j)}, \ldots, u_m^{(j)}), j = 0,1, \tag{13}$$

set $u^{(r)} = ru^{(1)} + (1 - r)u^{(0)}$. Then one uses the inequality

$$(ra + (1 - r)b)^2 \leq ra^2 + (1 - r)b^2, \quad 0 < r < 1, \tag{14}$$

which is valid for any two numbers a and b, to show that

$$F(u^{(r)}) \leq rF(u^{(1)}) + (1 - r)F(u^{(0)}) \tag{15}$$

and

$$S(u^{(r)}) \leq rS(u^{(1)}) + (1 - r)S(u^{(0)}). \tag{16}$$

Therefore, all the Pareto-optimal graduated sets are obtained by the usual Whittaker-Henderson procedure, i.e., minimizing $k_1F(u) + k_2S(u)$.

4. The Problem of Risk Exchange

Consider n insurance companies whose surplus at the end of the year will be $X_1, \ldots, X_n$, respectively. These are n random variables with known joint distribution. The decision to be made is the selection of a risk exchange. A risk exchange is best characterized by its effect on the distribution of the surplus among the n companies. In this sense a risk exchange is a random vector

$$Y = (Y_1, \ldots, Y_n), \tag{17}$$

where Y_i should be interpreted as the modified surplus of company i at the end of the year. Since the combined surplus before and after the exchange is the same, we must have

$$Y_1 + \ldots + Y_n = X_1 + \ldots + X_n. \tag{18}$$

We want to allow for the possibility that some of the companies are not willing to pay more than a certain amount towards the losses of the other companies. For this purpose assume n constants $c_1, \ldots, c_n$ with $0 \leq c_i \leq \infty$. Then only risk exchanges are admissible for which

$$Y_i \geq X_i - c_i, \; i = 1, \ldots, n. \tag{19}$$

We shall exclude the case where $c_1 = \ldots = c_n = 0$, because in that case only the trivial "exchange" (no exchange) is possible. To summarize, a risk exchange is a random vector of the form (17) that satisfies conditions (18) and (19) with probability one.

To evaluate the different risk exchanges, assume n utility functions $u_1(x), \ldots, u_n(x)$, $-\infty < x < \infty$. Suppose that these functions are twice differentiable, with

$$u_i'(x) > 0, \; u_i''(x) \leq 0. \tag{20}$$

For simplicity, we shall also assume that at most one of these utility functions is linear and that all of the others have the property that their derivative decreases from ∞ to 0 as the argument increases from $-\infty$ to ∞. Then a problem of the following type has a unique solution: given a number λ and positive numbers $k_1, \ldots, k_n$, find numbers $z_1, \ldots, z_n$ such that

$$k_i u_i'(z_i) \text{ is independent of } i \tag{21}$$

and

$$z_1 + \ldots + z_n = \lambda. \tag{22}$$

This solution $z = (z_1, \ldots, z_n)$ has a geometric interpretation: it corresponds to the point on the surface

$$F_\lambda = \{x = (x_1, \ldots, x_n)/x_i = u_i(t_i), t_1 + \ldots + t_n = \lambda\} \tag{23}$$

where the tangential plane is perpendicular to the vector $(k_1, \ldots, k_n)$. In the case of exponential utility functions,

$$u_i(x) = \alpha_i(1 - \exp(-x/\alpha_i)), \tag{24}$$

where $x_1 > 0, \ldots, x_n > 0$, this problem can be solved explicitly. One finds that

$$z_i = \beta_i\lambda + \alpha_i(\log k_i - \sum_{j=1}^{n} \beta_j \log k_j), \tag{25}$$

where $\beta_i = \alpha_i/(x_1 + \ldots + x_n)$.

It is assumed that company i is only interested in the expected utility of its own surplus,

$$s_i(Y) = E[u_i(Y_i)], \tag{26}$$

$i = 1, \ldots, n$. In this sense we are faced with the problem of finding Pareto-optimal risk exchanges. Let us verify the validity of Condition C in this case. If $Y^{(0)}$, $Y^{(1)}$ are any two risk exchanges,

$$Y^{(j)} = (Y_1^{(j)}, \ldots, Y_n^{(j)}), j = 0,1, \tag{27}$$

define

$$Y^{(r)} = (Y_1^{(r)}, \ldots, Y_n^{(r)}), 0 < r < 1, \tag{28}$$

by setting

$$Y_i^{(r)} = rY_i^{(1)} + (1 - r)Y_i^{(0)}. \tag{29}$$

Since $Y^{(0)}$ and $Y^{(1)}$ satisfy (18) and (19), it follows that $Y^{(r)}$ satisfies these conditions. Thus $Y^{(r)}$ is a risk exchange. Since the function u_i is concave from below,

$$u_i(Y_i^{(r)}) \geq ru_i(Y_i^{(1)}) + (1 - r)u_i(Y_i^{(0)}). \tag{30}$$

Taking expected values, we get

$$s_i(Y^{(r)}) \geq r s_i(Y^{(1)}) + (1 - r) s_i(Y_i^{(0)}), \tag{31}$$

which shows that Condition C holds. Obviously, S is closed, so to find the Pareto-optimal risk exchanges it is enough to choose positive constants $k_1, \ldots, k_n$ and to try to maximize

$$\sum_{i=1}^{n} k_i s_i(Y) = \sum_{i=1}^{n} k_i E[u_i(Y_i)]. \tag{32}$$

In this paragraph we shall construct a risk exchange Y and then verify that it maximizes (32). Let

$$I^{(0)} = \{1, \ldots, n\},\ J^{(0)} = \varphi. \tag{33}$$

We define random vectors

$$Z^{(m)} = (Z_1^{(m)}, \ldots, Z_n^{(m)}) \tag{34}$$

and index sets $I^{(m)}$ and $J^{(m)}$ as follows. For $m = 1, 2, \ldots$ set

$$Z_i^{(m)} = X_i - c_i \text{ if } i \in J^{(m-1)} \tag{35}$$

and choose $Z_i^{(m)}$, $i \in I^{(m-1)}$, such that

$$k_i u_i'(Z_i^{(m)}),\ i \in I^{(m-1)}, \text{ is independent of } i \tag{36}$$

and

$$\sum_{I^{(m-1)}} Z_i^{(m)} = \sum_{I^{(m-1)}} X_i + \sum_{J^{(m-1)}} c_i. \tag{37}$$

Then

$$I^{(m)} = \{i / Z_i^{(m)} > X_i - c_i\} \tag{38}$$

and

$$J^{(m)} = \{i / Z_i^{(m)} \leq X_i - c_i\}. \tag{39}$$

From this recursive definition it follows immediately that

(i) $I^{(m)} \subseteq I^{(m-1)}$, $J^{(m)} \supseteq J^{(m-1)}$

(ii) $I^{(m)}$ is not empty.

Furthermore, if $M^{(m)}$ denotes the common value of the expressions in (36), one can show that

(iii) $M^{(m+1)} \geq M^{(m)}$

(iv) $k_i u_i'(Z_i^{(m)}) \leq M^{(m)}$

(v) $k_i u_i'(Z_i^{(m)}) < M^{(m)}$ implies $Z_i^{(m)} = X_i - c_i$.

Now let $\tilde{Y} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)$ be the limit of $Z^{(m)}$, $m \to \infty$. (Note that this limit is obtained after finitely many steps; as a matter of fact, $\tilde{Y} = Z^{(n)}$.) Observe that $\tilde{Y}$ is a risk exchange and has the following property:

Property B

Let $M = \max\{k_i u_i' (\tilde{Y}_i)/i = 1, \ldots, n\}$. Then $k_i u_i'(\tilde{Y}_i) < M$ implies that $\tilde{Y}_i = X_i - c_i$.

We shall now compare $\tilde{Y}$ with an arbitrary risk exchange $Y = (Y_1, \ldots, Y_n)$ as follows: since the function $u_i(\cdot)$ is concave from below, and since $k_i u_i'(Y_i) < M$ implies that $Y_i \geq \hat{Y}_i$,

$$k_i u_i(Y_i) \leq k_i u_i(\tilde{Y}_i) + k_i u_i' (\tilde{Y}) \cdot (Y_i - \tilde{Y}_i)$$
$$\leq k_i u_i(\tilde{Y}_i) + M \cdot (Y_i - \tilde{Y}_i). \qquad (40)$$

Thus

$$\sum_{i=1}^{n} k_i u_i(Y_i) \leq \sum_{i=1}^{n} k_i u_i(\tilde{Y}_i) \qquad (41)$$

and

$$\sum_{i=1}^{n} k_i E[u_i(Y_i)] \leq \sum_{i=1}^{n} k_i E[u_i(\tilde{Y}_i)]. \qquad (42)$$

Furthermore, the last inequality is strict unless $Y = \tilde{Y}$ (almost surely). Our findings can be summarized as follows.

Theorem

a) Given $k_1 > 0, \ldots, k_n > 0$, there is exactly one risk exchange that satisfies Property B. b) A risk exchange is Pareto-optimal if and only if it is of this form.

Special cases

1) If $c_1 = \ldots = c_n = \infty$, this result reduces to the classical *Theorem of Borch*, see [2], [3], or [4].

2) Consider the case, where $u_1(x) = x$, $c_1 = \infty$, $u_2(x) = u(x)$ (strictly concave from below), and $c_2 = P > 0$. We find that the Pareto-optimal risk exchanges are of the form

$$\tilde{Y}_1 = \begin{cases} X_1 + P & \text{if } X_2 \geq \alpha \\ X_1 + P - (\alpha - X_2) & \text{if } X_2 < \alpha \end{cases} \qquad (43)$$

$$\tilde{Y}_2 = \begin{cases} X_2 - P & \text{if } X_2 \geq \alpha \\ \alpha - P & \text{if } X_2 < \alpha \end{cases} \qquad (44)$$

where the parameter α, satisfying the equation $k_1 = k_2 u' (\alpha - P)$, plays the role of a deductible. This result (*optimality of a stoploss contract*) is due to Arrow, see [1].

3) Consider the case, where $u_1(x) = x$, $c_1 = 0$, $u_2(x) = u(x)$ (strictly concave from below), and $c_2 = \infty$. Thus $Y_1 = X_1 + D$, $Y_2 = X_2 - D$, where $D \geq 0$ is a *dividend* payable from company 2 to company 1. We find that Pareto-optimal dividends are of the form

$$\tilde{D} = \begin{cases} X_2 - \alpha & \text{if } X_2 > \alpha \\ 0 & \text{if } X_2 \leq \alpha \end{cases} \tag{45}$$

This result has been found in [7] in connection with dividend formulas in group insurance.

References

[1] ARROW, K., (1963), "Uncertainty and the Welfare Economics of Medical Care", American Economic Review, LIII, 941-73.
[2] BORCH, K, (1962), "Equilibrium in a reinsurance market", Econometrica 30, 424-44
[3] BÜHLMANN, H, 1970, Mathematical Methods in Risk Theory. New York: Springer
[4] DUMOUCHEL, W. H., (1968), "The Pareto-optimality of a n-company reinsurance treaty", Skandinavisk Aktuarietidskrift 51, 165-70
[5] DEGROOT, M. H., 1975 Probability and Statistics, Reading: Addison-Wesley.
[6] GREVILLE, T. N E., (1973), Part Five Study Notes on Graduation, Chicago: Society of Actuaries.
[7] JONES, D. A. and GERBER, H. U., (1974), "Dividend Formulas in Group Insurance", Transactions of the Society of Actuaries 26, 77-93.

OPTIMAL REINSURANCE AND DIVIDEND PAYMENT STRATEGIES*

PANTELIS M. PECHLIVANIDES
National Technical University, Athens

I. INTRODUCTION AND SUMMARY

This paper presents a normative model for the sequential reinsurance and dividend-payment problem of the Insurance Company (I.C.). Optimal strategies are found in closed form for a class of utility functions. In some sense the model studied can be viewed as an adaptation of Hakansson's investment-consumption model of the individual [3] or a generalization of Frisque's model for the dynamic management of an I.C. [2].

In Section 2 the model is formulated as a discrete time dynamic programming problem. The objective of the I.C. is assumed to be maximization of the expected utility of the dividend streams paid to stock/policy-holders (s/p-holders). The initial reserves level is assumed to be known. The premiums to be collected in each period for selling policies are known in advance. The losses due to claims from policy-holders are random variables independent from period to period. In each period the I.C. must decide on the portion of the reserves to be paid as dividends and on the form and level of reinsurance with a reinsurer that quotes prices for any contract.

Optimal strategies in closed form are found in Section 3 when the utility function of the I.C. is given by the discounted sum of one-period utilities of dividends; and when the one-period utilities belong to the linear risk-tolerance class, which is given by: (Ia) $u(x) = (ax + b)^{c+1}/a(c + 1)$; $ax + b > 0$, $ac < 0$. (Ib) $u(x) = \log(ax + b)$; $ax + b > 0$. (II) $u(x) = -e^{-\gamma x}$; $\gamma > 0$.

The results of Section 3 are discussed and interpreted in Section 4. The optimal dividend payments are found to be linear in the reserves level; while the optimal reinsurance treaty transforms the reserves level (as a function of the losses) in such a way that its *form* is independent of the prereinsurance total wealth of the I.C. It only depends on the I.C.'s utility function, the prices quoted

* This study is based on my Ph.D. thesis submitted to the University of California, Berkeley (1975). I am grateful to Professor W. S. Jewell (Chairman) as well as to Professors Nils Hakansson and David Gale for many helpful comments and criticisms.

by the reinsurer and the probability density function of the prereinsurance losses.

Finally, in Section 5 we discuss a generalization to include expenditures for promotion of sales and an extension to multiplicative utilities.

2. Formulation of the Model

2.1. *The description of the Insurance Company*

The I.C. is faced with a N-period problem. The periods are numbered backwards, thus the interval $(t, t - 1)$ is the t^{th} period. We will use the following notation:

p_t: premiums collected by selling policies during period t. They are assumed to be collected at the end of the period for simplicity and they are known in advance.

ξ_t: claims paid to policy-holders during period t — a random variable which takes values on the internal X_t and whose value will be denoted by x_t. For simplicity it is assumed that claims are paid at the end of the period and are independent from period to period.

C_t: dividends paid to s/p-holders at start of period t (decision variable).

R_t: level of reserves at start of period t before dividends are paid.

$\varphi_t(x)$: probability density function of the r.v. ξ_t.

2.2. *The utility function of the I.C.*

We will assume that the utility function of the I.C. over possible streams of dividends $C = C_N, \ldots, C_1, C_0$ is given by one of the three forms: *

(S) Discounted Sum:

$$U(C) = \sum_{k=0}^{N} \alpha^k u(C_{N-k})\,;\ 0 < \alpha < 1$$

(MP) Multiplicative Positive:

$$U(C) = \prod_{k=0}^{N} u(C_{N-k})\,;\ u(.) > 0$$

(MN) Multiplicative Negative:

$$U(C) = -\prod_{k=0}^{N} [-u(C_{N-k})]\,;\ u(.) < 0$$

* For justification and discussion of these forms see [4], [5].

In each case the objective of the I.C. is to maximize the expected value of $U(C)$.

In the following we will concentrate on the form (S). The forms (MP) and (MN) are briefly discussed in Section 5. For more details the interested reader is referred to [6].

2.3 *Reinsurance*

We assume that in each period t there is a reinsurer who accepts any risk for the appropriate premium. The way he quotes premium is the following.

For any claims random variable ξ_t (value denoted by $x_t \varepsilon X_t$) whose probability distribution he knows, the reinsurer assigns a *price function*. $P_{\xi_t}(x_t) > 0$ such that the premium for assuming a contract $Z_t(\xi_t)$, which promises to pay to the cedent \$ $Z_t(x_t)$ at the end of period t depending on the outcome x_t of the random variable ξ_t, is given by:

$$\tilde{P}_t[Z_t(\xi_t)] \equiv \int_{X_t} Z_t(x) P_{\xi_t}(x) dx \tag{1}$$

As a marginal case consider the contract $Z_t(x) = 1$; $\forall x \varepsilon X_t$ which pays \$1 to the cedent at the end of period t under any event. The premium or present value of \$1 asked by the reinsurer is

$$P_t[1] \equiv \int_{X_t} P_{\xi_t}(x) dx \equiv \pi_t < 1 \tag{2}$$

In other words, $\dfrac{1 - \pi_t}{\pi_t}$ is the interest rate for period t.

The description of the reinsurance process above implies that:

1) There are no transaction costs in reinsuring.
2) Borrowing and lending rates are the same.
3) Reinsurance contracts have a span of one period. That is at the end of each period when the risks realize (the value of ξ is observed) the contracts are fulfilled and then cease to exist.

In the following we will denote by $P_t(x)$ the price function of the claims r.v. ξ_t of period t to avoid the complexity of the notation $P_{\xi_t}(x_t)$.

2.4 *Dynamic Programming formulation*

At the start of period t the I.C.'s reserves level is R_t. It immediately pays dividends C_t thus remaining with $R_t - C_t$ which by the end of the period grow to $(R_t - C_t)/\pi_t$ where

$$\pi_t = \int_{X_t} P_t(x) dx \tag{3}$$

At the end of the period the I.C. collects premiums p_t and pays claims x (the value of ξ_t) and thus, if it conducted no reinsurance, the reserves level for the next period (t — 1) would be $R^\circ_{t-1}(x) = \frac{R_t - C_t}{\pi_t} + p_t - x$. With reinsurance, however, the I.C. *sells* to the reinsurer $R^\circ_{t-1}(x)$ and *buys* $R_{t-1}(x)$ so that the budget constraint

$$\int_{X_t} R_{t-1}(x)P_t(x) = \int_{X_t} \left[\frac{R_t - C_t}{\pi_t} + p_t - x\right] P_t(x)dx \tag{4}$$

is satisfied.

It will be useful to denote the premium demanded by the reinsurer for assuming the risk ξ_t by

$$\rho_t = \int_{X_t} xP_t(x)dx \tag{5}$$

Now let

$f_t(R_t)$: the maximum expected utility for a t-period problem with initial reserves level R_t.

Then the problem of an I.C., whose utility function is of the form (S) above, can be written as a Dynamic Programming problem:

$$f_t(R_t) = \max_{C_t,\ R_{t-1}} \{u(C_t) + \alpha E[f_{t-1}(R_{t-1}(\xi_t))]\};\ 0 < \alpha < 1 \tag{6}$$

subject to the budget constraint (4) and with boundary condition,

$$f_0(R_0) = u(R_0) \tag{7}$$

3. Closed form Solutions

The D.P. problem formulated by (4), (6) and (7) cannot in general be solved analytically. In this section we will find closed form solutions to the problem when we additionally assume that the one-period utility function of the I.C. belongs to the Linear Risk-Tolerance (LRT) class.

The quantity $-\frac{u''(x)}{u'(x)}$ is known as the absolute risk aversion index (Pratt [7]). The inverse, $-\frac{u'(x)}{u''(x)}$ is known as the risk-toleranre index. The LRT class is then defined as the solutions to the equation

$$\frac{u'(x)}{u''(x)} = \frac{ax + b}{g} \tag{8}$$

where g, a, b reals and $u''(x) < 0$ and $u'(x) > 0$.

It can be shown that the solutions to (8) are

$$u(x) = \frac{(ax+b)^{c+1}}{a(c+1)}\ ; c \neq -1,\ ax+b > 0,\ ac < 0 \qquad \text{(Ia)}$$

$$u(x) = \frac{1}{a}\log(ax+b);\ ax+b > 0,\ a > 0 \qquad \text{(Ib)}$$

$$u(x) = \frac{1}{\gamma}(1 - e^{-\gamma x});\ -\infty < x < +\infty,\ \gamma > 0 \qquad \text{(II)}$$

It will be useful later to split class Ia into:

$$a > 0,\quad c < -1 \quad \rightarrow u(\cdot) < 0$$

$$a > 0,\ -1 < c < 0 \rightarrow u(\cdot) > 0 \qquad (\text{Ia}_2)$$

$$a < 0,\quad c > 0 \quad \rightarrow u(\cdot) < 0 \qquad (\text{Ia}_3)$$

Theorem Ia (*Model Ia*)

If $u(x)$ belong to class (Ia) then the solution to the t-period problem as described by (6) subject to (4) and (7) is unique and is given by

$$f_t(R_t) = D_t u(A_t R_t + B_t) \qquad (9)$$

The optimal dividend strategy is

$$C_t^* = A_t R_t + B_t \qquad (10)$$

The optimal reinsurance strategy transforms the wealth of the I.C. to

$$R_{t-1}^*(\xi_t) = \frac{1}{A_{t-1}\alpha^{1/c}}\left[A_t R_t + B_t + \frac{b}{a}\right]\left[\frac{P_t(\xi_t)}{\varphi_t(\xi_t)}\right]^{1/c} + -\frac{b}{aA_{t-1}} - \frac{B_{t-1}}{A_{t-1}} \qquad (11)$$

as long as the initial reserves R_t satisfy the condition:

$$a(A_t R_t + B_t) + b > 0 \qquad (12)$$

where

$$D_t = 1 + D_{t-1}\frac{m_t}{\alpha^{1/c}},\ D_t \geq 1 \qquad (13)$$

$$A_t = \frac{1}{D_t},\ 0 \leq A_t \leq 1 \qquad (14)$$

$$B_t = A_t\left[p_t\pi_t - \rho_t + \frac{B_{t-1}\pi_t}{A_{t-1}} + \frac{b\pi_t}{aA_{t-1}} - \frac{b\ m_t}{a\alpha^{1/c}A_{t-1}}\right] \qquad (15)$$

can be calculated recursively starting with

$$D_0 = 1,\ A_0 = 1,\ B_0 = 0$$

and

$$m_t = \int_{x_t} \left[\frac{P_t(x)}{\varphi_t(x)}\right]^{1/c} P_t(x)dx \tag{16}$$

Proof: The proof is inductive showing the result to be valid for a 1-period problem and then proving the induction step from $t - 1$ to t.

One period problem ($t = 1$)

The DP relation (6) becomes for $t = 1$

$$f_1(R_1) = \max_{C_1, R_0} \{u(C_1) + \alpha E[u(R_0(\xi_1))]\} \tag{17}$$

subject to (4) which for $t = 1$ becomes,

$$\int_{x_1} R_1(x)P_1(x)dx = R_1 - C_1 + p_1\pi_1 - \rho_1 \tag{18}$$

Fix. C_1. To maximize the second term in (17) subject to (18) according to the calculus of variation $R_0^*(\cdot)$ must be chosen so that

$$u'(R_0^*(x)\varphi_1(x) = \lambda P_1(x) \tag{19}$$

where λ is to be determined by substituting in (18).

Using the fact that $u(\cdot)$ belongs to class Ia we solve (19) to find

$$R_0^*(x) = \frac{\lambda^{1/c}}{a}\left[\frac{P_1(x)}{\varphi_1(x)}\right]^{1/c} - \frac{b}{a} \tag{20}$$

Upon substitution of (20) in (18) we find

$$\lambda^{1/c} = \frac{a}{m_1}\left(R_1 - C_1 + p_1\pi_1 - \rho_1 + \frac{b}{a}\pi_1\right) \tag{21}$$

with ρ_1, m_1 defined in (5) and (16) respectively.

Substituting (20) and (21) in (17) we obtain after some algebra:

$$f_1(R_1) = \max_{C_1}\left\{u(C_1) + \frac{\alpha m_1}{a(c+1)}\left[\frac{a}{m_1}\left(R_1 - C_1 + p_1\pi_1 - \rho_1 + \frac{b}{a}\pi_1\right)\right]^{c+1}\right\} \tag{22}$$

where we have used the identity:

$$E\left[\left(\frac{P_1(x)}{\varphi_1(x)}\right)^{1+1/c}\right] \equiv \int_{x_1}\left[\frac{P_1(x)}{\varphi_1(x)}\right]^{1/c} P_1(x)dx \equiv m_1 \tag{23}$$

The second term in the RHS of (22) is strictly concave as long as

$$a(R_1 - C_1 + p_1\pi_1 - \rho_1) + b\,\pi_1 > 0 \tag{24}$$

while the first term, $u(C_1)$, is strictly concave as long as

$$aC_1 + b > 0 \tag{25}$$

Differentiating the maximand in (22) w.r.t. C_1 and equating to zero we obtain the unique optimal dividend strategy

$$C_1^* = A_1R_1 + B_1 \tag{26}$$

with A_1, B_1 as defined in (14) and (15).

Further, when C_1 is given by (26) the conditions (24) and (25) are equivalent and thus the only condition required is

$$a(A_1R_1 + B_1) + b > 0 \tag{27}$$

Finally, substituting (26) in (22) we obtain

$$f_1(R_1) = D_1u(A_1R_1 + B_1)$$

which is in the desired form.

The t-period problem

We assume that the theorem holds for a $(t - 1)$-period problem and we show that it holds for a t-period problem. The arguments are similar and we will thus be rather brief (a more detailed proof can be found in [6]).

We first fix C_t and we find that the optimal post-reinsurance wealth $R_{t-1}^*(\xi_t)$ must satisfy

$$R_{t-1}^*(x) = \frac{\lambda^{1/c}}{aA_{t-1}}\left[\frac{P({}_tx)}{\varphi_t(x)}\right]^{1/c} - \frac{b}{aA_{t-1}} - \frac{B_{t-1}}{A_{t-1}} \tag{28}$$

where

$$\frac{\lambda^{1/c}}{aA_{t-1}} = \frac{1}{m_t}\left[R_t - C_t + p_t\pi_t - \rho_t + \frac{B_{t-1}}{A_{t-1}}\frac{\pi_t}{m_t} + \frac{b\,\pi_t}{aA_{t-1}}\right] \tag{29}$$

Substitution of (28), (29) in (6) yields

$$f_t(R_t) = \max_{C_t}\left\{u(C_t) + \frac{\alpha D_{t-1}m_t}{a(c+1)}\left[\frac{aA_{t-1}}{m_t}\left(R_t - C_t + p_t\pi_t - \rho_t + \frac{B_{t-1}\pi_t}{A_{t-1}m_t} + \frac{b\,\pi_t}{aA_{t-1}}\right)\right]^{c+1}\right\} \tag{30}$$

Differentiating the maximand w.r.t. C_t and setting equal to zero we find the unique optimal dividend:

$$C_t^* = A_tR_t + B_t \tag{31}$$

as long as R_t is such that

$$a(A_tR_t + B_t) + b > 0 \tag{32}$$

Finally substituting (31) in (29) and using the definitions of A_t, B_t in (14) and (15) we obtain (11) and the Theorem is proved.

Remark: If for a t-period problem the initial reserves R_t are such that $a(A_tR_t + B_t) + b > 0$ and the optimal strategies (10) and (11) are followed, then at the start of period $t - 1$ the reserves R_{t-1} will again satisfy $a(A_{t-1} R_{t-1} + B_{t-1}) + b > 0$ To see this we only need to observe (11). This means that following the optimal strategies for a t-period problem we are guaranteed that we will be able to reapply them for a $t - 1$ period problem with no further conditions.

