Volume 17, No. 1 April 1987

ASTIN BULLETIN

A Journal of the International Actuarial Association

CONTENTS

EDITORIAL AND ANNOUNCEMENTS

Editorial 1

Obituary on Marcel Henry 5

Obituary on Karl Heinrich Borch 7

19th ASTIN Colloquium 9

ARTICLES

T PENTIKÄINEN
Approximative Evaluation of the Distribution Function of Aggregate Claims 15

BJØRN SUNDT
Two Credibility Regression Approaches for the Classification of Passenger Cars in a Multiplicative Tariff 41

THOMAS WITTING
The Linear Markov Property in Credibility Theory 71

WORKSHOP

C D DAYKIN, G. D. BERNSTEIN, S M. COUTTS, E R F. DEVITT, G B. HEY, D. I. W REYNOLDS and P D. SMITH
The Solvency of a General Insurance Company in Terms of Emerging Costs 85

Book Reviews 133

EDITORIAL POLICY

ASTIN BULLETIN started in 1958 as a journal providing an outlet for actuarial studies in non-life insurance Since then a well-established non-life methodology has resulted, which is also applicable to other fields of insurance For that reason ASTIN BULLETIN will publish papers written from any quantitative point of view—whether actuarial, econometric engineering, mathematical, statistical, etc —attacking theoretical and applied problems in any field faced with elements of insurance and risk

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

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

MEMBERSHIP

ASTIN is a section of the International Actuarial Association (IAA) Membership is open automatically to all IAA members and under certain conditions to non-members also Applications for membership can be made through the National Correspondent or, in the case of countries not represented by a national correspondent, through a member of the Committee of ASTIN

Members receive ASTIN BULLETIN free of charge

SUBSCRIPTION AND BACK ISSUES

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

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

The current subscription or back issue price per volume of 2 issues including postage is £19 ($36)

INDEX TO VOLUMES 1-13

The Cumulative Index to Volumes 1-13 is also published for ASTIN by Tieto Ltd at the above address and is available for the price of £5 00

TYPESET by MCS LTD, SALISBURY and
PRINTED by SHORT RUN PRESS LTD, EXETER
GREAT BRITAIN

EDITORIAL

Could ASTIN do better?

(Excerpts from Presidential Address, Tel Aviv, September 23, 1986)

Article 2 of the rules of our Association states that "ASTIN has as main objective the promotion of actuarial research, particularly in non-life insurance". From time to time the question "Is ASTIN doing well?" arises, either in the pages of this Bulletin or during our colloquia. This question is seldom answered. Indeed the greatest part of our time during Committee meetings and General Assemblies is devoted to the day-to-day life of our Association: finances, elections, Editors' reports, membership file, etc .. When the meeting is adjourned, no time is left to discuss more basic and far reaching problems. Yet, this is what we should all do, regularly. This is certainly what the Chairman of any association should do, when approaching his mid-term. I shall successively consider the evolution of ASTIN itself, its impact on teaching and its influence on actuarial research.

In my opinion one of the main achievements of ASTIN is simply its size, its continuous increase in membership. It seems obvious that the most necessary of all conditions to promote research, whatever that research may be, is that it should be accessible to the largest number of people. So it is a very positive fact that our membership is about to exceed 1,500. I am also pleased to report that the number of external subscriptions to the ASTIN BULLETIN, stagnant for so many years, is now definitely on the rise. We can thus proudly affirm that the ASTIN BULLETIN is the actuarial journal that has, by far, the largest circulation. Without a permanent secretariat, with a Committee that meets at most once a year, with an annual budget that consists of only a small fraction of the income of similar associations, the ASTIN section of the International Actuarial Association has managed to develop an internationally renowned scientific journal; this is considered by many as a permanent miracle, for which all past and present Editors of our Bulletin have to be warmly thanked.

This of course does not mean that we could not do much better. The number of external subscriptions, while increasing, remains at a low level; this indicates that our Bulletin is not sufficiently known outside actuarial circles. The membership is growing, but not evenly, only in selected countries, thanks to the efforts of some National Correspondents. In too many countries the number of members has not increased at all for many years; this means that the information transmission channels from ASTIN to young actuaries have not worked efficiently in some parts of the world.

This is an important point, which should be of concern to all of us: ASTIN needs to keep growing. We need to grow so that our Editors can keep telling prospective authors: "by publishing your paper in the ASTIN BULLETIN you will ensure the largest circulation to your ideas". We need to grow so that our membership and subscription fees will remain among the lowest. The membership dues have not been increased since July 1977! This can only go on if we continue to produce scale economies by becoming more numerous. Above all, we need to grow because there will never be enough of us to tackle the numerous research problems faced by non-life actuaries.

Considering now the teaching of non-life insurance, it is clear that we have witnessed a tremendous evolution in the last decade. Among our membership we have many young bright actuaries holding newly created Chairs in insurance. Many universities that had a degree in insurance have added a non-life position to their faculty; several countries that had been absent from the actuarial research scene for years are now flooding our actuarial journals with high quality papers.

Besides, most of those newly appointed professors have rightfully recognized the need to work in close collaboration with the industry and frequently decide to spend several months working for a company.

Another positive aspect is that those young stars not only teach excellent courses, they have also started writing textbooks. The time when we had only a handful of good books to recommend to our students is now gone, since every year several new textbooks appear.

In view of those favourable points, the question that arises for ASTIN is: what can we do to further enhance those developments? What can we do even to accelerate this trend, besides continuing to publish ASTIN BULLETIN and to offer a splendid forum for interaction, our colloquia? I think that maybe ASTIN has missed an opportunity to help the writing of textbooks. A look at the list of books written in the 80's will reveal that many of these have been published by small university presses or companies. While they certainly have to be congratulated for printing actuarial work, it is clear that they do not have the experience, the expertise of a large publisher; so the promotion of the book is largely left to the unexperienced writer, whose ideas are consequently not going to be read as much as they deserve. Maybe ASTIN should have helped; maybe ASTIN should help in the future. Quite a few of our members have been through this strenuous process of writing a book (often in a language we only practise occasionally) and presenting it to a large publisher. This experience has largely been left unused. Yet ASTIN, as a powerful association, could initiate contacts with a publisher and use all of its power to convince him of the use of actuarial books. ASTIN could possibly start and endorse a new series of textbooks, thereby providing mutual reinforcement to the selected works and presenting them in a unified way. To promote the books that have appeared recently under the name "ASTIN Series" could only have been beneficial to ASTIN and all authors. Some other associations are extremely efficient at sponsoring new books; often they manage to obtain the collaboration of the industry to solve financial problems. Shouldn't ASTIN think about it? At the very least the ASTIN BULLETIN, and possibly the IAA Bulletin, should develop their book review section and let the authors know that their books will be systematically and rapidly reviewed.

A few words about research, to conclude. Again, at first sight, we can be extremely proud of some of our achievements: the mathematics of motor insurance, credibility theory, the theory of premium calculation principles, to name but a few, have been developed and still continue to be developed by our members. Yet, we could perform much better; I am not satisfied at all to notice that quite large areas of research are not being tackled at all by actuaries, and that they grow outside ASTIN. Deductible selection, design of optimal insurance policies, moral hazard, adverse selection, analysis of underwriting cycles, for instance, are areas currently developed by economists and financial analysts, in non-actuarial journals. Those subjects are of extreme practical importance: practitioners know that the selection of a deductible is very often the most crucial decision to be made when designing a new policy; they know that the profitability of health insurance is much more influenced by adverse selection than by the failure to introduce a significant variable in the rate-making process, for instance. Yet those vast areas of research are nowadays explored by others; I am sure the actuarial community could produce very interesting models in these domains.

In this respect, I was pleased to notice that several of the papers presented at the Tel Aviv Colloquium tackle problems that are outside the traditional scope of the works discussed during our meetings; hopefully, many of these papers will be published in the ASTIN BULLETIN, enlarging the range of the subjects dealt with in our association journal.

A few thoughts and suggestions have been presented in this editorial. We should all try to think about other suggestions, other ways to develop further the activities of ASTIN. One such new idea came up two years ago, namely the organization of our first competition for young researchers, which proved to be a tremendous success. Since there are nearly 1,500 of us, it would be very surprising if no new suggestion comes up in the near future.

Jean LEMAIRE
Chairman

IN MEMORIAM

MARCEL HENRY

1900–1986

L'Institut des Actuaires Francais est en deuil: son Président d'Honneur, Monsieur Marcel Henry, est décédé le 14 octobre 1986, dans sa quatre vingt septième année.

Marcel Henry était né le 29 janvier 1900. Après des études au Lycée Condorcet, il fut reçu à l'Ecole Polytechnique en 1918 et incorporé aussitôt comme canonier de 2ème classe. Revenu à l'Ecole, dès la fin des hostilités, il y accomplit ses deux années d'études puis rejoint ensuite l'Ecole d'Application d'Artillerie à Fontainebleau.

Peu de temps après Marcel Henry—déjà attiré par les applications des sciences mathématiques aux problèmes économiques et sociaux—entre à la Statistique Générale de la France et participe ainsi à la naissance de l'organisme qui devait, trente ans plus tard, devenir l'INSEE.

Après quelques années consacrées à la statistique, il se tourne vers l'Assurance et entre à l'Urbaine Crédit.

Il devient membre diplômé de l'Institut des Actuaires Francais en 1924.

Mais dès 1936 il est fait appel à lui pour mener la réorganisation de La Preservatrice, à laquelle contribue l'ensemble de la Profession de l'Assurance. Il est admis comme membre agrégé de l'IAF en 1937.

Il devient rapidement Directeur Général de La Préservatrice IARD et de La Préservatrice VIE. Son action et sa renommée au sein de la Profession comme parmi les Associations d'Actuaires ne vont cesser de s'affirmer.

En 1943, il est désigné par ses collègues comme Secrétaire Général de l'IAF—poste qu'il occupera jusqu'à la retraite du Président Auterbe. Il est alors élu Président de l'IAF en 1958.

Il le restera jusqu'en 1975, souhaitant voir la Présidence occupée désormais par un collègue plus jeune. Il est alors nommé Président d'Honneur. Il continue à ce titre à siéger au Bureau et assiste à ses séances avec une grande régularité. Il prend une part active à la vie de l'Institut auquel il n'a cessé de manifester son intérêt et son dévouement depuis plus de 50 ans.

Marcel Henry était Membre d'Honneur de l'Institute of Actuaries, Membre de l'Association des Actuaries Suisses, et de l'Association Royale des Actuaires Belges.

Il a siégé pendant plus de 30 ans à l'Association Actuarielle Internationale. Il a été un des membres-fondateurs de l'ASTIN, aux travaux de laquelle il était tout particulièrement attaché: c'est lui qui insista pour que le premier Colloque de l'ASTIN se tienne à La Baule en 1955.

Dans la Profession de l'Assurance sa personnalité, sa compétence, son sens de la mesure étaient connus de tous. Il fut Président du Groupement Technique Accident de 1957 à 1960, et Président de la Prévention Routière. Il prit une part active à toutes les grandes réformes du marché de l'Assurance, au cours des années d'après-guerre.

Marcel Henry était Officier de la Légion d'Honneur et Commandeur de L'Ordre National du Mérite.

A toutes ses qualités professionnelles s'ajoutaient de très grandes qualités humaines: doué d'un sens de l'humour qu'il aimait à exercer avec talent en de nombreuses circonstances, il possédait une vaste culture et sa curiosité des hommes et des choses avait fait de lui un voyageur inlassable.

Le souvenir du Président Marcel Henry restera toujours vivant parmi les Actuaires.

JEAN LAMSON

IN MEMORIAM

KARL HENRIK BORCH

13th March 1919–2nd December 1986

Karl Borch's life was eventful—in its outer features adventurous. The realities behind were, however, not always lenient. His studies were interrupted by service in the Free Norwegian Forces in Great Britain during the war. This was a prelude to a cosmopolitan career. After he had received his actuarial degree from the University of Oslo in 1947, he was affiliated to international organizations for a period of 12 years, first to the UN with tasks in the Middle East, South Asia and Africa, then to OEEC as head of the Productivity Measurement Advisory Service in Paris. He received his doctor's degree from the University of Oslo in 1962. In 1963 he was appointed professor of insurance at the Norwegian School of Economics and Business Administration (NHH) in Bergen, a position he held until his untimely death just barely before retirement at pensionable age. In the course of his career Karl Borch stayed at a number of universities: as research associate in Chicago and Princeton, and as visiting professor in California, Vienna, Oxford, Ohio, Bonn, Stockholm, Ottawa and Texas. He died on December 2nd 1986—with his seven-league boots on — in Marbella, Spain.

The written production of Karl Borch is extensive. His major field of interest is indicated by the title of his book "The Economics of Uncertainty" (Princeton University Press, 1968), which has become a modern classic. About one hundred and fifty of his papers have been published in international journals. A selection of papers is collected in the book "The Mathematical Theory of Insurance" (Lexington, Mass., 1974). For his scientific work Karl Borch was awarded a number of prizes and proofs of honour.

Among actuaries and econometricians Karl Borch gained world fame by his contributions to the theory of the economics of uncertainty and, in particular, by invoking this theory in analyses of insurance problems. His pioneering work on Pareto-optimal risk exchanges in reinsurance opened a new area of actuarial science, which has been in continuous growth since. This research field gives a deeper understanding of the attitudes and behaviour of the parties in an insurance market. It is, therefore, of great theoretical import and must, in the end, have a corresponding practical significance. The theory raises and answers problems that could not even be put into shape by traditional actuarial handicraft: how can risk be optimally shared between economic agents, how should insurance treaties be designed, and — ultimately — how should the insurance industry be organized to best further social security and public welfare?

Karl Borch never filled official posts in ASTIN, but he will be recognized as one of the most enthusiastic and influential personalities in the membership. No single person has contributed more to the columns of the ASTIN BULLETIN.

Karl Borch will be remembered by colleagues and students at the NHH and in many other places as a guide and a source of inspiration, by successors in research as the character behind many key references, and by a multitude of people all over the world as a gentle and considerate friend who was full of concern both in their work and in their everyday life.

RAGNAR NORBERG

19th ASTIN COLLOQUIUM
TEL AVIV ISRAEL 20th–24th SEPTEMBER 1986

The 19th ASTIN Colloquium, which was held at the Hilton Hotel in Tel Aviv, was attended by some 175 participants, together with some 75 accompanying persons, from 23 countries. It began in the customary fashion with a welcoming cocktail party on the Saturday evening.

The opening ceremony the following morning was presided over by the Chairman of ASTIN, Jean LEMAIRE. After the ceremony, Professor Eitan BERGLAS of Tel Aviv University gave a lecture on inflation, stabilization, the government budget and the capital market. Professor Berglas, besides being the Chairman of the Bank Hapoalim, was also one of the chief architects of the economic policy which had recently been introduced in Israel and which had led to a reduction in the year-on-year rate of inflation, in the space of a year, from several hundred per cent to less than 20 per cent. As a professor of economics who had been tackling a severely practical problem, so far with apparent success, he was listened to with great interest

Those colloquium papers which had been sent in sufficiently early had been assembled by the organizing committee in a single volume and distributed to the participants in advance of the colloquium. Following the lecture by Professor Berglas, the whole of the remainder of that day was devoted to Subject 1: The company environment of the non-life actuary. The papers were divided into three groups, as set out in the list at the end of this note, corresponding to the three working sessions.

The first four papers considered ratemaking and related issues.

HARWAYNE's paper refers to ratemaking in the United States, in the sense of determining the overall level of the rates to be charged for a class of insurance business rather than determining the premium relativities for the various rating cells. Using an example based on worker's compensation insurance, he illustrates an approach based on projecting paid losses, as an alternative to using figures of incurred losses which include a large element of subjective estimating.

DE PRIL's paper, in which he postulated a portfolio of independent life assurance policies for which the distribution by amount at risk is known and which is subject to known mortality rates, developed recursive expressions for computing the distribution of aggregate claims. The relevance of this paper to non-life insurance was not explained.

The first of SUNDT's papers describes experimental work carried out in the use of credibility regression models to derive results intended as an aid to the persons who have the task of classifying car models into groups for the purpose of insurance rating. Although the paper contains a formidable array of mathematics, the method was nevertheless over-simplified in that whilst it took account of such factors as cubic capacity of the engine, price, weight, etc., it did not incorporate any specific allowance for the interactions with other risk factors such as age of the car and the characteristics of the policyholder or driver. In his second paper he discusses, in the context of a credibility model, the extent to which data derived from the experience of earlier years may be used to yield improved estimators.

Among the papers on reserving techniques, the one by GATH and LUBITCH relates to dental expenses insurance as practised in Israel. Reserving for claims for the basic forms of dental treatment does not present any problems since the claims arise quickly and are settled quickly. The paper describes the authors' approach to reserving for claims for prosthetic treatment, where, although the period from reporting to settlement is quite short, it is necessary to allow not only for claims incurred but not reported, but also for claims for which the treatment has been authorized but not yet carried out.

STANARD, in his paper, describes the use of simulated figures of claim costs by year of accident and year of development, with the aim of estimating the variance of the errors in the estimates of ultimate losses. The paper by FALLQUIST and JONES set out ways in which the data typically available in the form of a loss development triangle can be used to obtain derived figures which may be used in various methods of estimating reserves for outstanding claims and the expenses associated with them.

The remaining two papers on Subject 1 related to profitability management and planning. BOHMAN and LEVY criticize the preoccupation, in insurance company accounting, with the single figure representing the purported profit or loss in the latest completed year, and advocate a form of presentation, at least for the information of management, which acknowledges the stochastic nature of the business. They illustrate their ideas by means of a much over-simplified model in which they separate the underwriting and investment aspects of the insurance business. In TAPIERO's paper, he describes a project to develop a computer expert system for insurance and reinsurance. Much of the paper is occupied with the author's comments on different forms of reinsurance treaty.

On the Monday the working session was confined to the morning, for discussion of the papers on Subject 2: Financial aspects of general insurance. Again, these came under three main headings.

In the first of the papers on solvency, COUTTS and DEVITT advocate the use of cash flow models for reporting on the financial strength of an insurance enterprise, the models embodying probability distributions and sets of decision rules defining future strategy. These two authors are also among those responsible for the paper by DAYKIN *et al.*, which presents the approach developed by a working party which was set up to extend, in the context of the solvency regime in the UK, the work carried out by the Finnish group of actuaries led by PENTIKAINEN and RANTALA. They recommend the simultaneous consideration of assets and liabilities, and put forward a case for a system of actuarial reporting on the financial strength of insurance companies, so that the supervisory authority can have available, within a solvency regime based on crude minimum solvency margins, an assessment which reflects the level of risk for the individual company.

There were two papers dealing with mathematical aspects of reinsurance. The short paper by KASS, GOOVAERTS and BAUWELINCKX derives upper and lower bounds for stop-loss premiums in terms of the claim number distribution, the mean claim, the mean claim less than the retention, and the probability of a claim less than the retention. LEMAIRE and QUAIRIERE generalize earlier work by Gerber on chains of reinsurance, using a theorem of Borch.

The miscellaneous group of papers on other financial aspects of general insurance includes two highly theoretical papers, one by BRIYS and the other by BRIYS, KAHANE and KROLL, on optimal insurance demand by individuals. The paper by EDEN and KAHANE presents a model of an insurance market with three players: an insured population, a local insurer and an international reinsurer, and discusses the allocation of risk among the three parties. MASTERSON presented the latest sets of figures in his customary review of price indices in various countries and indices of the costs of different types of non-life insurance claims in the US.

After lunch Mr. Gideon PATT, the Minister of Science and Development, gave an address in which he vigorously supported current policies of the Israeli government.

The final working day was the Tuesday, starting with Subject 3: The actuarial treatment of catastrophe — target risks and special lines.

ARIAV, KAHANE and TAPIERO discuss the possible advantages to a group of companies of establishing a back-up computer centre, which might, for example, be owned by a separate company which would sell the services when required to the members participating in the pool.

AJNE and WIDE discuss the question of defining catastrophe claims and assessing their expected cost. For their purpose, catastrophe claims are large individual claims, rather than large groups of claims arising out of a single catastrophic incident. The Pareto family of curves was found to give the best fit to the claim distributions for the four classes of business which were treated, and the paper includes comparisons between actual and expected numbers and amounts of catastrophe claims over a four-year period.

KAHANE's paper gives a wide-ranging survey of the problems of insurance against earthquake risks, with particular reference to Israel, and sets out a number of proposals as to how these problems should be tackled.

Subject 3 was followed by the customary Speakers' Corner session. Besides the presentations for which papers were submitted — these are listed at the end of this note — there were presentations by S. BENJAMIN on the reserving methods used at Lloyd's, by E. KREMER on premium calculation for largest claims reinsurance covers and on robust premium principles, and by R. NORBERG on life insurance rating.

The business proceedings ended with the General Assembly of ASTIN, at which the plans were announced for the 20th Colloquium, to be held in Scheveningen, Netherlands, 30th August–3rd September 1987.

The Colloquium dinner was held on the Tuesday evening at the Hilton Hotel.

An important feature of any colloquium is the opportunity it gives for the participants to discuss matters of mutual interest before and after the business meetings. The social arrangements provided by our Israeli hosts included a piano recital by David Levy, a talented young musician from the Royal Northern College of Music in Manchester; a visit to the Diaspora Museum, followed by a tour of the old city of Jaffa; and a full-day guided tour of Jerusalem and Bethlehem. For those able to stay for a further day, there was a tour of Galilee.

The excellent arrangements for both the business meetings and the social events were a tribute to the organizing committee led by Yitzchak GOLDSTEIN and Eddy LEVY and the scientific committee led by Yehuda KAHANE.

PETER JOHNSON

LIST OF PAPERS

Opening Lecture

Prof. E. BERGLAS
    Inflation, stabilization, government budget and capital market.

Subject 1: The Company Environment of the Non-Life Actuary

Ratemaking and Related Issues

F. HARWAYNE
    A comparison of practical ratemaking models: paid vs. incurred claims and claims adjustment expense experience.

N. DE PRIL
    On the exact computation of the aggregate claims distribution in the individual life model.

B. SUNDT
    Some credibility regression models for the classification of individual passenger car models.

B. SUNDT
    Credibility and old estimates.

Reserving techniques

Y. GATH and A. LUBITCH
    Reserving in dental insurance.

J. N. STANARD
    Bootstrap estimation of the variance of prediction errors of loss reserving methods.

R. J. FALLQUIST and B. A. JONES
    Loss reserving in the microcomputer environment.

Profitability management and planning

H. BOHMAN and E. J. LEVY
    Diagnostic tools for profitability.

C. S. TAPIERO
    Executive insurance system: the computer insurance expert.

Subject 2: Financial Aspects of General Insurance

Solvency

S. M. COUTTS and E. R. F. DEVITT
The assessment of the financial strength of insurance companies: a generalized cash flow model.

C. D. DAYKIN, G. D. BERNSTEIN, S. M. COUTTS, E. R. F. DEVITT, G. B. HEY, D. I. W. REYNOLDS and P. D. SMITH
The solvency of a general insurance company in terms of emerging costs.

Reinsurance

R. KAAS, M. J. GOOVAERTS and T. BAUWELINCKX
Some elementary stop-loss inequalities.

J. LEMAIRE and J. P. QUAIRIERE
Chains of reinsurance revisited.

Other financial aspects

E. BRIYS
On the theory of rational insurance purchasing in a continuous-time model.

E. BRIYS, Y. KAHANE and Y. KROLL
The demand for voluntary insurance coverage in the presence of compulsory insurance, speculative risk and risky–riskless portfolio opportunities.

Y. EDEN and Y. KAHANE
Moral hazard and insurance market structure.

N. E. MASTERSON
Inflation: an economic factor in non-life insurance.

Subject 3: The Actuarial Treatment of Catastrophes: Target Risks and Special Lines

A. ARIAV, Y. KAHANE and C. S. TAPIERO
A pooled computer center as a risk management tool.

B. AJNE and H. WIDE
On the definition and the expected cost of catastrophe claims: an application to the business insurance portfolio of a direct insurance company.

Y. KAHANE
Earthquake risk and insurance in Israel: policy considerations.

Speakers' Corner

E. P. PEREZ
Structure de la prime et rendements financiers des provisions techniques des assurances privées.

B. ZEHNWIRTH and G. BARNETT
Fundamental principles of forecasting and chain ladder model.

G. BENKTANDER
Claim size distributions: Pareto as an approximation to other continuous distributions in the tail areas.

Y. KAHANE, C. S. TAPIERO and L. JACQUE
Concepts and trends in the study of insurers' solvency.

G. SZPIRO
Risk aversion in the not-so-small: beyond mean and variance.

Y. EDEN and Y. KAHANE
Optimal reinsurance contracts and the insurance capacity problem.

R. KAAS and M. J. GOOVAERTS
On the use of Quadpack for the calculation of risk theoretical quantities.

H. BOHMAN
ELCED (A program which performs an algorithm representing a hypothetical arrangement between a ceding company and a reinsurer).

H. BOHMAN, A. NETTERMAN and E. LEVY
Non-life insurance premium using factorial method.

APPROXIMATIVE EVALUATION OF THE DISTRIBUTION FUNCTION OF AGGREGATE CLAIMS[1]

BY T. PENTIKAINEN

Helsinki

ABSTRACT

A formula, originally presented by HALDANE (1938)[2], for the evaluation of the distribution of aggregate claims is examined and compared with some other approaches. The idea is to apply a symmetrizing transformation to the original variable in order to make it susceptible to be approximated by the normal distribution.

KEYWORDS

Aggregate claim distribution; approximate evaluation; NP method; Haldane approximation; Wilson–Hilferty approximation

1. INTRODUCTION

1.1. A problem frequently faced in application-orientated risk theory is the numerical evaluation of the distribution function $F(X)$ of the aggregate amount X of claims. It is conventionally expressed by the formula

$$F(X) = \sum_{k=0}^{\infty} p_k S^{k*}(X), \tag{1.1}$$

where p_k is the probability that the number of claims is equal to k, S is the distribution function of the individual claim sizes and S^{k*} its kth convolution (see BPP, p. 51; this and similar quotations in the sequel refer to the Risk Theory book by BEARD, PENTIKAINEN and PESONEN, 1984).

The claim numbers are often assumed to follow a simple Poisson distribution (in order to define notations some well-known basic formulae are recapitulated)

$$p_k(n) = e^{-n} n^k / k! \tag{1.2}$$

where n, the Poisson parameter, is the expected number of claims. A more general approach is to adopt the mixed Poisson distribution (BPP, p.33):

$$p_k(n) = \int_0^{\infty} e^{-nq} \frac{(nq)^k}{k!} \, dH(q). \tag{1.3}$$

1 Presented originally at the Risk Theory Seminar in Oberwolfach 1984 and in an extended form at the Risk Theory Seminar of the American Risk and Insurance Association in Nashville 1985

2 J B S HALDANE (1892–1964) first studied mathematics and later became Professor of Biometrics at University College, London, before moving to India in 1957

The distribution function H introduces the so-called structure variation of the claim number probabilities into the model, indicating the time variation of the risk exposure, or the heterogeneity of the risks inside the collective, or both. A popular H is the gamma function resulting in the negative binomial distribution (BPP, p. 40):

$$p_k(n) = \binom{h+k-1}{k}\left(\frac{h}{n+h}\right)^h\left(\frac{n}{n+h}\right)^k \tag{1.4}$$

where h is a shape parameter. This alternative is usually called the *Polya case* in distinction from the *Poisson case* (1.2).

Note that for the approximation formulae only the lowest moments of H are necessary. Therefore, it is sufficient merely to estimate (or assume) them, not bothering about the analytic formulation of this function.

1.2. Need to have approximation methods. The construction (1.1) is unfortunately so intricate that the direct computation of F is tractable only in special cases, even though the recently developed recursive methods (ADELSON, PANJER, GERBER, JEWELL & SUNDT, description of the method and references see BPP, Section 3.8) as well as the Fourier transformation technique (BERTRAM, 1981) have made major progress in solving this problem. Nevertheless, they have not removed the need also to have rapid and reasonably comfortable, even though approximate, approaches. This is due to the fact that the number of the computation steps needed for the recursive calculation grows quite massive in cases where the risk portfolio is large (as most insurer's portfolios are) and/or when the claim size distribution has a long tail. This can be a major handicap, in particular in a sophisticated problem complex such as the analysis of long-term processes, simulations, etc. where the computation of F is needed frequently, say 1000, 10000 or 100000 times for one single procedure. Then the problem is, above all, to minimize the computation time in terms of milliseconds (rather than in seconds!) to make the usual present day personal computers operational.

Approximate methods can also have the merit of providing an analytic, often perspicuous, relationship between the main variables controlling the processes.

2. SOME EARLIER APPROACHES

2.1. Normal approximation. A classic approach, based on the central limit theorem, is to approximate $F(X)$ by the normal distribution:

$$F(X) \approx N(x) \tag{2.1}$$

where, denoting the mean and the standard deviation of X by m_X and σ_X

$$x = (X - m_X)/\sigma_X. \tag{2.2}$$

This expression is asymptotically correct in the Poisson case but not generally, e.g. not in the Polya case. Its major weakness is that it may crudely underestimate the risk of large aggregate claims (see BPP, p.105). This is due to the fact that N as a symmetric function cannot successfully approximate any distribution which is notably skew.

2.2. Symmetrization. A way of overcoming the weakness of the normal approximation is to transform the original variable X into an auxiliary variable y_{norm} by using a suitably chosen function v

$$y_{\text{norm}} = v(X) \tag{2.3}$$

so that it makes the distribution (at least approximately) symmetric. Then providing that y_{norm} is standardized to have a zero mean and standard deviation unity, one can expect that it can be satisfactorily approximated by the normal distribution:

$$F(X) = \bar{F}(y_{\text{norm}}) = N(y_{\text{norm}}). \tag{2.4}$$

Depending on the choice of the transformation v a family of approximation methods is constituted including those dealt with in this paper. An analysis of some of these transformations can be found for example in BOX and COX (1964).

2.3. NP approximation is obtained assuming $v^{-1}(y_{\text{norm}})$ as a polynomial:

$$x = (X - m_X)/\sigma_X = y_{\text{norm}} + \gamma(y_{\text{norm}}^2 - 1)/6 \tag{2.5}$$

where $\gamma = \gamma_X$ is the skewness of the original distribution (BPP, Chap. 3.11).

The transformation (2.5) is applicable only for the long tail $X > m_X$ of the distribution and therefore needs a modifying extension (BPP, Chap. 3.11) resulting in a three-piece formula.

2.4. Other methods. There are a number of approaches based on the principle of replacing the original distribution by some suitable approximating function, which is conveniently computable. Most of them are obtained by equating the lowest moments, as is also the case in the above items 2.2 and 2.3. For instance the three-parameter gamma function $\Gamma(ax + b, c)$ (BOHMAN and ESSCHER, 1964) or the Pearson functions (LAU, 1984) are suggested. OSCHWALD (1984) has recently presented an analogous transformation to (2.5) using the gamma function instead of the normal function.

Unfortunately the range of applicability of most of these methods has been examined only by means of very few (and often "easy") examples, as yet. So far as is known, those methods which meet the demand of reasonable convenience and the requirement for computation speed, do not have the accuracy or the other merits which would not prevent the approaches to be dealt with in the sequel to be competitive. Further studies would be desirable, but are, however, beyond the scope of this work.

One of the known approximations is still worthy of special mention. ESSCHER (1932) introduced a method which makes use of the whole range of the claim size distribution, not only of some of its moments. BOHMAN and ESSCHER (1964) gave a number of tests, which proved that the merits of this method may not be very superior to those of the gamma approximation. However, a recent (unpublished) work of PUSA (1985) seems to indicate a good fit also in some cases where the other methods fail. A drawback of the Esscher approach is that it employs an auxiliary variable h and the relevant quantitites F and X are available only as functions of this variable. To get the matching F and X it is necessary first to find the corresponding h, which seems to need an iteration process impairing the speed of the method.

2.5. Moment problem. One should keep in mind the fact that most of the methods referred to above give, when applied to the mixed compound Poisson function, the same approximating function to all those original distributions, which have the same three (or four) lowest moments determined by the mean, standard deviation, skewness (and the kurtosis) of the claim size distribution. However, these moments do not fully determine the flow a function, hence there is "a funnel of doubt" $[F_l(x), F_u(x)]$ inside which the values of the original distribution functions $F(x)$ are positioned for varying x values. F_l and F_u are the upper and the lower envelope curve, respectively, of all those original distribution functions which fulfil the specifications of the problem setting concerned. If the funnel is large, then there are always cases which markedly deviate from the approximating values, whichever of the methods is used.

GOOVAERTS & KAAS (1986) have recently provided a meritorious method of evaluating the range of the variation subject to the condition that the lowest moments are fixed and the variable x is limited to some interval, e.g. $[0, b]$. Some examples showed that the range in which the permitted $F(x)$ values may be positioned is rather wide, in fact reducing the prospects of finding suitable approximations based on the moments. Fortunately, this result does not wreck the prospects of finding reasonably useful approximations, if the basic condition is taken of fixing a sequence of the moments of *the claim size function S* (not of the aggregate distribution as Goovaerts and Kaas propose) and of limiting the claim size Z (not the aggregate X) to some finite interval. In fact, this is the proper problem setting for the NP method as well as for the methods to be discussed in the sequel. We will return to the moment problem in Section 6.9.

3. HALDANE APPROXIMATION

3.1. The idea. The approach we are going to deal with was originally presented by WILSON and HILFERTY (1931) for an approximate evaluation of the gamma function. HALDANE (1938) extended it to the function classes which have suitably convergent sequences of cumulants. In what follows we apply the method to the mixed compound Poisson function specified in Section 1.1, even though the most part of the derivation is valid more generally.

The method makes use of the symmetrization as described in Section 2.2 above. Haldane first adopted a power expression

$$y = (X/m_X)^h$$

where h is an auxiliary parameter. Then (truncated) expansions are derived for the mean m_y, standard deviation σ_y and skewness γ_y of y. The symmetrization is achieved by assigning for the auxiliary parameter h a value which equates the skewness γ_y with zero. Putting

(3.1) $$y_{\text{norm}} = (y - m_y)/\sigma_y = [(X/m_X)^h - m_y]/\sigma_y$$

the transformation aimed at, corresponding to v in (2.3), is obtained.

The derivation of the Haldane formulae is notably laborious, even though rather elementary methods only are needed. Therefore, we shall not give more than some intermediate expressions in the chain of treatments in order to provide a conception of how the results are found, the more so because the Haldane derivations do not result in any strictly rigorous estimates for the accuracy of the approximations nor for clear rules of their applicability, but rather only justify the expectation that in a certain environment the procedure may lead to acceptable outcomes. In fact, the discussion about the appropriateness of the approach is mainly based on tests where a number of distributions are calculated exactly and in parallel, by using the approximations that will be presented in the subsequent sections. Readers who are mainly interested in the practical results may well skip over to Section 3.4, at least at the first reading.

3.2. Derivation of the formula. We aim to preserve, as far as possible, the original procedures and notations of Haldane even though some of the results, e.g. the value of the parameter h, could be obtained by more straightforward ways.

The technique to be assumed operates partially the so-called cumulants, which are, as is well known from textbooks on statistics (e.g. KENDALL and STUART, 1979, Section 3.12), the coefficients in the expansion of the cumulant-generating function $\log \varphi(t)$ in the terms of (it): $\varkappa_1(it) + \varkappa_2(it)^2/2! + \dots$, where φ is the characteristic function and i the imaginary unit. For the convenience of the reader, we recall the connecting equations between the lowest cumulants and the more commonly applied central moments, denoted by $\mu_i (i = 1, 2, \dots)$:

$$\varkappa_1 = m_X,\ \varkappa_2 = \mu_2,\ \varkappa_3 = \mu_3,\ \varkappa_4 = \mu_4 - 3\mu_2^2.$$

Another auxiliary variable is introduced:

(3.2) $$x' = X - \varkappa_1$$

and substituted into (3.1) after which expansions for the moments of y about zero, denoted by β_r, are obtained as follows

(3.3) $$\beta_r = Ey^r = E\left\{\left[\left(1 + \frac{x'}{\varkappa_1}\right)^h\right]^r\right\} = E\left\{\left(1 + \frac{x'}{\varkappa_1}\right)^{rh}\right\}$$

$$= 1 + f(r,2)\,\frac{\mu_2}{2!\varkappa_1^2} + f(r,3)\,\frac{\mu_3}{3!\varkappa_1^3} + f(r,4)\,\frac{\mu_4}{4!\varkappa_1^4} + \dots$$

where, for brevity,

(3.4) $$f(r,j) = rh(rh-1)(rh-2)\ \dots\ (rh-j+1)$$

and the Taylor series of the power function

$$(1+x)^h = 1 + hx + h(h-1)x^2/2! + \dots$$

was applied. The μ's are moments of x' and, according to (3.2), central moments of the original variable X as well. It is now useful to adopt the cumulants $\varkappa$ of X as the current characteristics instead of the moments μ, because they give an essentially better convergence behaviour to the expansions. By using the relations referred to above between these sets of characteristics, equation (3.3) is transformed as follows:

$$\beta_r = 1 + f(r,2)\frac{\varkappa_2}{2\varkappa_1^2} + f(r,3)\frac{\varkappa_3}{6\varkappa_1^3} + f(r,4)\frac{3\varkappa_2^2+\varkappa_4}{24\varkappa_1^4} + f(r,5)\frac{\varkappa_2\varkappa_3}{12\varkappa_1^5} + f(r,6)\frac{\varkappa_2^3}{48\varkappa_1^6} + \ldots \tag{3.5}$$

The sequence

$$\rho_i = \varkappa_i/\varkappa_1^i \tag{3.6}$$

is assumed to be reasonably convergent when i increases (in the Poisson case ρ_i is of order n^{-i+1}, c.f. Section 5.1).

Finally, the central moments of y are the characteristics which are necessary for the transformation aimed at. They are calculated by means of the well-known general relationship between the central moments and the moments about zero ($\mu_2 = \beta_2 - \beta_1^2$, etc.). After elementary but quite tedious operations, and observing that the expression (3.5) can be expressed in terms of the cumulant ratios ρ_i, the following expansions result:

$$\begin{aligned} m_y = Ey = {} & 1 + \tfrac{1}{2}h(h-1)\rho_2 + \tfrac{1}{24}h(h-1)(h-2)[4\rho_3 + 3(h-3)\rho_2^2] \\ & + \tfrac{1}{48}h(h-1)(h-2)(h-3)[2\rho_4 + 4(h-4)\rho_2\rho_3 + (h-4)(h-5)\rho_2^3] \end{aligned} \tag{3.7}$$

$$\begin{aligned} \mu_2(y) = \varkappa_2(y) = \sigma_y^2 = {} & h^2\rho_2 + \tfrac{1}{2}h^2(h-1)[2\rho_3 + (3h-5)\rho_2^2] \\ & + \tfrac{1}{12}h^2(h-1)[(7h-11)\rho_4 + 4(h-2)(7h-12)\rho_2\rho_3 + 2(h-2)(7h^2-30h+32)\rho_2^3] \end{aligned}$$

$$\begin{aligned} \mu_3(y) = \varkappa_3(y) = {} & h^3[\rho_3 + 3(h-1)\rho_2^2] \\ & + \tfrac{1}{2}h^3(h-1)[3\rho_4 + 3(7h-10)\rho_2\rho_3 + (17h^2-55h+44)\rho_2^3] \end{aligned}$$

$$\varkappa_4(y) = \mu_4(y) - 3\mu_2^2(y) = h^4[\rho_4 + 12(h-1)\rho_3\rho_2 + (h-1)(16h-20)\rho_2^3].$$

Terms only having the order three or less were accepted. (I am grateful to my colleague Mr. H. Simberg for the correct $\varkappa_4$.)

We are now enabled to fix the parameter h. For the aimed at symmetrization of the transformed distribution its skewness should be made to vanish, or, what is the same, $\varkappa_3(y)$ should be equal to 0. For the sake of computational convenience, only its leading term will be equated to zero. Hence

$$h = 1 - \tfrac{1}{3}\rho_3/\rho_2^2 = 1 - \frac{\varkappa_1\varkappa_3}{3\varkappa_2^2}. \tag{3.8}$$

3.3. Transformation A. Haldane now states that, when the above value of h is substituted into equations (3.7), the standardized variable

$$y_{\text{norm}} = (y - m_y)/\sigma_y \tag{3.9}$$

is "almost normally distributed" with a zero mean and standard deviation unity. Haldane calls this formula "Transformation A" as distinguished from another transformation to be dealt with shortly.

3.4. For the risk theory applications, where the mixed compound Poisson function is to be approximated, it is convenient to take the basic characteristics (see BPP, p.54):

$$\begin{aligned} \text{Mean} &= m_X = \varkappa_1 \\ \text{Standard deviation} &= \sigma_X = \sqrt{\varkappa_2} \\ \text{Skewness} &= \gamma_X = \varkappa_3/\sigma_X^3 \end{aligned} \tag{3.10}$$

of the original distribution as the entries of the calculations. Accepting the terms of the order of at most two and introducing as an auxiliary quantity

$$s = \sigma_X/m_X \tag{3.11}$$

the Haldane approach A can be written in an operative form

$$\begin{aligned} h &= 1 - \tfrac{1}{3}\gamma_X/s \\ m_y &= 1 - \tfrac{1}{2}h(1-h)[1 - \tfrac{1}{4}(2-h)(1-3h)s^2]s^2 \\ \sigma_y &= hs\sqrt{[1 - \tfrac{1}{2}(1-h)(1-3h)s^2]} \\ y_{\text{norm}} &= [(1+xs)^h - m_y]/\sigma_y \\ F(X) &= N(y_{\text{norm}}). \end{aligned} \tag{3.12}$$

The lower case x refers to the standardized variable (2.2). The third degree quantities $\varkappa_3$ (and γ_3) were eliminated for computational convenience by using (3.8).

Note that the moment $\mu_4(y)$ was not needed in this context. It was derived in (3.7) because it will be useful in later sections.

The above formulas are fairly comfortable for computer programming. Examples will be given later and the applicability discussed, but before that we will make some further remarks and present an extended version of the transformation.

3.5. Negative h values. Haldane limited the range of validity to positive h values only. Examples show that the formula also works in cases where h turns negative. However, negative values seem to appear in the area where the skewness is excessive and the goodness of fit is unsatisfactory.

3.6. The case $h = 0$. Special attention is called to the case when $h \to 0$. Then y_{norm} has the limit

$$y_{\text{norm}} = [\ln(1+xs) + \tfrac{1}{2}s^2 - \tfrac{1}{4}s^4]/[s\sqrt{(1 - \tfrac{1}{2}s^2)}]. \tag{3.13}$$

3.7. Wilson and Hilferty applied, as mentioned already, essentially the same transformation as Haldane for the evaluation of the gamma function and arrived at a constant value of 1/3 for the parameter h. On the other hand we know that the original distribution can be approximated by the gamma function which is obtained by equating the mean, standard deviation and skewness with the corresponding characteristics of the distribution to be approximated. Hence we can expect that the Wilson–Hilferty formula may also approximate the original distribution. The result can be written as follows (see BPP, p.71)

$$F(X) \approx \Gamma(\alpha + x\sqrt{\alpha}, \alpha) \approx N[c_1 + c_2(x + c_3)^{1/3}] \tag{3.14}$$

where x is the standardized variable (2.2) and

$$c_1 = \frac{\gamma}{6} - \frac{6}{\gamma} \qquad c_2 = 3\left(\frac{2}{\gamma}\right)^{2/3}; \qquad c_3 = \frac{2}{\gamma}.$$

This formula is very comfortable for computer programming as is also its inverse. Therefore, it is tested in parallel with the Haldane and NP approximations in what follows.

3.8. A link to the NP formula can be found by expanding y_{norm} in (3.12) in terms of (xs) as a Taylor series and then expressing h by means of γ:

(3.15)

$$\begin{aligned} y_{\text{norm}} &= [(1 + xs)^h - m_y]/\sigma_y \\ &= [1 + hxs + \tfrac{1}{2}h(h-1)x^2s^2 + \ldots - 1 - \tfrac{1}{2}h(h-1)s^2 - \ldots]/(hs - \ldots) \\ &= x + \tfrac{1}{2}(h-1)s(x^2-1) + \ldots \\ &= x - \tfrac{1}{6}\gamma(x^2-1) + \ldots \end{aligned}$$

But this is just what is also obtained if y in (2.5) is expanded in the terms of x (see BPP, p. 117, eq. (3.11.14)). Hence it can be expected that the Haldane and the NP formulas are close to each other at least in the area of the best convergence. This will be confirmed by examples given later.

4. HALDANE'S TRANSFORMATION B

Haldane also experimented with another formula, which is derived introducing *two parameters h and g* (instead of only one, h, above). They are assigned values which minimize both the skewness and the kurtosis of the transformed variable. The new parameter g is chosen so that the original variable X is first transformed to another variable which has g as its mean and consequently also as $\varkappa_1$:

$$X'' = X + g - \varkappa_1. \tag{4.1}$$

Then the transformation $y = (X/m_X)^h$ is replaced by

$$y = (X''/E[X''])^h = [1 + (X - \varkappa_1)/g]^h. \tag{4.2}$$

The transformation (4.1) does not affect higher cumulants than $\varkappa_1$. Hence all the results are still valid, if $\varkappa_1$ is replaced by g everywhere.

The parameters h and g are now determined by equating the leading terms of $\varkappa_3$ and $\varkappa_4$. This implies that

$$(4.3) \qquad g = \frac{12\varkappa_2^2\varkappa_3}{20\varkappa_3^2 - 9\varkappa_2\varkappa_4}, \qquad h = \frac{16\varkappa_3^2 - 9\varkappa_2\varkappa_4}{20\varkappa_3^2 - 9\varkappa_2\varkappa_4}.$$

By using the characteristics (3.10) and, in addition, the kurtosis γ_{2X} ($= \varkappa_4/\sigma^4 = \mu_4/\sigma^4 - 3$) of the original variable X and further introducing, for brevity, the auxiliary coefficients

$$(4.4) \qquad b = \tfrac{5}{3}\gamma_X - \tfrac{3}{4}\gamma_{2X}/\gamma_X; \qquad c = \tfrac{4}{3}\gamma_X - \tfrac{3}{4}\gamma_{2X}/\gamma_X$$

the *Haldane transformation B* can be written as follows

$$(4.5) \qquad \begin{aligned} h &= c/b \\ m_y &= 1 - \tfrac{1}{2}c(b-c)[1 + \tfrac{1}{4}(2b-c)(3c-b)] \\ \sigma_y &= |c|\sqrt{[1 + \tfrac{1}{2}(b-c)(3c-b)]} \\ y_{\text{norm}} &= [(1+bx)^h - m_y]/\sigma_y \\ F(X) &= N(y_{\text{norm}}) \end{aligned}$$

where the lower case x again refers to the standardized aggregate claim (2.2).

Haldane made some reservations concerning the applicability of this transformation, mainly providing for the positivity of h. If the denominators of (4.3) are vanishing, the formulas become invalid.

5. ON THE APPLICABILITY OF THE HALDANE EXPANSION

5.1. General conditions for convergence of the expansions. Haldane assumed that the variable X is inherent from a collective, the risk volume of which can be described by a parameter n. In our risk theory applications, n can be just the expected number of claims as provided in Section 1.1. Furthermore, Haldane assumed that the cumulants $\varkappa_i$ for $i = 1, 2, 3$ and 4 are of the order n when n grows large and that cumulants for $i > 4$ are of the order n^{i-4} or less. Haldane states that the expansions concerned are asymptotically convergent when n is large enough, i.e. the transformations can be applied in large collectives.