Theorem Ib (Model Ib)

If $u(x)$ belongs to class (Ib) then the solution to the t-period problem as described by (6) subject to (4) and (7) is unique and is given by

$$f_t(R_t) = D_t u(A_tR_t + B_t) + E_t \tag{33}$$

The optimal dividend strategy is

$$C_t^* = A_tR_t + B_t \tag{34}$$

The optimal reinsurance strategy transforms the wealth of the I.C. to

$$R_{t-1}^*(\xi_t) = \frac{\alpha}{A_{t-1}}\left(A_tR_t + B_t + \frac{b}{a}\right)\frac{\varphi_t(\xi_t)}{P_t(\xi_t)} - \frac{b}{aA_{t-1}} - \frac{B_{t-1}}{A_{t-1}} \tag{35}$$

as long as the initial reserves R_t satisfy the condition:

$$a(A_tR_t + B_t) + b > 0 \tag{36}$$

where

$$D_t = 1 + \alpha D_{t-1}, \quad D_t \geq 1 \tag{37}$$

$$A_t = \frac{1}{D_t}, \quad 0 \leq A_t \leq 1 \tag{38}$$

$$B_t = A_t\left[p_t\pi_t - \rho_t + \frac{B_{t-1}}{A_{t-1}}\pi_t + \frac{b\,\pi_t}{aA_{t-1}} - \frac{\alpha}{A_{t-1}}\frac{b}{a}\right] \tag{39}$$

$$E_t = \frac{\alpha}{a} D_{t-1}[\log \alpha + q_t] + \alpha E_{t-1} \tag{40}$$

can be calculated recursively starting with

$$D_0 = 1,\ A_0 = 1,\ B_0 = 0,\ E_0 = 0 \tag{41}$$

and

$$q_t = E\left[\log\left(\frac{\varphi_t(\xi_t)}{P_t(\xi_t)}\right)\right] \tag{42}$$

Proof: is similar to that of Theorem Ia and is deleted. For more details see [6].

Remark 1: Except (33), (40), (42) all the results of Theorem Ib can follow from Theorem Ia by letting $c \to -1$ and $m_t \to 1$.

Remark 2: The Remark at the end of Theorem Ia again holds as it can be checked by observing (35).

Theorem II (Model II)

If $u(x)$ belongs to class (II) then the solution to the t-period problem as described by (6) subject to (4) and (7) is unique and is given by

$$f_t(R_t) = D_t u(A_t R_t + B_t) + E_t \tag{43}$$

The optimal dividend strategy is

$$C_t^* = A_t R_t + B_t \tag{44}$$

The optimal reinsurance strategy transforms the wealth of the I.C. to

$$R_{t-1}^*(\xi_t) = \frac{1}{A_{t-1}}[A_t R_t + B_t] - \frac{B_{t-1}}{A_{t-1}} + \frac{\log \alpha}{\gamma A_{t-1}} - \frac{1}{\gamma A_{t-1}} \log\left(\frac{P_t(\xi_t)}{\varphi_t(\xi)}\right) \tag{45}$$

where

$$D_t = 1 + \pi_t D_{t-1}, \quad D_t \geq 1 \tag{46}$$

$$A_t = \frac{1}{D_t},\ 0 \leq A_t \leq 1 \tag{47}$$

$$B_t = A_t\left[p_t\pi_t - \rho_t + \frac{B_{t-1}}{A_{t-1}}\pi_t + \frac{w_t}{\gamma A_{t-1}} - \frac{\pi_t}{\gamma A_{t-1}}\log \alpha\right] \tag{48}$$

$$E_t = \frac{D_{t-1}}{\gamma}(\alpha - \pi_t) + \alpha E_{t-1} \tag{49}$$

can be calculated recursively starting with

$$D_0 = 1,\ A_0 = 1,\ B_0 = 0,\ E_0 = 0$$

and

$$w_t = \int_{X_t} \log \left(\frac{P_t(x)}{\varphi_t(x)}\right) P_t(x)\, dx \tag{50}$$

Proof: Similar to that of Theorem Ia. An outline of the proof appears in [6].

4. Interpretation of the Optimal Strategies

4.1 *The dividend strategy*

In all Models the optimal dividend strategy is linear in the reserves level at the start of the period. In our formulation the dividends were not restricted to be positive. Negative dividends would, of course, mean that the s/p-holders agree that an increase in the reserves *now* is desirable for better profits in the *future*. If, however, we insist that dividends should be non-negative we can easily achieve it by restricting to Models Ia_1, Ia_2, Ib with $-b/a > 0$. In the case of Model II, a sufficient condition to guarantee the non-negativity of dividends for a N-period problem is $A_N R_N + B_N \geq 0$ and $\alpha \geq \dfrac{P_t(x)}{\varphi_t(x)}$; $x \varepsilon X_t$, $t = N, \ldots, 1$. This can be seen by looking at (45). A necessary condition for the latter is $\alpha \geq \pi_t$ for all t.

4.2 *The reinsurance strategy*

We can interpret $\left(\dfrac{P_t(\phi_t}{\varphi_t(\xi_t)}\right)^{1/c}$ as a *unit* of *post-reinsurance risky asset* for Model Ia. The name is suggested by observing (11) since $\left(\dfrac{P_t(\xi_t)}{\varphi_t(\xi_t)}\right)^{1/c}$ is the only quantity which is a function of the outcome of the random variable ξ_t and its *form* is independent of the initial wealth of the I.C. In this sense, m_t can be interpreted as the *cost of a unit of post-reinsurance risky asset*. Similarly, in Model Ib (35) the unit of post-reinsurance risky asset is $\dfrac{\varphi_t(\xi_t)}{P_t(\xi_t)}$ and its cost is 1. In Model II (45) the unit of risky asset is $\log \dfrac{P_t(\xi_t)}{\varphi_t(\xi_t)}$ and its cost is w_t.

In Models Ia, Ib the amount of risky asset increases linearly with the initial reserves level, while in Model II the amount of risky asset is fixed independent of the reserves level.

If $\frac{P_t(x)}{\varphi_t(x)}$ is non-decreasing in x then the post-reinsurance wealth of the I.C. is non-increasing in x in all Models. This of course means that the I.C. participates positively in the risk. That is, the larger the claims x paid to the policy-holders, the less the wealth of the I.C. after reinsurance. We can think of $\frac{P_t(x)}{\varphi_t(x)}$ as the *loading factor*. An increasing loading factor then means that the reinsurer asks for a greater loading to a certificate that guarantees final reserves of \$1 to the cedent when the claims x paid to the policy-holders are large than when they are small.

Further, in Models Ia and Ib the post-reinsurance wealth $R_{t-1}(\xi_t)$

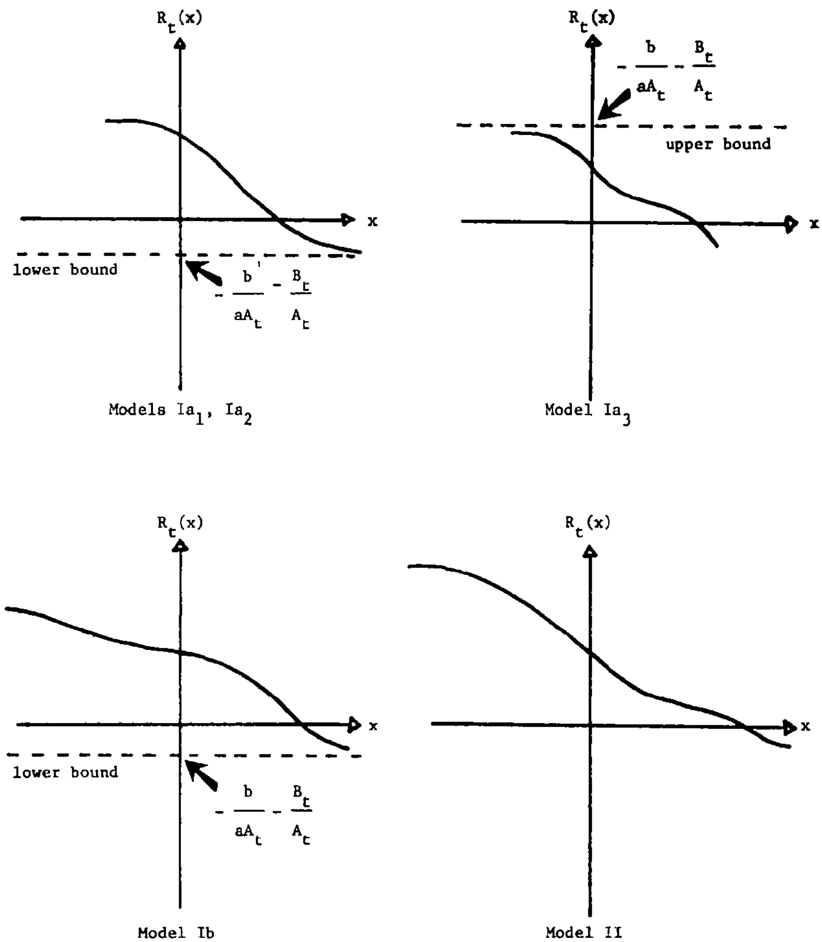


Figure 1. The post-reinsurance wealth $R_t(x)$ as a function of the claims x.

of the I.C. satisfies the condition $a(A_{t-1}R_{t-1}(\xi_t) + B_{t-1}) + b > 0$. This condition imposes upper or lower bounds on $R_{t-1}(\xi_t)$ depending on the sign of a which is negative for Model Ia_3 and positive for Models Ia_1, Ia_2 and Ib (see Figure 1).

The negativity of a makes Class Ia_3 the only one with an increasing risk-aversion index (Classes Ia_1, Ia_2, Ib have decreasing while Class II has a constant risk-aversion index). Thus Class Ia_3 (to which also the quadratic utility function belongs) must be applied with caution as it is doubtful whether it has meaning in real life (for a discussion of this point see Arrow [1]).

5. Generalizations - Extensions

(a) All Models can be easily extended to an infinite horizon by simply letting the number of periods N tend to infinity. The optimal strategies remain qualitatively the same.

(b) All Models can be generalized to include a decision on expenditures to promote sales if we assume that the sales volume is a concave function of the money spent. The optimal dividend and reinsurance strategies remain essentially the same. This is intuitively expected by observing that the quantity p_t (premiums collected from policy-holders) appears only in the constant B_t and not in A_t or D_t or E_t.

(c) Multiplicative Utilities. If instead of the form (S) we assume that the I.C.'s utility over dividend streams is given by (MN) or (MP) (Section 2.2) we can again find closed form solutions but only when (MN) is coupled with the Class Ia_1 of utility functions or (MP) with Class Ia_2. The results are similar in nature with those of Section 3. Again the optimal dividend strategy is linear in the reserves while the *form* of the post-reinsurance wealth of the I.C. is independent of its initial wealth. It only depends on the price function, the probability density function of the claims, the one-period utility function of the I.C. and the number of periods remaining.

These extensions-generalizations are treated in detail in [6].

References

[1] Arrow, K., 1971 *Essays in the Theory of Risk Bearing*, North-Holland
[2] Frisque, A, 1974 "Dynamic Model of Insurance Company's Management," *ASTIN*, Vol 8, part 1
[3] Hakansson, N, 1970 "Optimal Investment and Consumption Strategies under Risk for a Class of Utility Functions," *Econometrica*, Vol. 38, No. 5
[4] Koopmans, T., 1960. "Stationary Ordinal Utility and Impatience," *Econometrica*, Vol 28, No 2.

[5] MEYER, R., 1969. "On the Relationship Among the Utility of Assets, the Utility of Consumption and Investment Strategy," Proceedings of the 5th International Conference of Operations Research

[6] PECHLIVANIDES, P., 1975 "Reinsurance Market Mechanisms and Dividend Strategies for an Insurance Company," Ph.D. Thesis published as an Operations Research Center Report, ORC 75-17, University of California, Berkeley

[7] Pratt, J., 1964 "Risk Aversion in the Small and in the Large," *Econometrica*, Vol 32, No. 1-2

FROM AGGREGATE CLAIMS DISTRIBUTION TO PROBABILITY OF RUIN

HILARY L. SEAL,
Ecole Polytechnique Fédérale de Lausanne

INTRODUCTION

When the distribution of the number of claims in an interval of time of length t is mixed Poisson and the moments of the independent distribution of individual claim amounts are known, the moments of the distribution of aggregate claims through epoch t can be calculated (O. Lundberg, 1940, ch. VI). Several approximations to the corresponding distribution function, $F(\cdot, t)$, are available (see, e.g., Seal, 1969, ch. 2) and, in particular, a simple gamma (Pearson Type III) based on the first three moments has proved definitely superior to the widely accepted "Normal Power" approximation (Seal, 1976). Briefly,

$$F(t + z\sqrt{\varkappa_2}, t) \approx \frac{1}{\Gamma(\alpha)} \int_0^{\alpha + z\sqrt{\alpha}} e^{-y} y^{\alpha - 1} \equiv P(\alpha, \alpha + z\sqrt{\alpha}) \qquad (1)$$

where the P-notation for the incomplete gamma ratio is now standard and α, a function of t, is to be found from

$$\alpha = \frac{4}{\varkappa_3^2/\varkappa_2^3} = \frac{4}{\gamma_1^2}$$

the kappas being the cumulants of $F(\cdot, t)$. An excellent table of the incomplete gamma ratio is that of Khamis (1965).

The problem that is solved in this paper is the production of an approximation to $U(w, t)$, the probability of non-ruin in an interval of time of length t, by using the above mentioned gamma approximation to $F(\cdot, t)$.

THE PROBABILITY OF NON-RUIN IN A PERIOD OF LENGTH T

In Seal (1974) it was shown that when the distribution of the number of claims in an arbitrary interval of time is generated by a stationary point process the probability of non-ruin in an interval which the insurance company enters with a risk-reserve of w and operates throughout with a risk-premium loading of η, is $U(w, t)$ given by

$$U(w, t) = F(w + \pi_1 t, t) - \pi_1 \int_0^t U(0, \tau) f(\overline{w + \pi_1 t - \tau, t - \tau})\, d\tau \qquad (2)$$

where π_1 is the risk-loaded pure premium rate and $f(\cdot, t)$ is the density corresponding to $F(\cdot, t)$. This is the formula which we will use for our numerical approximations.

The only stationary point processes that have been utilized by actuaries in practical applications are those that lead to ordinary or mixed Poisson distributions (O. Lundberg, *l.c.*) and in these circumstances the Prabhu-Benes-Takács formula (Seal, 1974)

$$U(0, t) = \frac{1}{\pi_1 t} \int_0^{\pi_1 t} F(y, t)\, dy \tag{3}$$

may be used to produce the first factor in the integrand of (2).

Application of Relation (1)

Considering (1) as applied to (2) we note that if the distribution of the number of claims is Poisson with mean t and the density of individual claim amounts, $b(\cdot)$, has mean

$$\mu = 1 \text{ so that } \pi_1 = 1 + \eta,\ F(w + \overline{1 + \eta} \cdot t, t) \approx P(\alpha, \alpha + z\sqrt{\alpha})$$

where

$$\alpha = \alpha(t) = \frac{4\, \varkappa_2^3}{\varkappa_3^2} = \frac{4(tp_2)^3}{(tp_3)^2} = \frac{4tp_2^3}{p_3^2}$$

p_2 and p_3 being the second and third moments about zero of the $b(\cdot)$-distribution of individual claim amounts (Seal, 1969, 2.41). In order to evaluate z we have

$$t + z\sqrt{\varkappa_2} = t + z\sqrt{(tp_2)} = w + (1 + \eta)t$$

so that

$$z = (w + \eta t)\,(tp_2)^{-1/2}$$

Further, by differentiation of (1) with respect to z,

$$f(\tau + z\sqrt{\varkappa_2}, \tau) \approx \frac{\beta}{\Gamma(\alpha)} \exp\left[-\alpha - z\sqrt{\alpha}\right]\,(\alpha + z\sqrt{\alpha})^{\alpha - 1} \tag{4}$$

where

$$\alpha = \alpha(\tau) = \frac{4\tau\, p_2^3}{p_3^2} \qquad \beta = \sqrt{(\alpha/\varkappa_2)}$$

and, when $\tau + z\sqrt{\varkappa_2} = w + (1 + \eta)\,\tau$,

$$z = (w + \eta\,\tau)\,(\tau\, p_2)^{-1/2} \quad 0 < \tau < t$$

Finally, by (3),

$$U(0, \tau) = \frac{1}{(1 + \eta)\tau} \int_0^{(1+\eta)\tau} F(y, \tau)\, dy$$

$$= \frac{\sqrt{\kappa_2}}{(1 + \eta)\tau} \int_{-\tau/\sqrt{\kappa_2}}^{\eta\tau/\sqrt{\kappa_2}} F(\tau + z\sqrt{\kappa_2}, \tau)\, dz$$

$$\approx \frac{\sqrt{\kappa_2}}{(1 + \eta)\tau} \int_{-\tau/\sqrt{\kappa_2}}^{\eta\tau/\sqrt{\kappa_2}} P(\alpha, \alpha + z\sqrt{\alpha})\, dz \qquad \text{by (1)}$$

$$= \frac{\sqrt{(\kappa_2/\alpha)}}{(1 + \eta)\tau} \int_{\alpha - \tau\sqrt{(\alpha/\kappa_2)}}^{\alpha + \eta\tau\sqrt{(\alpha/\kappa_2)}} P(\alpha, u)\, du$$

$$= \frac{1}{(1 + \eta)\tau\beta} \int_{\alpha - \tau\beta}^{\alpha + \eta\tau\beta} \frac{du}{\Gamma(\alpha)} \int_0^u x^{\alpha - 1} e^{-x} dx$$

$$= \frac{1}{(1 + \eta)\tau\beta\Gamma(\alpha)} \left\{ \int_0^{\alpha + \eta\tau\beta} - \int_0^{\alpha - \tau\beta} \right\} du \int_0^u x^{\alpha - 1} e^{-x} dx$$

$$= \frac{1}{(1 + \eta)\tau\beta\Gamma(\alpha)} \Big[\int_0^{\alpha + \eta\tau\beta} (\alpha + \eta\tau\beta - x)\, x^{\alpha - 1} e^{-x} dx +$$

$$- \int_0^{\alpha - \tau\beta} (\alpha - \tau\beta - x)\, x^{\alpha - 1} e^{-x} dx \Big]$$

$$= \frac{1}{(1 + \eta)\tau\beta} [(\alpha + \eta\tau\beta) P(\alpha, \alpha + \eta\tau\beta) - \alpha\, P(\alpha + 1, \alpha + \eta\tau\beta)$$

$$- (\alpha - \tau\beta)\, P(\alpha, \alpha - \tau\beta) + \alpha\, P(\alpha + 1, \alpha - \tau\beta)] \qquad (5)$$

where

$$\beta = \sqrt{(\alpha/\kappa_2)} \quad \text{and} \quad \alpha = \alpha(\tau).$$

A remarkable feature of the approximation (1) is that only the first three moments of the distribution of individual claim amounts are involved. If, therefore, a two-parameter distribution is successfully fitted to the observational distribution of claim amounts by means of the mean and variance it implies that the appropriateness of the chosen functional form has been determined by the approximate equivalence of the third moments of the observational and theoretical distributions of individual claims. For example, if the gamma distribution (Johnson & Kotz, 1970, ch. 17) were fitted the third central moment (or cumulant) would necessarily be twice the variance.

Now only two functional forms for $b(\cdot)$, the density function of individual claim amounts, result in explicit results for $F(x, t)$ when the distribution of the number of claims in an interval of length t is Poisson with mean t (Seal, 1969, p. 31, referring to Hadwiger, 1942). These are the gamma and the inverse Gaussian distributions and it would be convenient to use one or other of these forms for $b(.)$ so that direct checks may be made of our numerical approximations using (1).

The Inverse Gaussian Distribution

According to Seal (1969, p. 30) by far the greatest number of graduations of observed individual claim amounts have been based on the lognormal distribution, namely where the logarithm of the claim amount (the latter possibly increased or decreased by some constant) has a Normal distribution.

On the other hand the inverse Gaussian density (Tweedie, 1957)

$$b(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right] \quad x > 0,\ \mu > 0,\ \lambda > 0 \qquad (6)$$

which has the distribution function

$$B(x) = \Phi\left(\sqrt{\frac{\lambda}{x}} \cdot \overline{\frac{x}{\mu} - 1}\right) + e^{2\lambda/\mu}\left\{1 - \Phi\left(\sqrt{\frac{\lambda}{x}} \cdot \overline{\frac{x}{\mu} + 1}\right)\right\} \qquad (7)$$

as shown by Shuster (1968) (but misprinted in Johnson & Kotz, 1970), where $\Phi(\cdot)$ is the standardized Normal distribution function, can be made to start at the same claim amount (which we take as the origin) as the lognormal and be given the same mean μ and variance μ^3/λ. Although the Inverse Gaussian has never been used to graduate a set of individual claim amounts it may produce nearly the same γ_1-value as that possessed by the corresponding lognormal distribution and would then lead to approximately the same distribution of aggregate claims as provided by (1).

When individual claims are distributed according to the inverse Gaussian,

$$f(x, t) = e^{-t} \sum_{k=1}^{\infty} \frac{t^k}{k!} \left(\frac{\Lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\frac{\Lambda(x-M)^2}{2M^2 x}\right] \qquad (8)$$

where $\Lambda = k^2\lambda$ and $M = k\mu$, and from (6) and (7)

$$F(x, t) = e^{-t} + e^{-t} \sum_{k=1}^{\infty} \frac{t^k}{k!} \left[\Phi\left(\sqrt{\frac{\Lambda}{x}} \cdot \left(\frac{x}{M} - 1\right)\right) + e^{2\Lambda/M} \left\{1 - \Phi\left(\sqrt{\frac{\Lambda}{x}} \cdot \left(\frac{x}{M} + 1\right)\right)\right\}\right] \qquad (9)$$

We mention that $\beta(s)$, the Laplace transform of (6), is given by

$$ln\ \beta(s) = \frac{\lambda}{\mu}\left\{1 - \left(1 + \frac{2\mu^2 s}{\lambda}\right)^{1/2}\right\} \qquad (10)$$

Choice of Parameters

Upwards of 50 actual individual claim distributions have been fitted by the lognormal (Seal, 1969, p. 30). The γ_1-values for 45 of these were calculated *, using the formulas provided by Johnson & Kotz (1970) applicable to the constants of the linear transform, and compared with the corresponding γ_1's calculated for (6) using the calculated mean and variance. 60% of the γ_1 pairs were approximately equal implying that the lognormal and inverse Gaussian distributions would produce nearly the same value for (1). Among the 27 distributions was Cannella's (1963) costs of 124, 279 "specialty" pharmaceutical prescriptions in the province of Rome during 1960. The two γ_1's were ·355 and ·354, respectively, but the mean and variance of the distribution were stated to be 786.4 and 280582.09 after lognormal fitting. Unfortunately this mean and variance produce γ_1's of 2.326 and 2.021, respectively, for the lognormal and inverse Gaussian indicating that, in fact, the latter distribution is not in this case a very good approximation to the lognormal. This error of Cannella was not discovered until too late and we had already chosen $\mu = 1$ and $\lambda = (786.4)^2/280582.09 = 2.20408$ for the inverse Gaussian. In order to apply this to (1) we have (Tweedie, *loc. cit.*) $p_2 = \mu^2 + \mu^3\lambda^{-1} = 1.453704$ and $p_3 = \mu^3 + 3\mu^4\lambda^{-1} + 3\mu^5\lambda^{-2} = 2.978654$ so that $\alpha(t) = 1.384993\, t$.

Results

The following Table compares the results obtained for $f(10 + t, t)$ by (4) and (8) and for $F(10 + t, t)$ by (1) and (9). In the first set of comparisons the gamma approximation is only in error by a few units in the fifth decimal place. In the second set the gamma

* It is not always easy to decide whether an author is using natural or common logarithms for his transform.

approximation is never more in error than by two units in the fourth decimal place. These are very good results.

TABLE 1

Values of $f(10 + t, t)$, $F(10 + t, t)$ and $U(10, t)$

t	$f(10+t, t)$ (4)	$f(10+t, t)$ (8)	$F(10+t, t)$ (1)	$F(10+t, t)$ (9)	$U(10, t)$ (1) to (5)	$U(10, t)$ method of 1974 paper
1	.00004	.00003	.99996	.99997	.9999	1.0000
2	.00019	.00016	.99978	.99983	.9997	1.0000
3	.00049	.00045	.99937	.99945	.9991	.9993
4	.00095	.00090	.99866	.99879	.9980	.9981
5	.00154	.00150	.99764	99779	9964	.9964
6	.00226	.00222	.9963	.9965	.9943	.9943
7	.00305	.00303	.9947	.9948	.9916	.9915
8	.00390	.00390	9927	.9929	.9884	.9883
9	.00479	.00479	.9906	.9908	9847	.9846
10	.00569	.00570	.9882	.9884	.9807	.9804
11	.00658	.00661	.9857	.9858	9762	.9759
12	.00747	.00750	.9830	.9831	.9715	.9711
13	.00833	.00837	.9801	.9803	.9665	.9660
14	.00916	.00921	.9772	.9774	.9613	.9607
15	.00997	.01002	9742	.9744	.9559	.9552
16	.01074	.01079	.9711	9713	.9503	.9495
17	01147	.01153	.9680	.9682	.9447	.9438
18	.01217	.01223	.9649	.9651	.9369	.9380
19	.01283	.01289	.9618	.9619	.9331	.9321
20	.01346	.01352	.9586	.9588	.9273	.9262
21	.01406	.01411	.9554	.9556	.9214	.9202
22	.01462	01467	.9523	.9524	9155	.9143
23	.01515	.01520	.9491	.9493	.9097	.9083
24	.01564	.01570	.9460	.9462	.9038	.9024
25	.01611	.01617	.9429	.9431	.8980	.8965

The approximate values of f, F and $U(0, t)$ (by relation (5)) were then inserted into (2) with $w = 10$ and $\eta = 0$ using repeated Simpson at unit steps in t for the value of the integral. When t was odd the last three panels were approximated by the three-eighths rule; $U(0, 1)$ was obtained by the trapezoidal. There is no "exact" result for $U(10, t)$ but the Laplace transform inversion methods described in Seal (1974) were used to produce results supposedly correct to three decimals. These, together with our new approximations appear in the last two columns of the Table. The new method appears to be producing values of $U(10, t)$ "nearly" correct to three decimals.

CONCLUSION

The proposed new approximation to $U(w, t)$ using the gamma approximation to $F(x, t)$ produces reasonably accurate results. Is it easy to apply? The writer confessed in his 1974 paper that steps in t at greater intervals than unity tended to harm the efficiency of the approximation to the integral in (2). For example, by using steps of five instead of unity in (2) we obtained, with the new approximations, the following values which are barely correct to two

	$U(10, t)$	
t	*Unit steps* (Table 1)	*Quinquennial steps*
5	.996	.994
10	.981	.977
15	.956	947
20	.927	918
25	.898	.887

decimals. Nevertheless this may be considered sufficient if a computer is not being used and desk calculations are the order of the day.

REFERENCES

CANNELLA, S., 1963. "Variation de la prime d'assurance de l'assistance pharmaceutique en fonction de la participation de l'assuré au coût de l'assistance" *ASTIN Bull.* *2*, 30-44

HADWIGER, H. 1942. "Wahl einer Naherungsfunktion fur Verteilungen auf Grund einer Funktionalgleichung" *Bl. Versich. Math.* 5, 345-352.

JOHNSON, N. L, and KOTZ, S, 1970. *Distributions in Statistics. Continuous Univariate Distributions*-- 1. Houghton Mifflin, Boston.

KHAMIS, S H. and RUDERT, W., 1965 *Tables of the Incomplete Gamma Function Ratio* Von Liebig, Darmstadt

LUNDBERG, O, 1940. *On Random Processes and their Application to Sickness and Accident Statistics.* Almqvist and Wiksell, Uppsala.

SEAL, H. L., 1969 *Stochastic Theory of a Risk Business* John Wiley, New York.

—— 1974 "The numerical calculation of U(w, t), the probability of non-ruin in an interval (o, t)". *Scand Actu. J.* 1974, 121-139

—— 1976. "Approximations to Risk Theory's F(x, t) by means of the gamma distribution" *ASTIN Bull* 9, 213-218.

SHUSTER, J., 1968. "On the inverse Gaussian distribution function". *J Amer. Statist. Assn.* *63*, 1514-1516.

TWEEDIE, M. C K, 1957. "Statistical properties of inverse Gaussian distributions" *Ann Math. Statist* *28*, 362-377 and 696-705

LARGEST CLAIMS REINSURANCE (LCR). A QUICK METHOD TO CALCULATE LCR-RISK RATES FROM EXCESS OF LOSS RISK RATES

G. BENKTANDER
Switzerland

Let us denote by $E(x)$ the pure risk premium of an unlimited excess cover with the retention x and by $H(x)$ and $m(x)$ the corresponding expected frequency and severity.

We thus have $E(x) = H(x) \cdot m(x)$.

$H(x)$ is a non-increasing function of x and for practical purposes we can assume that it is decreasing; $H'(x) < 0$. The equation $H(x) = n$ has then only one solution x_n, where n is a fixed integer.

Let E_n denote the risk premium for a reinsurance covering the n largest claims from the bottom.

Let us define $E'_n = nx_n + E(x_n) = n(x_n + m(x_n))$. Intuitively we feel that E'_n is a good approximation for E_n.

We shall first show that when the claims size distribution is Pareto and the number of claims is Poisson distributed, E'_n is a good approximation for E_n, being slightly on the safe side. We further include a proof given by G. Ottaviani that the inequality $E_n < E'_n$ always holds.

In the Pareto case we have

$$H(x) = t(1 - F(x)) = t \cdot x^{-\alpha}$$

where the Poisson parameter t stands for the expected number of claims in excess of 1 (equal to a suitably chosen monetary unit) and

$$m(x) = \frac{x}{\alpha - 1}.$$

The retention x_n over which we expect n claims should satisfy

$$n = H(x_n) = t \cdot x_n^{-\alpha}$$

which gives

$$t = n \cdot x_n^{\alpha}$$

or

$$x_n = \left(\frac{t}{n}\right)^{1/\alpha}.$$

According to B. Berliner [2] we have, when the number of claims is Poisson distributed

$$E_n = t^{1/\alpha} \sum_{i=1}^{n} \frac{1}{\Gamma(i)} \cdot \Gamma_t \left(i - \frac{1}{\alpha}\right)$$

where

$$\Gamma_t(n) = \int_0^t e^{-u} \cdot u^{n-1} \, du.$$

Replacing the incomplete Gamma function Γ_t by $\Gamma_\infty = \Gamma$ we arrive at

$$\bar{E}_n = t^{1/\alpha} \cdot \frac{\alpha}{\alpha - 1} \cdot \frac{1}{\Gamma(n)} \cdot \Gamma \left(n + 1 - \frac{1}{\alpha}\right)$$

which formula was given by H. Ammeter already in 1964 [1]. Obviously $E_n < \bar{E}_n$

In all cases when t is large compared to n, we have

$$\frac{E_n}{\bar{E}_n} (n, \alpha; t) \quad \text{very close to } 1.$$

If in a practical situation t is too small we can always increase t by decreasing the monetary unit, in other words by enlarging to the left the range of the Pareto distribution.

Inserting $t = nx_n^\alpha$, as deduced above, in $\bar{E}_n$, we obtain

$$\bar{E}_n = n^{1/\alpha} \cdot x_n \cdot \frac{\alpha}{\alpha - 1} \cdot \frac{1}{\Gamma(n)} \Gamma \left(n + 1 - \frac{1}{\alpha}\right).$$

However

$$E'_n = n(x_n + m(x_n)) = n \cdot x_n \cdot \frac{\alpha}{\alpha - 1} = x_n \cdot \frac{\alpha}{\alpha - 1} \frac{\Gamma(n + 1)}{\Gamma(n)}.$$

Thus we have

$$\frac{\bar{E}_n}{E'_n} = \frac{n^{1/\alpha} \cdot \Gamma\left(n + 1 - \frac{1}{\alpha}\right)}{\Gamma(n + 1)}$$

Tabulation of

$$\frac{\bar{E}_n}{E'_n}$$

n	$\alpha = 2$	$\alpha = 2.5$	$\alpha = 3$
1	0.886	0.894	0.903
2	0.940	0.943	0.948
3	0.959	0.961	0.964
4	0.969	0.971	0 973
5	0.975	0.976	0.978
.			
.			
10	0.988	0.988	0.989

The figures illustrate

that the approximation is good,
that the approximation is on the safe side,
and *that* the approximation is rather invariant to variations of the parameter alpha within the given interval.

The safety margin in the approximation—E'_n replacing $\bar{E}_n$—is roughly of the form constant/n.

This is illustrated below for alpha = 2.5

n	$\frac{\bar{E}_n}{E'_n}$	$n \cdot \frac{E'_n - \bar{E}_n}{E'_n}$
1	0.894	0.11
2	0.943	0.11
3	0.961	0.12
4	0.971	0.12
5	0.976	0.12
.		
.		
10	0.988	0.12

We have thus shown that in the Pareto case

$$\frac{E_n}{E'_n} \sim 1$$

and

$$E_n < \bar{E}_n < E'_n = nx_n + E(x_n) = nx_n + n\,\frac{x_n}{\alpha - 1} =$$

$$= nx_n \cdot \frac{\alpha}{\alpha - 1} = \alpha \cdot E(x_n).$$

Thus

$$\frac{E_n}{E(x_n)} \sim \alpha.$$

This means that the LCR risk premium is approximately equal to alpha times the risk premium of an XL cover with a retention chosen in such a way that the expected number of claims is equal to the number of LCR-claims protected.

In the Poisson-Pareto case E'_n gives a handy and fairly good approximation of E_n. The reader is invited to examine other claims size distributions $F(x)$ which are of importance in the practice.

Most such distributions will for all $x > x_0$ have $m''(x) < 0$. We believe that $m''(x) < 0$ will guarantee that E'_n will be a good approximation of E_n with $E'_n > E_n$.

We now give a proof by G. Ottaviani that the inequality $E_n < E'_n$ is valid for any n and for arbitrary distribution functions of the number of claims and of the claim size.

We do not even need the condition of section 2 that the equation $H(x) = n$ has only one solution since the proof will be valid for any X_n, such that $H(x_n) = n$.

Let s denote the total number of claims which occur and $N = \min(s, n)$. We thus allow for the possibility that less than n claims occur.

Let X_n be the set consisting of the N largest claims.

Let

$$\nu(X_n) = E(N)$$
$$\nu(X_n) \leq n \qquad (1)$$

Let $\mu(X_n) = E_n/\nu(X_n)$ be the expected value of a claim in the set X_n.

Analoguesly we denote by X'_n the set consisting of all claims exceeding x_n, the expected number of claims exceeding x_n by $\nu(X'_n)$ and the expected value of a claim in the set X'_n by $\mu(X'_n)$.

We thus have

$$\nu(X'_n) = n \qquad (2)$$

and

$$\mu(X'_n) = x_n + m(x_n).$$

Let

$$Y_n = X_n \cap X'_n$$
$$Z_n = (X_n \cup X'_n) - X'_n$$
$$Z'_n = (X_n \cup X'_n) - X_n$$

$\nu(Y_n)$, $\mu(Y_n)$, $\nu(Z_n)$, $\mu(Z_n)$, $\nu(Z'_n)$, $\mu(Z'_n)$ are defined analoguesly to $\nu(X_n)$ and $\mu(X_n)$. From the above definition it follows directly that

$$\mu(Z_i) < x_n \text{ and} \tag{3}$$

$$\mu(Z'_i) \geq x_n. \tag{4}$$

Thus

$$E_n = \nu(X_i) \cdot \mu(X_i) = \nu(Y_i)\,\mu(Y_i) + \nu(Z_i)\,\mu(Z_i) \tag{5}$$

and

$$E'_n = \nu(X'_i) \cdot \mu(X'_i) = \nu(Y_i)\,\mu(Y_i) + \nu(Z'_i)\,\mu(Z'_i). \tag{6}$$

From (1) and (2) it follows that

$$\nu(Y_i) + \nu(Z_i) = \nu(X_i) \leq n = \nu(X'_i) = \nu(Y_i) + \nu(Z'_i).$$

Thus

$$\nu(Z_i) \leq \nu(Z'_i). \tag{7}$$

From (3) and (4) it follows that

$$\mu(Z_i) < \mu(Z'_i) \tag{8}$$

and from (7) and (8)

$$\nu(Z_i) \cdot \mu(Z_i) < \nu(Z'_i)\,\mu(Z'_i). \tag{9}$$

Adding $\nu(Y_i) \,.\, \mu(Y_i)$ to both sides of (9) and using (5) and (6) leads to

$$E_n < E'_n \qquad \text{q.e.d.}$$

References

[1] AMMETER H., The Rating of "Largest Claim" Reinsurance Covers. Quarterly Letter from the Algemeene Reinsurance Companies Jubilee Number 2, July 1964

[2] BERLINER B., Correlations between Excess of Loss Reinsurance Covers and Reinsurance of the n Largest Claims. The ASTIN Bulletin Vol. VI, Part III, May 1972.

THE EFFICIENCY OF A BONUS-MALUS SYSTEM

NELSON DE PRIL
Belgium

ABSTRACT

The concept efficiency of a bonus-malus system was defined, apparently in a totally different way, consecutively by Loimaranta (1972) and Lemaire (1975, 1976). In this paper we start with a more general model that leads us to a definition of efficiency that contains both earlier ones as special cases Further we introduce the definition of efficiency over a finite planning horizon and consider the efficiency not only for a single risk but also for the entire risk group As a consequence of our approach we can also generalize the concepts excess premium and central value as they were introduced by Loimaranta

1. THE BONUS-MALUS SYSTEM AS MARKOV CHAIN

The basis of a fair tarification in insurance, in our case motorcar insurance, consists in the fact that each policyholder is charged a premium that is proportional to the risk that he actually represents. This risk is determined by a great number of risk factors. Some of them, such as type and use of the car, can be taken into consideration a priori for the tarification and they enable us to split up the heterogeneous collectivity of risks into a number of risk groups which have a more homogeneous risk structure. Other factors cannot be taken into account a priori since they are too difficult to observe, or for social and psychological reasons, or just because one doesn't know all the factors which influence the risk. Due to these factors there will still be accident proneness differentials within a risk group. In the course of time these differentials will be reflected by the individual claim experience of the risk. Therefore one can bring into account a posteriori the earlier neglected risk factors by means of an individual experience rating method, such as a bonus-malus system.

From a point of time $t = 0$ we consider such a risk group in which the tarification is based on a bonus-malus system that is determined by the following factors.

— The length of an insurance period is 1, which means nothing else than that the length of a period is choosen as unit of time.

— The number of classes is n.

— The premium which a risk of class j has to pay at the moment t

to be insured for the period $[t, t+1[$ is $b_t(j)$; $j \in \{1, \ldots, n\}$, $t \in \{0, 1, \ldots\}$.

— The initial class in which a risk is placed at $t = 0$ is the class s.

— The transition rules are given in the form of probabilities $t_{ij}(k)$; $i, j \in \{1, \ldots, n\}$, $k \in \{0, 1, \ldots\}$; where $t_{ij}(k) = 1$ if a risk of class i moves to class j when k claims have occurred in the past period, and $t_{ij}(k) = 0$ if such a risk goes to a class different from j. In order that the transition rules be complete and free of contradictions we must have: for each (i, k) there is one and just one j so that $t_{ij}(k) = 1$.

We assume that the accident proneness of a risk of the considered risk group can be represented by a risk parameter λ, which is the claim frequency of the risk, i.e. the expected number of claims per period for that risk. The value of the risk parameter is regarded as a realization of a random variable Λ, whose distribution function $U(\lambda)$ represents the risk structure of the group. We take that the value of the risk parameter is independent of time. Further we assume that for a given risk λ the random variables which give the number of claims for the successive periods are mutually independent and identically distributed with common probability distribution $p_k(\lambda)$, which depends explicitely and uniquely on the parameter λ.

These assumptions permit us to describe the evolution of a given risk through the bonus-malus system by a Markov chain with constant transition matrix. The probability $p_{ij}^{(t)}(\lambda)$ that a risk λ which is in the class i will be in the class j t periods later, is given by the recursion formula

$$\begin{cases} p_{ij}(\lambda) = \sum_{k=0}^{\infty} p_k(\lambda)\, t_{ij}(k) & (1.a) \\ p_{ij}^{(t)}(\lambda) = \sum_{r=1}^{n} p_{ir}(\lambda)\, p_{rj}^{(t-1)}(\lambda) \quad t = 2, 3, \ldots & (1.b) \end{cases}$$

2. The Efficiency of a Bonus-Malus System

One notices that each country and in some countries even each insurance company has its own bonus-malus system. However all this systems have the same purpose, viz. to come to a fair tarification by adjusting the premiums of each individual policyholder as good as possible to the risk that he actually represents. To measure how good a system fulfils this requirement the concept efficiency is introduced

We denote by $X_\tau(\lambda)$ a random variable that gives the discounted value of all premiums that will be paid by a risk λ in the time interval $[0, \tau[$, $\tau \varepsilon \{1, 2, \ldots\}$.

These premiums are the ones paid at the moments $0, 1, \ldots, \tau - 1$; where the premium at a moment t equals $b_t(j)$ if the risk is in class j at the moment t, and is zero if the risk has by that time left the system. The expectation $E[X_\tau(\lambda)]$ of the discounted value of these premium payments, which is determined by the used bonus-malus system, can be called the *bonus-malus premium for a risk* λ *in* $[0, \tau[$. By $Y_\tau(\lambda)$ we denote a random variable that gives the discounted value of all claim costs of a risk λ in $[0, \tau[$. The expectation $E[Y_\tau(\lambda)]$ of the discounted value of these claim costs represents the *risk premium for a risk* λ *in* $[0, \tau[$.

To verify how good the premium of a certain policy holder corresponds with the risk that he represents we measure the sensibility of the bonus-malus premium by changing risk premium. Therefore we compare a relative variation $\frac{dE[Y_\tau(\lambda)]}{E[Y_\tau(\lambda)]}$ in the risk premium with the relative variation $\frac{dE[X_\tau(\lambda)]}{E[X_\tau(\lambda)]}$ in the bonus-malus premium that it implies. By definition we call *efficiency of a bonus-malus system for a risk* λ *in* $[0, \tau[$ the ratio of these two quantities

$$e_\tau(\lambda) = \frac{\dfrac{dE[X_\tau(\lambda)]}{E[X_\tau(\lambda)]}}{\dfrac{dE[Y_\tau(\lambda)]}{E[Y_\tau(\lambda)]}} = \frac{d \ln E[X_\tau(\lambda)]}{d \ln E[Y_\tau(\lambda)]} \tag{2}$$

The efficiency in $[0, \tau[$ is thus the elasticity of the bonus-malus premium in $[0, \tau[$ with respect to the risk premium in $[0, \tau[$. Put into words this means that for a risk λ a variation of 1% in the expectation of the discounted claim costs in $[0, \tau[$ causes a variation of $e_\tau(\lambda)\%$ in the expectation of the discounted premium payments in $[0, \tau[$.

When we take in (2) the limit for $\tau \to \infty$ we get the efficiency in $[0, \infty[$, viz.

$$e(\lambda) = \lim_{\tau \to \infty} e_\tau(\lambda) \tag{3}$$

A first analysis of the definition of efficiency enables us to make the following observations

A reasonable bonus-malus system got to have a separation effect, so that in an average sense good risks pay lower premiums than bad ones. This means that the relative variations in bonus-malus and risk premium got to have the same sign, so that for each (λ, τ) holds $e_\tau(\lambda) \geq 0$.

The limitcase of a bonus-malus system in which the bonus-malus premium in $[0, \tau[$ remains the same for each risk, corresponds with $e_\tau(\lambda) = 0$ for each λ. This case shows up during the first period when each risk is in the initial class, so that we have $e_1(\lambda) = 0$ for each λ.

The ideal case in which for each risk and for each interval the bonus-malus premium equals the risk premium corresponds with $e_\tau(\lambda) = 1$ for each (λ, τ). In particular $e(\lambda) = 1$ corresponds with an asymptotical correct tarification for a risk λ. The conditions of an ideal system can in general never be met.

In practice a relative increase in the risk premium will generally cause a smaller relative increase in the bonus-malus premium, which means that the good risks have to pay for the bad ones. In general $e_\tau(\lambda)$ will thus lie between the values zero and one. Theoretically we can have $e_\tau(\lambda) > 1$ but such a case of overefficiency in which an increase in the expectation of the claim costs is more than compensated by the increase of the expectation of the premium payments is rarely found.

Because of:

$$E[X_\tau(\lambda)] > 0 \quad , E[Y_\tau(\lambda)] \to 0 \text{ for } \lambda \to 0 \qquad (4)$$
$$E[X_\tau(\lambda)] \text{ bounded}, E[Y_\tau(\lambda)] \to \infty \text{ for } \lambda \to \infty$$

We have in general that for each τ:

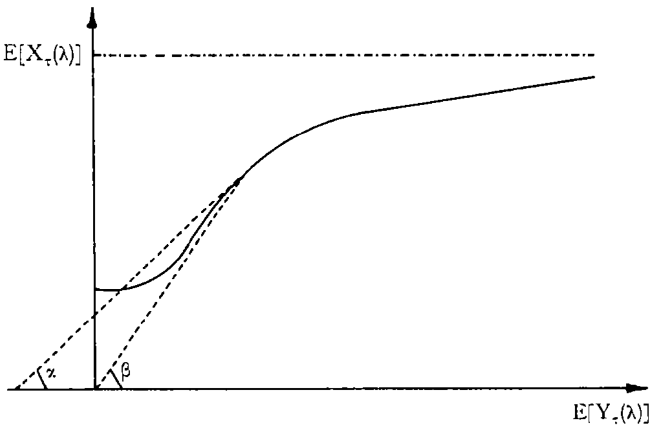
$$\lim_{\lambda \to 0} e_\tau(\lambda) = 0 \text{ and } \lim_{\lambda \to \infty} e_\tau(\lambda) = 0 \qquad (5)$$

Geometrically the definition-formula (2) can be interpreted in the following way

$$e_\tau(\lambda) = \frac{E[Y_\tau(\lambda)]}{E[X_\tau(\lambda)]} \frac{dE[X_\tau(\lambda)]}{dE[Y_\tau(\lambda)]} = \frac{tg\alpha}{tg\beta} \qquad (6)$$

So far the efficiency was defined for a risk with given and known risk parameter λ. The assumption that the risk parameter is known is useful for the development of the theory but is never fulfilled in practice. On the other hand the distribution function of the risk parameter, viz. the structure function $U(\lambda)$, is more likely known, so that it is natural to define the efficiency over the considered risk group. We call *efficiency of a bonus-malus system over a given risk group in* $[0, \tau[$ the expression:

$$e_\tau = \int_\Lambda e_\tau(\lambda)\, dU(\lambda) \qquad (7)$$



and we get for $\tau \to \infty$:

$$e = \int_{\Lambda} e(\lambda)\, dU(\lambda) \tag{8}$$

This averaged efficiency over the risk group enables us to compare the different bonus-malus systems in an objective way.

Further we notice that in our definition of efficiency we could also take into account the so called "bonus-hunger effect" (cfr. Lemaire). This can be done by changing the definition of $X_\tau(\lambda)$ and $Y_\tau(\lambda)$ in an appropriate way.

Finally we remark that our concept of efficiency is not only valid for bonus-malus systems but can be applied to other experience rating systems.

3. Calculation of the Efficiency under Different Assumptions Concerning the Risk Process.

3.1. We consider a risk λ which is placed in class s at $t = 0$ and assume that at the end of each period this risk can either take an insurance for the next period or leave the system. By $w_t(\lambda)$ we denote the probability that the risk λ is insured for the period $[t, t + 1[$. We take that a risk λ which left the system cannot reenter it, so that $w_0(\lambda) = 1 \geq w_1(\lambda) \geq w_2(\lambda) \geq \ldots$. Further we suppose that the average cost of a claim is independent of the number of claims and we denote by $C_t(\lambda)$ the average cost of a claim for a risk λ in the period $[t, t + 1[$. Finally we denote by $\beta \leq 1$ a discount factor.

Under these assumptions we get for the bonus-malus premium of a risk λ in $[0, \tau[$:

$$E[X_\tau(\lambda)] = \sum_{t=0}^{\tau-1} \beta^t w_t(\lambda) \sum_{j=1}^{n} p_{sj}^{(t)}(\lambda)\, b_t(j) \text{ with } p_{sj}^{(0)} = \delta_{sj} \tag{9}$$

and we have for the risk premium

$$E[Y_\tau(\lambda)] = \lambda \sum_{t=0}^{\tau-1} \beta^{t+1/2} w_t(\lambda)\, C_t(\lambda) \tag{10}$$

Using these formulae the efficiency can in principle be calculated. However, additional assumptions concerning the earlier mentioned elements of the risk process seem desirable in order to come to a more manageable expression.

3.2. We suppose now that for each period $[t, t+1[$, both the probabilities $w_t(\lambda)$ and the average claim costs $C_t(\lambda)$ are the same for all risks of the considered risk group, this is that they are independent of the parameter λ.

Under these assumptions formula (2) is reduced to

$$e_\tau(\lambda) = \frac{dE[X_\tau(\lambda)]}{E[X_\tau(\lambda)]} \frac{\lambda}{d\lambda} \tag{11}$$

More explicitly we have

$$e_\tau(\lambda) = \frac{\sum_{t=0}^{\tau-1} \beta^t w_t \sum_{j=1}^{n} \frac{dp_{sj}^{(t)}(\lambda)}{d\lambda} b_t(j)}{\sum_{t=0}^{\tau-1} \beta^t w_t \sum_{j=1}^{n} \frac{p_{sj}^{(t)}(\lambda)}{\lambda} b_t(j)} \tag{12}$$

where the derivatives are determined by the recursion formula

$$\begin{cases} \dfrac{dp_{ij}(\lambda)}{d\lambda} = \displaystyle\sum_{k=0}^{\infty} \frac{dp_k(\lambda)}{d\lambda} t_{ij}(k) & (13.a) \\[2ex] \dfrac{dp_{ij}^{(t)}(\lambda)}{d\lambda} = \displaystyle\sum_{r=1}^{n} \left[\frac{dp_{ir}(\lambda)}{d\lambda} p_{rj}^{(t-1)}(\lambda) + p_{ir}(\lambda) \frac{dp_{rj}^{(t-1)}(\lambda)}{d\lambda}\right] & (13.b) \end{cases}$$

We remark that in the case that the number of claims is Poisson distributed formula (13.a) becomes

$$\frac{dp_{ij}(\lambda)}{d\lambda} = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} [t_{ij}(k+1) - t_{ij}(k)] \tag{14}$$

3.3. Moreover we make an assumption concerning the evolution in time of the premiums $b_t(j)$ and the probabilities w_t. We assume that the premium of each class will increase with the same percentage for each period, this is $b_t(j) = \alpha^t b_j$ with $b_j = b_0(j)$ the value of the premium of class j at constant price and $\alpha \geq 1$ the price index of premiums. Further we assume that the probability to leave the system at the end of a certain period is independent of the considered period and equals ρ, this means $w_t = (1 - \rho)^t$ where $\rho \varepsilon [0, 1]$ is the rate of exit. We shall put $\theta = \beta\alpha (1 - \rho)$, in which we take that $\theta \leq 1$ what is satisfied in practical cases.

Finally we suppose that the Markov chain which is associated to the bonus-malus system is regular. Then the limit probabilities

$$a_j(\lambda) = \lim_{t\to\infty} p_{sj}^{(t)}(\lambda) \tag{15}$$

exist and are independent of the initial class. They are uniquely defined by the system of equations

$$\begin{cases} a_j(\lambda) = \sum_{i=1}^{n} a_i(\lambda)\, p_{ij}(\lambda) & (16.a) \\ \sum_{j=1}^{n} a_j(\lambda) = 1 & (16.b) \end{cases}$$

Under these assumptions equation (9) is reduced to

$$E[X_\tau(\lambda)] = \sum_{t=0}^{\tau-1} \theta^t \sum_{j=1}^{n} p_{sj}^{(t)}(\lambda)\, b_j \tag{17}$$

and if we put

$$b(\lambda) = \sum_{j=1}^{n} a_j(\lambda)\, b_j \tag{18}$$

$$g_{s,\tau-1}(\lambda) = \sum_{t=0}^{\tau-1} \theta^t \sum_{j=1}^{n} [p_{sj}^{(t)}(\lambda) - a_j(\lambda)]\, b_j \tag{19}$$

we get

$$E[X_\tau(\lambda)] = \begin{cases} \dfrac{1-\theta^\tau}{1-\theta}\, b(\lambda) + g_{s,\tau-1}(\lambda) & \theta < 1 \quad (20.a) \\ \tau b(\lambda) + g_{s,\tau-1}(\lambda) & \theta = 1 \quad (20.b) \end{cases}$$

We remark that

$$\lim_{\tau\to\infty} \frac{E[X_\tau(\lambda)]}{\tau} = \begin{cases} 0 & \theta < 1 \quad (21.a) \\ \sum_{j=1}^{n} a_j(\lambda)\, b_j & \theta = 1 \quad (21.b) \end{cases}$$

so that $b(\lambda)$ represents for a risk λ the limit value of the premium per period in the case $\theta = 1$. (e.g. $\alpha = \frac{1}{\beta}$, $\rho = 0$).

The first term in (20) is the discounted expectation of the premium payments in $[0, \tau[$ for a risk λ in the case that the premium to be insured for the period $[t, t+1[$ equals $b(\lambda)\alpha^t$, irrespective of the class in which the risk is placed at the moment t. The second term $g_{s,\tau-1}(\lambda)$ represents the discounted expectation of the extra premium (positive or negative) that has to be paid in $[0, \tau[$, since the premium that a risk λ has to pay to be insured for the period $[t, t+1[$ isn't $b(\lambda)\alpha^t$ but $b_j\alpha^t$, with j the class in which the risk is placed at the moment t. This correction term depends on the initial class s and $g_{s,\tau-1}(\lambda)$ is called the *excess premium of the class s for a risk* λ *in* $[0, \tau[$. The advantage of the introduction of the excess premiums lies in the fact that they simplify to a great extent the calculation of $E[X_\tau(\lambda)]$ and thus of $e_\tau(\lambda)$. It is easy to verify that for the excess premiums the following recursion formula is valid

$$\begin{cases} g_{i,0}(\lambda) = b_i - b(\lambda) & (22.a) \\ g_{i,\tau}(\lambda) = b_i - b(\lambda) + \theta \sum_{j=1}^{n} p_{ij}(\lambda)\, g_{j,\tau-1}(\lambda) \quad \tau = 1, 2, \ldots & (22.b) \end{cases}$$

so that it is no longer necessary to calculate $g_{s,\tau-1}(\lambda)$ from (19), which would require the preliminary computation of all appearing $p_{sj}^{(t)}$.

Further we have that for each (λ, τ) the following relation hold

$$\sum_{i=1}^{n} a_i(\lambda)\, g_{i,\tau}(\lambda) = 0 \qquad \tau = 0, 1, \ldots \tag{23}$$

According to (11) and (20) we get then for the efficiency in $[0, \tau[$

$$e_\tau(\lambda) = \begin{cases} \lambda \dfrac{(1-\theta^\tau)\dfrac{db(\lambda)}{d\lambda} + (1-\theta)\dfrac{dg_{s,\tau-1}(\lambda)}{d\lambda}}{(1-\theta^\tau)\, b(\lambda) + (1-\theta)\, g_{s,\tau-1}(\lambda)} & \theta < 1 \quad (24.a) \\ \lambda \dfrac{\tau \dfrac{db(\lambda)}{d\lambda} + \dfrac{dg_{s,\tau-1}(\lambda)}{d\lambda}}{\tau\, b(\lambda) + g_{s,\tau-1}(\lambda)} & \theta = 1 \quad (24.b) \end{cases}$$

where $b(\lambda)$ can be calculated from (16) and (18), while $g_{s,\tau-1}(\lambda)$ is given by (22) in which (23) is useful for control purposes.