In the Poisson case the cumulants are $\varkappa_i = na_i$, a_i being the ith moment about zero of the claim size variable. Hence, the Haldane conditions are satisfied in so far as the moments a_i are finite. On the other hand the asymptotic behaviour of the Polya case does not fulfil the conditions. Moreover, the volume parameter n is always finite, often rather small, in practical applications. Then the Haldane criterion does not suit, because the convergence of the relevant expansions may be poor, even though they may asymptotically converge. Furthermore, one should appreciate that the convergence of the expansion itself does not guarantee full accuracy because there are other, deeper, aspects involved, e.g. those which we already discussed preliminarily in Section 2.5 about the moment problem. Fortunately, and unexpectedly, the tests given in the following sections seem to prove that the approximation outcomes are also fairly satisfactory in numerous cases where the convergence criterion would suggest failure. Therefore, we do not feel that it is useful to explore the problem of the convergence of the Haldane expansions other than that a convergence indicator will be introduced in Section 5.3 below. Otherwise the original paper is again referred to.

5.2. Measures of deviation from normality were suggested by Haldane, making use of the residue skewness and kurtosis which remain as they are reduced by the symmetrization procedure. Hence (cf. (3.7))

$$\text{(5.1)} \qquad \begin{aligned} \text{Measure } 1 &= \gamma' = \varkappa_3(y)/\sigma_y^3 \\ \text{Measure } 2 &= \gamma_2' = \varkappa_4(y)/\sigma_y^4 \end{aligned}$$

where the value (3.8) is to be assigned to h.

The same formulas are valid also for the transformation B, when h is taken from (4.3) and $\varkappa_1$ is replaced by g; it is also obtained from (4.3).

These measures will be illustrated by examples in Section 6.

5.3. Cumulant ratios. A crucial condition for the convergence of the expansions (3.7) is a rapid convergence of the sequence of the cumulant ratios ρ_i, defined by (3.6). Therefore, the author experimented with the indicator

$$\text{(5.2)} \qquad \vartheta = \rho_4/\rho_3$$

as an alternative measure for applicability. If ϑ is small, it implies that the higher cumulant ratios can be expected to be negligible. Values of ϑ are given in the context of test examples and an overall view is provided by Figure A.7 (Appendix 2).

6. EXAMPLES

6.1. Tests. The approximation methods dealt with in the previous sections are tested by calculating a great number of numerical examples on the one hand by using the exact recursive formula, and on the other hand the NP, Wilson–Hilferty (briefly WH), Haldane-A (HA) and the Haldane-B (HAb) approaches. Both the Poisson case and the Polya case, having differing shape parameters h, were examined experimentally. The claim size distribution was the truncated Pareto or log-normal or their mixtures or could also be freely chosen (and given manually to the computer). Because the recursive technique (see details in BPP, Section 3.8) is applicable merely for discrete distributions, the claim sizes were discretized permitting only integer values $Z = 1 + id (i = 0, 1, 2, \ldots, I)$ where d and I are freely eligible positive integer parameters.

The tested distributions, 54 in total, were chosen to cover broadly the area that is usually applied in risk theory considerations, and also to provide comparisons between the approaches. Regretably, it is not possible to print all the data.

Typical cases only were picked for the tables and diagrams given in Appendixes 1 and 2. A comprehensive collection of the data will be deposited in the Library of the Actuarial Society of Finland (Address: Bulevardi 28, 00120, Helsinki 12). Copies are available upon request.

6.2. Appendix 1 exhibits exact F values and the approximated ones in parallel. Numerical values for the convergence criterions proposed above are also given in the side column of each distribution box of each of the tables. Discussion of the outcomes will be deferred to Section 7.

6.3. Figures A.1 and A.2 (Appendix 2) graphically present two of the distributions of Appendix 1 to provide a clearer illustration. The deviations between the relevant curves are so slight that they are scarcely discernible in cases where the distribution is not markedly skew. Therefore, the tails of Figure A.1(*a*) are plotted in a magnified scale in Figure A.1(*b*).

6.4. The effect of discretization. The deviations between the exact and the approximated values are partially due to the fact that the approximating functions always, more or less, deviate from the exact one and partially to the fact that the "exact" F is discrete but the approximating functions are continuous. This is clearly seen in Figure A.1(*b*). In order to eliminate the effect of this discrepancy from the tabulated outcomes, such as given in Appendix 1, the discrete F curves were replaced by a broken line which connected the midpoints of the upward steps. If this kind of smoothing of the discrete results is not made the comparison deviations depend on where, for the purpose of comparison, the selected values of the x variable are positioned on the x-axis. As seen in Figure A.1(*b*) the effect may be larger than the "genuine" deviations are, and depends on whether the test point happens to fall immediately before or after a step.

Information about the steps of F is provided in the last columns of the tables of Appendix 1 where the half of the step height ($= dF\%$) is given. It proved to be mostly larger than the approximation errors in the preceding columns as long as the skewness remained moderate. It depends on the actual relevant problem setting as to whether or not it should be regarded as appropriate to add both the errors.

Note that the discretization results in inaccuracy also if a continuous original claim size distribution, such the Pareto one, is replaced by a step function. This feature was discussed in BPP (Section 3.8c). Because both the exact method and the approximating methods were based on the same discretized claim size distribution, this inaccuracy did not appear in our tests.

6.5. Figure A.2 represents an extreme case where the skewness is large. Then all the approximations turn out irregular and lose their applicability particularly at the tails of the curves.

6.6. Figure A.3 attempts to provide a summarizing survey over a sample of distributions. The relative errors (dNP%, dWH%, etc. in Appendix 1) are grouped according to different skewness ranges and argument values $x = -2, 2$ and 3 respectively and then shown in the specified diagram boxes. For example all the relative errors of the Haldane-A approximation for $x = 3$ and the skewness less than 0.3 are placed upon the "point" HA-a in the top-most right-hand box. The points inherent from the same distribution are connected by the lines in between.

This figure is meant to provide a rapid visual comparison of the tested approaches. A narrow bundle of the connecting lines indicates a good overall fit of the formula concerned.

6.7. The measures of deviation (5.1) are investigated in Figure A.4. This and the remaining figures are limited to the Haldane-A method only.

The tested cases for $x = -2$ and $x = 3$, respectively, are displayed in the diagram by using the measures 1 and 2 as coordinates. The relative errors $|dHA\%|$ are indicated by symbols, as shown in the figure.

As expected, the fit is good for small measure values. Another useful observation is that the measure values are well correlated, i.e. the points are clustered at a straight line. This suggests that it is sufficient to use only one of the measures, preferably the measure 1.

6.8. Convergence properties are studied in Figures A.5, A.6 and A.7. The tested cases were first placed in Figure A.5 by using the standardized x and the skewness as coordinates. The tests were made only for a sequence of discrete x values $= -2, -1.5, -1, \ldots, 4$. For clarity, the points, such as in Figure A.4, were not plotted in the final diagram, but the zones where the errors $|dHA\%|$ having some specified magnitudes are positioned were used instead. For example, in the area below the zone boundary designated by 1 only cases that have $|dHA\%|$ less than 1% are found, and below the 3-boundary only cases having $|dHA\%|$ less than 3%, etc. More exactly, the points of the boundary numbered by N ($N = 1, 3, 5, 10$ or 25) were determined according to that sample case for which the relevant x-value had the error $|dHA\%| \geqslant N\%$ and the lowest skewness. Note that cases having $|dHA\%| < N\%$ may be also found above the N-boundary, even though they are mainly clustered below.

The fit is good as long as the skewness is relatively small. This is a well-known feature about, for example, the applicability of the NP method (BPP, Section 3.11c). Note that $1 - F$ is very small for $x > 3.5$ if the skewness is not excessive Hence the poor relative accuracy in the lower right-hand corner is seldom harmful in applications.

The somewhat zigzag course of the zone boundaries is due to fact that the goodness of fit is sensitive to the selection of tested distributions. Of course, if another set of distributions were chosen, a more or less differing course for the boundaries would result. However, the number of tests, 54, was already so large and the selections so variable that it is not likely that any very essential differences would appear.

Figure A.6 represents the dependence of the error $|dHA\%|$ on the measure 1 using the same display technique as Figure A.5. This diagram contains the same information as Figure A.4, but in another shape and extended to more x values.

Finally, Figure A.7 describes the effect of the cumulant ratio convergence depicted by the same technique as applied in Figures A.5 and A.6 and by using the ratio ϑ (cf. (5.2)) as a measure candidate.

6.9. The moment problem that was discussed in Section 2.5 was explored by varying the claim size distribution subject to the conditions that its mean, standard deviation and skewness:

(6.2) (1) m_Z, σ_Z, and γ_Z are fixed,
(2) the claim sizes are limited to integer values $1, \ldots, Z_{\max}$ and
(3) the claim number distribution $p_k(n)$ is fixed.

These conditions determine a family of the mixed compound distribution functions that all are approximated by one and the same NP, WH or HA. We illustrated the problem in Section 2.5 by saying that the functions to be approximated and fulfilling the conditions (6.2) are spread in a more or less wide "funnel of doubt" confined by the upper and lower envelope curve. If the funnel is broad for the relevant argument values x, this implies that there is no single curve which could approximate well all of the original curves, i.e. the approximation problem based on the characteristics of (6.2) has no satisfactory solution despite the method used. Unfortunately, evaluation of the envelope curves proved intractable. However, in order to get a grasp of the magnitude of the funnel at the tails of the distribution a lower limit was experimented with as follows.

Some of the test cases presented in Appendix 1 were chosen as examples. The values of the mixed compound Poisson function F were then calculated for those two distributions that fulfil the conditions (6.2) and have maximal and minimal kurtosis respectively, or what is an equivalent provision, maximal or minimal fourth moment of the claim size distribution. Then also the kurtosis of the aggregate claim distribution is maximized and minimized respectively. It can be reasonably expected that these are the extreme distributions at the tails among all those permitted by the conditions.

It proved that, in the exemplified cases, these distributions were the most dangerous and least dangerous respectively in the meaning defined by GOOVAERTS *et al.* (1984), Section 4.4 (suitably choosing their limit constant β).

Table 1 exhibits two examples, one connected to case 3 of Table A.1 and another connected to case 7. Among all of the distributions having the same characteristics as the selected case those two that have the minimal and maximal kurtosis respectively were sought, and then the exact F was calculated for them also.

It proved that the funnel of doubt is very narrow as long as the skewness is moderate and the individual claim sizes have a reasonably low upper limit. This confirms the earlier experience that the mixed compound Poisson distribution is robust under these provisos. On the other hand the funnel is rather large for large skewness values. This confirms the fact that there cannot be any approximation

TABLE 1

$F(x)$ resp. $1 - F(x)$

x	-2	2	3	4	
Min	0.01612	0.02881	0.00309	0.00021	
Case 3	0.01620	0.02880	0.00312	0.00022	$\gamma_X = 0.24$
Max	0.01632	0.02879	0.00318	0.00023	
Min	0.00105	0.04327	0.01142	0.00243	
Case 7	0.00330	0.03758	0.01253	0.00417	$\gamma_X = 1.08$
Max	0.00507	0.03089	0.01236	0.00571	

based on the three lowest characteristics (6.2) which would fit in all cases and for the whole relevant range of the variable x.

Note that the three curves representing the parallel distributions, as given in the table, intersect each other. That means, for example, that the most dangerous curve is most dangerous only for rather high values of x. The requested funnel of doubt is obviously constituted as an area between the envelope curves in a rather complicated way. Further study of this interesting problem was deferred to a later date.

The smoothing mentioned above in 6.4 has also some effect on the breadth of the funnel, although not an essential one. If for example the height of the step is regarded in the numbers of Table 1 for the case 3 and $x = 4$, the minimum and the maximum should be replaced by 0.00020 and 0.00024 respectively. Similarly, the numbers corresponding to $x = 4$ of the latter example should be replaced by 0.00238 and 0.00580.

7. DISCUSSION

7.1. On accuracy. It would be, of course, highly desirable to find ways to determine rigorously the accuracy of the proposed approximations. Unfortunately, this has not been tractable as yet. Therefore, we have to collect experience by testing various distributions. If a method turns out to have consistently acceptable accuracy in numerous and relevant areas of application well covered by the tests, then the use of the method may be justified in practical calculations. The Figures A.4–A.7 are aimed to provide a survey in concentrated form of the expected accuracy. Three alternative indicators were introduced: the skewness, the Haldane's measure and the convergence of the cumulants. Obviously the skewness is most convenient, because it has to be calculated as one of the entries of the approximation calculations.

7.2. How accurate should the method be? In deeming the usefulness of the approximation one should also appreciate the fact that in many cases the basic data are highly uncertain. In particular this concerns the structure function H (1.3) and its parameters. These choices may have a great effect on the process to be evaluated. If the initial data are inaccurate, then it is meaningless to demand essentially greater accuracy from the calculation technique, at least if this can be done only at the cost of greatly complicating the calculations. On the other hand, the collectives concerned are often fairly large and the top risks are cut away by reinsurance. Then the skewness may seldom exceed 0.1 or 0.2 and the inaccuracy involved with the approximation formulas obviously scarcely spoils the outcomes.

The situation is different for problems where long chains of computations are needed, e.g. in the calculation of integrals having F in the integrant. Then one should beware of an accumulation of errors.

7.3. An appropriate tool for simulations. Before proceeding further with the discussion about approximation methods attention is called to an attractive feature of the formula of type (2.4). It can be of special benefit for simulations where random numbers are generated, which are distributed according to the mixed compound Poisson law. This is a problem frequently appearing in advanced model building. The approach is simply first to generate normally distributed random numbers r and then to transform them by the inverse of the symmetrizing function (2.3): $X = v^{-1}(r)$ (see BPP, Section 6.8).

The number of the necessary random numbers can be very great. Then it is important that the inverse transformation v^{-1} is convenient to program and is fast. We proposed the NP formula in BPP (Section 6.8.3). The present experience suggests either the WH or HA-A transformations. In particular the WH formula is very handy (which was already recognized in exercise 6.8.1 of BPP).

7.4. Observations. Appendix 1 and Figure A.3 are the most convenient for the evaluation and comparison of the four tested methods.

If the skewness is moderate, i.e. no more than 0.3, and if an inaccuracy of some ± 2 per cent is tolerable, then all four methods are acceptable. However, the Haldane-B showed, by far, the narrowest range of the relative error, the Haldane-A being obviously the next best.

The situation is greatly different for the skewness values 0.3–1. Then the Haldane-B fails for $x = 3$ (note the different scale for the different lines of Figure A.3!). By the way, a similar observation was also made when the long version of the NP formula was investigated (PENTIKAINEN, 1977). These approaches, which are based on four characteristics, kurtosis included, instead of three characteristics (mean, standard deviation and skewness), proved to have superior accuracy for slightly skewed distributions but do not tolerate markedly skewed cases.

When the skewness exceeds unity, then all of the methods already show great irregularities and soon turn out to be useless. The lower example of Table 1 (Section 6.9) suggests that no method that is based only on the three lowest characteristics can be good for all greatly skewed distributions.

A general observation is that the short tail ($x < 0$) shows considerably worse results than the long tail.

For the reasons referred to above it seems doubtful whether the Haldane variant B is useful, in particular regarding the fact that it is markedly more complicated than the A variant. Haldane himself also observed that it gives sometimes poorer results than the simpler A formula.

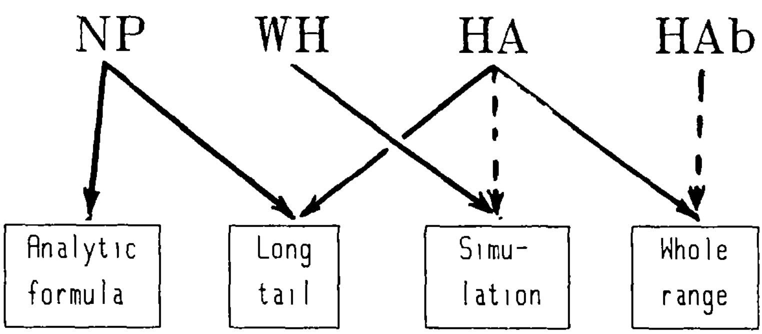
Some rules of thumb are sometimes proposed to guide the use of the approximations (e.g., BPP, 3.11e). Our latest studies do not suggest any such simple rules. Instead, it is much more effective to use Figure A.5 (or Figures A.6 or A.7) as a kind of "map" where the possible accuracy can be evaluated, and just in the environment of concern. If, for example, only positive x values are needed for some particular interval, then the area of applicability is wider than if negative x values are also needed.

Note that even though the general shapes of the error zones in Figures A.5, A.6 and A.7 are similar, it does not imply a full similarity in the test outcomes. For instance, the ϑ indicator would suggest a poor accuracy in case 4 of Appendix 1, but the skewness and Haldane measures still indicate acceptability as seen in the side column of the table.

7.5. The Wilson–Hilferty formula is clearly simpler and also somewhat faster than the Haldane-A. However, its tolerance for medium size and large skewnesses is poorer, as seen from Figure A.3. If the skewness is moderate, this formula may be appropriate at least in cases where very great speed is necessary.

7.6. The NP method has as its special merit the analytic form (2.5) for the long tail. It is of frequent use in many risk theory considerations (see e.g., BPP, Chapter 4). If only the long tail is of concern, then the NP method is the simplest and is also fairly competitive with the other methods concerning the accuracy, with the proviso that the distribution is not very skew.

7.7. In conclusion we summarize our present conception about the usefulness of the studies' approaches by means of a diagram as follows:



7.8. Finally let us note that the exact and approximate methods complement each other in a happy way. The exact methods (and possibly direct simulations, see BPP, p.239) are most appropriate for small collectives, and the approximate formulas for the large ones.

REFERENCES

BEARD, R E, PENTIKAINEN, T. and PESONEN, E (1984) *Risk Theory* Chapman and Hall, London (referred as BPP in the text)
BERTRAM, J. (1981) Numerische Berechnung von Gesamtschadenverteilung. *Blatter der Deutschen Gesellschaft fur Versicherungsmathematik*
BOHMAN, H and ESSCHER, F (1964) Studies in risk theory with numerical illustrations *Scandinavian Actuarial Journal*
BOX, G E B and COX, D. R (1964) An analysis of transformations *Journal of the Royal Statistical Society* **B26**, 211–252
ESSCHER, F (1932) On the probability function *Scandinavian Actuarial Journal.*
GOOVAERTS, M J, DE VYLDER, F and HAEZENDONCK, J (1984) *Insurance Premiums* North-Holland, Amsterdam
GOOVAERTS, M J, and KAAS, R (1986) Best bounds for positive distributions with fixed moments. *Insurance Mathematics & Economics* **5**, 1
HALDANE, J B S (1938) The approximate normalization of a class of frequency distributions *Biometrica* **29**, 392–404
KENDALL M and STUART, A (1979) *The Advanced Theory of Statistics* Vol I Charles Griffin, London
LAU, H-S (1984) An effective approach for estimating the aggregate loss *Journal of Risk and Insurance*, **LI.1**
OSCHWALD, M. (1984) Gamma Power-Entwicklung zur Berechnung der Verteilungsfunktionen des Gesamtschadens *Mitteilungen der Vereinigung Scweizerischer Versicherungsmathematiker*
PENTIKAINEN, T (1977) On the approximation of the total amount of claims *ASTIN Bulletin* **9**, 281–289.
PUSA, O (1985) Tests of the fit of the Esscher and NP formulas, Examination Paper, as yet available in Finnish only
WILSON, E B and HILFERTY, MARGARET (1931) The distribution of chi-square *Proceedings of National Academy of Science, USA* **17**, 684–688

APPENDIX 1

EXAMPLES

n = Poisson parameter.
h = Polya parameter.
ha = Haldane parameter (3.8).

$m, \sigma, \gamma, \gamma_2$ are the mean, standard deviation, skewness and kurtosis of the aggregate claim X.

r_i ($i = 1, 2$) is the risk indexes a_i / a_1^i indicating the heterogeneity of the claim size distribution (see BPP, p.54).
F = d.f. of X for $x < 0$ and $1 - F$ for $x > 0$ (1.1).
x = standardized aggregate claim size (2.2).
dF% = half of the step of F in per cent at the points where the discretized probability mass is concentrated ($= 50\star[F(x+) - F(x-)]$ divided by $F(x+)$ or $1 - F(x-)$).
NP = F approximated by the NP formula, dNP% its deviation from F in per cent. WH, HA and HAb are the corresponding outcomes for the Wilson–Hilferty, Haldane A and Haldane B formulas
Ln(.,.,.), the log normal claim size distribution having the mean, standard deviation and the skewness given in parentheses.

Pareto(.) Pareto claim size distribution with index given in parentheses (see BPP, p.74).
M is the greatest value of the discretized claim size and d is the interval between the consecutive non-zero points (see Section 6.1).
1A and 2A are the measures of deviation defined by (5.1) for the Haldane A and 1B and 2B for the Haldane B.
$\upsilon = \rho_4/\rho_3$ = an indicator for the speed of convergence of the cumulant.