The derivative $\frac{db(\lambda)}{d\lambda} = \sum_{j=1}^{n} \frac{da_j(\lambda)}{d\lambda} b_j$ can be computed from the system of equations obtained by derivating (16)

$$\begin{cases} \frac{da_j(\lambda)}{d\lambda} = \sum_{i=1}^{n} \left[\frac{da_i(\lambda)}{d\lambda} p_{ij}(\lambda) + a_i(\lambda) \frac{dp_{ij}(\lambda)}{d\lambda}\right] & (25.a) \\ \sum_{j=1}^{n} \frac{da_j(\lambda)}{d\lambda} = 0 & (25.b) \end{cases}$$

whereby the derivative of $p_{ij}(\lambda)$ is given by (13.a) or (14).

Finally the derivative of $g_{s,\tau-1}(\lambda)$ is determined by the recursion formula

$$\begin{cases} \frac{dg_{i,0}(\lambda)}{d\lambda} = - \frac{db(\lambda)}{d\lambda} & (26.a) \\ \frac{dg_{i,\tau}(\lambda)}{d\lambda} = - \frac{db(\lambda)}{d\lambda} + \theta \sum_{j=1}^{n} \left[\frac{dp_{ij}(\lambda)}{d\lambda} g_{j,\tau-1}(\lambda) + p_{ij}(\lambda) \frac{dg_{j,\tau-1}(\lambda)}{d\lambda}\right] & \\ \qquad \tau = 1, 2, \ldots & (26.b) \end{cases}$$

and the controlling equations (23) become

$$\sum_{i=1}^{n} \left[\frac{da_i(\lambda)}{d\lambda} g_{i,\tau}(\lambda) + a_i(\lambda) \frac{dg_{i,\tau}(\lambda)}{d\lambda}\right] = 0 \quad \tau = 0, 1, \ldots \qquad (27)$$

To calculate the efficiency in $[0, \infty[$, we first extend the concept excess premium to $[0, \infty[$

$$g_s(\lambda) = \lim_{\tau \to \infty} g_{s,\tau}(\lambda) = \sum_{t=0}^{\infty} \theta^t \sum_{j=1}^{n} [p_{sj}^{(t)}(\lambda) - a_j(\lambda)] b_j \qquad (28)$$

Since the $p_{sj}^{(t)}(\lambda)$ converge geometrically fast to the limit probabilities $a_j(\lambda)$ the series (28) is absolute convergent. When we take in (22.b) and (23) the limit for $\tau \to \infty$ we get the following system of equations for the excess premiums in $[0, \infty[$ of the different classes

$$\begin{cases} g_i(\lambda) = b_i - b(\lambda) + \theta \sum_{j=1}^{n} p_{ij}(\lambda) g_j(\lambda) & (29.a) \\ \sum_{i=1}^{n} a_i(\lambda) g_i(\lambda) = 0 & (29.b) \end{cases}$$

We remark that, in this system of $n + 1$ equations in n unknown, for $\theta < 1$ the equation (29.b) is a consequence of the relations (29.a), while for $\theta = 1$ (29.b) is independent of (29.a) but in this case we have that the relations (29.a) are linear dependent.

In an appendix we shall prove that the $g_i(\lambda)$ are determined in an unique way by the system (29).

For the efficiency in $[0, \infty[$ we have now

$$e(\lambda) = \begin{cases} \lambda \dfrac{\dfrac{db(\lambda)}{d\lambda} + (1-\theta)\dfrac{dg_s(\lambda)}{d\lambda}}{b(\lambda) + (1-\theta)\, g_s(\lambda)} & \theta < 1 \qquad (30.a) \\[2ex] \dfrac{\lambda}{b(\lambda)} \dfrac{db(\lambda)}{d\lambda} & \theta = 1 \qquad (30.b) \end{cases}$$

in which $g_s(\lambda)$ can be calculated from (29), while its derivative is determined in an unique way (cfr. appendix) by the system of equations

$$\begin{cases} \dfrac{dg_i(\lambda)}{d\lambda} = -\dfrac{db(\lambda)}{d\lambda} + \theta \displaystyle\sum_{j=1}^{n} \left[\dfrac{dp_{ij}(\lambda)}{d\lambda}\, g_j(\lambda) + p_{ij}(\lambda)\dfrac{dg_j(\lambda)}{d\lambda}\right] & (31.a) \\[2ex] \displaystyle\sum_{i=1}^{n} \left[\dfrac{da_i(\lambda)}{d\lambda}\, g_i(\lambda) + a_i(\lambda)\dfrac{dg_i(\lambda)}{d\lambda}\right] = 0 & (31.b) \end{cases}$$

So we find as a special case (30.b) the definition of efficiency given by Loimaranta.

4. The Central Value

We consider the equation

$$E[X_\tau(\lambda)] = E[Y_\tau(\lambda)] \qquad (32)$$

which expresses the equality between the bonus-malus premium and the risk premium for a risk λ in $[0, \tau[$. Because of the relations (4) equation (32) has at least one solution λ_τ^*. We call a solution λ_τ^* of (32) a *central value of the bonus-malus system in* $[0, \tau[$.

We assume now that $w_i(\lambda)$ and $C_i(\lambda)$ are independent of λ and we shall show that the central value in $[0, \tau[$ is unique if $e_\tau(\lambda) < 1$ for all λ.

From (11) we have

$$d\ln E\,[X_\tau(\lambda)] = e_\tau(\lambda)\; d\ln\lambda$$

which gives if we integrate with λ_τ^* as initial value

$$E[X_\tau(\lambda)] = E[X_\tau(\lambda_\tau^*)]\ e^{\int_{\lambda_\tau^*}^{\lambda} e_\tau(\lambda)\, d\ln\lambda}$$

where

$$E[X_\tau(\lambda_\tau^*)] = E[Y_\tau(\lambda_\tau^*)] = \frac{\lambda_\tau^*}{\lambda} E[Y_\tau(\lambda)]$$

so that

$$E[X_\tau(\lambda)] = \begin{cases} E[Y_\tau(\lambda)]\ e^{\int_{\lambda}^{\lambda_\tau^*} [1 - e_\tau(\lambda)]\, d\ln\lambda} & \lambda \leq \lambda_\tau^* \qquad (33.a) \\ E[Y_\tau(\lambda)]\ e^{-\int_{\lambda_\tau^*}^{\lambda} [1 - e_\tau(\lambda)]\, d\ln\lambda} & \lambda \geq \lambda_\tau^* \qquad (33.b) \end{cases}$$

If $e_\tau(\lambda) < 1$ for all λ we have thus that λ_τ^* is unique and that

$$E[X_\tau(\lambda)] \gtreqless E[Y_\tau(\lambda)] \qquad \text{if } \lambda \lesseqgtr \lambda_\tau^*$$

Now we make some assumptions that will permit us to rewrite equation (32) in an easier form. As in section 3.3. we assume that $b_t(j) = \alpha^t b_j$, $w_t = (1-\rho)^t$ and $\theta \leq 1$, where $\theta = \beta\alpha\ (1-\rho)$. Further we assume that the evolution in time of the average claim cost can be given in the form $C_t(\lambda) = \gamma^t C$, with $C = C_0$ the average cost at constant price and $\gamma \geq 1$ the price index of claims. Hereby we put $\varepsilon = \beta\gamma\ (1-\rho)$ and take that $\varepsilon \leq 1$. The central value in $[0, \tau[$, λ_τ^*, is then the solution of the equation

$$b(\lambda) \sum_{t=0}^{\tau-1} \theta^t + g_{s,\tau-1}(\lambda) = \lambda C \sum_{t=0}^{\tau-1} \varepsilon^t \qquad (35)$$

In particular we have that $\lambda_1^* = \dfrac{b_s}{C}$.

We call central value in $[0, \infty[$

$$\lambda^* = \lim_{\tau\to\infty} \lambda_\tau^* \qquad (36)$$

and we distinguish the following cases. In the case $\theta < 1$, $\varepsilon < 1$ we have from (35) that λ^* is the solution of the equation

$$\frac{b(\lambda)}{1-\theta} + g_s(\lambda) = \frac{\lambda C}{1-\varepsilon} \qquad (37)$$

For $\theta = 1$, $\varepsilon < 1$ we have that $\lambda^* \to \infty$, while for $\theta < 1$, $\varepsilon = 1$ holds $\lambda^* \to 0$. Finally in the case $\theta = \varepsilon = 1$ we obtain that λ^* is the solution of

$$b(\lambda) = \lambda C \qquad (38)$$

The solution of this last equation corresponds with the concept central value as introduced by Loimaranta.

5. Appendix

We shall prove that the systems (29) and (31) of $n + 1$ equations in n unknown have an unique solution.

5.1. We make use of the following two lemmas, in which 0 denotes the zero matrix and I the unit matrix.

Lemma 1

If Q is a square matrix and Q^k tends to 0 as k tends to infinity, then

$$\det (I - Q) \neq 0$$

and $(I - Q)^{-1} = I + Q + Q^2 + \ldots = \sum_{k=0}^{\infty} Q^k$

Proof

see e.g. Kemeny and Snell p. 22.

Lemma 2

If to the x-th row (column) of the blocks of a partioned matrix Q we add the y-th row (column) multiplied on the left (right) by a rectangular matrix R of the corresponding dimensions, then the rank of Q remains unchanged under this transformation and, if Q is a square matrix, the determinant of Q is also unchanged.

Proof

see e.g. Gantmacher p. 45.

We introduce the following matrix notations

A : $1 \times n$ matrix with elements $a_i(\lambda)$
B : $n \times 1$ matrix with elements b_i
G : $n \times 1$ matrix with elements $g_i(\lambda)$
P : $n \times n$ matrix with elements $p_{ij}(\lambda)$
E : $n \times 1$ matrix with all elements equal to 1
$D = b(\lambda) E$: $n \times 1$ matrix with all elements equal to $b(\lambda)$
$M = EA$: $n \times n$ matrix whose rows are all identical and equal to A

According to (15), (16) and (18) we have then

$$\lim_{k\to\infty} P^k = M$$
$$AP = A,\ AE = 1$$
$$AB = b(\lambda)$$

5.2. We now prove that the system (29) has an unique solution. In matrix notations this system becomes

$$\begin{cases} (\theta P - I)G = D - B \\ AG = 0 \end{cases} \tag{39}$$

The necessary and sufficient conditions for an unique solution are

$$\text{rank} \begin{bmatrix} \theta P - I \\ \\ A \end{bmatrix} = n,\ det \begin{bmatrix} \theta P - I & D - B \\ \\ A & 0 \end{bmatrix} = 0 \tag{40}$$

According to lemma 2 we have

$$\text{rank} \begin{bmatrix} \theta P - I \\ \\ A \end{bmatrix} = \text{rank} \begin{bmatrix} \theta(P - M) - I \\ \\ A \end{bmatrix}$$

where we have substracted from the first row the last row multiplied on the left by θE. Since for each power q holds $M^q = M$ we have $(P - M)^k = P^k - M$. From $\lim_{k\to\infty} P^k = M$ it follows then that $\lim_{k\to\infty} [\theta(P - M)]^k = 0$ for each $\theta \leq 1$ and we have that $\det\,[\theta(P - M) - I] \neq 0$ according to lemma 1. This shows that the coefficient-matrix has rank n.

To prove that the determinant in (40) is zero we make the following transformations:

$$\det \begin{bmatrix} \theta P - I & D - B \\ \\ A & 0 \end{bmatrix} = \det \begin{bmatrix} \theta(P - M) - I & D - B \\ \\ A & 0 \end{bmatrix}$$

$$= \det \begin{bmatrix} \theta(P - M) - I & D - B \\ \\ 0 & -A\,[\theta(P - M) - I]^{-1}\,(D - B) \end{bmatrix}$$

$$= -A\,[\theta(P - M) - I]^{-1}\,(D - B).\ \det[\theta(P - M) - I]$$

We have now

$$-A\,[\theta(P - M) - I]^{-1} = A\,[I - \theta(P - M)]^{-1}$$
$$= A \sum_{k=0}^{\infty} [\theta(P - M)]^k$$

$$= A[I + \sum_{k=1}^{\infty} \theta^k (P^k - M)]$$

$$= A + \sum_{k=1}^{\infty} \theta^k A(P^k - M)$$

$$= A$$

with $A(D - B) = 0$, which completes the proof.

5.3. For the system (31) we have

$$\begin{cases} (\theta P - I)G' = D' - \theta P'G \\ AG' = - A'G \end{cases} \tag{41}$$

where a quote indicates derivation with respect to λ.

The necessary and sufficient conditions for an unique solution are

$$\text{rank} \begin{bmatrix} \theta P - I \\ A \end{bmatrix} = n, \ \det \begin{bmatrix} \theta P - I & D' - \theta P'G \\ A & - A'G \end{bmatrix} = 0 \tag{42}$$

The first condition is the same as in section 5.2. and is thus satisfied. For the second condition we obtain after some transformations

$$\det \begin{bmatrix} \theta P - I & D' - \theta P'G \\ A & - A'G \end{bmatrix}$$

$$= \{- A'G - A[\theta(P - M) - I]^{-1} [D' - \theta P'G + \theta EA'G]\} .$$
$$\det [\theta(P - M) - I]$$

The proof follows now from

$$- A'G + AD' - \theta AP'G + \theta A'G = - A'G + AD' + \theta A'PG$$
$$= A' [(\theta P - I)G + B]$$
$$= A'D = 0.$$

6. References

GANTMACHER, F. R., (1960), The Theory of Matrices, Vol. I, Chelsea Publishing Company, New York.

KEMENY, J. G. and SNELL J L., (1960), Finite Markov Chains, D. Van Nostrand Company, New York

LEMAIRE, J., (1975), Si les assurés connaissaient la programmation dynamique. . Bulletin de l'Association Royale des Actuaires Belges, N° 70, pp. 54-63.

LEMAIRE, J., (1976), Driver versus Company—Optimal Behaviour of the Policy Holder, Scandinavian Actuarial Journal, N° 1976-4, pp. 209-219.

LOIMARANTA, K., (1972), Some asymptotic properties of bonus systems. Astin Bulletin, Vol. VI, Part 3, pp. 233-245.

ON A FORM OF AUTOMOBILE LIABILITY INSURANCE WITH A PREPAID DISCOUNT

RICCARDO OTTAVIANI

1. Recently two types of insurance policy covering automobile liability* have become available. With this kind of policy, the insured party receives a prepaid discount on the annual premium; however, he must make an additional payment to the insurer on first report of an accident.

We shall examine only one of these two types of policy, since the second is very similar to the first.

On stipulation of the contract, the insured party pays a premium equal to 78 percent of that currently in force for complete coverage, depending on the various limits. However, at the same time at which the insured reports his first accident in the course of the year (and only in this case), he must make an additional payment equal to 35 percent of the premium due in the case of complete coverage. This additional payment, considered as a deposit, is repaid to the insured party if the accident has no follow up within four months, however, it may be requested again by the insurer if the case is subsequently reopened and leads to payment of damages. This additional payment will thus become a part of the premium only in the case of payment of damages for the first accident reported.

In this way, the premium paid by the insured is equal to 78 percent or to 113 percent of the current premium, depending on whether or not the insured reports accidents.

2. Let us consider an insurance policy for complete coverage, excepting limits, and let us assume that

P = insurance premium
M = quota necessary to cover purchase and administrative expenses and profits
S = pure premium, i e., the quota necessary to cover accident payments.

Therefore, premium P can be expressed as $P = M + S$.

In Italy, if the insurance premium is equal to 100, the part covering pure premium S is equal to 75 lire.

* In Italy termed R.C.A , or "Responsabilità Civile Auto"

On this basis let us now analyze the prepaid discount policy and the integration of the premium at first report of an accident.

We shall indicate by

q : the probability that the insured has no accident during the insurance period

p_1: the probability that the insured has at least one accident but reports none

p_2: the probability that the insured has at least one reported accident.

Obviously $q + p_1 + p_2 = 1$.

We shall further indicate by:

S_1 the average cost of unreported accidents for each insured party

S_2 the average cost of reported accidents for each insured party.

Therefore

$$S_1 + S_2 = S = 75.$$

We note that the accidents leading to average accident S_1 are not only those whose probability is p_1, but also some cases included in probability p_2, because the insured party might not report a first accident and subsequently report another more serious accident.

We shall indicate by M_1 the margin for expenses and profits relative to the prepaid discount policy. This can be expressed by:

$$1)\ M_1 = 78\,(p_1 + q) + 113p_2 - S_2$$

and thus, considering that

$$S_2 = 75 - S_1$$

we have

$$2)\ M_1 = 3 + 35p_2 + S_1.$$

If we overlook the differences in expenses between a normal policy and a prepaid discount one (in fact, in the former consideration must be given to possibly greater administrative costs due to considering all accidents and in the latter to the possibly greater administrative costs due to considering the double payment of the premium), from a technical point of view, an insurer will benefit by offering a prepaid discount policy rather than a complete coverage policy if the margin for expenses and profit of the discounted policy is greater than the margin for a complete coverage policy, i.e., if

$$M_1 \geqslant M = 25.$$

Let us now determine the value of M_1 based on A.N.I.A.* statistics for automobile accidents.

In order to determine margin M_1 it is necessary to make several hypotheses concerning the behavior of the insured parties in the presence of accidents that may or may not be reported. In fact, probabilities p_1 and p_2 depend on the insured party's greater or lesser inclination to report minor accidents.

Let us calculate this margin on the basis of the following hypotheses.

a) All insured parties stipulate prepaid discount policies and report all accidents regardless of their amount. In this case, probability p_2 will be equal to the probability that the insured has at least one accident, and this probability will be slightly less than the frequency of accidents (in fact, some insured parties may cause and report several accidents during the period considered).

This frequency, on the basis of A.N.I.A. automobile statistics, was equal to 33.35% in 1972 and 32.37% in 1973.

These values include settled and unsettled accident cases over the year, and thus include also accident cases that may prove to have no follow up in subsequent years.

Value S_1 is equal to zero, since, for the hypotheses made, all accidents are reported.

Then the margin for expenses and insurer's profits, on the basis of hypothesis (a) and relative to 1973 data, is

$$M_1 \leqslant 3 + 35 \cdot 0.3237 = 14.33$$

which proves to be less than margin $M = 25$ which the insurer obtains in the case of a normal policy.

Hence the insurer, on the basis of these hypotheses and from a technical point of view, would have no interest in offering prepaid discount policies.

b) Let us now hypothesize that all insured parties stipulate prepaid discount policies, but that they report accidents only if they have already reported at least one other previous accident, or, if not, if the first accident represents a presumable value greater than 70,000 lire, which value, being exclusive of the technical expenses of verification and settlement, would become approximately 100,000 lire should the insurer have to pay damages.

In this case, p_2 will prove to be slightly less than the frequency

* A.N.I.A.: Associazione Nazionale tra le Imprese Assicuratrici (National Association of Insurance Companies)

of accidents greater or equal to 100,000 lire, which, on the basis of A.N.I.A. statistics, proved to be 26.01% in 1972 and 28.63% in 1973.

Average cost S_1 of unreported accidents, i.e., of those accidents causing damages of less than 100,000 not preceded by accidents causing damages greater than or equal to 100,000 lire, is obviously slightly less (the difference can be overlooked) than the average cost of accidents having a value of less than 100,000 lire. This average cost, on the basis of A.N.I.A. statistics, for each 75 lire of pure premium, was equal to 16.69 lire in 1972 and 15.11 lire in 1973.

In this case, on the basis of A.N.I.A. statistics for 1973, margin M_1 for the insurer will be

$$M_1 \leqslant 3 + 35 - 0.2863 + 15.11 = 28.13.$$

Consequently, the insurer might benefit technically be offering prepaid discount policies, since it is probable that $M_1 \geqslant M = 25$.

The insured having the possibility of choosing between complete coverage policies and prepaid discount policies, we note that those who select the latter are those who expect to have fewer accidents than the majority of automobile drivers, i.e., those for whom probability p_2 should be lower than the probability deduced from the A.N.I.A. statistics.

It is obvious that in this case margin M_1 obtained by the company is still lower than that which it would obtain if all insured parties held prepaid discount policies.

We note moreover that, if all insured parties decide to report accidents having a value of less than the additional payment, which is 35% of the premium, then the hypotheses for case (b) refer only to those insured parties who pay a premium for the entire insurance coverage greater than 200,000 lire, and this group includes only a minority of insured parties.

Unfortunately, it has not been possible to study cases of the type given in hypotheses (b) where the limit of 70,000 lire for unreported accidents is lowered. In fact, in the A.N.I.A. statistics of accident number by value, the first class considered is that of accidents having a value of up to 100,000 lire.

The considerations set forth so far refer to a case in which all the insurance companies offer exclusively the conditional discount policy in place of the normal policy. We can note that, should all the companies offer a choice between a conditional discount policy and a normal policy, there would be a selection of the insured parties choosing the discount policy.

These insured parties would obviously be those who expected to gain by this type of policy, i.e., those who expect to cause on an average fewer accidents in the course of the year than the majority of automobile drivers. Consequently, for these drivers, the value of probability p_2^* of having at least one reported accident is less than corresponding value p_2 for the majority of the insured. From this, substituting in (2) value p_2^* for value p_2, we obtain a margin $M_1^* < M_1$.

Moreover, with the transfer of the "good" insured parties from a normal policy to a discounted policy, there is a prior selection of the insured who continue to use a normal policy; and therefore pure premium S may not be sufficient to cover the accidents of the normal policies, and thus margin M would have to be lowered as well.

In a case where a single company (or a limited number of companies) offered the conditional discount policy as an alternative to the normal policy, while the conditions we have just described for a case where all companies offered both types of policy would still hold, we can note that there would be a request for discount policies also on the part of "good" insured parties who previously held normal policies with other companies; and thus the company (or companies) offering the prepaid discount policy would show an increase in business.

TESTING GOODNESS-OF-FIT OF AN ESTIMATED RUN-OFF TRIANGLE

G. C. TAYLOR
Sydney, Australia

1. THE RUN-OFF TRIANGLE — ACTUAL AND EXPECTED

By the term *actual run-off triangle* we shall mean the two-way tabulation—according to year of origin and year of payment—of claims paid to date, which has the following form:

Development Year

Year of origin	0	1	2	.	.	.	k
0	C_{00}	C_{01}	C_{02}	.	.	.	C_{0k}
1	C_{10}	C_{11}	C_{12}	.	.	.	
2	C_{20}	C_{21}	C_{22}	.	.		
.	.			.			
.	.		.				
.	.	.					
k	C_{k0}						

where C_{ij} is the amount paid during development year j in respect of claims whose year of origin is i.

The information relating to the area below and/or to the right of this triangle is unknown since it represents the future development of various cohorts of claims.

Now in seeking to use this triangle as a basis for projection of claims in future development years for each of the years of origin 0, 1, 2, etc., we must recognise that the entries C_{ij} in the above triangle, being random variables, contain random deviations from their expected values μ_{ij}. It is the corresponding triangle of these expected values in which we are interested, and which shall be called the *expected run-off triangle*.

Explicitly, it is:

μ_{00}	μ_{01}	μ_{02}	.	.	.	μ_{0k}
μ_{10}	μ_{11}	μ_{12}	.	.	.	
μ_{20}	μ_{21}	μ_{22}	.	.		
.			.			
.		.				
.	.					
μ_{k0}						

2. The Requirement of a Test of Goodness-of-fit.

One method of projecting future claims is to identify some internal structure within the expected run-off triangle and hence extrapolate outside it. In this respect, a commonly made assumption is the following:

Assumption 1

In the absence of any disturbing influences, e.g. claims cost inflation, changing rate of growth of volume of business etc., the distribution of expected claim delays remains constant over varying years of origin.

We can represent this assumption symbolically. If R_{ij} is the observed proportion of all claim payments in respect of year of origin i made in development year j after removal of the "disturbing influences" referred to above, then $E(R_{ij}) = r_j$ independent of i. Examples of estimation procedures based on this assumption can be found in Beard (1974) and Taylor (1977).

Naturally, if a model based on Assumption 1 is to be used for projection of future claims, it is necessary to check at some stage that this model accords with experience (i.e. that the expected run-off triangle based on the model accords with the actual run-off model) within statistically reasonable limits. Hence the need for a test of goodness-of-fit.

Suppose that the "disturbing influences" in the triangle have been determined so that it is possible to remove them from the data. Let C'_{ij} be the result of adjusting C_{ij} for removal of these influences. Then, according to Assumption 1,

$$\mu'_{ij} (= E(C'_{ij})) = C'_i r_j,$$

where C'_i denotes total claims (some still to be paid) in respect of year of origin i after removal of disturbing influences.

Estimation procedures based on Assumption 1 will produce estimates $\hat{\mu}'_{ij}$ of μ'_{ij}, where $\hat{\mu}'_{ij} = C'_i \hat{r}_j$ and $\hat{r}_j$ is an estimate of r_j. It is then necessary to apply a significance test to the deviations $(C'_{ij} - \hat{\mu}'_{ij})$.

One tempting possibility is to set up a contingency table containing the cells as displayed below:

3. A Contingency Table test?

<table>
<tr><td>(0, 0)</td><td>(0, 1)</td><td>(0, 2)</td><td colspan="2">. . .</td><td>(0, k)</td><td>(0, k +)</td></tr>
<tr><td>(1, 0)</td><td>(1, 1)</td><td>(1, 2)</td><td>. . .</td><td>(1, k — 1)</td><td colspan="2">(1, (k — 1) +)</td></tr>
<tr><td>(2, 0)</td><td>(2, 1)</td><td>(2, 2)</td><td>. . .</td><td colspan="3">(2, (k — 2) +)</td></tr>
<tr><td>⋮</td><td></td><td colspan="5"></td></tr>
<tr><td>(k, 0)</td><td colspan="6">(k, 0 +)</td></tr>
</table>

Here the $(i, (k-i)+)$ cell relates to data for year of origin i and development years $k-i+1$, $k-i+2$, etc. combined. The standard chi-square test might then be applied to this table as in the theorem in Section 30.3 of Cramér (1946, 426-7).

There are, however, several points to be noted in connection with this suggestion.

Firstly, the triangle of previous sections has been augmented with extra cells to form a square. This has been done in conformity with the theorem quoted above which requires that for a given year of origin, the probability of a randomly chosen unit of claim payment being found in some cell of the table should be unity. This augmentation of the triangle can cause difficulties because data may not be available in respect of the extra cells. This point receives further comment in the later section dealing with numerical examples.

Secondly, and more importantly, it is implicit in the theorem quoted above (see both the statement of it on P. 427 and the proof on P. 429) that the marginal distribution of each C'_{ij} is binomial. In the present circumstances this is not true and, in fact, is sufficiently untrue to have important consequences for the contingency table test, as will be dealt with in the next section.

Thirdly, an examination of the theorem stated by Cramér reveals that the chi-square test is strictly applicable only when the expected cell frequencies have been determined by the modified χ^2 minimum method of estimation. When this method has not in fact been used, some consideration should be devoted to the closeness of this and the method actually used. For example, the "separation method" used by Taylor (1977) is not always equivalent to the modified χ^2 minimum method, but is, as shown in Section 6 of that paper, identical in certain cases to the maximum likelihood method which, as pointed out by Cramér (1946, 426), is in turn equivalent to the modified χ^2 minimum method.

4. Modification of the Standard Chi-square Test of a Contingency Table.

The most important of the objections raised against the standard chi-square test is the second which concerns the marginal distributions of the individual cell frequencies. As noted there, the standard test requires that the (i, j) - cell frequency be binomial. The parameters of this binomial distribution would be C'_i and r_j, and hence the variance would be

$$v_{ij} = C'_i r_j (1 - r_j) = \mu'_{ij} (1 - r_j) \tag{1}$$

As also noted in the previous section, the distribution of C'_{ij} will not be binomial in fact. In order to approximate its correct form we make two further assumptions.

Assumption 2

The number of claims pertaining to the (i, j) - cell is a stationary Poisson variable.

Assumption 3

The sizes of the individual claims pertaining to the (i, j) - cell are i. i. d. random variables.

It follows from these two assumptions that C'_{ij} is a compound Poisson variable with variance:

$$\sigma^2_{ij} = \mu'_{ij} \times \frac{\alpha_{2j}}{\alpha_{1j}} \tag{2}$$

where α_{1j}, α_{2j} are the first and second moments (about the origin) respectively of individual claim size in development year j.

It is now evident that in those cases where μ'_{ij} is not too small the compound Poisson distribution of C'_{ij} and the binomial distribution with the same mean and variance (1) will be rather similar except that the former will have a variance greater than that of the latter by a factor of

$$\frac{\sigma^2_{ij}}{v_{ij}} = \frac{\alpha_{2j}}{\alpha_{1j}(1 - r_j)}. \tag{3}$$

Thus, if the standard chi-square statistic,

$$\sum_{\text{all cells}} (C'_{ij} - \mu'_{ij})^2 / \mu'_{ij}$$

is replaced by:

$$\chi^2 = \sum_{\text{all cells}} (C'_{ij} - \mu'_{ij})^2 \, v_{ij} \, / \, \sigma^2_{ij} \, \mu'_{ij}$$

$$= \sum_{\text{all cells}} (1 - r_j) \left(\frac{\alpha_{1j}}{\alpha_{2j}}\right) (C'_{ij} - \mu'_{ij})^2 \, / \, \mu'_{ij}, \tag{4}$$

then χ^2 can be assumed to have an approximate chi-square distribution with an appropriate number of degrees of freedom.

Suppose that it is desired that a significance test be applied to the

Null Hypothesis: $r_j = \hat{r}_j$ for each j.

Then it follows from (4) and the hypothesis that

$$\hat{\chi}^2 = \sum_{\text{all cells}} (1 - \hat{r}_j) \left(\frac{\alpha_{1j}}{\alpha_{2j}}\right) (C'_{ij} - \hat{\mu}'_{ij})^2 \, / \, \hat{\mu}'_{ij} \tag{5}$$

is a chi-square statistic and can be tested as such for significance.

5. Applying the Modified Test in Practice.

All quantities appearing in statistic (5) are immediately available with the exception of the ratio $(\alpha_{1j}/\alpha_{2j})$. If the investigation is being carried out by an individual company in respect of its own experience, then this ratio can be estimated by means of a cost-band analysis of claims.

On the other hand, if the test is being applied by a supervisory authority, it is unlikely that any cost-band information will be available for estimation of $(\alpha_{1j}/\alpha_{2j})$. The authority will however have returns from each company and may, therefore, consider ways of estimating the ratio from this data.

The slender evidence to which the author had access (a confidential report) suggested that α_{1j}/α_{2j} was *not* independent of company, but that, for a given class of insurance, the coefficient of variation, $w_j = \alpha_{2j}/\alpha^2_{1j}$, varied comparatively little between different companies. This suggests estimating w_j by $\hat{w}_j$, based on data from all companies and replacing $\hat{\chi}^2$ by the alternative statistic:

$$\hat{\hat{\chi}}^2 = \sum_{\text{all cells}} \left[\frac{C'_{ij} - \hat{\mu}'_{ij}}{\hat{\mu}'_{ij}}\right]^2 \hat{n}_{ij} \, \frac{1 - \hat{r}_j}{\hat{w}_j} \tag{6}$$

where n_{ij} is the expected number of claims paid in development year j of year of origin i, and $\hat{n}_{ij}$ estimates n_{ij}.

The difficulty now is, of course, the estimation of w_j. For this purpose, let

r_{jt} denote the value of r_j in the t-th company (for a particular class of insurance);
C'_{ijt} denote the random variable C'_{ij} in the t-th company;
$C'_{i\cdot t}$ denote the constant C'_i in the k-th company.

Let us suppose that, for fixed j, the r_{jt} are realizations of a random variable with mean p_j and variance z_j. Suppose also that $C_{i_1jk_1}$ and $C_{i_2jk_2}$ are stochastically independent whenever $(i_1, k_1) \neq (i_2, k_2)$.

Then it is not difficult to show that, for each i, j,

$$\begin{aligned} Var\,[C'_{ijt} / C'_{i\cdot t}] &= E_{r_{jt}}[Var\,[C'_{ijt} / C'_{i\cdot t} \mid r_{jt}]] \\ &\quad + Var_{r_{jt}}[E\,[C'_{ijt} / C'_{i\cdot t} \mid r_{jt}]] \\ &= E_{r_{jt}}[w_j\, r^2_{ijt}] + Var_{r_{jt}}[r_{jt}]. \end{aligned}$$

i.e.

$$w_j = \frac{Var\,[C'_{ijt} / C'_{i\cdot t}] - Var\,[r_{jt}]}{E[r^2_{jt}]}. \tag{7}$$

A reasonable estimate $\hat{w}_j$ of w_j can be obtained by replacing each of the three terms on the right of (7) by an estimator. The first term of the numerator can be estimated from the sample variance of the ratios $(C'_{ijt} / C'_{i\cdot t})$ for fixed j. However, the other two terms present difficulties, since the corresponding sample statistics depend upon the observed values of r_{jt} for companies other than the one to which the significance test is being applied. These r_{jt} are neither known nor the subject of our hypothesis.

The simplest way out of the difficulty appears to be as follows:

1. Use some method which is known to be generally fairly reliable to obtain an estimate of r_{jt} for each j and t.
2. Use these estimates to calculate the sample statistics corresponding to the quantities appearing in (7).
3. Use these sample statistics to obtain an estimate of w_j as already described.

A second practical difficulty arises from the appearance of the quantities C'_i in our formulas. These quantities, being total payments after run-off has been completed, are of course unknown for any cohorts not fully developed.

However, this situation is not quite as serious as it might at first appear. Let us consider the impact of the C_i on each of the terms of (6) in turn.

Firstly,

$$\sum_j (C'_{ij} - \hat{\mu}'_{ij}) = C'_i [\sum_j R_{ij} - \sum_j \hat{r}_j]$$
$$= 0, \tag{8}$$

since both summations yield unity. Thus,

$$C'_{i,(k-i)+} - \hat{\mu}'_{i,(k-i)+} = - \sum_{j=0}^{k-i} (C'_{ij} - \hat{\mu}'_{ij}), \tag{9}$$

and so the terms $[(C'_{ij} - \hat{\mu}'_{ij}) / \hat{\mu}'_{ij}]^2$ are all fully determined.

Secondly, the term $C'_{i \cdot t}$ appears in $\hat{w}_j$ (see (7)). Here it is possible to use equation (8) again and obtain

$$C'_{i \cdot t} = \sum_j C'_{ijt} = \sum_j \hat{\mu}'_{ijt}. \tag{10}$$

Finally the value n_{ij} can be estimated by $\hat{n}_{ij}$, the *actual number of claims* pertaining to the (i, j) - cell.

All of the terms appearing in (6) are now determined.

6. A Practical Simplification of the Test Statistic.

The procedure outlined in the previous section for estimating w_j is complicated and involves lengthy computations. Moreover, no idea of the stability of the estimate of w_j has been obtained.

However, experience indicates that, even in the relatively stable class of business such as private motor insurance, w_j tends to be rarely less than unity. These occasions on which it is < 1 are usually just those on which r_j is relatively large. The result of this is that usually (always ?) we have

$$\frac{1 - r_j}{w_j} < 1. \tag{11}$$

combining (5) and (11) we see that

$$\hat{\chi}^2 < \sum_{\text{all cells}} \left[\frac{C'_{ij} - \hat{\mu}'_{ij}}{\hat{\mu}'_{ij}}\right]^2 \hat{n}_{ij}, \tag{12}$$

and so deduce that treating the right side of (12) as a chi-square statistic amounts to applying a somewhat too stringent test to the hypothesis. The overstringency is not too great, at least for motor portfolios, as typical values of the factor $(1 - r_j) / w_j$ appear to lie in the range 0.3 to 0.7.

7. A Numerical Example.

Let us apply the simplified test developed in Section 6 to the run-off triangle dealt with in Example 1 of Taylor (1977). The actual triangle with each C_{ij} divided by $10^{-3} \times$ numbers of claims for year of origin i, is:

50.4	28.2	9.0	4.8
58.0	29.2	9.7	
59.5	33.2		
66.2			

Multiplying by claim numbers to obtain the C_{ij} gives:

2481	1387	441	237
2899	1463	485	
3126	1744		
3538			

The calculations in Taylor (1975) yield the following C'_{ij}'s:

2481	1217	374	180
2533	1239	368	
2648	1323		
2684			

and the following array of $\hat{\mu}'_{ij}$'s:

2480	1223	368	179	234
2522	1244	374		420
2647	1306			833
2695				2178

There is a certain degree of arbitrariness in the values of $\hat{\mu}_{i,(3-i)+}$ which were not determined by Taylor (1975). These will not affect the result materially, however.

Finally the triangle of $\hat{n}_{ij}$'s is:

30034	13309	960	393	164
30678	12974	1216		458
31461	15417			1783
31386				22045

From these figures we readily obtain:

$$\sum_{\text{all cells}} \left[\frac{C'_{ij} - \hat{\mu}'_{ij}}{\hat{\mu}'_{ij}}\right]^2 \hat{n}_{ij} = 6.25.$$

Now a value of 6.25 for χ^2_3 is not significant at the 5% level and so, recalling that the true χ^2_3 statistic would be appreciably less than 6.25, we should have no hesitation in accepting that the model produced by the separation technique and leading to the above $\hat{\mu}'_{ij}$'s is quite plausible statistically.

8. Acknowledgements.

The author acknowledges the assistance of Macquarie University, Sydney, Australia and Government Actuary's Department, London, U.K. in the preparation of this paper. Several discussions of the subject of the paper with Professor R. E. Beard proved invaluable.

References

Beard, R. E., (1974). Some historical, theoretical and practical aspects Appears in *Claims provisions for non-life insurance business* Institute of Mathematics and its Applications

Cramér, H., (1946). *Mathematical methods in statistics*, edition. Princeton University Press.

Taylor, G. C., (1977). Separation of inflation and other effects from the distribution of non-life insurance claim delays The ASTIN Bulletin, 9, 219-30.

AN INVESTIGATION OF THE USE OF WEIGHTED AVERAGES IN THE ESTIMATION OF THE MEAN OF A LONG-TAILED CLAIM SIZE DISTRIBUTION

G. C. TAYLOR

Sydney, Australia.

ABSTRACT

The paper discusses the problem of estimating the mean of a long-tailed claim size distribution when the investigator's knowledge of the distribution is only vague

One method of dealing with this problem, the method developed by Johnson and Hey, is examined and found to produce strongly biased estimators

The situation in which a sufficient statistic (but nothing else) for the claim size distribution is known is examined, and an approximately unbiased estimator developed This estimator is substantially more efficient than the arithmetic mean in some cases. It appears to be quite successful when the sufficient statistic is real-valued. It is of limited use when the sufficient statistic is vector-valued.

1. THE PROBLEM OF LONG-TAILED CLAIM SIZE DISTRIBUTIONS

For the purposes of this paper we can take a *long-tailed distribution* to be one whose density converges to zero less rapidly than the simple exponential family. Such distributions occur relatively frequently in the field of nonlife insurance. They are particularly prevalent among the distributions of individual claim sizes in respect of fire policies and liability policies.

Since the mean of a distribution is one of its most important properties—and indeed in the context of claim size distributions, usually the most important property—it is necessary that one have as reliable a method as possible for the estimation of this parameter.

In nonlife insurance this estimating problem can prove quite troublesome, because of the fact that standard statistical techniques are of limited applicability. This statement deserves some explanation particularly as the majority of this paper is concerned with methods which lie outside the scope of "standard" methods.

The statistician faced with the problem of estimating the mean of a long-tailed (or any other) distribution would begin by defining the family of likelihoods which are admissible as a representation of the distribution under consideration. He would then select estimates of the unknown parameters according to some optimization criterion, e.g. maximum likelihood, minimum-variance unbiasedness, etc.

The difficulty for the actuary involved with nonlife insurance arises at the very first stage, i.e. in deciding the admissible likelihood functions. In practice, he may have only the vaguest notion of the shape of the distribution. For example, he may be prepared to assert that it is within the exponential family of likelihoods. The exponential family is an extremely large one, so that although the requirement of delimiting the admissible likelihoods has been satisfied technically, the practical benefit of this stage of the procedure is doubtful.

It is basically for this reason that alternative methods of approaching the estimation of mean claim size are necessary. Of course, one can estimate this parameter with the sample mean. This has the advantage of ensuring unbiasedness, but, as is well-known, the sample mean from some long-tailed distributions has rather a large variance. Since unbiasedness and small variance are properties which one would usually like an estimator to possess simultaneously, the need for considering estimators other than the sample mean is immediate.

2. The Johnson-Hey Method of Weighted Averages

Hey (1970), concerned by the disturbance to the sample mean of claim sizes resulting from a few but substantial large claims, suggested that the difficulty might be alleviated by using a *weighted average* of the sample claim sizes, the weights tending to decrease with increasing claim size. This suggestion was followed up by Johnson and Hey (1972).

To state this in mathematical terms, they were concerned that the sample mean claim size, though an unbiased estimator of the true mean, had too large a variance. Their solution was to estimate the true mean claim size m by means of the statistic:

$$M = (\sum_{i=1}^{n} S(C_i)/n) \times G, \qquad (1)$$

where

$C_1, C_2, \ldots, C_n$ are the sample values of claim size;
$S(\cdot)$ is a weight function which is nondecreasing but whose first derivative is nonincreasing;
G is a "grossing-up factor" which is so chosen that M is an unbiased estimate of m.

3. Purpose of the Paper

The purpose of the present paper is three-fold:

(i) to indicate certain dangers arising from use of the Johnson-Hey (*J-H*) method;

(ii) to point out that there are sound theoretical reasons for introducing the transformation S;

(iii) to investigate ways other than the *J-H* method of producing an estimate of m from the statistics $C_1, \ldots, C_n$.

4. Some Comments on the Johnson-Hey Method

It is clear from a brief scrutiny of formula (1) that the problematic factor is G. Hey himself (1970, p. 81) noted that "we have no knowledge of the sensitivity of the grossing-up factor". Other difficulties arising from the manner in which G is estimated are mentioned by Johnson and Hey (1972, pp. 227-8).

In this section, however, we shall ignore these difficulties by assuming that, for a single given m, it is possible to choose G exactly correctly. We shall see that difficulties still arise in the use of estimator M.

Let us deal with nonzero claims only and assume that their sizes are sampled from a lognormal distribution. It is to be emphasised that this particular distribution has been chosen for illustrative purposes only, though, as Hey (1970, pp. 62-3) and others remark, it is not far from the truth for some classes of motor insurance.

Thus, we assume that $C_1, C_2, \ldots, C_n$ is a random sample in which each $\log C_i$ has a normal distribution with mean μ and variance σ^2. Then, as in well-known (see e g. Kendall and Stuart, 1961, p. 68),

$$m = E[C_i] = \exp\{\mu + \tfrac{1}{2}\sigma^2\}. \qquad (2)$$

Also

$$E[\sum_{i=1}^{n} \log C_i/n] = \mu. \qquad (3)$$

Thus, if we choose $S(\cdot)$ as $\log(\cdot)$, then it follows from (1), (2) and (3) that, for M to be unbiased, it is necessary that

$$G = \mu^{-1}\exp\{\mu + \tfrac{1}{2}\sigma^2\}. \qquad (4)$$

A difficulty arises here due to the fact that G is dependent (often quite strongly) on μ and σ^2. This means that, if G is appropriate to some particular μ and σ^2, it may not be appropriate to some other choice of these parameters. This is the reason for the phenomenon noted by Johnson and Hey (1972, p. 228) that G appears to vary between different risk categories.

In order to appreciate the extent of the difficulty, it is necessary to understand that the *J-H* method provides that G be calculated from the *aggregation of data from all risk categories* in such a way that the estimate of m for the risk category of *an individual chosen at random from the whole portfolio* is unbiased. Note that, despite this type of unbiasedness, the resulting estimators may be biased in respect of each separate risk category, and the bias will of course be worse for the more extreme categories.

A number of simulations were carried out to illustrate this point and some of the results are given in Tables 1 and 2. The sampling distribution for claim size was taken to be lognormal with parameters μ and σ^2, though, as is fairly obvious, the point being illustrated here is valid for other distributions too. This was confirmed by other simulations whose results are not reproduced here. The portfolio was assumed to consist of five different risk categories. In each case $S(\cdot)$ was taken as log $(\cdot)$.

TABLE 1

Risk Category	μ	σ^2	True Mean $\exp\{\mu + \frac{1}{2}\sigma^2\}$	Arithmetic Mean *sample size = 10	sample s. = 50	sample s. = 250	*J-H* estimate sample s. = 10	sample s. = 50	sample s. = 250
1	4.000	1.20	99	94	99	99	128	128	130
2	4.250	1.10	122	127	123	120	137	136	138
3	4.500	1.00	148	139	146	149	144	144	146
4	4.625	0.95	164	164	162	165	148	149	150
5	4.750	0.90	181	188	179	183	154	152	154

* sample size in each risk category

TABLE 2

Risk Category	μ	σ^2	True Mean $\exp\{\mu + \frac{1}{2}\sigma^2\}$	Arithmetic Mean *sample size = 10	sample s. = 50	sample s. = 250	*J-H* estimate sample s. = 10	sample s. = 50	sample s. = 250
1	4.48	1.00	145	149	147	147	151	149	148
2	4.49	1.00	147	147	147	148	150	149	148
3	4.50	1.00	148	157	150	148	153	149	149
4	4.51	1.00	150	152	150	151	152	149	149
5	4.52	1.00	151	152	151	151	151	150	149

* sample size in each risk category

The main effect of the *J-H* method appears clearly in Table 1 where it can be seen that, although the true mean varies over risk categories by a factor of 1.82, the *J-H* estimates vary by a factor of the order of only 1.2 approximately. Generally, the *J-H* estimates for the various risk categories are "squashed together", with high-risk categories underestimated and low-risk categories overestimated.

This "squashedness" of the *J-H* estimates has obvious implications for tariff-splitting.

The same phenomenon becomes apparent upon a scrutiny of Johnson and Hey's own results presented on p. 227 of their paper. However, it is not quite so obvious there because their simulated portfolio is rather like that represented in Table 2 of this paper, i.e. risk categories are all quite close together.

Thus, as the portfolio becomes more homogeneous, so is the bias in the *J-H* method reduced. But then so also is the need for recognizing different risk categories. Regrettably, we must conclude that the *J-H* method attains reasonable effectiveness only when it is least needed.

5. Theoretical Justification for Weighted Average

Let us consider the family of likelihoods, dependent upon some parameter θ, which have the form:

$$f(x \mid \theta) = c(\theta)\, h(x) \exp\left[\sum_{j=1}^{p} \pi_j(\theta)\, t_j(x)\right]. \tag{5}$$

This is the so-called *exponential family of likelihoods*. It is very rich in the sense that, for most of our practically occurring distributions, we can find a member of the family which will serve as a good approximation.

Moreover, the exponential family has a number of attractive properties which make it relatively easy to work with. In particular (see e.g. Ferguson (1967, pp. 125-37)):

1. The statistic $T = \left(\sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_p(X_i)\right)$ is a sufficient statistic, i.e. contains just as much information as does the whole vector of observations $X_1, \ldots, X_n$ in a sample of size n.
2. The likelihood of T is also a member of the exponential family, with the same π_j's as in $f(x \mid \theta)$.
3. Under rather weak conditions which will usually be met by an insurance portfolio, it is possible to conclude that, if $g(T)$ is an unbiased estimator of a function of θ, then it has the smallest variance among all unbiased estimators.

Since the object of Johnson and Hey's quest was stability of the estimator, Property 3 is particularly suggestive, although it must be mentioned that this property does not preclude the existence of *more stable but biased* estimators.

Now if a claim size distribution is a member of the exponential family, then, by Property 1,

$$T = \left(\sum_{i=1}^{n} t_1(C_i), \ldots, \sum_{i=1}^{n} t_p(C_i) \right)$$

is a sufficient statistic. We thus have in Property 3 a theoretical justification for basing our estimate of θ on the average (or, equivalently, the sum) of *transformed claim sizes*. Furthermore, the transformation to be used is by no means arbitrary, but is determined by (5).

The usefulness of this observation is seen fully when viewed against the background of the actuary's vague knowledge of the shape of the distribution, as described in Section 1. If the situation is slightly better than described there and the actuary is willing to assert that $p = 1$ and $t_1(.) = \log(.)$, then from this none too definitive assertion, we may deduce that θ should be estimated by some function of $\sum_{i=1}^{n} \log C_i$.

6. An Example of the Use of Transformed Claim Sizes

Suppose that C has a lognormal distribution with parameter $\theta = (\mu, \sigma^2)$, then

$$f(C \mid \theta) = (\sqrt{2\pi}\, \sigma C)^{-1} \exp[-(\log C - \mu)^2/2\sigma^2]$$

$$= c(\theta)\, h(C) \exp[\pi_1(\theta)\, t_1(C) + \pi_2(\theta)\, t_2(C)],$$

$$c(\theta) = (\sqrt{2\pi}\, \sigma)^{-1} \exp[-\mu^2/2\sigma^2],$$

$$h(C) = C^{-1},$$

$$\pi_1(\theta) = \mu/\sigma^2,\ t_1(C) = \log C,$$

$$\pi_2(\theta) = -1/2\sigma^2,\ t_2(C) = (\log C)^2.$$

Thus we lose no information from our claim size observations if we reduce them to the two values,

$$T_1 = \frac{1}{n} \sum_{i=1}^{n} \log C_i \quad \text{and} \quad T_2 = \frac{1}{n} \sum_{i=1}^{n} (\log C_i)^2 - T_1^2.$$

It is not immediately clear how an unbiased estimator is to be constructed from T_1 and T_2. However, in the case of the lognormal distribution, it was shown by Finney (1941) that an unbiased estimator of $E[C]$ is

$$\exp(T_1)\, g(\tfrac{1}{2} T_2), \tag{6}$$

where

$$g(x) = 1 + x + \frac{n-1}{n+1}\frac{x^2}{2!} + \frac{(n-1)^2}{(n+1)\ (n+3)}\frac{x^3}{3!} + \dots \tag{7}$$

For large n, $g(x)$ does not differ by too much from e^x, so that (6) becomes approximately:

$$\{\prod_{i=1}^{n} C_i\}^{1/n} \exp\left[\tfrac{1}{2} T_2\right]. \tag{8}$$

This is approximately unbiased, and so, by Property 3 above, has small variance.

We have thus constructed an unbiased estimator with small variance in terms of transformed claim size, where the transformation is:

$$C \rightsquigarrow (\log C, (\log C)^2).$$

7. Further Development of the Use of Transformed Claim Sizes

It is apparent that the method used in the previous section for estimating $E[C]$ when C is lognormally distributed differs considerably from the *J-H* method. It was also pointed out that the methods used there lead to minimum-variance unbiased estimators.

Unfortunately, however, the actuary may not be in a position to make as strong an assertion as that claim size is lognormally distributed. Possibly the strongest assertion he can make with any confidence is that claim sizes, after some prescribed transformation (e.g. log) are roughly exponentially distributed. This really amounts to asserting something like the order of convergence of the probability density of claim size.

Under these circumstances, it is natural to seek some extension of the method used in Section 6. This aim is pursued in this section, but it should be stated at the outset that the success achieved in this direction is limited, and perhaps the main result emerging from the study is that, when knowledge of the claim size distribution is as vague as above, the simple arithmetic mean is surprisingly efficient.

Let us suppose that the sample of claim sizes, $C_1, C_2, \dots, C_n$, is drawn from a distribution belonging to the exponential family with t_1 a one-to-one transform. Henceforth we denote t_1 by just t. The statistic,

$$T_n = \frac{1}{n}\sum_{i=1}^{n} t(C_i), \tag{9}$$

is a minimum-variance unbiased estimator of $E[t(C)]$, by Properties 2 and 3 given in Section 5. It is therefore reasonable to assume that the statistic $t^{-1}(T_n)$, after approximate correction for bias will provide an estimator of $E[C]$ of relatively small variance.

Let us write

$$\mu = E[t(C)], \quad \sigma^2 = \text{Var}\,[t(C)].$$

From Section 2,

$$m = E[C].$$

Now, we know that

$$E[t^{-1}(T_1)] = E[t_1^{-1}(t_1(C))] = m. \tag{10}$$

We therefore need to estimate the difference,

$$E[t^{-1}(T_n)] - E[t^{-1}(T_1)],$$

occasioned by increase of sample size from 1 to n. This change represents the bias in $t^{-1}(T_n)$ as an estimator of m.

Let us now write Z_n for the standardized version of T_n, i.e.

$$Z_n = \frac{T_n - \eta}{\sigma\sqrt{n}}.$$

Let the d.f. of Z_n be expanded in an Edgeworth series,

$$\sum_{k=0}^{\infty} c_k^n \, \Phi^{(k)}(z),$$

where, as usual, $\Phi^{(k)}$ is the k-th derivative of the standard normal d.f. Then

$$E[t^{-1}(T_n)] = \Sigma\, c_k^n\, E^{(k)}\,[t^{-1}(n^{-1/2}\,\sigma Z_n + \eta)], \tag{11}$$

where $E^{(k)}$ [function of Z_n] denotes the expected value of the argument on the assumption that Z_n has "distribution function" $\Phi^{(k)}$.

Now, if D denotes the differentiation operator, repeated integration by parts gives

$$E^{(k)}\,[t^{-1}\,(\eta + n^{-1/2}\,\sigma Z_n)] = n^{-k/2}\,(-\sigma)^k\, E^{(0)}\,[D^k t^{-1}(\eta + n^{-1/2}\,\sigma Z_n)], \tag{12}$$

under obvious regularity conditions on the functions t^{-1}, Dt^{-1}, $D^2 t^{-1}$, etc.

Thus, by (11) and (12),

$$E[t^{-1}(T_n)] = \sum_{k=0}^{\infty} c_k^n\, n^{-k/2}\,(-\sigma)^k E^{(0)}\,[D^k t^{-1}\,(\eta + n^{-1/2}\,\sigma Z_n)]. \tag{13}$$

It is apparent from (10) and (13) that

$$E[t^{-1}(T_n)] = m - \sum_{k=0}^{\infty} (-\sigma)^k \{c_k^1 E^{(0)} [D^k t^{-1} (\eta + \sigma Z)] - n^{-k/2} c_k^n E^{(0)} [D^k t^{-1} (\eta + n^{-1/2} \sigma Z)]\},$$

where the subscripts on Z have been suppressed since they are made irrelevant by the distributional assumptions implied by $E^{(0)}$. Hence an unbiased estimator of m is

$$U_n = t^{-1}(T_n) + \sum_{k=0}^{\infty} (-\sigma)^k \{c_k^1 E^{(0)} [D^k t^{-1} (\eta + \sigma Z)] - n^{-k/2} c_k^n E^{(0)} [D^k t^{-1} (\eta + n^{-1/2} \sigma Z)]\}, \qquad (14)$$

Now it is known that

$$c_0^n = 1, \; c_1^n = c_2^n = 0, \; c_3^n = -\tfrac{1}{6} n^{-1/2} \gamma_1, \; c_4^n = \tfrac{1}{24} n^{-1} \gamma_2,$$

where γ_1 and γ_2 are the coefficients of skewness and excess respectively of T_n. Moreover,

$$\gamma_1 \sigma^3 = K_3,$$
$$\gamma_2 \sigma^4 = K_4,$$

where K_j is the j-th cumulant of $t(C)$.