Case 1	n	h	m	σ	γ	γ_2	ha	r_2	r_3
	200 0	0 0	1998 4	145 9	0 080	0 007	0 633	1 07	1 25

Ln(10, 3, 5)											
$M = 31, d = 1$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = −0 0001	−2 0	0.0205	0 0205	0 0205	0 0205	0 0205	−0 1	0 1	−0 0	−0 0	0 9
2A = +0 0009	−1 5	0 0646	0 0646	0 0645	0 0646	0 0645	0 1	−0 0	−0 0	−0 0	0 7
1B = −0 0001	−1 0	0 1586	0 1588	0 1586	0 1587	0 1586	0 1	−0 0	0 0	−0 0	0 5
2B = −0 0000	1.0	0 1586	0 1587	0 1586	0 1587	0 1586	0 0	−0 0	0 0	−0 0	0 5
υ = 0 006	2 0	0 0249	0 0249	0 0249	0 0249	0 0249	0 1	0 0	−0 0	0 0	0 8
	3 0	0 0019	0 0019	0 0019	0 0019	0 0019	0 4	0 5	−0 1	0 0	1 0
	4 0	0 0001	0 0001	0 0001	0 0001	0 0001	1 0	1 7	−0 5	0 0	1 2

Case 2	n	h	m	σ	γ	γ_2	ha	r_2	r_3
	100 0	100 0	787 1	116 9	0 224	0 071	0 497	1 21	1 72

Mixture											
$M = 31, d = 1$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = −0 0008	−2 0	0 0164	0 0163	0 0164	0 0163	0 0164	−0 5	0 3	−0 3	−0 1	1 3
2A = +0 0039	−1 5	0 0599	0 0605	0 0599	0 0599	0 0599	0 9	−0 0	−0 0	−0 0	1 0
1B = −0 0002	−1 0	0 1582	0 1597	0 1580	0 1583	0 1582	0 9	−0 1	0 0	−0 0	0 7
2B = −0 0000	1 0	0 1583	0 1587	0 1581	0 1583	0 1583	0 2	−0 1	0 0	−0 0	0 6
υ = 0 048	2 0	0 0284	0 0286	0 0284	0 0284	0 0284	0 5	0 0	0 0	0 0	0 9
	3 0	0 0029	0 0029	0 0030	0 0029	0 0029	0 4	0 8	−0 4	0 1	1 1
	4 0	0 0002	0 0002	0 0002	0 0002	0 0002	−0 3	2 9	−1 3	0 1	1 3

Case 3	n	h	m	σ	γ	γ_2	ha	r_2	r_3
	100 0	200 0	141 3	19 4	0 238	0 100	0 424	1 39	3 59

Pareto(3)											
$M = 21, d = 1$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = −0 0023	−2 0	0 0162	0 0159	0 0160	0 0160	0 0161	−1 9	−1 1	−1 4	−0 5	8 1
2A = +0 0175	−1 5	0 0595	0 0601	0 0594	0 0594	0 0594	0 9	−0 2	−0 2	−0 2	6 2
1B = −0 0004	−1 0	0 1578	0 1598	0 1580	0 1581	0 1577	1 2	0 1	0 2	−0 1	4 3
2B = −0 0000	1 0	0 1579	0 1587	0 1581	0 1582	0 1578	0 4	0 1	0 1	−0 1	3 6
υ = 0 058	2 0	0 0288	0 0289	0 0288	0 0288	0 0288	0 4	−0 1	−0 1	−0 1	4 7
	3.0	0 0031	0 0031	0 0031	0 0030	0 0031	−2 1	−1 7	−2 3	−0 3	6 0
	4 0	0 0002	0 0002	0 0002	0 0002	0 0002	−8 2	−5 1	−7 2	−0 4	6 4

Case 4	n	h	m	σ	γ	γ_2	ha	r_2		r_3	
	10 0	150 0	50 0	17 9	0 463	0 279	0 569	1 21		1 87	
Ln(5, 3, 4)											
$M = 21, d = 1$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = −0 0133	−2 0	0 0088	0 0097	0 0094	0 0080	0 0085	10 3	7 4	−8 5	−2 5	14 2
2A = +0 0214	−1 5	0 0506	0 0532	0 0504	0 0499	0 0501	5 3	−0 2	−1 4	−0 9	8 5
1B = −0 0044	−1 0	0 1572	0 1630	0 1558	0 1575	0 1570	3 7	−0 9	0 2	−0 2	5 3
2B = −0 0005	1 0	0 1576	0 1587	0 1566	0 1578	0 1574	0 7	−0 6	0 1	−0 1	3 7
v = 0 215	2 0	0 0337	0 0343	0 0337	0 0338	0 0338	1 7	−0 2	0 2	0 1	4 8
	3 0	0 0050	0 0051	0 0051	0 0049	0 0050	3 1	3 2	−0 4	0 5	5 6
	4 0	0 0005	0 0006	0 0006	0 0005	0 0005	4 9	11 2	−2 4	1 0	6 2

Case 5	n	h	m	σ	γ	γ_2	ha	r_2		r_3	
	100 0	100 0	170 4	34 2	0 593	0 679	0 016	3 03		36 93	
Pareto(2)											
$M = 61, d = 3$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = −0 0250	−2 0	0 0078	0 0067	0 0059	0 0067	0 0071	−14 1	−25 1	−14 4	−9 1	6 2
2A = +0 0582	−1 5	0 0456	0 0492	0 0443	0 0447	0 0447	7 9	−2 8	−2 0	−1 1	4 5
1B = +0 0196	−1 0	0 1506	0 1658	0 1538	0 1522	0 1510	10 1	2 1	1 1	0 3	2 9
2B = −0 0035	1 0	0 1529	0 1587	0 1554	0 1541	0 1529	3 8	1 6	0 8	0 0	0 9
v = 0 006	2 0	0 0362	0 0372	0 0361	0 0359	0 0358	2 8	−0 1	−0 6	−0 9	2 3
	3 0	0 0068	0 0064	0 0064	0 0066	0 0069	−4 9	−5 4	−2 1	1 4	2 5
	4 0	0 0011	0 0009	0 0009	0 0010	0 0012	−17 4	−12 2	−0 8	11 9	2 7

Case 6	n	h	m	σ	γ	γ_2	ha	r_2		r_3	
	25 0	20 0	47 6	17 5	0 779	0 976	0 297	2 15		13 25	
Pareto(2)											
$M = 31, d = 1$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = −0 0092	−2 0	0 0028	0 0035	0 0018	0 0020	0 0022	28 7	−35 4	−26 5	−22 0	22 9
2A = +0 0115	−1 5	0 0354	0 0434	0 0341	0 0344	0 0346	22 6	−3 6	−2 7	−2 2	11 8
1B = +0 0052	−1 0	0 1478	0 1712	0 1495	0 1491	0 1489	15 9	1 2	0 9	0 8	6 5
2B = −0 0002	1 0	0 1526	0 1587	0 1533	0 1530	0 1528	3 9	0 4	0 2	0 1	3 3
v = 0 006	2 0	0 0394	0 0411	0 0392	0 0391	0 0391	4 3	−0 7	−0 9	−1 0	4 1
	3 0	0 0084	0 0084	0 0083	0 0083	0 0083	1 0	−1 2	−0 8	−0 6	4 3
	4 0	0 0015	0 0014	0 0015	0 0016	0 0016	−5 4	−0 0	1 8	2 9	4 9

Case 7	n	h	m	σ	γ	γ_2	ha	r_2		r_3	
	75 0	75 0	122 0	30 8	1 082	2 703	−0 429	3 77		84 37	
Pareto(2)											
$M = 121, d = 4$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = +0 4175	−2 0	0 0033	0 0009	0 0000	0 0006	0 0011	−73 1	−100 0	−83 1	−67 8	10 0
2A = +0 5406	−1 5	0 0322	0 0342	0 0145	0 0209	0 0204	6 0	−55 2	−35 1	−36 5	5 9
1B = +0 6058	−1 0	0 1357	0 1834	0 1384	0 1307	0 1207	35 2	2 0	−3 7	−11 1	3 8
2B = −0 2405	1 0	0 1376	0 1587	0 1491	0 1419	0 1285	15 3	8 4	3 1	−6 6	2 1
v = 0 630	2 0	0 0376	0 0470	0 0431	0 0407	0 0379	25 1	14 7	8 3	0 7	1 9
	3 0	0 0125	0 0119	0 0112	0 0116	0 0139	−4 9	−10 4	−7 3	11 3	1.6
	4 0	0 0042	0 0027	0 0028	0 0035	0 0068	−35 5	−34 0	−16 5	63 7	1 9

Case 8	n	h	m	σ	γ	γ_2	ha	r_2		r_3	
	25 0	50 0	42 7	16 6	1 628	5 801	−0 394	3 29		54 61	
Pareto(2)											
$M = 91, d = 3$	x	F	NP	WH	HA	HAb	dNP%	dWH%	dHA%	dHAb%	dF%
1A = +1 4047	−2 0	0 0004	0 0000	0 0000	0 0000	0 0000	−95 8	−100 0	−100 0	−99 4	41 8
2A = +1 0048	−1 5	0 0178	0 0197	0 0000	0 0024	0 0026	10 7	−100 0	−86 5	−85 5	16 2
1B = +2 2843	−1 0	0 1190	0 2180	0 0950	0 0892	0 0702	83 2	−20 1	−25 0	−41 0	8 9
2B = −1 5625	1 0	0 1280	0 1587	0 1399	0 1248	0 0983	24 0	9 3	−2 5	−23 2	3 7
$v = 1$ 387	2 0	0 0370	0 0562	0 0473	0 0395	0 0291	51 8	27 7	6 7	−21 4	3 4
	3 0	0 0140	0 0184	0 0158	0 0139	0 0119	31 4	12 6	−0 9	−14 7	2 4
	4 0	0 0069	0 0057	0 0053	0 0054	0 0065	−18 1	−24 1	−21 4	−6 4	1 8

APPENDIX 2

FIGURES

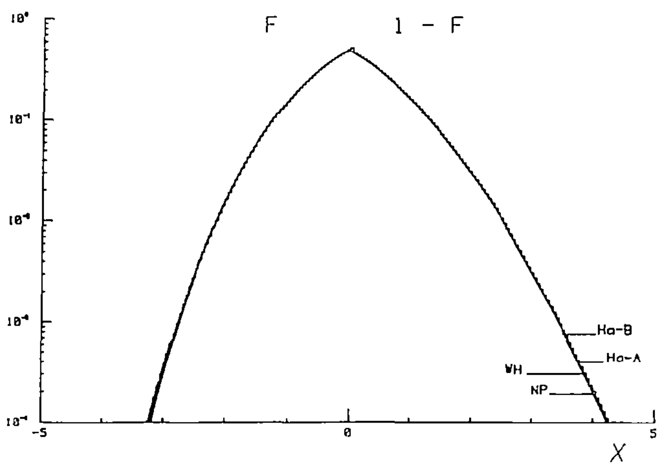


FIGURE A 1(a) The case 3 of Appendix 1 presented as a graph The step curve represents the exact F The key parameters are $n = 100$, $h = 200$, $\gamma = 0$ 24

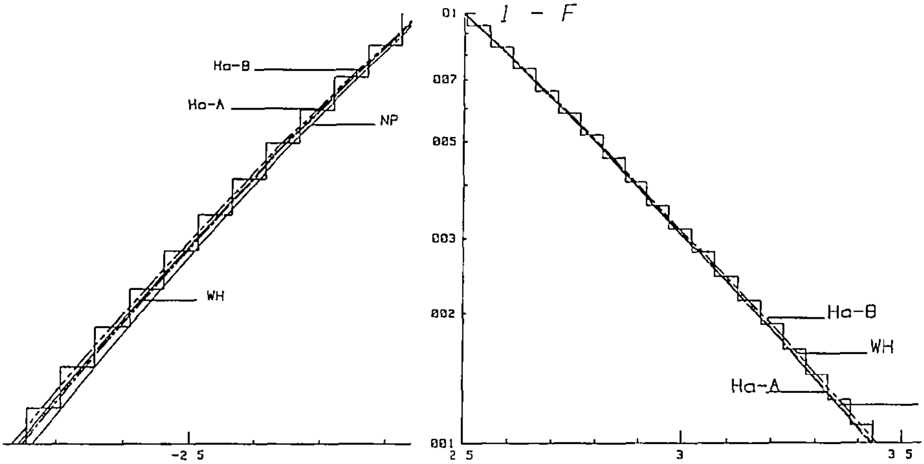


FIGURE A 1(*b*) The tails of the curves of Figure A 1(*a*) plotted in a magnified scale

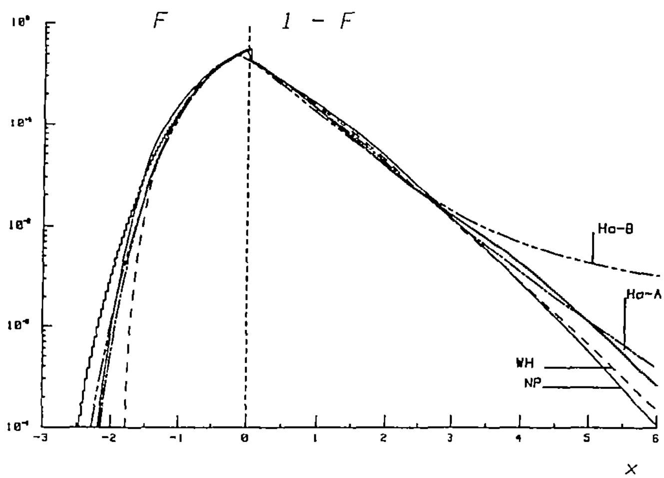


FIGURE A 2 The case 7 of Appendix 1 The key parameters are $n = 75$, $h = 75$, $\gamma = 1\ 08$

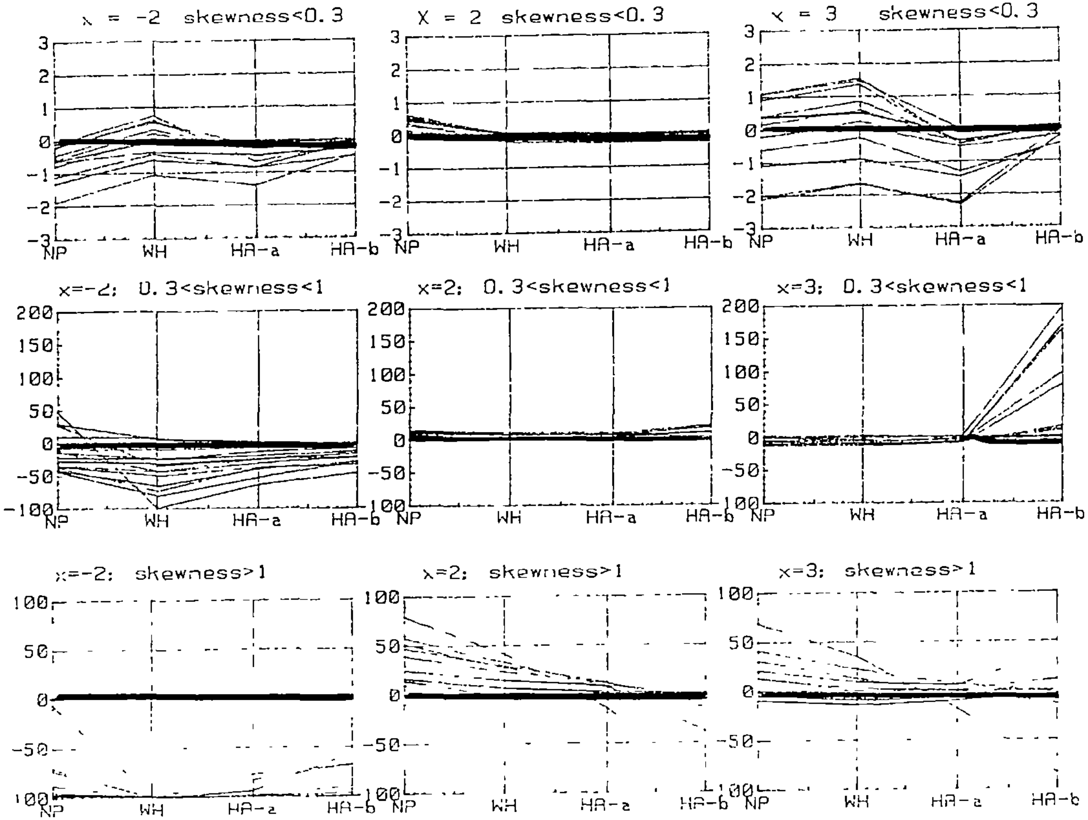
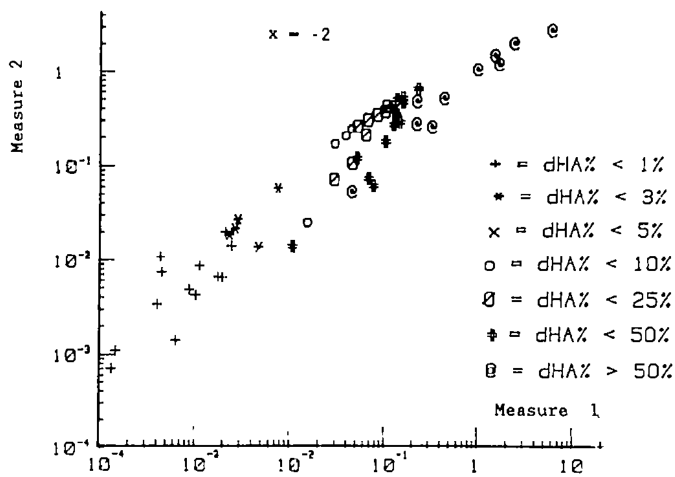


FIGURE A 3. The relative errors dNP%, dWH%, dHA% and dHAb% (see Appendix 1) grouped according to specified skewness and x values (see Section 6.5 for further explanations) calculated for a selected sample of distributions



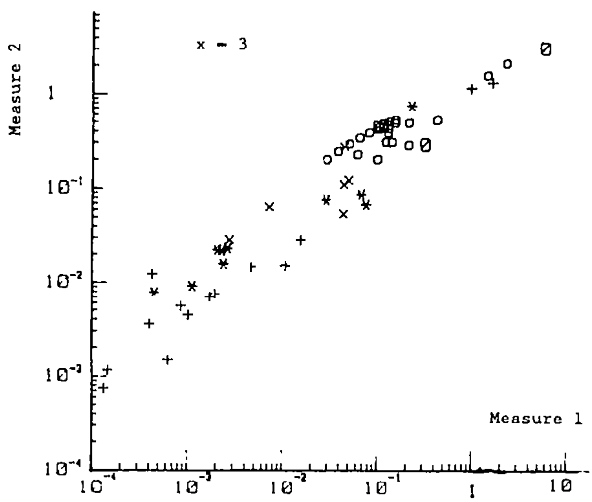


FIGURE A 4 The relative errors dHA% of the Haldane-A formula displayed according to absolute values of the measures 1 and 2 as defined by (5 1) The order of | dHA% | is indicated by the symbols given in the graph

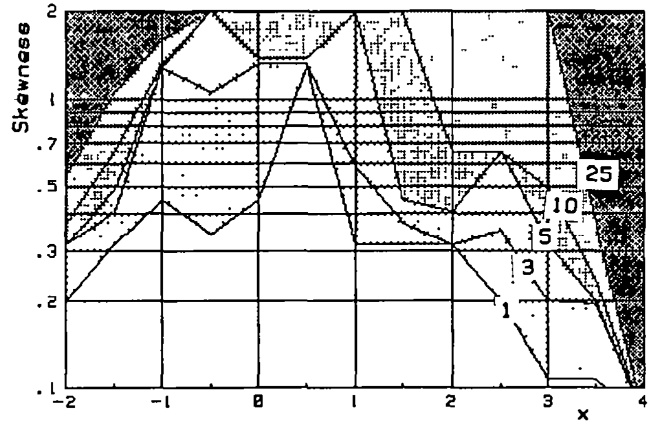


FIGURE A 5 The zones of the relative errors | dHA% | according to the argument x and the skewness For explanations see Section 6 8

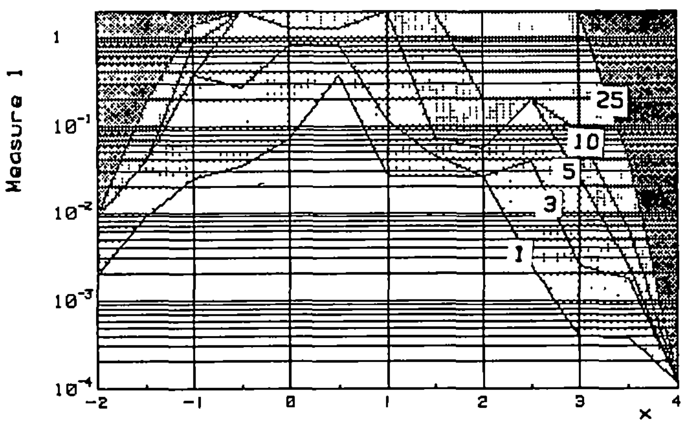


FIGURE A 6 The zones of the relative errors | dHA% | according to the argument x and the measure l

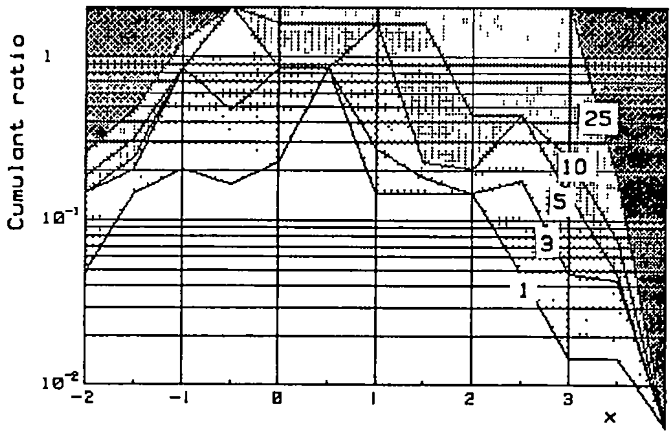


FIGURE A 7 The zones of the relative errors | dHA% | according to the argument x and the cumulant convergence indicator v (see the explanations of Appendix 1)

TEIVO PENTIKAINEN
Kasavuorentie 12 C 9, SF-02700 Kauniainen, Finland.

TWO CREDIBILITY REGRESSION APPROACHES FOR THE CLASSIFICATION OF PASSENGER CARS IN A MULTIPLICATIVE TARIFF[1]

By Bjørn Sundt

Storebrand Insurance Company Ltd, Oslo

Abstract

In the present paper we present two credibility regression models for the classification of passenger cars. As regressors we use technical variables like price, weight, etc. In both models we derive credibility estimators and find expressions for their estimation errors. Estimators for structure parameters are proposed. A numerical example based on real data is given. The second model is hierarchical with a level for make of car.

1. Background

In Norway there is no common passenger car tariff for all insurance companies, and thus there are several different tarriffs in the market. However, most of them seem to have about the same structure as the one used by Storebrand to be described below, but with different parameter values.

In this paper we are going to discuss vehicle damage insurance for passenger cars. The tariff structure is multiplicative with factors for bonus-malus, mileage/district, deductibles, age of car, and car model. We shall concentrate on the factor for car model. There are 65 classes numbered from 30 to 94, and the factor for class c is equal to 1.04^{c-30}.

Until the present research was started, the classification of individual car models was performed rather subjectively. There was one person classifying new car models. When a new car model appeared on the market, he looked at its specifications and tried to find out to which cars it was comparable. Then he looked at the factors for these cars, both by Storebrand and by the competing companies. When the car had been in the market for some time and claims statistics had become available, the rating factor was reassessed, taking into account the observed claims ratio, the observed volume of exposure (premium), the old factor, and the premiums of the competing companies. This reassessment was also performed in a rather subjective way, but not by the same person who had made the initial classification of the car.

The procedure described in the previous paragraph has obvious advantages compared to an objectively based statistical procedure. It would be impossible to build into a mathematical model all the experience, knowledge, and intuition of a skilled person. How could the model incorporate, say, the person's opinion of

[1] Paper presented at the XIXth ASTIN Colloquium, Tel Aviv, 20–24 September 1986

the importance of the shape of the car (a limousine and a coupé are bought by quite different sorts of people)? And even if one should succeed in creating a model which to a great extent incorporated the knowledge of the skilled person, this model would probably be too complicated for practical use. However, the advantage of the subjective procedure is also a disadvantage. The procedure is too dependent on the person performing it. As it is impossible to build the knowledge of the skilled person into a statistical model, it is also impossible to give an adequate documentation of the procedure. And what then if the person leaves the company?

This was the background that motivated the present research. One wanted an objective method for classification of cars, and in this paper we are going to describe the models and methods that were considered. We are also going to comment upon the difficulties that occurred during the work. As should be well known to everyone who has worked on modelling insurance data, these data are very seldom what you want them to be.

When the project was started, it was decided that this time we should concentrate only on the determination of the factor for car model. Ideally, one should of course have developed models and methods for simultaneous determination of all the factors in the multiplicative model, but that would have been a much more ambitious and time-consuming project. It was discussed whether one should concentrate only on the classification of new car models, for which we have no claims data, but in the present author's opinion, classification of new models is only a special case of reclassification (i.e. the case with exposure volume equal to zero). It would therefore be unnatural not to treat these two situations together, and it was decided to follow this line.

As was argued above, the subjective approach has great advantages compared to a statistically based procedure, and it would be wrong to throw this system away completely. It is the author's intention that the methods presented in this paper should not replace the skilled person, but rather be an aid to him. The system proposes a class to the person, but he should himself decide whether to follow this proposal or not. In particular, this is important for reclassification of cars that have already been in the market for some time, and for which we know the rating of the competitors. It would be too ambitious to build a model that also incorporates the premiums of competing companies. For marketing reasons, it could also be desirable to make smaller changes by the reclassification than those proposed by the statistical procedure.

Furthermore, in the statistical investigations it became clear that some car models behaved so strangely, relative to the model studied, that they ought to be considered as outliers in the present context. For such cars one should not use the factor suggested by the system, and perhaps even more important, these cars should be left out when estimating the model parameters. The most striking example in our investigations was Volkswagen Golf GTI, and the parameter estimates changed considerably when this car was taken out of the estimation procedures. It is important that the person doing the classification identifies such cars and sees to it that they are left out of the statistical analysis. One could of course argue that the model assumptions should also embrace such cars, but it is the opinion of the present author that it is preferable to have a relatively simple model giving satisfactory results for "normal" cars, than a complicated model that could be used for all cars. In particular, as he believes that in practice the "outliers" would usually be easy to identify.

For the numerical computations we used the program package SAS, which in particular was very convenient for the matrix calculus.

2. PRELIMINARIES

2.1. *Optimality Criterion for Estimators*

Let m be a random variable. We shall say that an estimator $m^{(1)}$ of m is better than another estimator $m^{(2)}$ if

$$E(m^{(1)}-m)^2 < E(m^{(2)}-m)^2,$$

that is, we use the quadratic loss function.