Using these facts, we can simplify (14) somewhat to give:

$$\begin{aligned} U_n = {} & t^{-1}(T_n) \\ & + \{E^{(0)} [t^{-1} (\eta + \sigma Z)] - E^{(0)} [t^{-1} (\eta + n^{-1/2} \sigma Z)]\} \\ & + \tfrac{1}{6} K_3 \{E^{(0)} [D^3 t^{-1} (\eta + \sigma Z)] - n^{-2} E^{(0)} [D^3 t^{-1} (\eta + n^{-1/2} \sigma Z)]\} \\ & + \tfrac{1}{24} K_4 \{E^{(0)} [D^4 t^{-1} (\eta + \sigma Z)] - n^{-3} E^{(0)} [D^4 t^{-1} (\eta + n^{-1/2} \sigma Z)]\} \\ & + \dots . \end{aligned}$$

Since we do not have true values of μ, σ, K_3 and K_4, we replace them by estimates. The obvious choices are (see Cramer, 1946, 352)

$$\hat{\eta} = T_n, \quad \hat{\sigma} = \sqrt{\frac{n}{n-1}\, a_2},$$

$$\hat{K}_3 = \frac{n^2}{(n-1)(n-2)}\, a_3,$$

$$\hat{K}_4 = \frac{n^2}{(n-1)(n-2)(n-3)} \left[(n+1) \frac{a_4}{a_2^2} - 3(n-1)\right],$$

where a_ν is the ν-th sample central moment of $t(C)$.

Thus we finally adopt as our estimator of n the statistic:

$$\begin{aligned}\hat{m} = {} & t^{-1}(T_n) \\ & + \{E^{(0)}[t^{-1}(T_n + \hat{\sigma}Z)] - E^{(0)}[t^{-1}(T_n + n^{-\frac{1}{2}}\hat{\sigma}Z)]\} \\ & + \tfrac{1}{6}\hat{K}_3\{E^{(0)}[D^3t^{-1}(T_n + \hat{\sigma}Z)] - n^{-2}E^{(0)}[D^3t^{-1}(T_n + n^{-\frac{1}{2}}\hat{\sigma}Z)]\} \\ & + \tfrac{1}{24}\hat{K}_4\{E^{(0)}[D^4t^{-1}(T_n + \hat{\sigma}Z)] - n^{-3}E^{(0)}[D^4t^{-1}(T_n + n^{-\frac{1}{2}}\hat{\sigma}Z)]\} \\ & + \ldots \qquad (15)\end{aligned}$$

It is of course apparent that $\hat{m}$ is not in general unbiased. However, the inclusion of the corrective terms should remove the majority of the bias which would be present if $t^{-1}(T_n)$ alone were taken as estimator of m.

8. Numerical Results

Although the development of $\hat{m}$ as an estimator of m began with considerations which rested on sound theory (see Section 5), a number of subsequent approximations have led to the position in which the bias and stability of $\hat{m}$ are not entirely clear. For this reason, a number of simulations were carried out in order to compare the estimator $\hat{m}$ with the simple arithmetic mean for bias and stability. The most informative results are summarized in Tables 3 and 4 below.

In Table 3 the sampling distribution for claim size was taken to be log-Laplacian, i.e. $\log C (= L$, say) was taken to have a likelihood function, dependent upon parameter k, equal to

$$\tfrac{1}{2}k \exp[-k\,|\,L\,|\,], \quad -\infty < L < \infty.$$

Table 3

Risk Category	k	True Mean $k^2/(k^2-1)$	Arithmetic Mean			$\hat{m}$		
			sample *size = 10	sample size = 50	sample size = 250	sample size = 10	sample size = 50	sample size = 250
1	1.10	5.8	2.3(4.2)	3.7(40.1)	4.0(27.5)	6.3(884)	3.5(10.2)	3.3(1.8)
2	1.30	2.4	2.6(46.8)	2.2(1.7)	2.4(1.4)	5.1(679)	2.2(1.4)	2.2(0.2)
3	1.49	1.8	1.7(1.8)	1.8(1.0)	1.8(0.1)	2.1(10.3)	1.8(0.8)	1.7(0.1)
4	1.70	1.5	1.5(0.9)	1.6(0.3)	1.5(0.02)	1.7(3.9)	1.6(0.2)	1.5(0.02)
5	1.89	1.4	1.4(0.3)	1.4(0.1)	1.4(0.02)	1.5(0.5)	1.4(0.1)	1.4(0.01)

*sample size in each risk category

In Table 4, the sampling distribution was taken to be lognormal as in Tables 1 and 2. As in Tables 1 and 2, the portfolio is assumed to consist of five risk categories, and $t(\cdot)$ is taken to be log (.). The figures for "arithmetic mean" and $\hat{m}$ are simulated values of these estimators. The figures in parentheses are the corresponding simulated values of the variances of the estimates.

TABLE 4

Risk Cate-gory	μ	σ^2	True Mean $\exp\{\mu+\frac{1}{2}\sigma^2\}$	Arithmetic Mean sample *size = 10	sample size = 50	sample size = 250	$\hat{m}$ sample size = 10	sample size = 50	sample size = 250
1	4.000	1.20	99	94(1800)	99(550)	100(110)	98(2200)	101(630)	100(110)
2	4.250	1.10	122	127(3100)	123(700)	120(100)	135(5100)	124(750)	121(110)
3	4.500	1.00	148	139(3200)	146(750)	149(160)	143(4100)	147(800)	149(160)
4	4.625	0.95	164	164(5600)	162(740)	165(150)	173(10000)	163(780)	165(150)
5	4.750	0.90	181	188(5900)	179(870)	183(140)	195(9600)	181(930)	183(140)

* sample size in each risk category

9. CONCLUSIONS

The theme of the paper has been the estimation of mean claim size in the light of only vague information about the claim size distribution. When this information includes knowledge of a sufficient statistic, it is tempting to base the estimator on this statistic.

One such estimator is provided by the Johnson-Hey method, but Section 4, and particularly Table 1 therein, reveals that there are quite common situations in which this estimator gives poor results.

The estimator $\hat{m}$ developed in Section 7 attempts to improve on the *J-H* method. Indeed, Table 3 indicates that for some long-tailed claim size distributions, this estimator is largely unbiased and achieves a significant reduction in variance as compared with a simple arithmetic mean. The longer the tail, the larger is the reduction in variance.

The usefulness of $\hat{m}$ as an estimator is limited, however, as is evidenced by Table 4 where the variance of $\hat{m}$ is slightly *greater* than the variance of the arithmetic mean. The reason for this is, presumably, that the sufficient statistic for the distribution involved here is an ordered pair rather than a single real value (as in the case of Table 3), and in such a case the transformation (9) makes only partial use of our knowledge of the sufficient statistic.

Perhaps the estimator $\hat{m}$ can be refined to make fuller use of the sufficient statistic?

Perhaps also the main conclusion to be drawn from this investigation is that, in the possession of only the vague knowledge outlined in Section 1, it is often very difficult to improve upon the simple arithmetic mean as an estimator of mean claim size.

10. ACKNOWLEDGEMENT

The author gratefully acknowledges the use of computing and other facilities provided by the University of Essex, Colchester, England and Heriot-Watt University, Edinburgh, Scotland.

References

CRAMÉR, H., 1946. *Mathematical methods in statistics* Princeton University Press.

FERGUSON, J. S., 1967 *Mathematical statistics: a decision-theoretic approach* Academic Press.

FINNEY, D. J., 1941. On the distribution of a variate whose logarithm is normally distributed. *Supplement of the Journal of the Royal Statistical Society*, 7, 155.

HEY, G. B., 1970. Statistics in non-life insurance *Journal of the Royal Statistical Society (Series A)*, 133, 56-85

JOHNSON, P. D. and HEY, G. B., 1972 Statistical review of a motor insurance portfolio. *Astin Bulletin, VI*, 222-32.

KENDALL M. G. and STUART, A., 1963. *The theory of advanced statistics, Vol. I (2nd. ed.)*. Charles Griffin and Co Ltd.

PARAMETER ESTIMATION IN CREDIBILITY THEORY

FL. DE VYLDER
University of Louvain, Belgium

ABSTRACT.

The problem of distribution-free parameter estimation in recent credibility theory is discussed in the papers [1], [3] and [4] of the bibliography. Here, we consider a multiclass model with regression assumption. In that case, already treated by Ch. Hachemeister, [3], this author obtains an unsymmetrical matrix as an estimator of a covariance matrix. Of course, for practical use, this matrix is symmetrized in the obvious way. We show that this procedure can be avoided and that a lot of symmetrical unbiased estimators can be obtained at once.

By particularisations to the 1-rank model, we find the estimators given by Buhlmann and Straub, [1], [4].

In the multirank case, a generalization of the minimumvariance principle (minimization of the trace of the covariance matrix) leads to an optimal estimator of the mean regression vector. A noteworthy conclusion of our discussion is that there is no difference at all between the various credibility formulae (the inhomogenous formula, the homogeneous formula, the mean-free formula) if the mean regression vector is estimated optimally.

Finally we show that it must not be hoped to find, in a large set of unbiased estimators of the covariance matrix, one estimator furnishing, always, a semidefinite positive estimate

1. THE MULTICLASS MODEL WITH REGRESSION ASSUMPTION.

1.1. *Description of the model*

We consider the array of observable random variables

$$\begin{array}{cccccccc}
{}_1X_1 & {}_2X_1 & \ldots & {}_jX_1 & \ldots & {}_kX_1 \\
{}_1X_2 & {}_2X_2 & \ldots & {}_jX_2 & \ldots & {}_kX_2 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
{}_1X_s & {}_2X_s & \ldots & {}_jX_s & \ldots & {}_kX_s \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
{}_1X_t & {}_2X_t & \ldots & {}_jX_t & \ldots & {}_kX_t
\end{array}$$

In the notation ${}_jX_s$ ($j = 1, 2, \ldots, k; s = 1, 2, \ldots, t$), the left index j is the class index, the right index s is the year index. For example ${}_jX_s$ might be the claim rate of treaty j in year s in a reinsurer's portfolio, but other interpretations are possible. The column ${}_jX = ({}_jX_1, {}_jX_2, \ldots, {}_jX_t)'$, will be called the *class j*. To ${}_jX$ is associated the structure variable ${}_j\Theta$. We abbreviate:

$$\Theta = ({}_1\Theta, {}_2\Theta, \ldots, {}_k\Theta).$$

The numbers k (number of classes) and t (number of observation years) are fixed. As variable class-indices we use $i, j = 1, 2, \ldots, k$ and as variable time indices $r, s = 1, 2, \ldots, t$.

Before we specify the assumptions relating the observable and structure variables, we make some general remarks about the matrix notation used throughout the text. A $\frac{n}{m}$ matrix is one with m rows and n columns. Then $\frac{n}{m}$ is the *dimension* of the matrix. Rows, columns, scalars are particular matrices. The dimension $\frac{1}{1}$ is also denoted more simply by 1. Some relations are completed by the dimensions of the displayed matrices. The same symbol (without indices) is used for a matrix and for its elements (with indices). The inferior right index is the row-index. The superior right index is the column-index. *Matrix rules are applied to indices written on the right only.*

The following assumptions are made.

(i) Independence of classes: ${}_1X, {}_2X, \ldots, {}_kX$ are independent.

(ii) In each class, irrelevancy of other parameter values than that one of the given class: For each class-index j and function $f(\,.\,)$,

$$E(f({}_jX)/\Theta) = E(f({}_jX)/{}_j\Theta).$$

(iii) Independence of parameters: ${}_1\Theta, {}_2\Theta, \ldots, {}_k\Theta$ are independent.

(iv) Equidistribution of the parameters ${}_1\Theta, {}_2\Theta, \ldots, {}_k\Theta$.

(v) There exist functions $\mu_s(.)$ satisfying

$$E({}_jX_s/{}_j\Theta) = \mu_s({}_j\Theta).$$

(The assumption is in the fact that $\mu_s(\,.\,)$ does not depend on j.)

(vi) There exist symmetrical definite positive $\frac{t}{t}$ matrices ${}_jv$ and a scalar function $\sigma^2(\,.\,)$ satisfying

$$\mathrm{COV}({}_jX_s, {}_jX_r/{}_j\Theta) = \sigma^2({}_j\Theta)\, {}_jv_r^s.$$

(vii) Regression assumption: For each j, the $\frac{1}{t}$ column $\mu({}_j\Theta)$ of elements $\mu_s({}_j\Theta)$, can be written as

$$\underset{\substack{t\\1}}{\mu({}_j\Theta)} = \underset{\substack{g\\t}}{y}\,\underset{\substack{1\\g}}{\beta({}_j\Theta)},$$

where y is a known $\frac{g}{t}$ matrix and $\beta(\,.\,)$ a $\frac{1}{g}$ vector of elements $\beta_p(\,.\,)$. It is assumed moreover that y is of rank g and that $g < t$.

1.2. *Special assumptions*

Occasionally, (vi), (vii) will be specified in the following manner.

(BS): ${}_jv$ is a diagonal matrix with diagonal elements

$$1/{}_jp_1,\ 1/{}_jp_2,\ \ldots,\ 1/{}_jp_t.$$

(BS1): (BS) is true and moreover, y is the $\overset{1}{t}$ column $y = (1, 1, \ldots, 1)'$. (Of course, then $\beta(\,.\,)$ is a scalar function.)

The case (BS) is introduced in Bühlmann and Straub, [1] and is further used in Hachemeister [4]. In [1], the number ${}_jp_s$ is the premium volume underlying treaty j in year s in a reinsurer's portfolio. In [4], each class is related to an American state and ${}_jp_s$ is a number of claims in state j in the observation period s.

Assumption (BS1) is a stationarity in time assumption, since then $\mu_s(.)$ does not depend on s.

In the sequel we assume (i) to (vii). The matrices ${}_jv$, y are supposed to be known. Assumptions (BS), (BS1) are mentioned explicitly if they are used.

1.3. *Summary of credibility theory results*

The following credibility approximations to the vector $\beta({}_j\Theta)$ are known.

— The inhomogeneous approximation ([3], [5], [2])

$$\underset{\substack{1\\g}}{{}_j\hat{B}} = \underset{\substack{g\\g}}{(1 - {}_jz)}\ \underset{\substack{1\\g}}{b} + \underset{\substack{g\\g}}{{}_jz}\ \underset{\substack{1\\g}}{{}_jB}.$$

— The homogeneous approximation ([3], [2])

$$\underset{\substack{1\\g}}{{}_j\tilde{B}} = \underset{\substack{g\\g}}{(1 - {}_jz)}\ \underset{\substack{1\\g}}{b}\ \underset{1}{S} + \underset{\substack{g\\g}}{{}_jz}\ \underset{\substack{1\\g}}{{}_jB}.$$

— The homogeneous mean-free approximation ([5], [2])

$$\underset{\substack{1\\g}}{{}_j\ddot{B}} = \underset{\substack{g\\g}}{(1 - {}_jz)}\ \underset{\substack{1\\g}}{B} + \underset{\substack{g\\g}}{{}_jz}\ \underset{\substack{1\\g}}{{}_jB}.$$

In these formulae:

$$\underset{\substack{g\\g}}{{}_jz} = \underset{\substack{g\\g}}{a}\ \underset{\substack{t\\g}}{y'}\ \underset{\substack{t\\t}}{{}_jd^{-1}}\ \underset{\substack{g\\t}}{y}\ \Big(\underset{\substack{g\\g}}{1} + \underset{\substack{g\\g}}{a}\ \underset{\substack{t\\g}}{y'}\ \underset{\substack{t\\t}}{{}_jd^{-1}}\ \underset{\substack{g\\t}}{y}\Big)^{-1},$$

$$\underset{\substack{1\\g}}{{}_jB} = \Big(\underset{\substack{t\\g}}{y'}\ \underset{\substack{t\\t}}{{}_jd^{-1}}\ \underset{\substack{g\\t}}{y}\Big)^{-1}\ \underset{\substack{t\\g}}{y'}\ \underset{\substack{t\\t}}{{}_jd^{-1}}\ \underset{\substack{1\\t}}{{}_jX},$$

$$\underset{\frac{1}{g}}{B} = (\sum_i \underset{\frac{g}{g}}{{}_iz})^{-1} \sum_i \underset{\frac{g}{g}}{{}_iz}\, \underset{\frac{1}{g}}{{}_iB},$$

$$\underset{1}{S} = \sum_i \underset{\frac{g}{1}}{b'}\, \underset{\frac{g}{g}}{a^{-1}}\, \underset{\frac{g}{g}}{{}_iz}\, \underset{\frac{1}{g}}{{}_iB} \Big/ \sum_i \underset{\frac{g}{1}}{b'}\, \underset{\frac{g}{g}}{a^{-1}}\, \underset{\frac{g}{g}}{{}_iz}\, \underset{\frac{1}{g}}{b}$$

and further:

1) b is the $\frac{1}{g}$ vector of elements

$$b_p = E\ (\beta_p({}_j\Theta)),\ \text{independent of } j.$$

2) ${}_jd$ is the $\frac{t}{t}$ matrix of elements

$${}_jd_r^s = E\ \mathrm{COV}\ ({}_jX_r, {}_jX_s \,/\, {}_j\Theta) = E(\sigma^2({}_j\Theta))\ {}_jv_r^s = s^2\ {}_jv_r^s,$$

where

$$s^2 = E(\sigma^2({}_j\Theta)),\ \text{independent of } j.$$

3) a is the $\frac{g}{g}$ matrix of elements

$$a_p^q = \mathrm{COV}(\beta_p({}_j\Theta), \beta_q({}_j\Theta)),\ \text{independent of } j.$$

1.4. *Problem*

Our problem is to find unbiased estimators for b, s^2, a. For brevity, these quantities will be called, respectively, the *mean vector*, the *variance*, the *covariance matrix*.

2. Fixed-Class Estimators.

In this section we consider a fixed class ${}_jX$ and we make inferences based only on the variables in that class.

2.1. *Estimation of the mean vector*

2.1.1. *Theorem*

For the estimator

$${}_j\hat{b} = (y'\ {}_jv^{-1}y)^{-1}\ y'\,{}_jv^{-1}\ {}_jX,\ (\tfrac{1}{g}), \tag{1}$$

we have

$$E({}_j\hat{b}/{}_j\Theta) = \beta({}_j\Theta), \tag{2}$$

$$E({}_j\hat{b}) = b. \tag{3}$$

Demonstration. Follows from the fact that

$$E({}_jX/{}_j\Theta) = \mu({}_j\Theta) = y\beta({}_j\Theta)$$

and the definition of b.

2.1.2. *Remark.* The arguments in favor of the estimator $_j\hat{b}$ are the same as those justifying the identically constructed estimator in multivariate regression theory. We shall not repeat them here, but we note however that such an estimator can be obtained as well by least-squares theory as under normal assumptions.

It is seen that $_j\hat{b}$ is $_jB$ defined in 1.3.

2.2. *Estimation of the variance*

2.2.1. *Lemma*

For any symmetrical $\frac{t}{t}$ matrix r:

$$E({}_jX'\, r\, {}_jX/{}_j\Theta) = \sigma^2({}_j\Theta) tr(r\, {}_jv) + \mu'({}_j\Theta)\, r\mu({}_j\Theta).$$

Demonstration. We have, dropping everywhere the fixed class-index j,

$$E(X'rX/\Theta) = \sum_{rs} r_r^s E(X_r X_s/\Theta)$$

$$= \sum_{rs} r_r^s [E(X_r X_s/\Theta) - E(X_r/\Theta)\, E(X_s/\Theta)] + \sum_{rs} r_r^s E(X_r/\Theta) E(X_s/\Theta)$$

$$= \sum_{rs} r_r^s \, \mathrm{COV}\,(X_r, X_s/\Theta) + \sum_{rs} \mu_r(\Theta)\, r_r^s\, \mu_s(\Theta)$$

$$= \sigma^2(\Theta) \sum_{rs} r_r^s v_r^s + \mu'(\Theta)\, r\mu(\Theta)$$

$$= \sigma^2(\Theta)\, tr(rv) + \mu'(\theta)\, r\mu(\Theta).$$

2.2.2. *Theorem*

For the estimator

$$_j\hat{s}^2 = \frac{1}{t-g} ({}_jX - y_j\hat{b})'\, {}_jv^{-1} ({}_jX - y_j\hat{b}), \tag{4}$$

we have

$$E({}_j\hat{s}^2/{}_j\Theta) = \sigma^2({}_j\Theta), \tag{5}$$

$$E({}_j\hat{s}^2) = s^2 \tag{6}$$

Demonstration

We drop everywhere the class-index j. Using (1), we have, after obvious simplifications:

$$(t-g)\, \hat{s}^2 = X'\, r\, X,$$

where

$$r = v^{-1} - v^{-1}y(y'v^{-1}y)^{-1}y'v^{-1}.$$

Therefore

$$\begin{aligned} tr(rv) &= tr(1^{\circ}) - tr(v^{-1}y(y'v^{-1}y)^{-1}y') \\ &= tr(1^{\circ}) - tr(y'v^{-1}y(y'v^{-1}y)^{-1}) \\ &= tr(1^{\circ}) - tr(1^{\circ\circ}) = t - g, \end{aligned}$$

where 1°, $1^{\circ\circ}$ are respectively the ${}^{t}_{t}$ and the ${}^{g}_{g}$ unit matrix.

Also,

$$\mu'(\Theta)\, r\mu(\Theta) = \mu'(\Theta)\, ry\beta(\Theta) = 0,$$

since $ry = 0$. Therefore (5) follows from the lemma. Then (6) is evident from the definition of s^2.

2.3. *Relation for the covariance matrix*

2.3.1. *Remark*

The covariance matrix a cannot be estimated from observations in one class. However, the following relation (8) is the first step in the construction of unbiased estimators for a. Observe that, as is indicated, the relations (7), (8) are ${}^{g}_{g}$ matrix relations.

2.3.2. *Theorem*

$$E(({}_j\hat{b} - b)\ ({}_j\hat{b} - b)'/{}_j\Theta) = \sigma^2({}_j\Theta)\ (y'\ {}_jv^{-1}y)^{-1} + \\ + (\beta({}_j\Theta) - b)\ (\beta({}_j\Theta) - b)',\ ({}^{g}_{g}), \qquad (7)$$

$$E(({}_j\hat{b} - b)\ ({}_j\hat{b} - b)') = s^2(y'\ {}_jv^{-1}y)^{-1} + a,\ ({}^{g}_{g}). \qquad (8)$$

Demonstration

We drop everywhere j. First we prove:

$$E(XX'/\Theta) = \sigma^2(\Theta)\, v + y\beta(\Theta)\beta'(\Theta)y'. \qquad (9)$$

Indeed, the ${}^{s}_{r}$ element of the first member of (9) is

$$\begin{aligned} E(X_rX_s/\Theta) &= \mathrm{COV}(X_r, X_s/\Theta) + E(X_r/\Theta)\, E(X_s)/\Theta) \\ &= \sigma^2(\Theta)v_r^s + \mu_r(\Theta)\, \mu_s(\Theta). \end{aligned}$$

So we have (9) since the last expression is the ${}^{s}_{r}$ element of the matrix

$$\sigma^2(\Theta)\, v + \mu(\Theta)\mu'(\Theta) = \sigma^2(\Theta)\, v + y\beta(\Theta)\beta'(\Theta)\, y'.$$

By (1): $\hat{b}\hat{b}' = (y'v^{-1}y)^{-1}y'v^{-1}\, XX'\, v^{-1}y(y'v^{-1}y)^{-1}$.

By an application of $E(./\Theta)$, using (9):

$$E(\widehat{\hat{b}\hat{b}'}/\Theta) = \sigma^2(\Theta)\ (y'v^{-1}y)^{-1} + \beta(\Theta)\ \beta'(\Theta). \qquad (10)$$

From (2) and the relation

$$(\hat{b} - b)\,(\hat{b} - b)' = \widehat{\hat{b}b'} - b\hat{b}' - \hat{b}b' + bb'$$

it follows that

$$E(\,(\hat{b} - b)\,(\hat{b} - b)'/\Theta) = E(\widehat{\hat{b}b'}/\Theta) - b\beta'(\Theta) - \beta(\Theta)\,b' + bb'.$$

Combining this relation with (10), we have (7). Then (8) follows.

3. Global Estimators.

Here we use the statistical material of all the classes.

3.1. *Estimation of the mean vector*

3.1.1. *Theorem*

Whatever be the $\frac{g}{g}$ matrices ${}_j\pi$ satisfying $\sum\limits_j {}_j\pi = 1$, the vector

$$\hat{b} = \sum_j {}_j\pi\,{}_j\hat{b} \tag{11}$$

$$\frac{1}{g} = \quad \frac{g}{g}\ \frac{1}{g}$$

is an unbiased estimator of b.

Demonstration

Use (3).

3.1.2. *Natural estimator*

In the (BS) case, the estimator

$$\hat{b} = \sum_j {}_jp\,{}_j\hat{b}, \tag{12}$$

where the scalars ${}_jp$ are defined by

$${}_jp = \sum_s {}_jp_s / \sum_{is} {}_ip_s, \tag{13}$$

will be called the *natural estimator* of b. The natural estimator is a particular estimator (11) obtained by taking for ${}_j\pi$ the diagonal matrix with each diagonal element equal to ${}_jp$. The numbers ${}_jp$ will be called the *natural weights*. The matrices ${}_j\pi$ in (11) can be considered as generalized weights.

The natural estimator $\hat{b}$ is used (at least implicitly) in Buhlmann and Straub, [1], in the (BS1) case.

3.2. *Estimation of the variance*

3.2.1. *Theorem*

Whatever be the scalar weights ${}_j\rho$ satisfying $\sum_j {}_j\rho = 1$,

$$\hat{s}^2 = \sum_j {}_j\rho \; {}_j\hat{s}^2 \tag{14}$$

is an unbiased estimator of s^2.

Demonstration

Use (6).

3.2.2. *Natural and unweighted estimators*

In the (BS) case, the estimator

$$\hat{s}^2 = \sum_j {}_jp \; {}_j\hat{s}^2 \tag{15}$$

will be called the *natural estimator of* s^2.

In the general case, the estimator

$$\hat{s}^2 = \frac{1}{k} \sum_j {}_j\hat{s}^2 \tag{16}$$

will be called the *unweighted estimator of* s^2.

The unweighted estimator is considered in Bühlmann and Straub, [1] in the (BS1) case and also in Hachemeister, [3] in the more general (BS) case.

3.3. *Estimation of the covariance matrix*

3.3.1. *Theorem*

Let ${}_{ij}k$ be weights satisfying ${}_{ij}k = {}_{ji}k$, $\sum_{ij} {}_{ij}k = 1$ and set ${}_{i\Sigma}k = \sum_j {}_{ij}k$. Let $\hat{s}^2$ be an unbiased estimator of s^2. Then the $\substack{g \\ g}$ matrix $\hat{a}$ defined by the relation

$$\sum_{ij} {}_{ij}k \, ({}_i\hat{b} - {}_j\hat{b}) \, ({}_i\hat{b} - {}_j\hat{b})' =$$

$$2(1 - \sum_i {}_{ii}k) \, \hat{a} + 2 \, \hat{s}^2 \sum_i ({}_{i\Sigma}k - {}_{ii}k) \, (y'_i \, v^{-1} y)^{-1}, \; \binom{g}{g}, \tag{17}$$

is an unbiased estimator of a.

Demonstration

$$({}_i\hat{b} - {}_j\hat{b}) \, ({}_i\hat{b} - {}_j\hat{b})' = (({}_i\hat{b} - b) - ({}_j\hat{b} - b)) \, (({}_i\hat{b} - b) - ({}_j\hat{b} - b))' =$$

$$({}_i\hat{b} - b) \, ({}_i\hat{b} - b)' + ({}_j\hat{b} - b) \, ({}_j\hat{b} - b)' - ({}_i\hat{b} - b) \, ({}_j\hat{b} - b)' -$$

$$({}_j\hat{b} - b) \, ({}_i\hat{b} - b)'.$$

Therefore, by the assumption of independence of classes, by (3), (8), writing momentarily $_iw$ for $s^2\,(y'\,_iv^{-1}y)^{-1}$:

$$E(\sum_{ij} {}_{ij}k\,({}_i\hat{b} - {}_j\hat{b})\,({}_i\hat{b} - {}_j\hat{b})') =$$

$$\sum_{ij} {}_{ij}k({}_iw + a) + \sum_{ij} {}_{ij}k\,({}_jw + a) - 2\sum_{ij} {}_{ij}k\,\delta_{ij}\,({}_iw + a) =$$

$$2\,a\,(1 - \sum_i {}_{ii}k) + 2\sum_i ({}_{i\Sigma}k - {}_{ii}k)_iw.$$

From this the theorem is clear.