Let $\boldsymbol{m} = (m_1, \ldots, m_s)'$ be an unknown random vector and $\boldsymbol{m}^{(1)} = (m_1^{(1)}, \ldots, m_s^{(1)})'$ and $\boldsymbol{m}^{(2)} = (m_1^{(2)}, \ldots, m_s^{(2)})'$ two estimators of $\boldsymbol{m}$. Then we shall say that $\boldsymbol{m}^{(1)}$ is a better estimator of $\boldsymbol{m}$ than $\boldsymbol{m}^{(2)}$ if

$$E(m_i^{(1)} - m_i)^2 \leqslant E(m_i^{(2)} - m_i)^2 \quad i = 1, \ldots, s$$

with strict inequality for at least one i.

We implicitly assume that second-order moments exist for all random variables to be considered.

2.2. *Credibility Estimators*

Let $\boldsymbol{x}$ and $\boldsymbol{m}$ be random vectors, $\boldsymbol{x}$ observable and $\boldsymbol{m}$ unknown. We shall call $\dot{\boldsymbol{m}}$ a linear estimator of $\boldsymbol{m}$ (based on $\boldsymbol{x}$) if $\dot{\boldsymbol{m}}$ may be written in the form $\dot{\boldsymbol{m}} = \boldsymbol{a} + A\boldsymbol{x}$, where $\boldsymbol{a}$ is a non-random vector and A a non-random matrix. The credibility estimator of $\boldsymbol{m}$ (based on $\boldsymbol{x}$) is defined as the best linear estimator of $\boldsymbol{m}$. We summarize some results about credibility estimators in the following theorem.

THEOREM 2.1. (i) *There always exists a unique credibility estimator of* $\boldsymbol{m}$.

(ii) *Let* $\dot{\boldsymbol{m}}$ *be a linear estimator of* $\boldsymbol{m}$. *Then* $\dot{\boldsymbol{m}}$ *is a credibility estimator of* $\boldsymbol{m}$ *if and only if* $\dot{\boldsymbol{m}}$ *satisfies the two conditions*

$$E\dot{\boldsymbol{m}} = E\boldsymbol{m} \tag{2.1}$$

$$\operatorname{Cov}(\dot{\boldsymbol{m}}, \boldsymbol{x}') = \operatorname{Cov}(\boldsymbol{m}, \boldsymbol{x}'). \tag{2.2}$$

(iii) *Let* $\tilde{\boldsymbol{m}}$ *be the credibility estimator of* $\boldsymbol{m}$. *Then we have*

$$\operatorname{Cov}(\boldsymbol{m}, \tilde{\boldsymbol{m}}') = \operatorname{Cov}(\tilde{\boldsymbol{m}}, \boldsymbol{m}') = \operatorname{Cov}(\tilde{\boldsymbol{m}}) = \operatorname{Cov}(\boldsymbol{m}) - \operatorname{Cov}(\boldsymbol{m} - \tilde{\boldsymbol{m}}). \tag{2.3}$$

For proof of (i) we refer to DE VYLDER (1976), for proof of (ii) to SUNDT (1980), and for proof of (iii) to SUNDT (1981).

3. A NON-HIERARCHICAL APPROACH

3.1. *Model*

Consider a group of K different car models. These could be all passenger cars (station wagons included) that are rated in Storebrand, or a well-defined subgroup of these (e.g. diesel cars, cars with four-wheel drive, all Volkswagen models, all cars produced after 1982, etc.). For the parameter estimation described in Subsection 3.3 it could be reasonable to take a representative sample from the group considered.

For car model k we have observed I_k risk units (policies). Let X_{ki} be the total claim amount in the exposure time for unit i of model k, and let p_{ki} be the earned premium. We want to use earned premium as a measure of risk volume, but this premium also contains the car model factor which we are going to reassess, and this old value should not be included in the risk measure. Hence, let

$$w_{ki} = p_{ki}/f_k, \tag{3.1}$$

where f_k is the old factor, be our measure of risk volume. We assume that for fixed k, the X_{ki}'s are independent of the corresponding data from other car models, and that $X_{k1}, \ldots, X_{kI_k}$ are conditionally independent given Θ_k, a random parameter characterizing car model k. It is assumed that $\Theta_1, \ldots, \Theta_K$ are independent and identically distributed.

Let

$$Y_{ki} = X_{ki}/w_{ki}.$$

It is assumed that

$$E[Y_{ki} \mid \Theta_k] = m_k(\Theta_k)$$

$$\operatorname{Var} m_k(\Theta_k) = \lambda \tag{3.2}$$

$$\operatorname{Var}[Y_{ki} \mid \Theta_k] = s^2(\Theta_{ki})/v_{ki} \tag{3.3}$$

with $v_{ki} = w_{ki}$ (the reason for introducing v_{ki} will become clear in subsection 3.6, where we modify the present assumptions), and

$$\mu_k = Em_k(\Theta_k) = x_k'\beta,$$

where x_k is a known $q \times 1$ design vector based on the technical data of the car and β is an unknown $q \times 1$ regression vector. We further introduce

$$\phi = Es^2(\Theta_k) \qquad \varkappa = \phi/\lambda$$

$$X_k = \sum_{i=1}^{I_k} X_{ki} \qquad v_k = \sum_{i=1}^{I_k} v_{ki} \qquad w_k = \sum_{i=1}^{I_k} w_{ki}$$

$$Y_k = X_k/w_k. \tag{3.4}$$

We note that in the special case when

$$Em_k(\Theta_k) = \mu,$$

independent of k, the conditions of the Buhlmann–Straub model (BUHLMANN and STRAUB (1970)) are satisfied.

It is also interesting to relate our present model to HACHEMEISTER'S (1975) regression model. In that model one assumes that

$$E[Y_{kt} \mid \Theta_k] = x'_{kt} b(\Theta_k),$$

where x_{kt} is a known $q \times 1$ design vector and b is a $q \times 1$ vector function. To correspond to our present set-up we assume that $x_{kt} = x_k$ independent of t. We introduce

$$\Lambda = \text{Cov } b(\Theta_k) \qquad \beta = Eb(\Theta_k)$$

and get

$$\text{Var } E[Y_{kt} \mid \Theta_k] = x'_k \Lambda x_k.$$

Thus this variance would typically vary between car models whereas in (3.2) we have assumed it to be constant. Let us now assume that the first element of x_k is equal to one, which will usually be the case. Then we obtain our present model by assuming that only the first element of $b(\Theta_k)$ is random, which makes all elements of Λ except the (1,1) element equal to zero. We note that this Λ is not positive definite.

3.2. *Credibility Estimation of* $m_s(\Theta_s)$

Let $\tilde{m}_s$ be the credibility estimator of $m_s(\Theta_s)$ based on the observed Y_{kt}'s. We also introduce the estimation error

$$\psi_s = E(m_s(\Theta_s) - \tilde{m}_s)^2$$

of $\tilde{m}_s$ From Theorem 2.1 we get

$$\tilde{m}_s = \zeta_s Y_s + (1 - \zeta_s)\mu_s \tag{3.5}$$

$$\psi_s = \phi / (v_s + \kappa) = \lambda(1 - \zeta_s)$$

with

$$\zeta_s = v_s / (v_s + \kappa).$$

3.3. *Parameter Estimation*

The structure parameters ϕ, λ, and β will in practical applications be unknown and have to be estimated.

We have that

$$\phi_k^* = (I_k - 1)^{-1} \sum_{t=1}^{I_k} v_{kt}(Y_{kt} - Y_k)^2 \tag{3.6}$$

satisfies $E[\phi_k^* \mid \Theta_k] = s^2(\Theta_k)$, and thus

$$\phi^* = \sum_{k=1}^{K} u_k \phi_k^*$$

is an unbiased estimator of ϕ for all weights u_k ($\sum_{k=1}^{K} u_k = 1$). In an earlier version of the paper we suggested that one should simply apply $u_k = K^{-1}$. This choice has been criticized by Ragnar Norberg, who suggests that one should apply

$$u_k = (I_k - 1) \Big/ \left(\sum_{r=1}^{K} I_r - K \right).$$

An optimal choice of weights is difficult, involving fourth-order moments (cf. e.g. NORBERG (1982)), and it was not within the scope of the present research to perform a profound analysis of this problem. Both our choice and Norberg's choice can be criticized; our choice because it gives too much weight to cars with low exposure; Norberg's choice because if I_s is much greater than the other I_k's for some s, then the value of Θ_s will have a too dominant influence on the estimate of ϕ. The present discussion also applies to the analogously weighted estimators in subsections 3.6 and 4.3. We note that in the special case $I_1 = I_2 = \ldots = I_K$ Norberg's choice and our choice are equal, and in this case ϕ^* is equal to the estimator proposed by BUHLMANN and STRAUB (1970) for the Buhlmann–Straub model.

We introduce

$$Y = (Y_1, \ldots, Y_K)' \qquad X = (x_1, \ldots, x_K)'$$

$$v = \sum_{k=1}^{K} v_k \qquad D = \operatorname{diag}(v_1/v, \ldots, v_K/v)$$

and get

$$EY = X\beta$$

$$\operatorname{Cov} Y = (\phi/v)D^{-1} + \lambda I_K \tag{3.7}$$

with I_K denoting the $K \times K$ identity matrix. It is assumed that X has rank q.

We trivially have that

$$\hat{\beta} = (X'DX)^{-1}X'DY$$

is an unbiased estimator of β. It seems reasonable to base an estimator of λ on the statistic

$$Q = (Y - X\hat{\beta})'D(Y - X\hat{\beta}), \tag{3.8}$$

and we therefore want to find the expectation of Q. In the deduction we use that for an $r \times s$ matrix A and an $s \times r$ matrix B we have

$$\operatorname{tr}(AB) = \operatorname{tr}(BA), \tag{3.9}$$

where "tr" denotes the trace of a quadratic matrix (i.e. the sum of its diagonal elements); this result is easily proved. We have

$$EQ = E(Y - X\hat{\beta})'D(Y - X\hat{\beta}) = \operatorname{tr}\{D\, E(Y - X\hat{\beta})(Y - X\hat{\beta})'\}$$
$$= \operatorname{tr}\{D \operatorname{Cov}(Y - X\hat{\beta})\}$$

as

$$E(Y - X\hat{\beta}) = 0.$$

We further get

$$EQ = \operatorname{tr}(D \operatorname{Cov}[\{I_K - X(X'DX)^{-1}X'D\}Y]) = \\ \operatorname{tr}(D\{I_K - X(X'DX)^{-1}X'D\}(\operatorname{Cov} Y)\{I_K - DX(X'DX)^{-1}X\}),$$

and insertion of (3.7) gives

$$EQ = \lambda\tau_1 + (\phi/v)\tau_2 \tag{3.10}$$

with

$$\tau_1 = \operatorname{tr}(D\{I_K - X(X'DX)^{-1}X'D\}\{I_K - DX(X'DX)^{-1}X'\}) \tag{3.11}$$

and

$$\tau_2 = \operatorname{tr}(D\{I_K - X(X'DX)^{-1}X'D\}D^{-1}\{I_K - DX(X'DX)^{-1}X'\}). \tag{3.12}$$

From (3.11) we get

$$\tau_1 = \operatorname{tr} D - \operatorname{tr}\{DX(X'DX)^{-1}X'D\} - \operatorname{tr}\{D^2X(X'DX)^{-1}X'\} \\ + \operatorname{tr}\{DX(X'DX)^{-1}X'D^2X(X'DX)^{-1}X'\}. \tag{3.13}$$

By repeated use of (3.9) we see that the three last terms in (3.13) are all equal to $\operatorname{tr}\{(X'DX)^{-1}X'D^2X\}$, and as in addition $\operatorname{tr} D = 1$, we get

$$\tau_1 = 1 - \operatorname{tr}\{(X'DX)^{-1}X'D^2X\}. \tag{3.14}$$

From (3.12) we obtain

$$\tau_2 = \operatorname{tr}(\{I_K - DX(X'DX)^{-1}X'\}\{I_K - DX(X'DX)^{-1}X'\}) \\ = \operatorname{tr} I_K - \operatorname{tr}\{DX(X'DX)^{-1}X'\} = \operatorname{tr} I_K - \operatorname{tr}\{(X'DX)^{-1}X'DX\} = \operatorname{tr} I_K - \operatorname{tr} I_q,$$

and as the trace of an identity matrix is equal to its dimension, we get

$$\tau_2 = K - q. \tag{3.15}$$

From (3.8), (3.10), (3.14), and (3.15) we get that

$$\hat{\lambda} = \{(Y - X\hat{\beta})'D(Y - X\hat{\beta}) - (K - q)\phi^*/v\}/[1 - \operatorname{tr}\{(X'DX)^{-1}X'D^2X\}]$$

is an unbiased estimator of λ. It has, however, the disadvantage that it can take negative values whereas λ is always non-negative. Therefore we replace it by

$$\lambda^* = \max(0, \hat{\lambda}).$$

However, by this adjustment we lose the unbiasedness. For simplicity, in the following we proceed as if $\lambda^* > 0$; the adaption to the case $\lambda^* = 0$ is trivial. To avoid having to take special care of the case $\lambda^* = 0$, one could instead of putting λ^* equal to zero when $\hat{\lambda} \leqslant 0$, put λ^* equal to some small positive number; one possible choice would be ε/K for some small ε, as we would then have asymptotic unbiasedness when K goes to infinity.

If Cov Y were known, the best linear unbiased estimator of β would be

$$\ddot{\beta} = \{X'(\mathrm{Cov}\ Y)^{-1}X\}^{-1}X'(\mathrm{cov}\ Y)^{-1}Y,$$

and as $Y_1, \ldots, Y_K$ are independent,

$$\begin{aligned}\beta &= \left(\sum_{k=1}^{K} x_k(\mathrm{Var}\ Y_k)^{-1}x_k'\right)^{-1} \sum_{k=1}^{K} x_k(\mathrm{Var}\ Y_k)^{-1}Y_k \\ &= \left(\sum_{k=1}^{K} x_k v_k(\phi + \lambda v_k)^{-1}x_k'\right)^{-1} \sum_{k=1}^{K} x_k v_k(\phi + \lambda v_k)^{-1}Y_k \\ &= \left(\sum_{k=1}^{K} \zeta_k x_k x_k'\right)^{-1} \sum_{k=1}^{K} \zeta_k x_k Y_k,\end{aligned}$$

and we propose to estimate β by

$$\beta^* = \left(\sum_{k=1}^{K} \zeta_k^* x_k x_k'\right)^{-1} \sum_{k=1}^{K} \zeta_k^* x_k Y_k$$

with

$$\zeta_k^* = v_k/(v_k + \varkappa^*) \qquad \varkappa^* = \phi^*/\lambda^*.$$

It should be noted that in the Buhlmann–Straub model the estimators λ^* and β^* reduce to the estimators studied by BUHLMANN and STRAUB (1970).

3.4. *Determination of the Tariff Class*

By inserting the estimators ζ_k^* and β^* in (3.5) we get the empirical credibility estimator

$$\tilde{m}_s^* = \zeta_s^* Y_s + (1 - \zeta_s^*)\mu_s^*$$

with

$$\mu_s^* = x_s'\beta^*.$$

The estimation error ψ_s is estimated by

$$\psi_s^* = \lambda^*(1 - \zeta_s^*).$$

The estimator $\tilde{m}_s^*$ cannot yet be used as the proposed rating factor for car model s; it still needs to be adjusted by some scaling factor. The approach used in our numerical investigations was to determine the scaling factor such that the total premium for the portfolio used for the estimations would be the same with the new values of the model factor as with the old ones.

Let γ^* be our scaling factor. Then the new model factor will be

$$\rho_s^* = \gamma^*\tilde{m}_s^*,$$

and thus the total "new" premium will be $\gamma^*\sum_{k=1}^{K} w_k\tilde{m}_k^*$ whereas the "old" premium is $\sum_{k=1}^{K} p_k$ with

$$p_k = \sum_{i=1}^{I_k} p_{ki}.$$

As these two premiums should be equal, we get

$$\gamma^* = \left(\sum_{k=1}^{K} p_k\right) \Big/ \sum_{k=1}^{K} w_k \tilde{m}_k^*.$$

In addition to our estimator for the class, we also want a confidence interval for the "correct" factor, by which we mean $\gamma m_s(\Theta_s)$, where γ denotes the mean of γ^*. To get such a confidence interval we need some additional assumption, and for simplicity we assume that the conditional distribution of $m_s(\Theta_s)$ given the observations is normal with mean $\tilde{m}_s$ and variance ψ_s. This assumption seems highly unrealistic, in particular for cars with low exposure, but we really did not need an exact confidence interval, only some measure of the uncertainty of the estimator, and for this purpose the assumption seems adequate. As a 1-ε confidence interval (in the Bayesian sense, cf. e.g. DEGROOT (1970, subsections 11.5–6)) for the factor we obtain $\tilde{m}_s \pm g_{1-\varepsilon/2}\sqrt{\psi_s}$, where $g_{1-\varepsilon/2}$ denotes the $1-\varepsilon/2$ fractile in the standard normal distribution $N(0, 1)$ and by insertion of estimators for unknown parameters we finally get the estimated confidence interval $\rho_s^* \pm \gamma^* g_{1-\varepsilon/2}\sqrt{\psi_s^*}$.

From the estimator and the confidence interval of the model factor, we can trivially derive an estimator and a confidence interval for the model class (cf. Section 1).

When a new car model t, for which we have no data, is to be classified, we have $v_t = \zeta_t^* = 0$, which gives

$$\rho_t^* = \gamma^* \mu_t^* = \gamma x_t' \beta^*,$$

that is,

$$\rho_t^* = x_t' \alpha^* = \sum_{j=1}^{q} \alpha_j^* x_{tj} \tag{3.16}$$

with

$$\alpha^* = (\alpha_1^*, \ldots, \alpha_q^*)' = \gamma^* \beta^*.$$

Thus, (3.16) is the formula to be used to find the model factor for car model t.

Let us for a moment call ρ_t^*, given by (3.16), $\rho_t^*(0)$ to stress that this is the factor estimate without exposure. When we get an observed exposure, we get the factor

$$\rho_t^*(1) = \zeta_t^* \gamma^* Y_t + (1 - \zeta_t^*)\rho_t^*(0),$$

that is, a weighted mean of the initial factor estimate and the empirical factor $\gamma^* Y_t$. We also note that with no exposure we have $\psi_t = \lambda$ and $\psi_t^* = \lambda^*$.

3.5. *Choice of Regressors*

In subsection 3.1 we said that x_k should be a design vector based on the technical data of car model k without giving any further indication of which regressors one should use. In our numerical investigations we registered for each car model in our test sample the four basic variables engine power, cylinder volume, weight, and price. Diesel cars and cars with four-wheel drive were not included in our sample; otherwise it would have been appropriate to include (0,1)-variables for these characteristics. As interesting regressors we concentrated on the four basic variables and ratios between them.

It should be noted that the estimator ϕ^* of ϕ does not depend on the chosen regressors. For λ^* and $\boldsymbol{\beta}^*$ we made several computations using different regressors. In each design we of course included a constant term, that is, the first element of $\boldsymbol{x}_k$ being equal to 1.

As

$$\lambda = E(m_k(\Theta_k) - \boldsymbol{x}_k'\boldsymbol{\beta})^2,$$

λ measures how close the prior mean is to $m_k(\Theta_k)$, and it was therefore felt that one should use a set of regressors making λ^* small. This is also consistent with our choice of the quadratic loss function; one could think of all the possible regressors being studied as included in a huge design, but that for most of them we estimate the corresponding element of $\boldsymbol{\beta}$ by zero.

An important point when choosing regressors is that we know something about monotonicity. To motivate this, let us look at an example. At an early stage of our research we wanted to classify some new car models for which the prices were still unknown. A design giving small λ^* under these circumstances was (1 power/weight weight/power)$'$. For two of the cars we got the following results:

Car	Weight	Power	Class
1	1200 kg	63 HP	59
2	1227 kg	86 HP	42

This seems of course very unreasonable. Car 2 has a slightly higher weight and a much higher power than Car 1, but should be rated lower!

In accordance with our opinion about monotonicity, several sets of regressors were rejected when looking at $\boldsymbol{\beta}^*$. It should be noted that the more regressors we include, the more difficult it would be to control that our opinion about monotonicity is satisfied as the different regressors could be strongly correlated; even if we mean that the factor should be increasing with cylinder volume, it need not be disturbing to get a negative coefficient for this regressor if engine power has a positive coefficient, as cylinder volume and engine power are strongly correlated. Under these considerations we conclude that q should not be too large, say, at most 4–5.

It should be noted that the monotonicity secured by the choice of regressors is not necessarily satisfied for the posterior estimates $\tilde{m}_s^*$ with positive exposures. This is reasonable as we then have more information apart from the technical data; the monotonicity is important only when we base the factor on only the technical data.

One should be aware that in one respect price is different from the other basic variables considered, as the price may change whereas the car model is still the same.

We conclude this subsection by briefly recapitulating the criteria that should be taken into consideration by the choice of regressors:

(i) small λ^*;
(ii) monotonicity;
(iii) small q.

3.6. *Some Practical Modifications*

In subsection 3.3 we described how we would have estimated ϕ if we had had the necessary data. Unfortunately, we did not have them. From (3.6) we see that for each policy we had to match the exposure with the total amount of the claims occurred during the exposure period. At present, the data of Storebrand are organized such that for each calendar year we have one claims file and one policy file. The claims file contains data for all claims *reported* during the year. As stated above, we really wanted the claims *occurred* during the year, but this does not seem to be a serious problem. The policy file contains data from the middle of the year. The registered premium is the premium at the latest renewal prior to the middle of the year, which means that these renewals range from the middle of the previous year until the middle of the present year. Thus a match between claims and policies would be awkward. We also have the problem that the total registered premium for a fixed car model is not really the premium we wanted it to be, but we decided to use it as an approximation. If the premium volume of the car model is relatively stable over time, this approximation should be acceptable. However, if the premium volume is growing, we would register too low a value for the exposure volume. This will in particular be the case when a new car model is introduced, most extremely for cars introduced in the second half of the year, for which we may have claims, but no premium. Such cars should not be included in the parameter estimation.

The following additional model assumptions and estimation method were applied. Let N_{ki} be the number of claims from risk unit i of car model k, and let Z_{kij} denote the claim amount of the jth of these claims. Then

$$X_{ki} = \sum_{j=1}^{N_{ki}} Z_{kij}.$$

We assume that given Θ_k, the Z_{kij}'s are conditionally independent and identically distributed and conditionally independent of the N_{ki}'s. It is further assumed that N_{ki} is conditionally Poisson distributed with parameter $w_{ki}r_k(\Theta_k)$ given Θ_k. It is well known that under these conditions

$$\mathrm{Var}[X_{ki} \mid \Theta_k] = w_{ki}r_k(\Theta_k)t_k(\Theta_k)$$

with

$$t_k(\Theta_k) = E[Z_{kij}^2 \mid \Theta_k],$$

and by using (3.3) we obtain

$$s^2(\Theta_k) = r_k(\Theta_k)t_k(\Theta_k). \tag{3.17}$$

We now have that

$$\phi_k^* = \left(\sum_{i=1}^{I_k} \sum_{j=1}^{N_{ki}} Z_{kij}^2\right) \Big/ w_k$$

satisfies $E[\phi_k^* \mid \Theta_k] = s^2(\Theta_k)$, and thus

$$\phi^* = \sum_{k=1}^{K} u_k \phi_k^*$$

is an unbiased estimator of ϕ for all weights u_k. (We stress that the quantities ϕ_k^* and ϕ^* defined in the present subsection are not the same as the quantities defined in subsection 3.3, hoping that this abuse of notation will not present any problems to the reader.)

The author is not quite happy with the introduction of the present compound Poisson assumption in our model. From (3.17) we see that the functions r_k and t_k depend on k whereas their product is independent of k. And r_k really should depend on k as an independence assumption would imply that technical data have no influence on the number of claims, which seems very unrealistic.