3.3.2. *Natural estimator*

In the (BS) case, let $_{ij}k = {}_ip\,{}_jp$. Then $\hat{a}$ defined by (17) will be called the *natural estimator of a*, for the given $\hat{s}^2$, even if the latter estimator is not the natural one.

If $\hat{b}$ is the natural estimator (12) of b, then

$$\sum_{ij} {}_ip\,{}_jp\,({}_i\hat{b} - {}_j\hat{b})\,({}_i\hat{b} - {}_j\hat{b})' =$$

$$\sum_{ij} {}_ip\,{}_jp\,({}_i\hat{b}\,{}_i\hat{b}' + {}_j\hat{b}\,{}_j\hat{b}' - {}_i\hat{b}\,{}_j\hat{b}' - {}_j\hat{b}\,{}_i\hat{b}') =$$

$$2\sum_i {}_ip\,{}_i\hat{b}\,{}_i\hat{b}' - 2\,\hat{b}\,\hat{b}' = 2\sum_i {}_ip\,({}_i\hat{b} - \hat{b})\,({}_i\hat{b} - \hat{b})'.$$

So the natural estimator $\hat{a}$ results from the relation

$$\sum {}_ip\,({}_i\hat{b} - \hat{b})\,({}_i\hat{b} - \hat{b})' =$$

$$(1 - \sum_i {}_ip^2)\,\hat{a} + \hat{s}^2 \sum_i {}_ip(1 - {}_ip)\,(y'\,_iv^{-1}y)^{-1},\ \left(\tfrac{g}{g}\right), \tag{18}$$

where $\hat{b}$ is the natural estimator (12) of b.

3.4. *The (BS1) case*

3.4.1. *Notations*

Here we consider the (BS1) case and use the notations

$$_jp_\Sigma = \sum_s {}_jp_s,\ {}_\Sigma p_\Sigma = \sum_j {}_jp_\Sigma = \sum_{js} {}_jp_s.$$

Then the natural weights are $_jp = {}_jp_\Sigma / {}_\Sigma p_\Sigma$.

We use the abbreviations

$$_jX_E = \sum_s \frac{_jp_s}{_jp_\Sigma}\,{}_jX_s,\ {}_EX_E = \sum_j \frac{_jp_\Sigma}{_\Sigma p_\Sigma}\,{}_jX_E = \sum_{js} \frac{_jp_s}{_\Sigma p_\Sigma}\,{}_jX_s.$$

3.4.2. *Estimation of the mean*

Now $_j\hat{b}$, $\hat{b}$ are scalars denoted by $_j\hat{m}$, $\hat{m}$. By particularisation of the general results we have ${}_j\hat{m} = {}_jX_E$ and the natural mean equals $\hat{m} = {}_EX_E$.

3.4.3. *Estimation of the variance*

The j-th class variance estimator is, by particularisation of (4):

$$ {}_j\hat{s}^2 = \frac{1}{t-1} \sum_s {}_jp_s \, ({}_jX_s - {}_jX_E)^2. \tag{19}$$

The unweighted estimator is

$$ \hat{s}^2 = \frac{1}{k(t-1)} \sum_{j,s} {}_jp_s \, ({}_jX_s - {}_jX_E)^2. \tag{20}$$

This is the estimator considered in Bühlmann and Straub, [1].

3.4.4. *Estimation of a*

The natural estimator $\hat{a}$, a scalar in this case, results from the relation

$$ (1 - \Sigma\, {}_jp^2)\, \hat{a} = \Sigma\, {}_jp({}_jX_E - {}_EX_E)^2 - (k-1)\, \hat{s}^2/{}_\Sigma p_\Sigma, \tag{21}$$

obtained from (18). The Bühlmann and Straub, [1] estimator $\hat{\hat{a}}$ results from the relation

$$ (1 - \sum_j {}_jp^2)\, \hat{\hat{a}} = \sum_{j,s} \frac{{}_jp_s}{{}_\Sigma p_\Sigma} ({}_jX_s - {}_EX_E)^2 - \frac{kt-1}{{}_\Sigma p_\Sigma}\, \hat{s}^2. \tag{22}$$

By the identity

$$ \sum_{j,s} {}_jp_s\, ({}_jX_s - {}_EX_E)^2 = \sum_{j,s} {}_jp_s({}_jX_s - {}_jX_E)^2 + \sum_j {}_jp_\Sigma\, ({}_jX_E - {}_EX_E)^2, \tag{23}$$

it is seen that $\hat{\hat{a}} = \hat{a}$ if $\hat{s}^2$ is the unweighted estimator (20).

4. Optimal Estimation of the Mean Regression Vector

4.1. *Optimal estimator*

An estimator $\hat{c}$ in a set E of vector estimators shall be called *optimal* in E if the trace of the covariance matrix of $\hat{c}$ is minimal, in comparison with the traces defined similarly for the other elements in E. If E is a set of scalar estimators, the principle invoked is that of minimum-variance.

We leave the question of an optimal $\hat{a}$ or $\hat{s}^2$ unsettled. We consider, here, the case of an optimal $\hat{b}$ given by (11). We prove that the optimal sequence $({}_1\pi, {}_2\pi, \ldots, {}_k\pi)$ is the sequence $({}_1z, {}_2z, \ldots, {}_kz)$ of credibility matrices (see 1.3), except for the constant pre-factor $(\Sigma_i z)^{-1}$.

4.2. *Lemma*

Let ${}_1m, {}_2m, \ldots, {}_km$ be definite positive symmetrical $\frac{g}{g}$ matrices. Let ${}_1x, {}_2x, \ldots, {}_kx$ be variable $\frac{g}{g}$ matrices. Then the minimum of the trace

$$tr(\sum_i {}_ix\, {}_im\, {}_ix'), \tag{24}$$

subject to the constraint $\sum_i {}_ix = 1$, is reached for

$${}_jx = (\sum_i {}_im^{-1})^{-1}\, {}_jm^{-1},\ (j = 1, 2, \ldots, k). \tag{25}$$

Demonstration

If ${}_ix$ is fixed and if y is an arbitrary $\frac{g}{g}$ vector, we have

$$y'\, ({}_ix\, {}_im\, {}_ix')\, y = (y'\, {}_ix){}_im(y'\, {}_ix)' \geq 0$$

since ${}_im$ is definite positive. Therefore ${}_ix\, {}_im\, {}_ix'$ is semidefinite positive and has a nonnegative trace. Thus, (24) is ≥ 0. It is a quadratic form in the kg^2 variables ${}_ix^q_p$. If we eliminate g^2 variables by the constraints $\Sigma_i x = 1$, we have a quadratic polynomial in $(k - 1)g^2$ independent variables that is never negative. Such a function is minimum for finite values of the variables. (See, for example, the lemma 2.5 in De Vylder, [2]). Now we shall apply Lagrange's method and we shall find a unique extremum. From the preceding discussion it follows that this extremum must be the minimum.

We introduce the g^2 Lagrange multipliers λ^β_α corresponding to the constraints

$$\sum_i {}_ix^\beta_\alpha = \delta^\beta_\alpha.$$

We minimize

$$L = tr\, (\sum_i {}_ix\, {}_im\, {}_ix') - 2 \sum_{i\alpha\beta} \lambda^\beta_\alpha\, {}_ix^\beta_\alpha$$

$$= \sum_{i\alpha\beta\gamma} {}_ix^\beta_\alpha\, {}_ix^\gamma_\alpha\, {}_im^\beta_\gamma - 2 \sum_{i\alpha\beta} \lambda^\beta_\alpha\, {}_ix^\beta_\alpha.$$

We must have

$$0 = \frac{1}{2} \frac{\partial L}{\partial_j x^q_p} = \sum_\gamma {}_jx^\gamma_p\, {}_jm^q_\gamma - \lambda^q_p,$$

or, in matrix form,

$${}_jx\, {}_jm = \lambda.$$

Then, successively,

$${}_jx = \lambda_j m^{-1},\ 1 = \sum {}_jx = \lambda \Sigma_j m^{-1},\ \lambda = (\sum_j {}_jm^{-1})^{-1},$$

and (25). Note that the existence of the inverse matrix of $\Sigma_j m^{-1}$ results from the assumptions.

4.3. *Lemma*

The covariance matrix of $\hat{b}$, given by (11), is

$$\Sigma\ {}_i\pi\ (s^2\ (y'\ {}_iv^{-1}y)^{-1} + a)\ {}_i\pi',\ \binom{g}{g}. \tag{26}$$

Demonstration

From (8) and from the independence of classes:

$$E((\Sigma_i\pi\ {}_i\hat{b} - b)\ (\Sigma_j\pi\ {}_j\hat{b} - b)') =$$
$$E(\Sigma_i\pi\ ({}_i\hat{b} - b)\ (\Sigma_j\pi({}_j\hat{b} - b))') =$$
$$\sum_{ij}\ {}_i\pi\ E(({}_i\hat{b} - b)\ ({}_j\hat{b} - b)')\ {}_j\pi'\ \delta_{ij} =$$
$$\sum_i\ {}_i\pi\ (s^2(y'\ {}_iv^{-1}y)^{-1} + a)\ {}_i\pi'.$$

4.4. *Theorem*

The optimal estimator $\hat{b}$ in the class of estimators (11) is

$$\hat{b} = \sum_j\ (\sum_i\ {}_iz)^{-1}{}_jz\ {}_jb, \tag{27}$$

where the ${}_jz$ are the credibility matrices defined in 1.3.

Demonstration

From the definition of ${}_jz$ follows the relation

$$(y'\ {}_jd^{-1}y) + a = {}_jz^{-1}\ a.$$

Then, since

$$s^2(y'\ {}_jv^{-1}y)^{-1} = (y'\ {}_jd^{-1}y)^{-1},$$

the theorem follows from the lemma's.

4.5. *Corollary*

If b is estimated optimally, there is no difference between the credibility approximations ${}_j\hat{B}$, ${}_j\tilde{B}$, ${}_j\ddot{B}$ to $\beta({}_j\Theta)$ given in 1.3.

4.6. *Remarks*

It seems that we are in a circular situation if we try to use the optimal $\hat{b}$, since this $\hat{b}$ depends on a and that $\hat{b}$ is needed in, for example, the natural estimator $\hat{a}$ of a.

However, this anomaly is only apparent, since the first member of (18) can be written without $\hat{b}$. In other words, in (18) $\hat{b}$ *must* be the natural estimator and not the optimal one. It is not excluded, however, that (18) can be optimized in some way by a method of successive approximations, using successively improved $\hat{b}$'s and, eventually, redefinitions of the numbers ${}_jp_s$.

5. Non-Negativity Considerations

The covariance matrix a is semidefinite positive. In particular, in the (BS1) case the number a is non-negative. It is known that the estimator $\hat{\hat{a}}$ can provide negative values. In such cases, Bühlmann and Straub, [1] estimate a by 0.

A similar method can be used if a is a matrix. For example, suppose that a° is an estimate of a and that a° is not semidefinite positive. Then make a° diagonal by an orthogonal transformation. Replace the negative diagonal elements (i.e. the negative characteristic values) by 0 and apply the inverse orthogonal transformation.

If all diagonal elements of a° are positive and if a° is not semidefinite positive the following method can also be used. Multiply all non-diagonal elements of a° by the same number x. Then if x decreases from 1 to 0 the matrix becomes necessarily semidefinite positive. Keep the largest possible x.

Of course, a justification of these methods is difficult to find. Moreover, the estimators redefined in such a way are no longer unbiased. But it must be kept in mind that it is preferable to have an estimate that might be bad, than no estimate at all. And also that the application of credibility formulae with wrong parameters introduces unfairnesses in the different classes, but that these counterbalance each other, at least if b is estimated correctly.

Finally, let we go back to the general formula (17) and let we consider the following question. Is a reasonable general choice of the weights ${}_{ij}k$ and ${}_j\rho$ (in $\hat{s}^2$) possible in such a way that the resulting $\hat{a}$ always is semidefinite positive? The answer is negative. Indeed, let we consider the (BS1) case with each ${}_jp_s = 1$. Then our general hypothetical rule for fixing the weights must lead to equal weights ${}_j\rho$ since we start from a symmetrical situation. For the same reason, we must have

$$ {}_{ij}k = \alpha \ (i \neq j), \ {}_{ii}k = \beta $$

for some α and β. Since we must have $\Sigma_{ij}k = 1$, there is one independent parameter, say α, left. But an inspection of (17) shows that this parameter simplifies in that relation. So we may take $\alpha =$

β. Then $\hat{a}$ is the natural estimator, given by (21). The particular case $k = 2$, $t = 2$ shows that $\hat{a} < 0$ for the values

$$_1X_1 = 1,\ _2X_1 = 1,$$

$$_1X_2 = 0,\ _2X_2 = 0.$$

Bibliography

[1] Buhlmann, H., and Straub, E., (1970) Glaubwurdigkeit fur Schadensatze, Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.

[2] De Vylder, Fl. (1976) Geometrical Credibility, Scandinavian Actuarial Journal.

[3] Hachemeister, Ch. Credibility for regression models with application to trend

[4] Straub, E., Credibility in practice

[5] Taylor, G. C., Abstract credibility

The papers [3], [4] were presented to the Actuarial Research Conference on Credibility Theory, Berkeley, 1974. They are published in the book "Credibility" edited by P. M Kahn, Academic Press, 1975.

The paper [5] will be published imminently.

CONTRIBUTION À L'ÉTUDE DU COÛT DES SINISTRES AUTOMOBILES

P. PICARD
Paris

SUMMARY

A Contribution to the Study of Automobile claims amounts

After having a look at the results obtained by adjustment of automobile claims amounts distribution, we research how the number and the time-configuration of past claims condition the claims law of probability.

We have statistics about a group of 471 000 cars which was followed for three years: 1970, 1971 and 1972. We use mathematical technics and among multi-dimensional analysis, we use factorial analysis of correspondance (A F C.) A F.C. permits us to show the link which exists between the claim amount of the third year and the number of claims during the two years before. A quantitative analysis of the corporal claims shows that, of the frequency of corporal claims during the third year growths up in function of the number of past claims, the expected corporal claims amount of the third year decreases as the square of the material claims number during the two first years

I. POSITION MATHÉMATIQUE

La notion de processus de risque est désormais bien connue des actuaires. On ne rappellera donc ici que les définitions et propriétés utiles pour la suite des calculs.

Soit S_t la somme des montants des sinistres pendant la période de temps (o, t). S_t est une variable aléatoire dépendant du temps, c'est un processus aléatoire que l'on décompose en:

— la probabilité P_n^m (t, s) pour que le nombre de sinistres passe de n à m pendant la période de temps (t, s);

— la fonction $F_t(x/y)$, probabilité pour que S_t soit inférieur à y sachant qu'à l'instant précédent t, il était égal à x et sachant que t est l'abscisse d'un saut du processus (un sinistre).

Cette fonction $F_t(x/y)$ est l'objet de cette étude. On a:

$F_t(x/y) = \text{Prob}\ [S_t < y/S_{t-\varepsilon} = x$ et $t =$ abscisse d'un sinistre]

soit en posant $z = y - x$ (montant du sinistre à l'instant t)

$F_t(z) = \text{Prob}\ [\Delta S_t < z/t =$ abscisse d'un sinistre]

où ΔS_t est l'accroissement de S_t à l'instant t.

II. Ajustements de Lois

II.1. *Précautions à prendre pour analyser des coûts de sinistres automobiles*

La base statistique est un ensemble de sinistres survenus au cours d'un certain laps de temps à un groupe de véhicules bien défini. Mais pour analyser ces chiffres, des précautions doivent être prises:

— Si l'on observe des sinistres récents, beaucoup d'entre eux ne sont réglés que partiellement et la partie évaluée est peu précise. Pour avoir une meilleure connaissance des coûts, il faudra attendre le moment où la proportion des dossiers restant en évaluation est faible.

— Dans une étude de ce type, surtout si la période d'observation est longue, on est amené à comparer des sommes à des instants différents et, donc, se pose le problème du choix (ou de la construction) du type d'actualisation.

II.2. *Résultats obtenus*

Monsieur Marcel Henry a montré que la fonction y, nombre de sinistres supérieurs à une garantie x, pouvait être représentée d'une façon assez satisfaisante par la fonction de Galton-MacAlister:

$$y = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{x} e^{\frac{-z^2}{2}} dz \text{ avec } z = a \operatorname{Log} x + b$$

Il apparaît toutefois nécessaire de donner à a deux valeurs différentes, l'une pour les x inférieurs à un certain montant, l'autre pour des valeurs de x plus élevées (le nombre de gros sinistres décroît très rapidement).

On obtient des résultats comparables avec la formule de Pareto:

$$\operatorname{Log} y = a \operatorname{Log} x + \mathrm{b} \quad \text{ou} \quad y = \frac{b}{x^a}$$

qui a l'avantage de conduire à des calculs plus simples. Mais, comme dans la loi proposée par Monsieur Marcel Henry, on doit ajuster plusieurs courbes suivant l'importance des sinistres.

Monsieur B. Almer a proposé d'ajuster la distribution des sinistres par un trinôme exponentiel:

$$\Psi(x) = n \left[a_1\beta_1 c^{-\beta 1x} + a_2\beta_2 c^{-\beta 2x} + a_1\beta_1 c^{-\beta 3x}\right]$$

$$\text{avec } a_1 + a_2 + a_3 = 1$$

Plus récemment, Monsieur Gaudibert, dans une thèse présentée devant l'Institut des Actuaires Français a ajusté, pour les *gros sinistres*, une fonction du type:

$$y = \frac{A}{x^a e^{bx}}$$

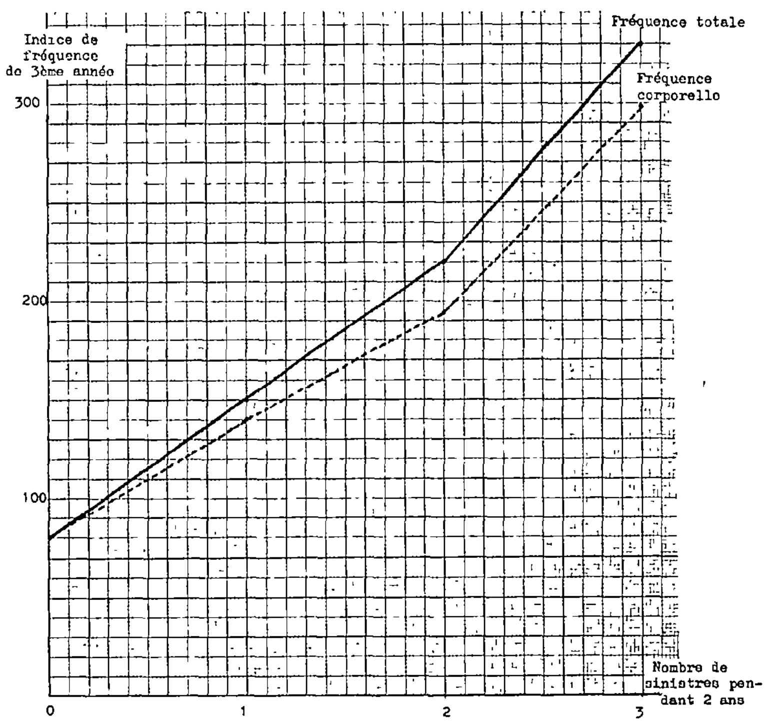
III. Liaisons des Résultats des Années Successives

III.1. *Rappel des résultats obtenus pour les fréquences des sinistres*

P. Depoid, dans son ouvrage "Applications de la statistique aux Assurances", fait apparaître la liaison entre les fréquences d'années successives de mêmes assurés. Monsieur Delaporte a formalisé le

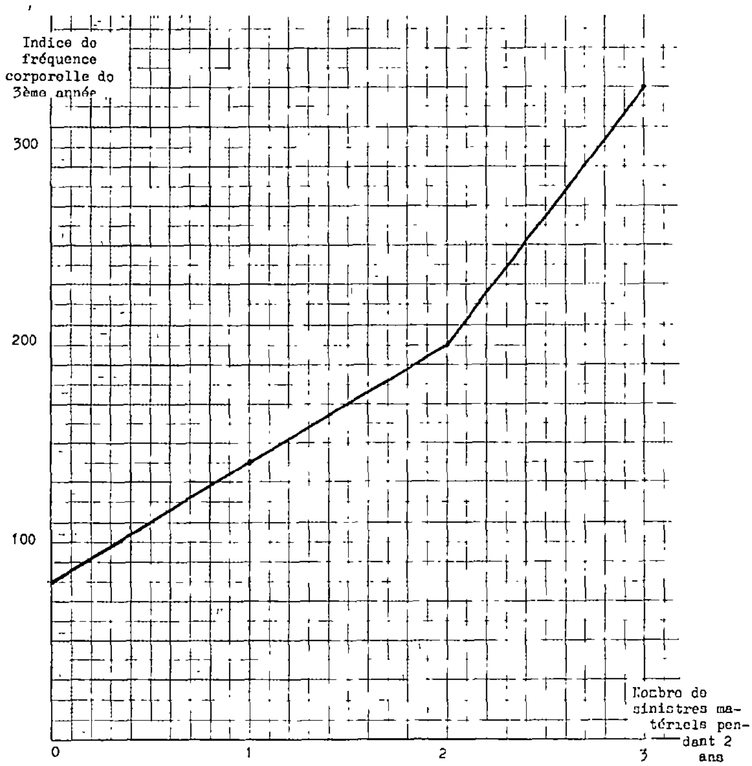
Graphique I

Sinisters Materiels et Corporels



Indice de Fréqence de 3ème Année en Fonction des Resultats de 1ère et 2ème Année.

GRAPHIQUE 2



Indice de Fréquence Corporelle de 3ème Année en Fonction des Resultats Materiels de 1ère et 2ème Année

problème. Monsieur M. Brichler a proposé une formule remarquablement simple:

$$f_{n/x} = F \frac{1 + x}{1 + nF}$$

où F est la fréquence d'ensemble et x le nombre de sinistres pendant n années.

Cette formule a ensuite été améliorée dans des travaux effectués à l'Association générale des Sociétés d'Assurance contre les Accidents.

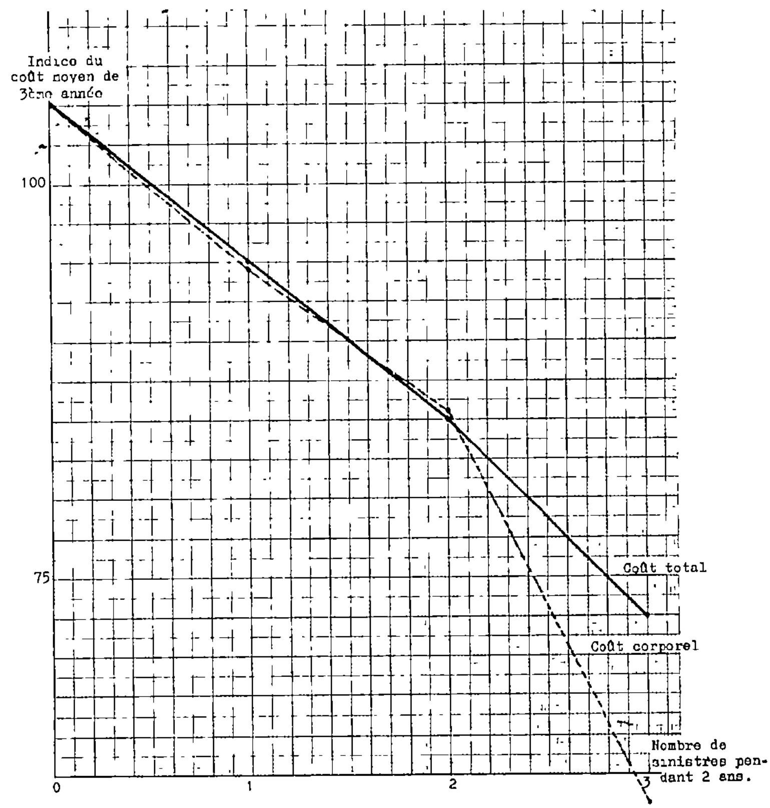
Pour illustrer ce phénomène, on se reportera au graphique n° 1 obtenu avec la "Statistique commune" de 1970, 1971 et 1972. On a porté, en ordonnée, un indice de fréquence: 100 = ensemble de la population pour l'année considérée et, en abscisse: le nombre de sinistres survenus dans les deux années précédentes. Le graphique n° 2 montre l'accroîssement du risque corporel en fonction du nombre de sinistres matériels passés.

III.2. *Liaisons entre les résultats des années précédentes et le coût moyen des sinistres automobiles*

Sur le graphique n° 3, on porte en abscisse, le nombre de sinistres des deux premières années et, en ordonnée, un indice du coût moyen des sinistres de troisième année. La décroissance du coût moyen en fonction de la gravité des antécédents est nette. En effet, pour 3 sinistres en deux ans, le coût moyen est diminué de 25%.

GRAPHIQUE 3

Sinistres Materiels & Corporels



Indice du cout Moyen de 3ème Année en Fonction du Nombre de Sinistres en 1ère et 2ème Année

III.3. *Choix d'une méthode de recherche*

Le but de la présente étude est de rechercher comment le nombre et la configuration temporelle des sinistres passés *conditionnent la distribution des coûts* des sinistres de dernière année.

La statistique classique permet d'ajuster une loi de probabilité de forme analytique donnée aux résultats empiriques. Pour étudier la liaison temporelle dans le coût des sinistres automobiles, il est intéressant de se détacher d'une hypothèse de loi: les "analyses multidimensionnées" et, plus particulièrement, l' "Analyse factorielle des correspondances" (A.F.C.) choisie dans cette étude le permettent.

IV. Analyse Factorielle des Correspondances (A.F.C.)

IV.1. *Théorie*

Cette analyse est généralement utilisée dans l'étude des tableaux de contingence, mais, par extension, cette méthode s'applique à tout tableau rectangulaire de nombres positifs ou nuls.

Soit p_{ij} l'élément de la ligne i et de la colonne j. On note:

$$f_{ij} = p_{ij} / \sum_{i,j} p_{ij};\ p_i = \sum_i p_{ij},\ p_j = \sum_j p_{ij}$$

$$f_i = \sum_j f_{ij};\ f_j = \sum_i f_{ij} \text{ et } \sum_i f_i = \sum_j f_j = 1$$

$$f_j^i = f_{ij}/f_i = p_{ij}/p_i;\ f_j^i = f_{ij}/f_j = p_{ij}/p_j$$

On associe à l'individu i (à la ligne i), la loi conditionnelle sur J: $\{f_1^i, \ldots, f_{J\max}^i\} = f_J^i$, muni de la masse f_i où $J\max = \text{card } (J)$.

De même, à l'élément j, on associe:
$\{f_1^j, \ldots, f_{I\max}^i\} = f_I^i$ muni de la masse f_i ou $I\max = \text{card } [I]$.

On a alors les deux nuages:
$N(I) = \{f_J^i$ de masse $f_i / i \in I\} \subset \mathbb{R}_J$ muni de la métrique du χ^2 de centre f_J:

$$d^2(i, i') = \| f_J^i - f_J^{i'} \|^2 = \sum_{j=1}^{J\max} (f_j^i - f_j^{i'})^2 / f_j$$

$N(J) = \{f_I^i$ de masse $f_j / j \in J\} \subset \mathbb{R}_I$ muni de la métrique du χ^2 de centre f_I.

Le meilleur espace de dimension k représentant $N(I)$ est engendré par k vecteurs orthonormés de $R_J \{(e_1)_1 \ldots (e_k)_J\}$, auquel correspondent les opérateurs de projection (facteurs): $\phi_1^J \ldots \phi_k$ appartenant à $\mathbb{R}^J = (\mathbb{R}_J)^*$ qui sont les vecteurs propres de $m \circ \sigma$ (m est la métrique et σ est la forme quadratique d'inertie) correspondant aux k plus grandes valeurs propres: $\lambda(\phi_1), \ldots, \lambda(\phi_k)$.