The fact that in our test sample ϕ_k^* was strongly correlated with the four basic technical variables, could be a consequence of the issue discussed in the previous paragraph. Let a_k denote the engine power of car model k. From our test sample consisting of 62492 policies distributed on 90 different car models, we found the correlations displayed in Table 3.1 by using a correlation procedure in SAS. As is seen from the table, the correlations become considerably lower if we divide ϕ_k^* by a_k. Therefore we replace assumption (3.3) by

$$\mathrm{Var}[Y_{ki} \mid \Theta_k] = a_k s^2(\Theta_k)/w_{ki} = s^2(\Theta_k)/v_{ki}$$

with $v_{ki} = w_{ki}/a_k$. Under this assumption (3.17) should be replaced by

$$s^2(\Theta_k) = r_k(\Theta_k) t_k(\Theta_k)/a_k.$$

We get that

$$\phi_k^* = \left(\sum_{i=1}^{I_k} \sum_{j=1}^{N_{ki}} Z_{kij}^2\right) \Big/ (a_k w_{ki})$$

satisfies $E[\phi_k^* \mid \Theta_k] = s^2(\Theta_k)$, and thus

$$\phi^* = \sum_{k=1}^{K} u_k \phi_k^*$$

TABLE 3.1

CORRELATION OF ϕ_k^* AND ϕ_k^*/a_k WITH THE FOUR BASIC RISK VARIABLES

	ϕ_k^*		ϕ_k^*/a_k	
	unweighted	weight w_k	unweighted	weight v_k
Weight	0 249	0 445	−0 009	0.078
Power	0 337	0 499	0 041	0 083
Cylinder volume	0 346	0 509	0 063	0 106
Price	0 286	0 486	0 060	0 187

is an unbiased estimator of $\phi = Es^2(\Theta_k)$ for all weights u_k; in the numerical example in Section 6 we applied $u_k = K^{-1}$.

The reason that we introduced a v_{kt} in subsection 3.1 should now be clear: The derivations made in the previous subsections are still valid under our revised assumptions; we have only changed the definition of some of the quantities.

3.7. *Introduction of Subjective Assessment*

Classification of individual car models by credibility has also been treated by CAMPBELL (1986). He computes a model factor by a pure Buhlmann–Straub model, that is, he makes no regression assumption about the technical attributes of the car. However, before performing the credibility analysis, he divides the cars by using cluster analysis into groups of cars that are similar with respect to technical attributes. The credibility analysis is then performed within each group of car models. Roughly speaking, one could say that in our set-up the regression assumption plays the role of the cluster analysis in Campbell's set-up.

After the Buhlmann–Straub analysis has been performed, Campbell lets the final value of the model factor be a weighted mean of the value found by the Buhlmann–Straub analysis and a subjective estimate based on a technical assessment of the car.

Let us now see how one could incorporate a subjective estimator in our model. We assume that when car model k is initially classified, a skilled person proposes a class C_k. His proposal is based on a technical assessment of the car. From the class C_k we find the factor

$$F_k = 1.04^{C_k - 30}.$$

This factor is not yet comparable to $m_k(\Theta_k)$ as it is differently scaled (cf. subsection 3.4). From (3.1) and (3.4) we get

$$Y_k = (X_k / p_k) f_k,$$

which motivates the scaling factor

$$N = \left(\sum_{k=1}^{K} X_k\right) \Big/ \sum_{k=1}^{K} p_k,$$

and we introduce the rescaled model factor

$$A_k = N F_k.$$

We now assume that A_k is independent of the data from the other car models, that it is conditionally independent of $Y_{k1}, \ldots, Y_{kI_k}$ given Θ_k, and that

$$E[A_k \mid \Theta_k] = m_k(\Theta_k) \qquad E \operatorname{Var}[A_k \mid \Theta_k] = \tau.$$

Now let $\tilde{m}_s$ be the credibility estimator of $m(\Theta_s)$ based on $Y_{s1}, \ldots, Y_{sI_s}$, and

A_s and let

$$\psi_s = E(m_s(\Theta_s) - \tilde{m}_s)^2.$$

From Theorem 2.1 we get

$$\tilde{m}_s = (v_s Y_s + \varepsilon A_s + \varkappa \mu_s)/(v_s + \varepsilon + \varkappa)$$

$$\psi_s = \phi/(v_s + \varepsilon + \varkappa)$$

with

$$\varepsilon = \phi/\tau.$$

We have that

$$\hat{\tau} = \sum_{k=1}^{K} u_k\{(Y_k - A_k)^2 - \phi^*/v_k\}$$

is an unbiased estimator of τ for all weights u_k, but as $\hat{\tau}$ can take negative values, we propose to estimate τ by $\tau^* = \max(\hat{\tau}, 0)$.

We can of course still estimate λ and $\boldsymbol{\beta}$ by the estimators previously found, but if we also want to include the A_k's in the estimation, we can easily modify the estimators presented in subsection 3.3 by using the following trick: We simply transform the subjective estimator A_k to an artificial risk unit $I_k + 1$ with "risk volume"

$$v_{k,I_k+1} = \varepsilon \tag{3.18}$$

and "claim amount"

$$X_{k,I_k+1} = \varepsilon A_k. \tag{3.19}$$

By adding the new risk units $X_{1,I_1+1}, \ldots, X_{K,I_K+1}$ to the statistics data, we can estimate λ and $\boldsymbol{\beta}$ in exactly the same way as in subsection 3.3. In (3.18) and (3.19) we estimate ε by

$$\varepsilon^* = \phi^*/\tau^*.$$

This author is for two reasons a bit reluctant about the introduction of the subjective estimator A_s in the credibility estimator $\tilde{m}_s$. Both reasons really have as a consequence that the model assumptions made about the A_k's are not fulfilled in practice.

Firstly, the person performing the assessment will probably gradually adapt himself to the statistical model. He will get a feeling of what class the statistical model will propose, and thus his assessment is no longer independent. This does not seem to be an important objection, but it means that after a while the attitude of the person is apt to change, and thus one should frequently update the estimate of τ.

The second objection is more serious. In a competitive market like the Norwegian one, not only the risk level of the car will influence the person performing the assessment, but also the classification of similar cars, not only by Storebrand, but also by the competing companies.

Thus this author is more attracted by the opinion expressed in Section 1, that the subjective assessment should be influenced by the statistical method instead of influencing the method itself.

4. A HIERARCHICAL APPROACH

4.1. *Model*

The make of the car is a characteristic that we have not mentioned yet, but it could contain valuable information about the risk of a car; the information that the car is a Mercedes Benz, may contain information about both the car and its driver that is not contained in characteristics like price, power, etc. It should be noted that make differs from the characteristics studied in subsection 3.5 in one important respect; whereas those characteristics were *quantitative,* make is *qualitative,* and thus we cannot directly include make in the set-up of Section 3. One possible approach would be to extend the regression analysis of Section 3 to a covariance analysis. Instead of following that line we are going to extend the non-hierarchical regression model of Section 3 to a hierarchical model with a new level representing the make of the car.

Consider a group of N different makes. For make n we have observed K_n different car models, and for model k of these we have observed I_{nk} risk units. Let X_{nki} denote the total claim amount in the exposure time for unit i of model k of make n, and let p_{nki} be the earned premium. We introduce

$$w_{nki} = p_{nki}/f_{nk},$$

where f_{nk} denotes the old factor for make of car.

We assume that claim amounts from cars of different makes are independent, and that from within one make n, claim amounts from different car models are conditionally independent given a random parameter H_n (capital Greek eta) characterizing make n. Within car model k of make n, the claim amounts from different risk units are assumed to be conditionally independent given $(\Theta_{nk}, \mathrm{H}_n)$, where Θ_{nk} is a random parameter characterizing car model k of make n. It is assumed that $\Theta_{n1}, \ldots, \Theta_{nK_n}$ are conditionally independent and identically distributed given H_n, and that their common conditional distribution depends on the make only through the value of H_n. We further assume that $\mathrm{H}_1, \ldots, \mathrm{H}_N$ are independent and identically distributed.

Let

$$Y_{nki} = X_{nki}/w_{nki}.$$

It is assumed that

$$\begin{aligned} E[Y_{nki} \mid \Theta_{nk}, \mathrm{H}_n] &= m_{nk}(\Theta_{nk}, \mathrm{H}_n) \\ E\,\mathrm{Var}[m_{nk}(\Theta_{nk}, \mathrm{H}_n) \mid \mathrm{H}_n] &= \lambda \\ \mathrm{Var}[Y_{nki} \mid \Theta_{nk}, \mathrm{H}_n] &= s^2(\Theta_{nk}, \mathrm{H}_n)/v_{nki} \end{aligned}$$

with $v_{nkt} = w_{nkt}/a_{nk}$, where a_{nk} is a known quantity which could be equal to one, engine power (cf. subsection 3.6), or something else. We further assume that

$$E[m_{nk}(\Theta_{nk}, \mathrm{H}_n) | \mathrm{H}_n] = \boldsymbol{x}'_{nk}\boldsymbol{b}(\mathrm{H}_n), \tag{4.1}$$

where $\boldsymbol{x}_{nk}$ is a known, non-random $q \times 1$ design vector based on the technical data of the car and $\boldsymbol{b}$ is a $q \times 1$ vector function. We introduce

$$\phi = Es^2(\Theta_{nk}, \mathrm{H}_n) \qquad \boldsymbol{\beta} = E\boldsymbol{b}(\mathrm{H}_n)$$

$$\boldsymbol{\Xi} = \mathrm{Cov}\ \boldsymbol{b}(\mathrm{H}_n) \tag{4.2}$$

$$\varkappa = \phi/\lambda$$

$$X_{nk} = \sum_{t=1}^{I_{nk}} X_{nkt} \qquad v_{nk} = \sum_{t=1}^{I_{nk}} v_{nkt} \qquad w_{nk} = \sum_{t=1}^{I_{nk}} w_{nkt}$$

$$Y_{nk} = X_{nk}/w_{nk}.$$

We note that for $\boldsymbol{\Xi} = \boldsymbol{0}$, the model reduces to the non-hierarchical model studied in Section 3.

4.2. *Credibility Estimators of $m_r(\Theta_{rs}, \mathrm{H}_r)$*

Let $\tilde{m}_{rs}$ and $\tilde{\boldsymbol{b}}_r$ denote the credibility estimators of $m_{rs}(\Theta_{rs}, \mathrm{H}_r)$ and $\boldsymbol{b}(\mathrm{H}_r)$ based on the observed Y_{nkt}'s. We introduce the estimation errors

$$\psi_{rs} = \mathrm{Var}(m_{rs}(\Theta_{rs}, \mathrm{H}_r) - \tilde{m}_{rs}) \qquad \boldsymbol{\Pi}_r = \mathrm{Cov}(\boldsymbol{b}(\mathrm{H}_r) - \tilde{\boldsymbol{b}}_r).$$

Then we have the following result.

THEOREM 4.1. *We have*

$$\tilde{m}_{rs} = \zeta_{rs} Y_{rs} + (1 - \zeta_{rs})\boldsymbol{x}'_{rs}\tilde{\boldsymbol{b}}_r \tag{4.3}$$

$$\psi_{rs} = (1 - \zeta_{rs})\{\lambda + (1 - \zeta_{rs})\boldsymbol{x}'_{rs}\boldsymbol{\Pi}_r\boldsymbol{x}_{rs}\} \tag{4.4}$$

with

$$\zeta_{rs} = v_{rs}/(v_{rs} + \varkappa).$$

PROOF As the coefficients of credibility estimators depend on only first- and second-order moments, it is sufficient to prove the result for a special case having the same first- and second-order moments as the general case. It is convenient to consider multinormal distributions as it is well-known that in that case the Bayes estimators are linear, and hence they are equal to the credibility estimators.

Let

$$W_{nkt} = v_{nkt}^{1/2}\{Y_{nkt} - m_{nk}(\Theta_{nk}, \mathrm{H}_n)\}$$

$$U_{nk} = m_{nk}(\Theta_{nk}, \mathrm{H}_n) - \boldsymbol{x}'_{nk}\boldsymbol{b}(\mathrm{H}_n).$$

We assume that the W_{nkt}'s are independent and identically distributed $N(0, \phi)$, the U_{nk}'s are independent and identically distributed $N(0, \lambda)$, the $\boldsymbol{b}(\mathrm{H}_n)$'s are independent and identically distributed $N(\boldsymbol{\beta}, \boldsymbol{\Xi})$, and that the W_{nkt}'s, the U_{nk}'s, and the $b(\mathrm{H}_n)$'s are independent. It is obvious that we have the same first- and second-order moments as in the distribution-free model. Furthermore, we have

$$E[m_{rs}(\Theta_{rs}, \mathrm{H}_r)|\, b(\mathrm{H}_r),\, Y_{rki} \forall (k, i)] = \zeta_{rs} Y_{rs} + (1 - \zeta_{rs}) x'_{rs} b(\mathrm{H}_r)$$

as under the conditional probability measure given $b(\mathrm{H}_r)$ we have the same first- and second-order moment structure for make r as in subsection 3.1 (cf. formula (3.5)). We get

$$\begin{aligned}\tilde{m}_{rs} &= E[m_{rs}(\Theta_{rs}, \mathrm{H}_r)|\, Y_{nki} \forall (n, k, i)]\\ &= E[E[m_{rs}(\Theta_{rs}, \mathrm{H}_r)| b(\mathrm{H}_r),\, Y_{nki} \forall (n, k, i)] \,|\, Y_{nki} \forall (n, k, i)]\\ &= E[E[m_{rs}(\Theta_{rs}, \mathrm{H}_r)|\, b(\mathrm{H}_r),\, Y_{rki} \forall (k, i)] \,|\, Y_{nki} \forall (n, k, i)]\end{aligned}$$

as different makes are independent, and thus

$$\begin{aligned}\tilde{m}_{rs} &= \zeta_{rs} Y_{rs} + (1 - \zeta_{rs}) x'_{rs} E[b(\mathrm{H}_r)|\, Y_{nki} \forall (n, k, i)]\\ &= \zeta_{rs} Y_{rs} + (1 - \zeta_{rs}) x'_{rs} \tilde{b}_r,\end{aligned}$$

which proves (4.3)

For ψ_{rs} we apply the same way of reasoning and get

$$\begin{aligned}\psi_{rs} &= E \operatorname{Var}[m_{rs}(\Theta_{rs}, \mathrm{H}_r)|\, Y_{nki} \forall (n, k, i)]\\ &= E \operatorname{Var}[m_{rs}(\Theta_{rs}, \mathrm{H}_r)|\, b(\mathrm{H}_r),\, Y_{nki} \forall (n, k, i)]\\ &\quad + E \operatorname{Var}[E[m_{rs}(\Theta_{rs}, \mathrm{H}_r)|\, b(\mathrm{H}_r),\, Y_{nki} \forall (n, k, i)] \,|\, Y_{nki} \forall (n, k, i)]\\ &= \lambda(1 - \zeta_{rs}) + (1 - \zeta_{rs})^2 x'_{rs} (\operatorname{Cov}[b(\mathrm{H}_r)|\, Y_{nki} \forall (n, k, i)]\, x_{rs}\\ &= (1 - \zeta_{rs})[\lambda + (1 - \zeta_{rs}) x'_{rs} \Pi_r x_{rs}],\end{aligned}$$

which proves (4.4).

This completes the proof of Theorem 4.1. Q.E.D.

We now want an expression for $\tilde{b}_r$. To reduce the dimension of the problem we first prove the following lemma.

LEMMA 4.1. *The credibility estimator $\tilde{b}_r$ depends on the Y_{nki}'s only through $Y_{r1}, \ldots, Y_{rK_r}$.*

PROOF Let $\tilde{b}_r^{(1)}$ be credibility estimator of $b(\mathrm{H}_r)$ based on $Y_{r1}, \ldots, Y_{rK_r}$. Then by Theorem 2.1(ii)

$$\begin{aligned}E\tilde{b}_r^{(1)} &= Eb(\mathrm{H}_r)\\ \operatorname{Cov}(\tilde{b}_r^{(1)}, Y_{rk}) &= \operatorname{Cov}(b(\mathrm{H}_r), Y_{rk}). \qquad k = 1, \ldots, K_r\end{aligned}$$

As

$$\begin{aligned}\operatorname{Cov}(Y_{rs}, Y_{rki}) &= \operatorname{Cov}(Y_{rs}, Y_{rk})\\ \operatorname{Cov}(b(\mathrm{H}_r), Y_{rki}) &= \operatorname{Cov}(b(\mathrm{H}_r), Y_{rk})\end{aligned}$$

for all (k, s, i), we get

$$\operatorname{Cov}(\tilde{b}_r^{(1)}, Y_{rki}) = \operatorname{Cov}(b(\mathrm{H}_r), Y_{rki}).$$

Furthermore, as different makes are independent, we have

$$\operatorname{Cov}(\tilde{b}_r^{(1)}, Y_{nki}) = \operatorname{Cov}(b(\mathrm{H}_r), Y_{nki}) = \mathbf{0}$$

for all $n \neq r$, and thus Lemma 4.1 follows from Theorem 2.1(ii). Q.E.D.

We want to find a matrix expression for $\tilde{b}_r$ and introduce

$$\mathbf{Z}_r = \operatorname{diag}(\zeta_{r1}, \ldots, \zeta_{rK_r})$$

$$X_r = (x_{r1}, \ldots, x_{rK_r})' \qquad Y_r = (Y_{r1}, \ldots, Y_{rK_r})'.$$

We write $\tilde{b}_r$ as

$$\tilde{b}_r = \gamma_r + \Gamma_r Y_r.$$

From (2.1) we get

$$\gamma_r = \beta - \Gamma_r X_r \beta,$$

that is,

$$\tilde{b}_r = \Gamma_r(Y_r - X_r\beta) + \beta.$$

From (2.2) we obtain

$$\Gamma_r \operatorname{Cov} Y_r = \operatorname{Cov}(b(\mathrm{H}_r), Y_r'). \tag{4.5}$$

We easily get

$$\operatorname{Cov} Y_r = \lambda \mathbf{Z}_r^{-1} + X_r \Xi X_r' \qquad \operatorname{Cov}(b(\mathrm{H}_r), Y_r') = \Xi X_r',$$

and insertion in (4.5) gives

$$\Gamma_r(\lambda \mathbf{Z}_r^{-1} + X_r \Xi X_r') = \Xi X_r'. \tag{4.6}$$

We multiply (4.6) by $\mathbf{Z}_r X_r$ from the right to obtain

$$\Gamma_r X_r(\lambda I_q + \Xi X_r' \mathbf{Z}_r X_r) = \Xi X_r' \mathbf{Z}_r X_r,$$

which gives

$$\Gamma_r X_r = \Xi X_r' \mathbf{Z}_r X_r(\lambda I_q + \Xi X_r' \mathbf{Z}_r X_r)^{-1}. \tag{4.7}$$

From (4.6) we get

$$\Gamma_r \lambda \mathbf{Z}_r^{-1} = (I_q - \Gamma_r X_r)\Xi X_r', \tag{4.8}$$

that is,

$$\Gamma_r = \lambda^{-1}(I_q - \Gamma_r X_r)\Xi X_r' \mathbf{Z}_r.$$

Insertion of (4.7) gives

$$\Gamma_r = (\lambda I_q + \Xi X_r' \mathbf{Z}_r X_r)^{-1} \Xi X_r' \mathbf{Z}_r.$$

If $X_r' \mathbf{Z}_r X_r$ is non-singular, we introduce

$$\hat{b}_r = (X_r' \mathbf{Z}_r X_r)^{-1} X_r' \mathbf{Z}_r Y_r.$$

Then we have

$$\Gamma_r Y_r = (\lambda I_q + \Xi X_r' Z_r X_r)^{-1} \Xi X_r' Z_r Y_r = (\lambda I_q + \Xi X_r' Z_r X_r)^{-1} \Xi X_r' Z_r X_r \hat{b}_r$$
$$= \Xi X_r' Z_r X_r (\lambda I_q + \Xi X_r' Z_r X_r)^{-1} \hat{b}_r = \Delta_r \hat{b}_r,$$

where we have introduced the credibility matrix

$$\Delta_r = \Xi X_r' Z_r X_r (\lambda I_q + \Xi X_r' Z_r X_r)^{-1},$$

and we get

$$\tilde{b}_r = \Delta_r \hat{b}_r + (I_q - \Delta_r)\beta.$$

We still have to find an expression for the estimation error matrix Π_r. By Theorem 2.1(iii) we get

$$\Pi_r = \operatorname{Cov} b(\mathrm{H}_r) - \operatorname{Cov} \tilde{b}_r = \Xi - \Gamma_r(\operatorname{Cov} Y_r)\Gamma_r' = \Xi - \Gamma_r(\lambda Z_r^{-1} + X_r \Xi X_r')\Gamma_r'.$$

We insert (4.6) and obtain

$$\Pi_r = \Xi - \Xi X_r' \Gamma_r' = \Xi (I_q - X_r' \Gamma_r'),$$

that is,

$$\Pi_r = \Xi (I_q - \Delta_r') = (I_q - \Delta_r)\Xi,$$

the last equality because Π_r and Ξ are symmetric.

We now have expressions for all the quantities that we need for the computation of $\tilde{m}_{rs}$ and ψ_{rs}.

4.3. *Parameter Estimation*

Corresponding to (3.6) we introduce

$$\phi_{nk}^* = (I_{nk} - 1)^{-1} \sum_{i=1}^{I_{nk}} v_{nki}(Y_{nki} - Y_{nk})^2,$$

for which we have

$$E[\phi_{nk}^* \mid \Theta_{nk}, \mathrm{H}_n] = s^2(\Theta_{nk}, \mathrm{H}_n)$$

if $v_{nki} = w_{nki}$, and in that case

$$\phi^* = \sum_{n=1}^{N} \sum_{k=1}^{K_n} u_{nk} \phi_{nk}^*$$

is an unbiased estimator of ϕ for all weights $u_{nk} (\sum_{n=1}^{N} \sum_{k=1}^{K_n} u_{nk}) = 1$.

It should be obvious how one could generalize the assumptions and estimators introduced in subsection 3.6 to the hierarchical model, and we shall not go any further into details on that matter.

In the following we just assume that we have got an unbiased estimator ϕ^* of

ϕ, and the following derivations do not depend on whether $v_{nki} = w_{nki}$ or not.

For the estimation of λ, β, and Ξ we shall also assume that X_n has full rank q for all n. In practice, this will mean that we exclude data from makes for which we have observed only a few car models, from the estimation procedures. It is of course questionable not to utilize these data, but the estimation procedures become much simpler.

We introduce

$$v_n = \sum_{k=1}^{K_n} v_{nk} \qquad v = \sum_{n=1}^{N} v_n$$

$$D_n = v_n^{-1}\operatorname{diag}(v_{n1}, \ldots, v_{nK_n})$$

$$b_n = (X_n' D_n X_n)^{-1} X_n' D_n Y_n.$$

Analogous to what we did in subsection 3.3, we get

$$E(Y_n - X_n\dot{b}_n)' D_n (Y_n - X_n\dot{b}_n) = \lambda[1 - \operatorname{tr}\{(X_n' D_n X_n)^{-1} X_n' D_n^2 X_n\}] + (K_n - q)\phi / v_n,$$

and thus

$$\hat{\lambda} = \frac{1}{v} \sum_{i=1}^{N} \frac{v_n (Y_n - X_n\dot{b}_n)' D_n (Y_n - X_n\dot{b}_n) - (K_n - q)\hat{\phi}}{1 - \operatorname{tr}\{(X_n' D_n X_n)^{-1} X_n' D_n^2 X_n\}}$$

is an unbiased estimator of λ. As $\hat{\lambda}$ may take negative values, we proceed as in subsection 3.3 to construct a modified estimator λ^* which is non-negative or positive. In the following, we assume for simplicity that λ^* is positive.

Let

$$W_n = \left(\sum_{r=1}^{N} X_r' Z_r X_r\right)^{-1} X_n' Z_n X_n$$

$$\hat{\beta} = \sum_{n=1}^{N} W_n \hat{b}_n = \left(\sum_{n=1}^{N} X_n' Z_n X_n\right)^{-1} \sum_{n=1}^{N} X_n' Z_n Y_n.$$

It seems reasonable to base our estimator of Ξ on

$$Q = \sum_{n=1}^{N} W_n (\hat{b}_n - \hat{\beta})(\hat{b}_n - \hat{\beta})'.$$

We have

$$\begin{aligned} EQ &= \sum_{n=1}^{N} W_n\, E(\hat{b}_n - \hat{\beta})(\hat{b}_n - \hat{\beta})' = \sum_{n=1}^{N} W_n \operatorname{Cov}(\hat{b}_n - \hat{\beta}) \\ &= \sum_{n=1}^{N} W_n [\operatorname{Cov} \hat{b}_n - \operatorname{Cov}(\hat{b}_n, \hat{\beta}') - \operatorname{Cov}(\hat{\beta}, \hat{b}_n') + \operatorname{Cov} \hat{\beta}] \\ &= \sum_{n=1}^{N} W_n [\operatorname{Cov} \hat{b}_n - \operatorname{Cov}(\hat{\beta}, \hat{b}_n')], \end{aligned}$$

that is,

$$EQ = \sum_{n=1}^{N} W_n (I_q - W_n) \operatorname{Cov} \hat{b}_n. \tag{4.9}$$

For all n we have

$$\operatorname{Cov} \hat{b}_n = (X_n' \mathbf{Z}_n X_n)^{-1} X_n' \mathbf{Z}_n (\operatorname{Cov} Y_n) \mathbf{Z}_n X_n (X_n' \mathbf{Z}_n X_n)^{-1}$$
$$= (X_n' \mathbf{Z}_n X_n)^{-1} X_n' \mathbf{Z}_n (\lambda \mathbf{Z}_n^{-1} + X_n' \Xi X_n) \mathbf{Z}_n X_n (X_n' \mathbf{Z}_n X_n)^{-1} = \lambda (X_n' \mathbf{Z}_n X_n)^{-1} + \Xi,$$

and insertion in (4.9) gives

$$EQ = \mathbf{T}\lambda + \left(\sum_{n=1}^{N} W_n (I_q - W_n) \right) \Xi$$

with

$$\mathbf{T} = \sum_{n=1}^{N} W_n (I_q - W_n)(X_n' \mathbf{Z}_n X_n)^{-1} = \sum_{n=1}^{N} (I_q - W_n) W_n (X_n' \mathbf{Z}_n X_n)^{-1}$$
$$= \left(\sum_{n=1}^{N} (I_q - W_n) \right) \left(\sum_{r=1}^{N} X_r' \mathbf{Z}_r X_r \right)^{-1} = (N-1) \left(\sum_{r=1}^{N} X_r' \mathbf{Z}_r X_r \right)^{-1}$$

that is,

$$EQ = (N-1) \left(\sum_{r=1}^{N} X_r' \mathbf{Z} X_r \right)^{-1} \lambda + \left(I_q - \sum_{r=1}^{N} W_n^2 \right) \Xi,$$

and

$$\Xi = \left(I_q - \sum_{n=1}^{N} W_n^2 \right)^{-1} \left\{ \sum_{n=1}^{N} W_n (\hat{b}_n - Y\hat{\beta})(\hat{b}_n - Y\hat{\beta})' - (N-1) \left(\sum_{r=1}^{N} X_r' \mathbf{Z}_r X_r \right)^{-1} \hat{\lambda} \right\}$$

is an unbiased estimator of Ξ. However, as Ξ is symmetric whereas Ξ does not in general have this property, we replace Ξ by

$$\hat{\Xi} = (\Xi + \Xi')/2.$$

When estimating λ, we had the problem that $\hat{\lambda}$ was not necessarily positive. The analogous problem when estimating Ξ is that $\hat{\Xi}$ is not necessarily positive semi-definite. As $\hat{\Xi}$ is symmetric, it can be written as

$$\hat{\Xi} = A' \mathbf{T} A,$$

where A is an orthonormal $q \times q$ matrix (i.e. $A'A = I_q$) and

$$\mathbf{T} = \operatorname{diag}(\tau_1, \ldots, \tau_q),$$

where $\tau_1, \ldots, \tau_q$ denote the eigenvalues of $\hat{\Xi}$. Let

$$\tau_i^0 = \max(\tau_i, 0) \qquad i = 1, \ldots, q$$
$$\mathbf{T}^0 = \operatorname{diag}(\tau_1^0, \ldots, \tau_q^0).$$

It can be shown (cf. BUNKE and GLADITZ (1974), RAO (1965)) that

$$\Xi^* = A' \mathbf{T}^0 A$$

satisfies

$$v' (\Xi^* - \hat{\Xi}) v \leqslant v' (P - \hat{\Xi})\, v$$

for all $q \times 1$ vectors v and all positive semi-definite $q \times q$ matrices P, and hence it seems reasonable to replace $\hat{\Xi}$ by Ξ^* to get a positive semi-definite estimator. To avoid having to take special care of the case when Ξ^* is not strictly positive definite, one could instead of replacing negative eigenvalues by zero, replace them by some small positive number; one possible choice would be ε/N for some ε as we would than have asymptotic unbiasedness when N goes to infinity.

The computation of Ξ^* from $\hat{\Xi}$, involving the construction of A and $\mathbf{T}$, may seem complicated. However, in SAS we had standard procedures for the computation of A and $\mathbf{T}$.

The procedure for estimation of Ξ depends on the parameters ϕ and λ, which were assumed to be unknown, and we therefore insert the estimators ϕ^* and λ^* for these parameters.

We have that

$$\boldsymbol{\beta}^* = \left(\sum_{n=1}^{N} \Delta_n\right)^{-1} \sum_{n=1}^{N} \Delta_n \hat{\boldsymbol{b}}_n \tag{4.10}$$

is the best linear unbiased estimator of $\boldsymbol{\beta}$. As $\boldsymbol{\beta}^*$ depends on the unknown parameters ϕ, λ, and Ξ, we insert the estimators ϕ^*, λ^*, and Ξ^* for these parameters in (4.10).

We have now found estimators for all the unknown parameters involved in the credibility estimators presented in subsection 4.2, and we are therefore able to construct empirical versions of these credibility estimators.

4.4. *A Disadvantage of the Hierarchical Model*

For a new car model s of make r (i.e. $w_{rs} = 0$) we have

$$\tilde{m}_{rs} = \boldsymbol{x}'_{rs} \tilde{\boldsymbol{b}}_r.$$

In the non-hierarchical model the corresponding formula was

$$\tilde{m}_{rs} = \boldsymbol{x}'_{rs} \boldsymbol{\beta}.$$

In subsection 3.5 on the choice of regressors, we said that we have some prior opinion on monotonicity, and that the regressors should be chosen such that this monotonicity was preserved. This was not too complicated in the non-hierarchical model. In the hierarchical model it is much more difficult. Whereas in the non-hierarchical model we could just look at the sign of the elements of $\boldsymbol{\beta}$, in the hierarchical model we have to look at the elements of $\tilde{\boldsymbol{b}}_r$ for all r.

In a parametric empirical Bayes analysis we could solve the problem by restricting the support of the distribution of $\boldsymbol{b}(\mathrm{H}_r)$ to a set $\mathscr{B}$ for which the monotonicity is preserved. Then of course also the posterior mean of $\boldsymbol{b}(\mathrm{H}_r)$ would be contained in $\mathscr{B}$. However, such a parametric model would presumably be complicated to handle, and we would probably have to leave the linearity of the estimators.

If our statistical models should be used as proposed in Section 1, that is, not as giving the final answer, but as an aid for the person who finally makes the decision, this author would recommend that this person receives the estimates from both the hierarchical and the non-hierarchical models, using the same regressors in both models. In his decision he should be aware that the hierarchical model utilizes information about the make of the car, information that is not used in the non-hierarchical model. On the other hand, for the assessment of new car models, the non-hierarchical model will preserve some monotonicity properties, which may be violated in the hierarchical model.

4.5. *A Modified Approach*

When the author gave a seminar on the present research, Ragnar Norberg suggested a modified approach that avoids the monotonicity problem discussed in the previous subsection. We replace the assumptions (4.1) and (4.2) by

$$E[m_{nk}(\Theta_{nk}, \mathrm{H}_n) | \mathrm{H}_n] = M_{nk}(\mathrm{H}_n) \qquad EM_{nk}(\mathrm{H}_n) = x'_{nk}\beta \qquad \mathrm{Var}\ M_{nk}(\mathrm{H}_n) = \xi.$$

One could say that these assumptions are more consistent with the assumptions made in the non-hierarchical model whereas (4.1) and (4.2) are more in line with HACHEMEISTER's (1975) regression model.

Under these new assumptions we obtain

$$\tilde{m}_{rs} = \zeta_{rs} Y_{rs} + (1 - \zeta_{rs})(x'_{rs}\beta + D_r) \tag{4.11}$$

with

$$D_r = \left\{\xi \sum_{p=1}^{K_r} \zeta_{rp}(Y_{rp} - x'_{rp}\beta)\right\} \Bigg/ \left(\lambda + \xi \sum_{p=1}^{K_r} \zeta_{rp}\right).$$

It is interesting to compare (4.11) to (3.5). We see that the only formal difference is that we have added a correction term D_r to the prior mean $x'_{rs}\beta$. For the case with no exposure for car model s this property is very attractive. We then get

$$\tilde{m}_{rs} = x'_{rs}\beta + D_r,$$

that is, we compute the prior mean $x'_{rs}\beta$ based on the technical data and add a correction term D_r as the car is of make r.

We hope to return to the present model in a subsequent paper.

5. SOME CARS ARE MORE EQUAL THAN OTHERS

As is well known, there are often several variants of one car model. In a Norwegian price list from 1984 (OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1984)) we found for instance 9 entries for Volkswagen Golf and 28 for Opel Ascona. The technical differences between such variants may be number of doors, engine, shape (coupé/sedan), etc. Such differences will of course in most cases also influence the price. In our investigations we have considered each variant as a separate model. However, variants of a car model usually have very much in common, and it is tempting to try to utilize this informaton in the estimation of the model factors.

One possible solution would be to extend our two-level hierarchical model (make, model) to three levels (make, model, variant) (for multi-level hierarchical models, cf. e.g. SUNDT (1980), NORBERG (1985)). This would be a more complicated model, and we would have to estimate more parameters.

Another possibility would be to drop the make level in the three-level hierarchical model to obtain a two-level model with levels for model and variant. For this model we could make the same assumptions as in Section 4, but the grouping of the cars would be different.

A third approach would be simply to consider different variants as one model. Then we have the difficulty that the different variants do not have the same technical specifications, but as design vector we could use a weighted mean of the design vectors for the different variants with weights proportional to the observed exposures. In this set-up, possible differences in risk characteristics of the variants now pooled together would be incorporated in $s^2(\Theta_k)$ (to use the notation of the non-hierarchical model). The present approach should be used with care as there exist variants with risk characteristics so different from other variants of the same model that they should definitely not be pooled together; a striking example is Volkswagen Golf GTI. Usually, one would be able to identify such "outliers" already before one obtains the risk statistics. However, this need not always be the case, and one should therefore, even if the variants are pooled together, always register the variant of each car in the statistics data so that one is able to detect an "outlier" and revise the pooling if necessary.

6 NUMERICAL EXAMPLE

6.1. *The Data*

We have already mentioned our numerical studies a couple of times. Our first studies were based on data from Storebrand for the year 1983, and in subsection 3.6 we presented some results based on these data. When our first studies had been performed, data from 1984 became available, and in our investigations on these data, we included a greater number of makes and car models than in our 1983 studies. In the present section we shall display figures found in our 1984 study; the 1983 data were analysed in the same way.

For each car model included in the study, we registered the technical variables weight, engine power, cylinder volume, and price. The price was the price given in a list from April 1984 (OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1984)), and we only included car models that were found in this list. This implies that we excluded car models that were no longer produced or imported to Norway. If one should also include older car models, one would have had to use older prices, which would have had to be adjusted to the price level of 1984. At the present stage of the development of models and methods, we decided to leave out this problem, but it is further discussed in SUNDT (1986). As already mentioned, for simplicity we also excluded diesel cars and cars with four-wheel drive.

In the following presentation we use the codes of Storebrand for make and

TABLE 6 1

Code	Name	K_n	Risk units	v_n
11	Audi	7	1112	9050
14	BMW	11	2754	15117
15	Citroen	10	1190	10429
16	Fiat	7	782	9541
17	Ford, British	7	2322	22326
18	Ford, German	24	13107	115557
24	Lancia	1	58	942
25	Mercedes Benz	7	1561	7444
31	Opel	33	8860	67880
33	Peugeot	14	1467	13017
34	Renault	1	950	13771
37	SAAB	6	3382	21261
39	Skoda	2	248	2549
45	Volkswagen	11	3145	35722
46	Volvo	19	3946	29881
47	Daihatsu	2	105	1555
53	Subaru	6	349	3076
54	Mitsubishi	14	1844	17962
66	Talbot	6	507	4976
93	Lada	5	3490	28800
94	Honda	9	3823	29963
96	Toyota	16	4034	35365
97	Nissan	14	3653	33067
98	Mazda	21	8069	69041
	Total	253	70758	598290

model. In Table 6.1 we give some summary policy data for our sample. For the headings of the table we have used the notation of Section 4, and in the following we use $v_{nk} = w_{nk}/$(engine power). As we see from the table, we have applied data from in all 253 different car models distributed on 24 different makes. We applied no such pooling of car models as described in Section 5.

It would obviously be too much to present the results for all 253 car models, and we therefore restrict ourselves to give more detailed data for a representative sample of 25 car models found by including each tenth model from our total sample, ordered by the codes for make and model. In Table 6.2 we display the exposure and the technical variables engine power, weight, price, and price/weight. Prices are given in NOK and weights in kg.

We estimated ϕ by the procedure described in subsection 3.6 and found $\phi^* = 651.1$.

6 2. *The Non-hierarchical Approach*

For the non-hierarchical model we computed from the 1983 data for several different sets of regressors the estimates λ^* and β^* as described in subsection 3.3. According to the criteria given in subsection 3.5, it seemed reasonable to use the two regressors cylinder volume and price/weight, giving $q = 3$. However, it was

TABLE 6.2

Make	Model	Name	Power	Weight	Price	Price/weight	Risk units	v_{nk}
14	541	BMW 320 I	125	1105	162540	147 10	173	928
15	313	Citroen Visa GT	80	830	79200	95 42	35	288
16	321	Fiat Panda 45	45	670	48400	72 24	179	3251
17	328	Ford Escort 1 6 L	79	880	87560	99 50	738	6437
18	451	Ford Sierra 2 0	105	1095	102100	93 24	77	628
18	741	Ford Sierra 1 6	75	1100	109260	99 33	19	206
25	504	Mercedes Benz 190 E	122	1100	199560	181 42	185	1043
31	327	Opel Corsa 1 2 ST Sedan	55	775	67270	86 80	144	1747
31	347	Opel Kadett 1 2 S Combi	60	870	72620	83 47	269	2770
31	421	Opel Rekord 2 0 S	100	1140	118290	103 76	2879	18532
33	354	Peugeot 305 GLS	74	930	88020	94 65	249	2146
33	892	Peugeot 505 Break	100	1295	146580	113 19	50	372
39	323	Skoda 120 GLS	58	910	50627	55 63	74	877
45	523	Volkswagen Santana 1 9 GX	115	1100	138810	126 19	17	120
46	506	Volvo 240 GLT B23A	129	1330	178900	134 51	21	110
46	907	Volvo 240 GLE B23	129	1300	178400	137 23	10	52
53	349	Subaru 1600 GL Swing-Back	71	885	74800	84 52	48	374
54	396	Mitsubishi Galant 1600 GL	75	1065	99900	93 80	206	2188
93	411	Lada 1600 S	78	1040	54570	52 47	716	5669
94	417	Honda Prelude EX	106	985	181400	184 16	36	306
96	433	Toyota Carina Coupé	75	1060	94000	88 68	321	2790
97	321	Nissan Stanza 1 6 GL	81	970	93800	96 70	49	408
97	832	Nissan Bluebird 1 8 GL	88	1150	108300	94 17	327	2804
98	353	Mazda 626 1 6 GLX Sedan	81	1035	93900	90 72	153	1351
98	474	Mazda 929 2 0 DX St Wagon	90	1200	108400	90 33	350	2835

argued that cylinder volume and engine power were strongly correlated, and that diesel cars and petrol cars were more comparable with respect to engine power than with respect to cylinder volume. Therefore it was felt that if we should later include also diesel cars in the analysis, it would be better to replace the regressor cylinder volume by engine power. We did this and got only a slightly higher value of λ^*. With the 1984 data we therefore concentrated on the design (1 power price/weight). We obtained

$$\lambda^* = 0.2063$$
$$\beta^* = (-0.4183\ 0.01238\ 0.01007)',$$

and from the values of ϕ^* and λ^* we found

$$x^* = \phi^*/\lambda^* = 3156.$$

In Table 6.3 we have displayed the observed Y_k, the estimated prior mean μ_k^*, the empirical credibility weight ζ_k^*, and the estimated estimation error ψ_k^* for each of the car models.

We see that Volkswagen Santana 1.9 GX and Volvo 240 GLE B23 have rather extreme values of Y_k. However, as these cars also have low exposure, $\tilde{m}_k^*$ does not differ much from μ_k^*.

We also computed estimates for tariff classes as described in subsection 3.4.

TABLE 6 3

Make	Model	Y_k	μ_k^*	$\bar{m}_k^*$	ζ_k^*	ψ_k^*
14	541	3 336	2 610	2 775	0 2272	0 1595
15	313	0 502	1 533	1 447	0 0836	0 1891
16	321	2.283	0 866	1 585	0 5075	0 1016
17	328	1 465	1 561	1 497	0 6711	0 0679
18	451	5 628	1 820	2.452	0 1660	0 1721
18	741	0 147	1 510	1 426	0 0614	0 1937
25	504	2 075	2 919	2 709	0 2485	0 1551
31	327	0 844	1 136	1 032	0 3563	0.1328
31	347	1 135	1 165	1 151	0 4675	0 1099
31	421	1 644	1 864	1 676	0 8545	0 0300
33	354	0 954	1 451	1 250	0 4048	0 1228
33	892	3 043	1 959	2 073	0 1054	0 1846
39	323	1 301	0.860	0 956	0 2175	0.1615
45	523	10 904	2 276	2 591	0 0365	0 1988
46	506	1 556	2 533	2 500	0 0337	0 1994
46	907	0.000	2 560	2 519	0 0161	0.2030
53	349	1 307	1 312	1 311	0 1060	0 1845
54	396	1 230	1 455	1 363	0 4094	0 1219
93	411	1 133	1 076	1 112	0 6424	0 0738
94	417	3 046	2 748	2 774	0 0883	0 1881
96	433	1 811	1 403	1 594	0.4693	0 1095
97	321	0 862	1 558	1 478	0 1145	0 1827
97	832	1 749	1 619	1 680	0 4705	0 1092
98	353	1 243	1 498	1.422	0 2998	0 1445
98	474	1 186	1 605	1 407	0 4732	0 1087

After having computed γ^*, we computed estimates for the classes based on both the credibility estimates and based on the prior means. For the estimates based on prior means, the deviations from the classes that were actually used in 1984, were in most cases quite small; for the estimates based on the credibility estimates, the deviations were somewhat larger. The explanation is probably that one has been a bit reluctant to alter the class of a car model. For the actual rating, one might feel that the procedure is too sensitive to the random variable Y_k, and one should pay attention to this in the final subjective determination of the class; the *statistical* procedures do not make *political* considerations.

6.3. *The Hierarchical Approach*

Also for the hierarchical model we used the design (1 power price/weight)'. The parameters λ, $\boldsymbol{\Xi}$, and $\boldsymbol{\beta}$ were estimated as described in subsection 4.3.

For λ we found the estimate

$$\lambda^* = 0.1913,$$

from which we obtained

$$\varkappa^* = \phi^*/\lambda^* = 3404.$$

It is reasonable that the value of λ^* is lower than in the non-hierarchical model.

When estimating Ξ, we obtained

$$\hat{\Xi} \cdot 10^5 = \begin{pmatrix} -299650 & -3113 & 4724 \\ -3113 & 75 & 18 \\ 4724 & -18 & -33 \end{pmatrix}.$$

This matrix is obviously not positive definite. It has one negative eigenvalue, and by replacing this eigenvalue by $10^{-6}/N = 4 \cdot 10^{-8}$ as described in subsection 4.3, we arrived at

$$\Xi^* \cdot 10^5 = \begin{pmatrix} 0.0488 & -2.1678 & 1.351 \\ -2.1678 & 107.385 & 66.820 \\ 1.3511 & 66.820 & 41.709 \end{pmatrix}.$$

As the value of λ^* was only slightly lower than in the non-hierarchical model whereas the difference between $\hat{\Xi}$ and Ξ^* is considerable, we presume that for practical purposes we would choose the non-hierarchical model, but we shall go on presenting some results for the hierarchical model for illustrative purposes. We mention that computations made on the same data with the modified model described in subsection 4.5, gave much more reasonable results.

For β we found

$$\beta^* = (-0.0587 \ 0.01228 \ 0.00687)'.$$

TABLE 6 4

Make	$\tilde{b}_n^{*\prime}$		
11	−0 05086	0 01078	0 00762
14	−0 05181	0 05778	−0 02127
15	−0 05134	0 03461	−0 00683
16	−0 05132	0 03384	−0 00638
17	−0 05091	0 01332	0 00631
18	−0 05102	0 01853	0 00276
24	−0 05090	0 01263	0 00662
25	−0 05058	−0 00309	0 01623
31	−0 05065	0 00014	0 01424
33	−0 05059	−0 00233	0 01643
34	−0.05100	0 01753	0.00355
37	−0 05078	0 00707	0 01019
39	−0 05093	0 01407	0 00577
45	−0 05106	0 02029	0 00170
46	−0 05043	−0 01066	0 02107
47	−0 05086	0 01094	0 00775
53	−0 05093	0 01400	0 00575
54	−0 05071	0 00327	0.01236
66	−0 05091	0 01307	0 00630
93	−0 05082	0 00877	0 00905
94	−0 05070	0 00284	0 01290
96	−0 05083	0 00927	0 00862
97	−0 05079	0 00733	0 01003
98	−0 05078	0 00682	0 01037

In Table 6.4 we have displayed the empirical credibility estimate $\tilde{b}_n^*$ for the 24 makes included in the study. The table illustrates the problem discussed in subsection 4.4; we see that for makes 25, 33, and 46 $x'_{nk}\tilde{b}_n^*$ will be decreasing in engine power, and for makes 14, 15, and 16 it will be decreasing in price/weight.