Le nuage $N(I)$ est approximé par sa projection sur la variété précédente, à l'individu i on associe ses coordonnées: $G(i, 1), \ldots, G(i, k)$, et on a avec le $n$ième facteur:

$$G(i, n) = \phi_n^J(f_J^i) = \sum_{j=1}^{Jmax} \phi_N^j f_j^i = \sqrt{\lambda(\phi_n)}\, \phi_n^i$$

De même pour $N(J)$, à j on associe:

$$F(j, 1) \ldots F(j, k)$$

avec

$$F(j, n) = \sqrt{\lambda(\phi_n)}\, \phi_n^j$$

et

$$G(i, n) = \sum_{j=1}^{Jmax} F(j, n)\, f_j^i / \sqrt{\lambda(\phi_n)}$$

$$G(i, n) = \sum_{j=1}^{Jmax} F(j, n)\, P_{ij}/P_i \sqrt{\lambda(\phi_n)}$$

IV.2. *Propriétés*

L'analyse factorielle des correspondance donne un rôle identique aux individus et aux caractères (Symétrie parfaite).

L'analyse des correspondances satisfait le principe *d'équivalence distributionnelle.*

IV.3. *Organisation des données*

Pour examiner la distribution des coûts des sinistres, on est conduit à discrétiser. Les abscisses des classes de coût sont les suivantes:

Classes de coût des sinistres automobiles

No	ABSCISSES	No.	ABSCISSES
C 01	Moins de 150 F	C 16	5 000 F. à 10 000 F.
C 02	150 F à 200 F.	C 17	10 000 F. à 15 000 F.
C 03	200 F. à 300 F.	C 18	15 000 F. à 20 000 F.
C 04	300 F à 400 F.	C 19	20 000 F. à 30 000 F.
C 05	400 F à 500 F.	C 20	30 000 F. à 40 000 F.
C 06	500 F. à 600 F.	C 21	40 000 F. à 50 000 F.
C 07	600 F à 700 F.	C 22	50 000 F. à 100 000 F.
C 08	700 F. à 800 F.	C 23	100 000 F. à 150 000 F
C 09	800 F. à 900 F.	C 24	150 000 F. à 200 000 F.
C 10	900 F. à 1 000 F.	C 25	200 000 F. à 500 000 F.
C 11	1 000 F. à 1 500 F.	C 26	Plus de 500 000 F.
C 12	1 500 F. à 2 000 F.		
C 13	2 000 F. à 3 000 F.		
C 14	3 000 F. à 4 000 F.		
C 15	4 000 F. à 5 000 F.		

Pour chaque configuration de sinistres durant les trois années consécutives observées, on note la distribution des coûts des sinistres de troisième année. C'est-à-dire que:

— pour tout triplet n° $i(a, b, c)$ avec $0 \leq a \leq 3$, $0 \leq b \leq 3$ et $1 \leq c \leq 3$, on a les nombres P_{ij} des sinistres de la classe de coût n° j.

	Configurations			Tranches de Couts					
No	1ère année	2ème année	3ème année	$C\ 01$	$C\ 02$	....	$C\ j$	....	$C\ N$
1	0	0	1	P_{11}	P_{12}	...	P_{1j}	....	P_{1N}
⋮	⋮	⋮	⋮	⋮	⋮		⋮		⋮
i	a	b	c	P_{i1}	P_{i2}	....	P_{ij}	....	P_{iN}
⋮	⋮	⋮	⋮	⋮	⋮		⋮		⋮
M	3	3	3	P_{M1}	P_{M2}	....	P_{Mj}	....	P_{MN}

M = 48 configurations N = 26 classes

REMARQUES: On a retenu comme nombre maximum de sinistres par an, le nombre trois, afin de posséder dans chaque cas un nombre d'observations suffisant pour l'analyse.

V. Analyses des Résultats

Les taux d'inertie des axes factoriels sont faibles. Cependant, les plans des axes 1 et 2 et des axes 1 et 3 possèdent des parts d'inertie suffisantes pour permettre une interprétation.

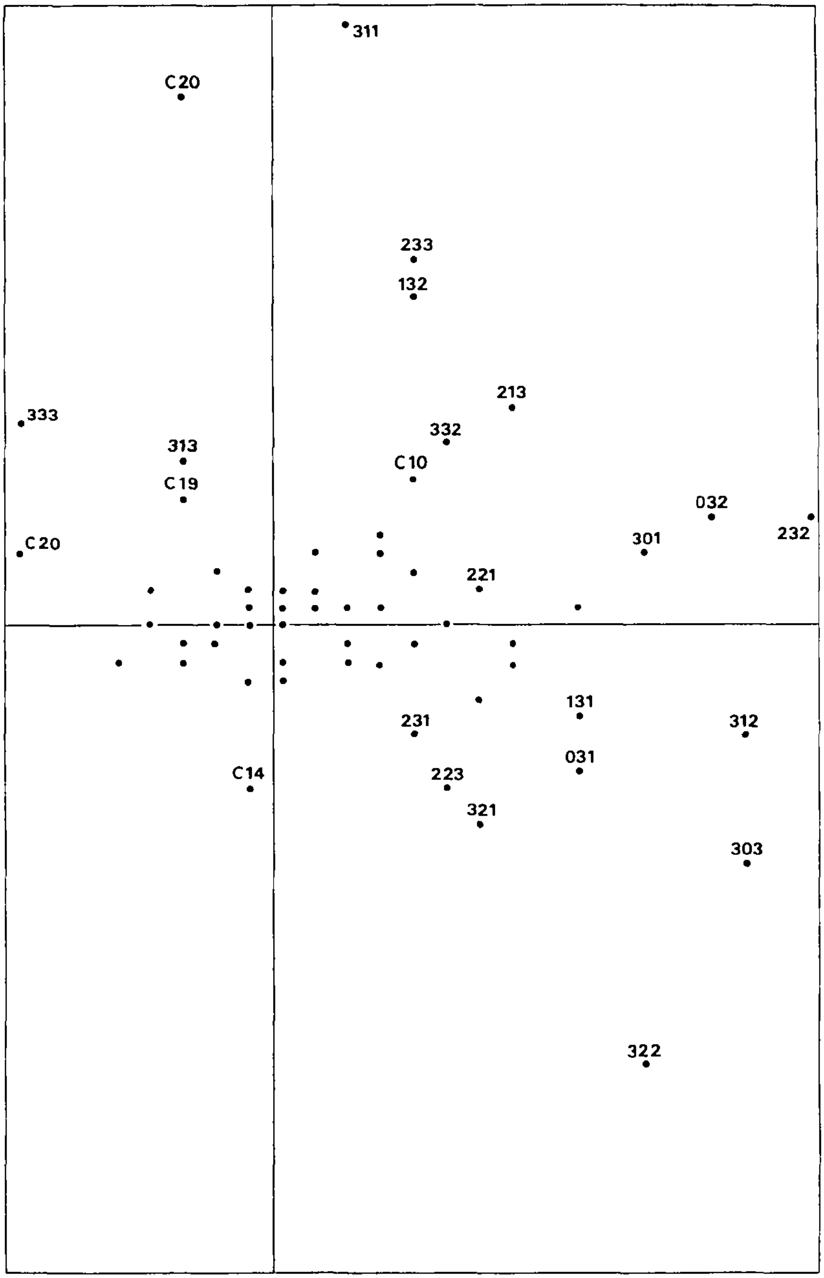
Pour améliorer la commodité de lecture des graphiques en cas de points superposés, on imprime un identificateur d'autant plus noir que la multiplicité est grande.

Les configurations 331 et 133 se distinguent particulièrement et rendent l'interprétation du reste des nuages difficile. Les figures n° 4 et n° 5 représentent les projections des nuages sur les plans (1, 2) et (1, 3) après suppression de points 331 et 133 (sans modification des actes factoriels).

L'étude du plan (1, 3), figure n° 5, permet de mettre en évidence un *effet Guttman*, c'est-à-dire que l'on peut disposer le tableau de

GRAPHIQUE 4

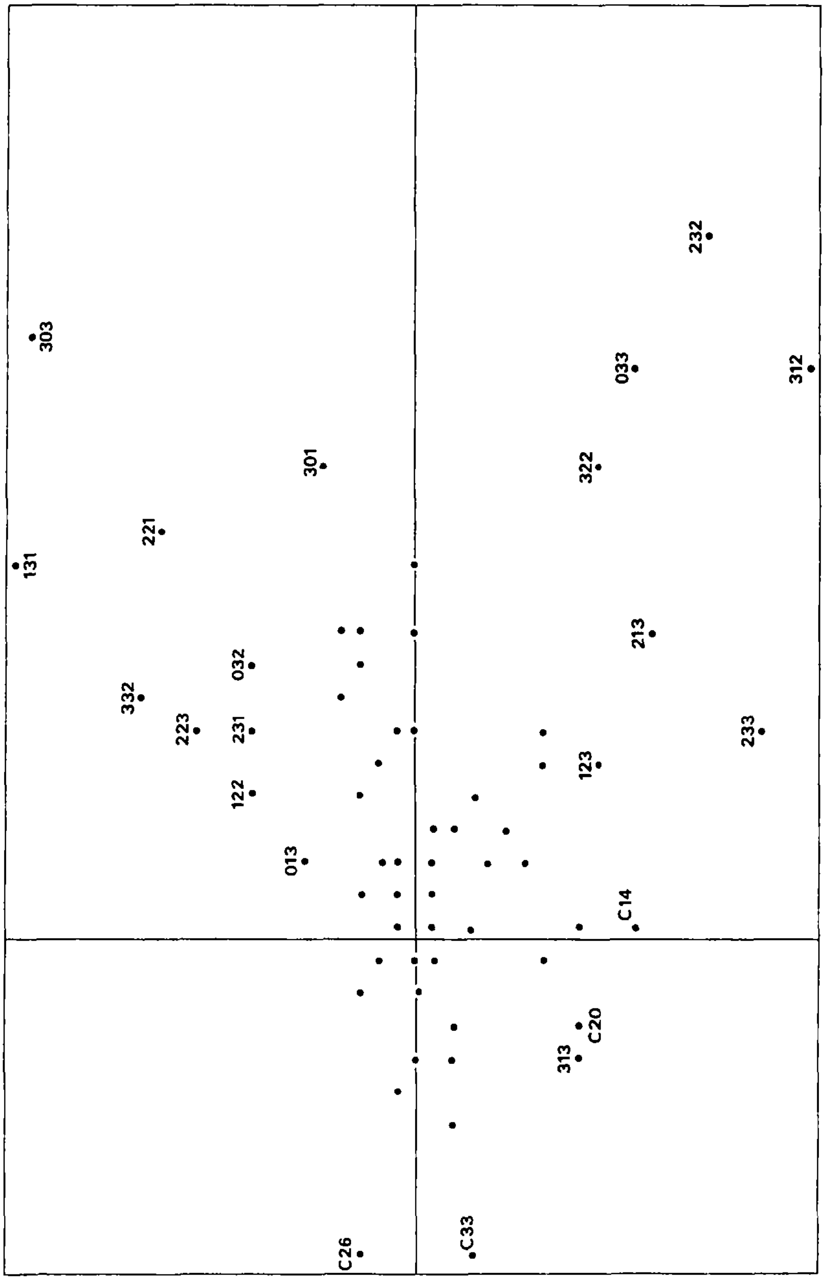
AXE 1 ET 2 AFC



Couts des Sinistres Automobiles

GRAPHIQUE 5

AXE 1 ET 3 AFC



Couts des Sinistres Automobiles

données, par permutation des lignes et des colonnes, sous une forme *bloc-diagonale* et donc il existe un *classement des types de configuration* qui induit un classement des coûts des sinistres de 3ème année. Cette remarque prouve l'existence d'un conditionnement de la distribution des coûts par les antécédents.

L'explication des liaisons entre les configurations et les coûts des sinistres se déduit de l'interprétation des premiers axes factoriels. Il est pratique d'examiner successivement les projections des nuages I et des nuages J (graphiques n° 6 et n° 7).

Sur le graphique n° 6, on repère les configurations par les figures suivantes:

Nombre de sinistres en deux ans	Figures
0	▭
1	○
2	△
3	▽

L'analyse des groupements ainsi obtenus permet de montrer que l'axe horizontal classe les configurations par leurs nombres de sinistres dans les deux premières années. On remarque que le nombre de sinistres de 3ème année ne semble pas avoir d'influence sur la distribution des coûts.

La forme triangulaire du graphique n° 7 est caractéristique d'éléments classés naturellement (tranches de coût). Pour l'interprétation, on doit prendre soin de se baser principalement sur les éléments ayant un poids important (population importante). A cette condition, on remarque le phénomène de classement sur l'axe horizontal suivant les coûts (coût élevé à gauche, moyen et faible à droite). La distinction entre tranches de coût moyen et tranches de coût faible peut être observée sur l'axe vertical.

Cette étude permet de mettre en évidence la liaison entre les survenances des sinistres passés et le coût des sinistres présents.

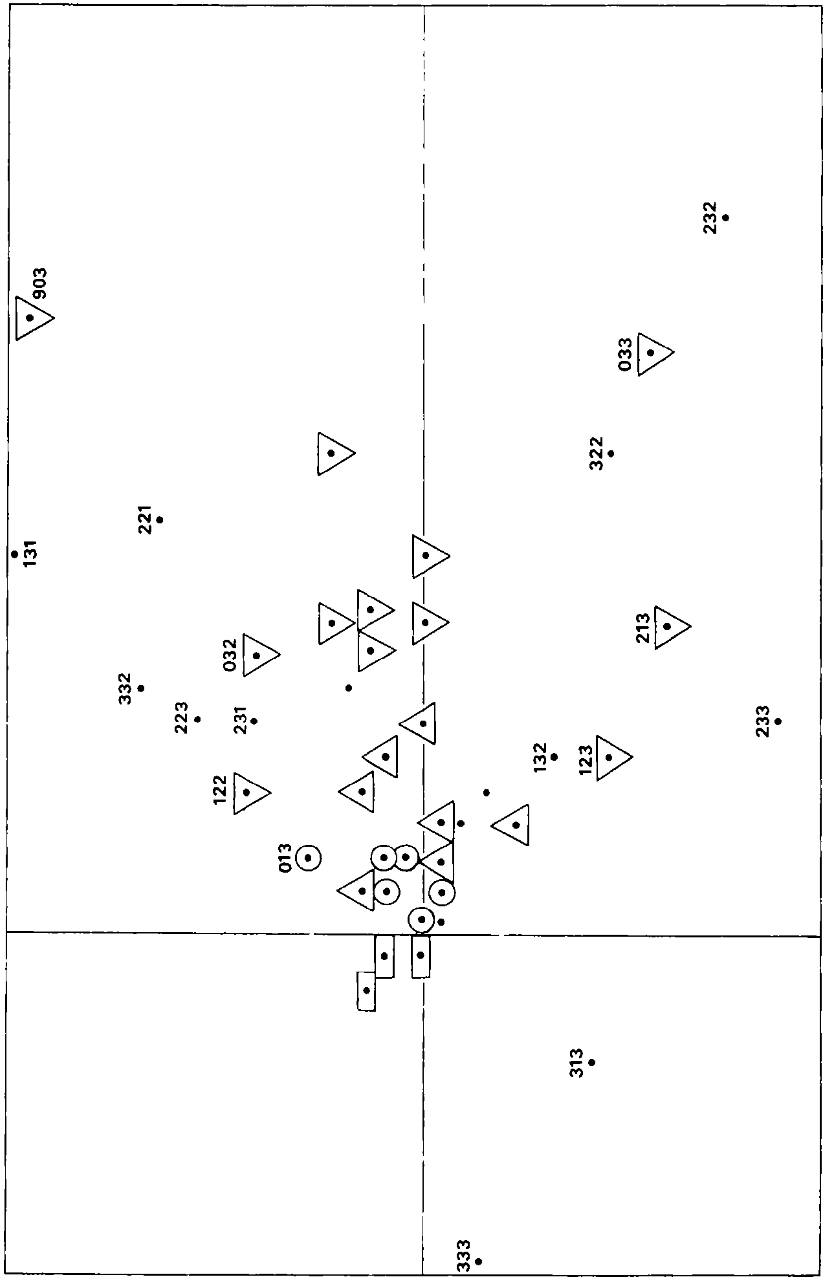
La distribution des coûts des sinistres de 3ème année est d'autant plus biaisée vers les classes inférieures que le nombre de sinistres pendant les deux premières années est important.

VI. Études Quantitatives

La mesure du phénomène observé réclamerait une période d'observation plus longue, car on constate que la sélection des assurés par leurs survenances est moins rapide vis-à-vis des coûts

GRAPHIQUE 6

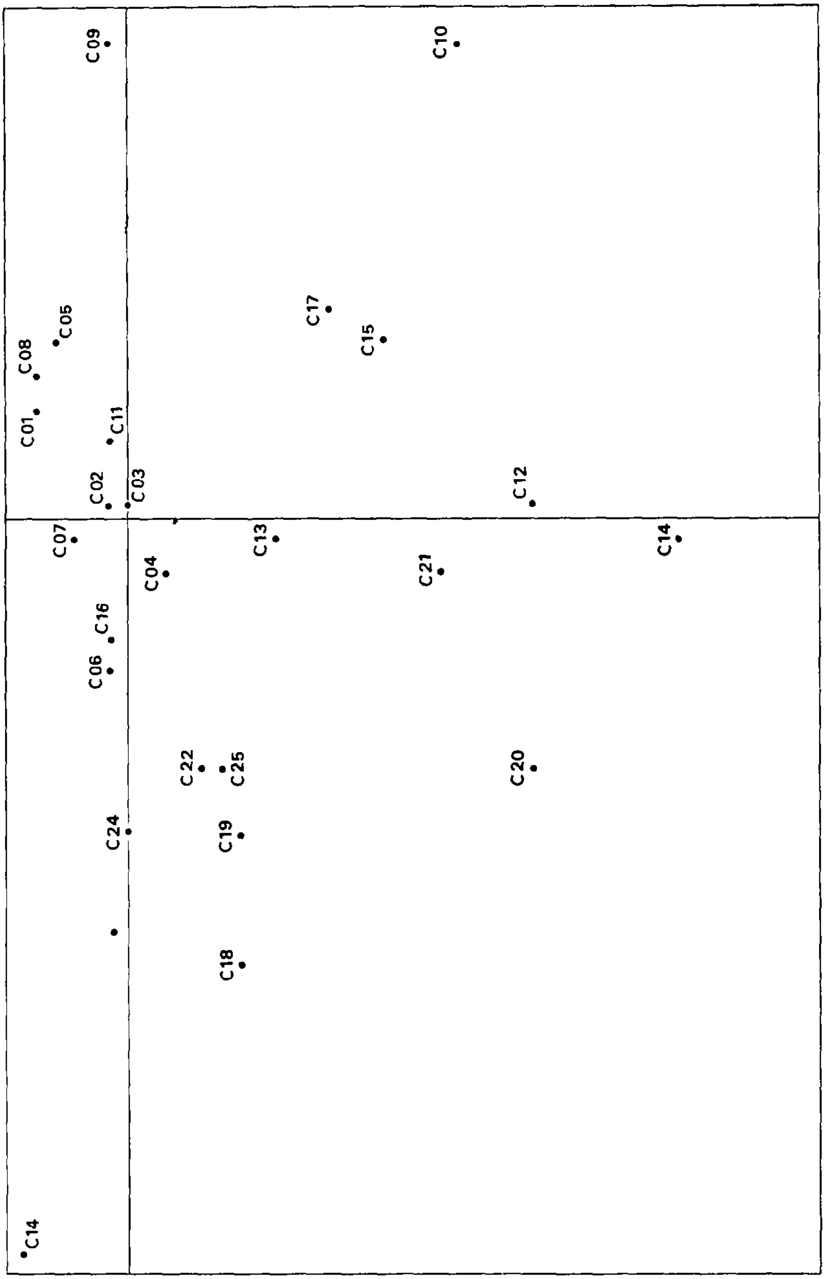
AXE 1 ET 3 A F C



Couts des Sinistres Automobiles

GRAPHIQUE 7

AXE 1 ET 3 AFC



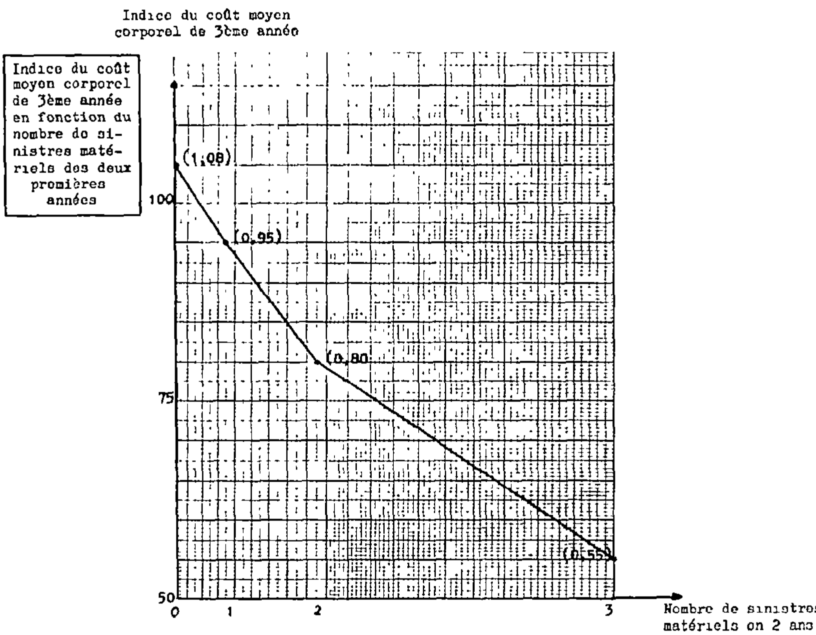
Couts des Sinistres Automobiles

que vis-à-vis des fréquences. Cependant, sur trois ans, l'influence est particulièrement nette sur les tranches élevées de coûts, c'est-à-dire sur les sinistres corporels.

On a vu (graphique n° 2) que la fréquence corporelle croît rapidement en fonction des antécédents. On va montrer que le *coût moyen corporel* décroît en fonction du nombre de sinistres passés.

Le graphique n° 8 montre que le coût moyen corporel de 3ème année décroît en fonction du nombre de sinistres matériels des deux années précédentes. On remarque que si l'assuré n'a aucun sinistre matériel pendant deux ans, son coût moyen corporel est supérieur d'environ 5% au coût moyen corporel de l'ensemble.

GRAPHIQUE 8



La baisse du coût moyen corporel de 3ème année est proportionnelle au carré du nombre de sinistres matériels des deux premières années.

Nombre de sinistres matériels en deux ans	N^2	Baisse du coût moyen corporel
1	1	5 %
2	4	20 %
3	9	45 %

VII. Conclusion

L'analyse factorielle des correspondances appliquée dans cette étude permet de mettre en évidence le conditionnement de la distribution des coûts des sinistres de dernière année par le nombre de sinistres passés. La probabilité d'un sinistre de coût élevé diminue très rapidement lorsque le nombre d'antécédents augmentent.

La probabilité de survenance d'un sinistre corporel est d'autant plus forte que le nombre de sinistres passés est important. Cependant, l'étude quantitative a permis de mesurer la décroissance de gravité des sinistres corporels. Le coût moyen corporel de troisième année semble être une *fonction quadratique* décroissante du nombre de sinistres matériels des deux années précédentes.

ERRATA

EARTHQUAKE INSURANCE IN JAPAN, by MASAO WAKURI AND YASUYUKI YASUHARA

The Astin Bulletin Vol. IX Part 3

(Page/Line)	(False)	(Correct)
332/6	(running from .	(a typical transform fault running from . .
332/19	7,9	7.9
332/30	280	2800
334/22	to have had the same	to have had nearly the same
335/13	. . . of Tokyo, off Shizuoka and off Shikoku Island	. . of Tokyo and off Shizuoka in the central part of Japan
335/15; 336/12; 338/24, 33; 339/3	crustal alteration(s)	crustal deformation(s)
336	Fig. 5 Should be replaced by:	

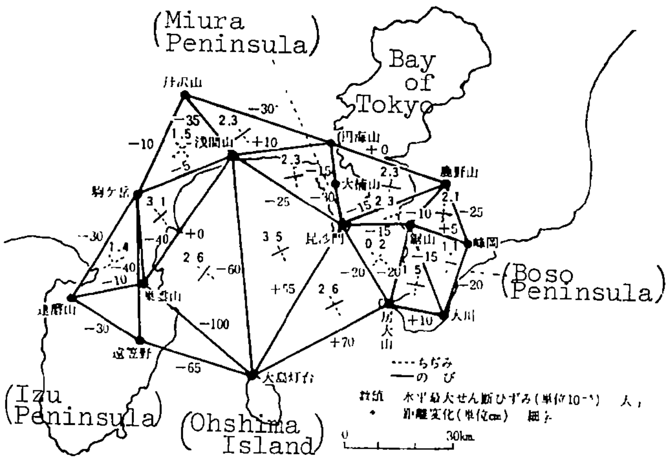


Fig. 5 Changes in distances in the area south of Tokyo (1925-1971) (After Geographical Survey Institute)

(Page/Line)	(False)	(Correct)
337/6	earth surface	earth's surface
338/27	the Meteorological Agency, universities	the Meteorological Agency, the Geographical Survey Institute, universities
339/3	the inclination of the earth	the changes in the inclination of the earth's surface
339/4	contraction of the earth	contraction of the earth's crust
347 at the foot of the table	In 1,000 million yen)	(In 1,000 million yen)
350/2, 13	Profits	Income
350/3	are also reserved	is also reserved
350/10	payed	paid
351/17	The net earthquake premium income	The total earthquake premium income
358/16	Seismic intensity or seismic coefficient	Seismic intensity and/or seismic coefficient
363/28	$f_{si} \cdot d_{si}$	$fs_i \cdot ds_i$

(The other similar expressions on pages 363 and 364 should be corrected in the same way.)

INFORMATION FOR CONTRIBUTORS

All correspondence on editorial matters should be addressed to the Editors:

P. ter Berg and F. Boermans, Groot Hertoginnelaan 8, P.O. Box 990, 2501 CZ The Hague, The Netherlands.

The following remarks are of special importance for manuscript preparing: Each paper should contain an introduction. As regards references, these should be reasonable complete; references in the text are given by the author's name in capitals followed by the year of publication between round brackets. At the end of the paper the references should be grouped alphabetically; for journal references give at least author(s), year, title, journal, volume and pages; for book references give at least author(s), year, title, city and publisher.
In mathematical expressions, authors are requested to minimize unusual or expensive typographical requirements. This may be achieved by using the slash in preference to built-up fractions and to write complicated exponentials in the form exp( ). Equation numbers must be at the left. 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. Figures should be designated by arabic numbers and must have a title. Any legends for figures must be typed on a single separate sheet rather than placed on the drawings. Tables should be numbered, should have a heading and should be prepared without vertical lines.
Material that has already been published or submitted for publication elsewhere, is not acceptable. So, a confirmation of the sole rights for the Bulletin is required.
It is advisable to minimize the size of the paper, maintaining the scientific quality and the clearness of the contents, in order to maximize the probability of acceptance.
No paper will be rejected until it has been reviewed by the Editorial Board.

* * *

SUBSCRIPTION

ASTIN (Actuarial Studies in Non-Life Insurance) 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 the Bulletin free of charge.

Non-members who want to receive the Bulletin as it comes out are invited to apply through the Treasurer:

J. Adam, 53 Boulevard Emile Jacqmain, 1000 Brussels, Belgium.

Anyone interested in one or several items of the series of The Astin Bulletin may send his order directly to:

E. J. Brill — Publisher, Oude Rijn 33a, Leiden, The Netherlands.

He will be invoiced accordingly.