As examples of the values found for Π_n^* we display the value for one make with low exposure (Skoda) and one with high exposure (Opel). We found

$$\Pi_{39}^* \cdot 10^5 = \begin{pmatrix} 0.0211 & -7.9151 & 4.979 \\ -7.9151 & 39.063 & 24.483 \\ 4.9790 & 24.483 & 15.463 \end{pmatrix}$$

$$\Pi_{31}^* \cdot 10^5 = \begin{pmatrix} 0.0071 & -0.095 & 0.068 \\ -0.0954 & 4.459 & 3.114 \\ 0.0678 & 3.114 & 2.255 \end{pmatrix}.$$

Table 6.5 is the hierarchical analogue to Table 6.3. The quantities displayed in the last three columns are the estimates of the quantities

$$\psi_{nk} = (1 - \zeta_{nk})[\lambda + (1 - \zeta_{nk})x'_{nk}\Pi_n x_{nk}]$$

$$\psi_{nk}^{(\text{make})} = \lambda + x'_{nk}\Pi_n x_{nk}$$

$$\psi_{nk}^{(0)} = \lambda + x'_{nk}\Xi X_{nk}.$$

TABLE 6 5

Make	Model	Y_{nk}	$x'_{nk}\tilde{b}_n^*$	$\tilde{m}_{nk}^*$	ζ_{nk}^*	$x'_{nk}\beta^*$	ψ_{nk}^*	$\psi_{nk}^{*(\text{make})}$	$\psi_{nk}^{*(0)}$
14	541	3 336	4 042	3 891	0 2141	2 495	0 1911	0 2573	1 4167
15	313	0 502	2 066	1 944	0 0779	1 587	0.2366	0 2621	0 6572
16	321	2 283	1 010	1 632	0 4885	0 998	0 0994	0 1973	0 1975
17	328	1 465	1 629	1 522	0 6541	1 603	0 0798	0 3050	0 5152
18	451	5 628	2 152	2 693	0 1558	1 880	0 1882	0 2288	2 5677
18	741	0 147	1 613	1 529	0.0572	1 553	0 1871	0 1989	0 3889
25	504	2 075	2 517	2 413	0 2346	2 694	0 1666	0 2258	0 3189
31	327	0.844	1 193	1 075	0 3391	1 221	0 1296	0.1987	0.2013
31	347	1 135	1 146	1 141	0 4486	1 260	0 1072	0 1969	0 2688
31	421	1 644	1 441	1 612	0 8448	1 890	0 0307	0 2336	1 5493
33	354	0 954	1 332	1 186	0 3867	1 508	0.1294	0 2235	0 4459
33	892	3 043	1 576	1 721	0 0985	1 955	0 2666	0 3072	1 1428
39	323	1 301	1.086	1 130	0 2049	1 044	0 2863	0 4036	0 7804
45	523	10 904	2 497	2 782	0 0339	2 229	0 5712	0 6053	1 6357
46	506	1 556	1 408	1 413	0 0313	2 458	0 2641	0 2753	2 4120
46	907	0 000	1 466	1 444	0 0150	2.477	0 2641	0 2692	2 2515
53	349	1 307	1 430	1 417	0 0990	1 402	0 3058	0.3557	0 5623
54	396	1 230	1 354	1 306	0 3912	1 515	0 1290	0 2252	0 4975
93	411	1 133	1 108	1 124	0 6248	1 268	0.0827	0 2688	2 3997
94	417	3 046	2 625	2 660	0 0824	2 517	0 2420	0.2703	0 3121
96	433	1 811	1 409	1 590	0 4505	1.480	0 1137	0 2199	0 6209
97	321	0 862	1.513	1 443	0 1071	1 609	0 2003	0 2283	0 6666
97	832	1 749	1 539	1 634	0 4517	1 677	0 1268	0 2641	1 1276
98	353	1 243	1 442	1 385	0 2841	1 567	0 1460	0 2089	0 8462
98	474	1 186	1.499	1 357	0 4544	1 675	0 1142	0 2244	1 4243

The quantity ψ_{nk} has already been defined as the estimation error of the credibility estimator $\tilde{m}_{nk}$. We have that $\psi_{nk}^{(\mathrm{make})}$ would be the estimation error of $x'_{nk}\tilde{b}_n$ as estimator of $m_{nk'}(\Theta_{nk'}, \mathrm{H}_n)$ for a car model k' with the same technical specifications as car model k, but for which we have no exposure. (To say that $\psi^{(\mathrm{make})}$ is the estimation error of $x'_{nk}\tilde{b}_n$ considered as estimator of $m_{nk}(\Theta_{nk}, \mathrm{H}_n)$ would be wrong as $\tilde{b}_n$ contains claims data from car model k.) Similarly, $\psi_{nk}^{(0)}$ would be the estimation error of $x'_{nk}\beta$ considered as estimator of $m_{n'k'}(\Theta_{n'k'}, \mathrm{H}_{n'})$ for a car model k' of make n', for which we have no exposure.

As a consequence of the fact that the value of λ^* was lower in the present model than in the non-hierarchical model, we see that the values of ζ_{nk}^* are also lower. This is intuitively reasonable as $\tilde{b}_{nk}$ in the hierarchical model would contain more information than β in the non-hierarchical model.

ACKNOWLEDGEMENTS

The present research project was performed by the present author, but of course it was discussed on several occasions in the actuarial department in Storebrand, and the author is grateful to his colleagues Walther Neuhaus, Johan Lindtvedt, and Hans Haugland for useful comments and suggestions. He is also grateful to Storebrand for letting him present this paper, and to Ragnar Norberg for useful suggestions and comments; the present proof of Theorem 4.1 was suggested by him.

REFERENCES

BUNKE, H and GLADITZ, J (1974) Empirical linear Bayes decision rules for a sequence of linear models with different regressor matrices *Mathematische Operationsforschung und Statistik* **5**, 234–244

BUHLMANN, H and STRAUB E (1970) Glaubwurdigkeit fur Schadensatze *Mitteilungen der Vereinigung schweizerischer Versicherungs mathematiker* **70**, 111–133

CAMPBELL, M (1986) An integrated system for estimating the risk premium of individual car models in motor insurance *ASTIN Bulletin* **16**, 165–183

DEGROOT, M H (1970) *Optimal Statistical Decisions* McGraw-Hill, New York

DE VYLDER, F (1976) Geometrical credibility. *Scandinavian Actuarial Journal*, 121–149.

HACHEMEISTER, C A (1975) Credibility for regression models with application to trend In *Credibility Theory and Applications* (ed P M KAHN), pp. 129–163. Academic Press, New York

NORBERG, R (1982) On optimal parameter estimation in credibility *Insurance* **1**, 73–89

NORBERG, R (1985) Hierarchical credibility Analysis of a random effect linear model with nested classification *Statistical Research Report*. Institute of Mathematics, University of Oslo

OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1984) Bilpriser 1984 – Vår Norsk Veiforening, Oslo

RAO, C R (1965) *Linear Statistical Inference and Its Applications* Wiley, New York

SUNDT, B (1980) A multi-level hierarchical credibility regression model *Scandinavian Actuarial Journal*, 25–32

SUNDT, B (1981) Recursive credibility estimation *Scandinavian Actuarial Journal*, 3–21

SUNDT, B (1986) Credibility and old estimates. *Storebrand-Norden Actuarial Research Report* **2**.

BJØRN SUNDT
Storebrand P.O. Box 1380 Vika, N-0114 Oslo 1, Norway.

THE LINEAR MARKOV PROPERTY IN CREDIBILITY THEORY

BY THOMAS WITTING

ETH Zurich

ABSTRACT

We study the linear Markov property, i.e. the possibility of basing the credibility estimator on data of the most recent time period without loss of accuracy. Necessary and sufficient conditions are derived generally. The meaning of the linear Markov property is also discussed in different experience rating and loss reserving models.

KEYWORDS

Linear Markov property; linear sufficiency; credibility.

1. INTRODUCTION

A fundamental question in credibility theory is that of upon which statistic of the available data the credibility estimator should be based. A very general treatment of this problem and a survey of other approaches can be found in NEUHAUS (1985). We consider the special case of data ordered with respect to time. Is it then possible to reduce the data to those of the last time period without diminishing the accuracy of the credibility estimator? If this is the case, then we have defined the linear Markov property. This principle is introduced generally and discussed in some important models of risk theory. We give some sufficient and necessary conditions which are useful in situations when the linear Markov property is not obvious. In most cases the linear Markov property results in a considerable reduction of the number of normal equations which it is necessary to solve to derive the credibility estimator explicitly.

This paper is in a way a summary of the first part of the author's PhD thesis which is taken sometimes as a reference. A copy of this thesis can be obtained from the author.

2. CREDIBILITY ESTIMATION AND LINEAR MARKOV PROPERTY

2.1. *General Assumptions and Notation*

In the present paper it is generally assumed that random variables are square integrable, i.e. all (mixed) second moments exist and are finite. The transpose of a matrix A is A^T. (Random) vectors are in boldface and have to be interpreted as column vectors, i.e. $\mathbf{x} = (x_1, \ldots, x_n)^T$ is a vector with n components.

I_n is the $n \times n$-unit matrix and δ_{ij} the Kronecker symbol.

For random vectors $X = (X_1, \ldots, X_n)^T$ and $Y = (Y_1, \ldots, Y_m)^T$ we use the following symbols and terminology:

P_X for the probability distribution of X.
$P_{X|Y=y}$ for the conditional probability distribution of X given $[Y = y]$ and $P_{X|Y}$ for the corresponding stochastic kernel.
$E[X] = (E[X_1], \ldots, E[X_n])^T$ for the expected value of X.
$E[X|Y] = (E[X_1|Y], \ldots, E[X_n|Y])^T$ for the conditional expected value of X given Y.
$C[X, Y] = E[(X - E[X])(Y - E[Y])^T]$ for the joint covariance matrix of X and Y.
$C[X] = C[X, X]$ for the covariance matrix of X.

It is generally assumed that all symmetric covariance matrices $C[X]$ appearing in the text are positive definite, i.e. the inverse $C[X]^{-1}$ exists. The regularity of $C[X]$ is equivalent to linear independence of the "vectors" $1, X_1, \ldots, X_n$ in the linear space $L_2(R)$ of all square integrable real random variables. For a proof see e.g. WITTING (1986). In particular all random variables appearing in the text are not degenerate. All equations between random variables should be understood in the sense of L_2-equivalence.

2.2. *Credibility Estimation*

We want to estimate the real random variable Y with help of the n-dimensional random vector X which represents the available data. It is well known that $g^*(X) = E[Y|X]$ is the optimal estimator in the sense of minimizing the expected squared loss $E[(g(X) - Y)^2]$ in the class of all measurable functions $g(x)$. Because $E[Y|X]$ can be calculated explicitly by a closed formula only in a few special cases the estimation problem is simplified: we look for the optimal estimator of Y only in the class of (inhomogeneous) linear estimators

$$g(X) = a_0 + \sum_{i=1}^{n} a_i X_i.$$

This optimal estimator exists, is uniquely determined and interpreted as the orthogonal projection of Y onto the $n+1$-dimensional subspace of $L_2(R)$ which is generated by $1, X_1, \ldots, X_n$. Therefore we denote it $\hat{E}[Y|X]$. $\hat{E}[Y|X]$ is called the *credibility estimator* of Y given X.

The orthogonal principle can be formulated in a probabilistic manner as follows:

$$\begin{aligned} E[\hat{E}[Y|X]] &= E[Y] \\ C[Y - \hat{E}[Y|X], X] &= \mathbf{0}. \end{aligned} \tag{1}$$

If the credibility estimator is written in the form

$$\hat{E}[Y|X] = a_0 + \sum_{i=1}^{n} a_i X_i$$

(1) is equivalent to

$$a_0 = E[Y] - \sum_{i=1}^{n} a_i E[X_i] \tag{2}$$

$$\sum_{i=1}^{n} a_i \operatorname{Cov}(X_i, X_k) = \operatorname{Cov}(Y, X_k) \qquad k = 1, \ldots, n.$$

This linear system of normal equations for determining the coefficients $a_0, a_1, \ldots, a_n$ has a unique solution because of our general assumption that $C[X]$ is positive definite. There is no guarantee for being able to calculate $C[X]^{-1}$ explicitly although this may be useful in theoretical situations. However, a recursive algorithm for the inversion of $C[X]$ exists always (see e.g. NORBERG (1985)).

If a random vector $Y = (Y_1, \ldots, Y_m)^T$ has to be estimated, we define $\hat{E}[Y \mid X] = (\hat{E}[Y_1 \mid X], \ldots, \hat{E}[Y_m \mid X])^T$ and confirm the property

$$E[(\hat{E}[Y \mid X] - Y)^T(\hat{E}[Y \mid X] - Y)] = \min_g E[(g(X) - Y)^T(g(X) - Y)].$$

The minimum is taken over the class of all functions $g(x) = a + Ax$ with m-dimensional vector a and $m \times n$-matrix A. The generalization of (1) to this case is obvious. Finally we get the well-known formula

$$\hat{E}[Y \mid X] = C[Y, X] C[X]^{-1}(X - E[X]) + E[Y]. \tag{3}$$

2.3. *Linear Sufficiency*

We consider again the problem of estimating Y by means of X. For many statistical problems one can restrict the investigation to decision functions which depend only through a "sufficient" statistic $T(x)$ on the original observation x. Here we call a statistic $T(x)$ *sufficient* if

$$P_{Y|X} = P_{Y|T(X)} \tag{4}$$

This corresponds with the Bayesian definition of sufficiency if Y is interpreted as a "prior variable".

In the credibility situation one should manage only with linear statistics and the knowledge of second-order moments. This fact suggests a slight change of the meaning of sufficiency in our case.

DEFINITION: The linear statistic $T(x)$ (which is formally a linear mapping $T: R^n \to R^r$ with $r < n$) is called *linear sufficient* if

$$\hat{E}[Y \mid X] = \hat{E}[Y \mid T(X)]. \tag{5}$$

REMARKS: (i) By comparing the system (X, Y) with the corresponding system $(\tilde{X}, \tilde{Y})$ which is normally distributed with the same second-order moment structure it can be proved that for linear statistics $T(x)$ the implication (4) $\Rightarrow$ (5) is valid.

(ii) We can restrict the investigation to homogeneous linear mappings $T(x)$, because a possible inhomogeneous part has to be adapted anyway afterwards when $\hat{E}[Y|T(X)]$ is calculated.
(iii) The concept of linear sufficiency has been already introduced into statistical literature, but only in the context of estimation for linear models; e.g. DRYGAS (1983, 1985).

LEMMA 1. *Let $r < n$ and A a full rank $r \times n$-matrix. The statistic $T(x) = Ax$ is linear sufficient if, and only if*

$$C[Y,X]\{I_n - A^T(AC[X]A^T)^{-1}AC[X]\} = 0.$$

PROOF. Without loss of generality we may assume $E[X] = \mathbf{0}, E[Y] = \mathbf{0}$. Then it follows that $T(x)$ is linear sufficient.

$$\underset{(3)}{\Leftrightarrow} \hat{E}[Y|X] = C[Y, T(X)]C[T(X)]^{-1}T(X)$$

$$\underset{(1)}{\Leftrightarrow} C[Y - C[Y, T(X)]C[T(X)]^{-1}T(X), X] = 0$$

$$\Leftrightarrow C[Y,X] = C[Y, T(X)]C[T(X)]^{-1}C[T(X), X] \tag{6}$$

$$\Leftrightarrow C[Y,X] = C[Y,X]A^T(AC[X]A^T)^{-1}AC[X]. \qquad \text{QED}$$

EXAMPLE. $(r = m = 1)$.
Let A be the $1 \times n$-matrix and E the $n \times n$-matrix whose elements are all equal to 1. We assume that the random variables $X_1, \ldots, X_n$ are exchangeable relative to Y, i.e. $P_{(X_1, \ldots, X_n, Y)} = P_{(X_{\pi(1)}, \ldots, X_{\pi(n)}, Y)}$ for all permutations π of $1, \ldots, n$.

In SUNDT (1979), Theorem 1, it is shown that this condition implies the linear sufficiency of the statistic $T(x) = Ax = \Sigma x_i$. This implication can also be derived from Lemma 1, for it follows from the exchangeability condition with appropriate constants c, d and e that:

$$C[X] = d(E + cI_n), C[Y,X] = eA.$$

This implies together with the simple relationships $EE = nE, AE = nA$ and $A^TA = E$ that:

$$AC[X]A^T = dn(n+c), A^TAC[X] = d(n+c)E.$$

From this follows

$$C[Y,X]\{I_n - A^T(AC[X]A^T)^{-1}AC[X]\} = eA(I_n - 1/nE) = e(A - A) = \mathbf{0}.$$

2.4. *Linear Markov Property*

In the present paper we consider mainly a special case of linear sufficiency, namely the linear Markov property.

Now, there are given n information **vectors** of dimension l

$$X_1 = (X_{11}, \ldots, X_{l1})^T, \ldots, X_n = (X_{1n}, \ldots, X_{ln})^T$$

from which the random vector $Y=(Y_1, \ldots, Y_m)^T$ shall be estimated. We patch the complete information together to the $n\times l$-dimensional vector $X=(X_1{}^T, \ldots, X_n{}^T)^T$.

DEFINITION. The sequence $X_1, \ldots, X_n, Y$ is called *linear Markovian (l.M.)* if $\hat{E}[Y|X]=\hat{E}[Y|X_n]$.

The linear Markov property is equivalent to the linear sufficiency of the statistic $T(x)=x_n$ and makes it possible to reduce the complete information to the information of the last period.

In the language of NEUHAUS (1985) it means that the secondary statistic $(X_1{}^T, \ldots, X_{n-1}{}^T)$ may be excluded from the basic statistic X without loss. The linear Markov property can be characterized by a relation between the second-order moments:

LEMMA 2. *The sequence* $X_1, \ldots, X_n, Y$ is l.M. if, and only if

$$C[Y,X_i]=C[Y,X_n]C[X_n]^{-1}C[X_n,X_i] \qquad \text{for } i=1,\ldots,n-1. \tag{7}$$

The proof follows as special case from Lemma 1 with $T(x)=x_n$, because (6) is then equivalent to (7).

Now we define the linear Markov property also for **processes:**

DEFINITION. Let X_i be a l-dimensional random vector for all $i\in N$. The stochastic (vector-)processes $(X_i)_{i\in N}$ is called *linear Markovian (l.M.)* if the sequences $X_1, \ldots, X_n, X_{n+k}$ are l.M. for all $n, k\in N$.

REMARKS. (i) We consider a 1-dimensional process $(X_i)_{i\in N}$.
Then $(X_i)_{i\in N}$ is l.M. if, and only if the following relation is valid with $c_{i,k}=\mathrm{Cov}(X_i, X_k)$:

$$c_{n+k,i}c_{n,n}=c_{n+k,n}c_{n,i} \qquad \text{for } i,k,n\in N \text{ with } i<n. \tag{8}$$

FELLER (1966) shows that the ordinary Markov property is characterized by (8) for a Gaussian process $(X_i)_{i\in N}$. In this special case the ordinary and the linear Markov property do coincide.

PAPOULIS (1965) shows the corresponding result for the optimal homogeneous linear estimation. In that case we would have to define $\hat{E}[Y|X_1, \ldots, X_n]$ as an orthogonal projection from Y onto the linear subspace generated by $X_1, \ldots X_n$. Then (8) is valid with $c_{i,k}=E[X_iX_k]$.

(ii) For a standard normal variable Z and arbitrary i.i.d. variables Z_1, Z_2 it follows that the sequence $X_1=Z^2, X_2=Z, Y=Z^2$ is Markovian in the ordinary sense but not l.M. The sequence $X_1=Z_1, X_2=Z_1+Z_2, Y=Z_1\cdot Z_2$ is l.M. but not Markovian in the ordinary sense.

These two examples show that the ordinary Markov property does not imply the linear Markov property and vice versa.

The following lemma gives some helpful necessary and sufficient conditions to detect the linear Markov property directly by inspection of the covariance structure of the process.

LEMMA 3. *The following conditions (7)', (9) and (10) are equivalent to the linear Markov property of the process* $(X_i)_{i\in N}$:

$$C[X_{n+k}, X_i] = C[X_{n+k}, X_n]C[X_n]^{-1}C[X_n, X_i] \tag{7)'}$$

for $i, k, n \in N$ *with* $i < n$.
There exists a sequence $(\Lambda_i)_{i\in N}$ *of regular* $l\times l$*-matrices with*

$$C[X_j, X_i] = \left(\prod_{k=i+1}^{j} \Lambda_k\right)C[X_i] \qquad \text{for } i \leqslant j. \tag{9}$$

There exist sequences $(A_i)_{i\in N}$ *and* $(B_i)_{i\in N}$ *of regular* $l\times l$*-matrices with*

$$C[X_j, X_i] = B_j A_i \qquad \text{for } i \leqslant j. \tag{10}$$

where

$$\prod_{k=r}^{m} C_k := \begin{cases} C_m \quad C_r & \text{for } r \leqslant m \\ \quad I_l & \text{for } r > m. \end{cases}$$

PROOF. Because of Lemma 2 the linear Markov property of the process $(X_i)_{i\in N}$ is equivalent to condition (7)'. Therefore it suffices to show:

(9) $\Rightarrow$ (10) $\Rightarrow$ (7)' $\Rightarrow$ (9);

(9) implies (10) with

$$B_j = \prod_{k=1}^{j} \Lambda_k \text{ and } A_i = \left(\prod_{k=1}^{i} \Lambda_k\right)^{-1} C[X_i];$$

(10) implies (7)' by means of the relation

$$\begin{aligned} C[X_{n+k}, X_n]C[X_n]^{-1}C[X_n, X_i] &= B_{n+k}A_nA_n^{-1}B_n^{-1}B_nA_i \\ &= B_{n+k}A_i \\ &= C[X_{n+k}, X_i]; \end{aligned}$$

(7)' implies (9) with $\Lambda_k = C[X_k, X_{k-1}]C[X_{k-1}]^{-1}$, for it follows with $i < j$:

$$\begin{aligned} C[X_j, X_i] &\underset{(7)'}{=} C[X_j, X_{j-1}]C[X_{j-1}]^{-1}C[X_{j-1}, X_i] \\ &= \Lambda_j C[X_{j-1}, X_i] \\ &\underset{\text{(induction)}}{=} \Lambda_j \ldots \Lambda_{i+1} C[X_i] \end{aligned}$$

QED

NOTATION. The sequence of $l\times l$-matrices $(A_i)_{i\in N}$ in (10) is called a *l.M.-factor*. A_i is fixed uniquely to the extent of multiplication from the left of a $l\times l$-matrix independent of i.

Analogously with the segment (10) $\Rightarrow$ (7)' in the proof of Lemma 3 the following result for finite sequences can be shown:

LEMMA 4. *Let (10) be valid for $1 \leqslant i \leqslant j \leqslant n$ whereby $A_1, \ldots, A_n$ is the beginning of the l.M.-factor and also*

(11) $\quad C[Y, X_i] = BA_i$ *for* $1 \leqslant i \leqslant n$ with a $l \times l$*-matrix* B.

Then the sequence $X_1, \ldots, X_n, Y$ is l.M.

2.5. *Componentwise Linear Markov Property*

We use the same notation as in 2.4.

DEFINITION. The sequence $X_1, \ldots, X_n, Y$ is called *componentwise linear Markovian (c.l.M.)* if the sequences $X_{k1}, \ldots, X_{kn}, Y$ are l.M. for all $k = 1, \ldots, l$.

If the components, i.e. the rows of the $l \times n$-matrix $(X_1, \ldots, X_n)$, are independent, it holds that:

$$X_1, \ldots, X_n, Y \text{ is l.M.} \Leftrightarrow X_1, \ldots, X_n, Y \text{ is c.l.M.}$$

This equivalence is evident by Lemma 2, because in the case of independence the matrices $C[X_n]^{-1}$ and $C[X_n, X_i]$ in (7) are diagonal.

Generally no direction of this equivalence is valid, for it holds with two independent real random variables Z_1 and Z_2:

The sequence $X_1 = (0, Z_2)^T, X_2 = (Z_1 + Z_2, 0)^T, Y = Z_1$ is c.l.M. but not l.M.

The sequence $X_1 = (Z_1, Z_2)^T, X_2 = (Z_2, Z_1)^T, Y = Z_1 + Z_2$ is l.M. but not c.l.M.

In the situation of insurance the independence of components is not always fulfilled. As an example one should imagine the components to be claim numbers and totals of claims. Reflecting on the better handling of the c.l.M.-property we are looking for an additional condition that the c.l.M.-property implies the l.M.-property even in the case of dependent components.

To solve this problem we consider two vector valued components. So let $X_i = (Z_i^T, N_i^T)^T$ with l_1-dimensional random vector Z_i and l_2-dimensional random vector N_i and $l_1 + l_2 = l$ $(1 \leqslant i \leqslant n)$.

LEMMA 5. *Let the following four conditions be valid:*

The sequence $Z_1, \ldots, Z_n, Y$ is l.M.
The sequence $N_1, \ldots, N_n, Y$ is l.M.
The sequence $Z_1, \ldots, Z_n, N_n$ is l.M.
The sequence $N_1, \ldots, N_n, Z_n$ is l.M.

Then the sequence $X_1, \ldots, X_n, Y$ is l.M.

To prove this lemma condition (7) has to be checked with help of inversion of the matrix $C[X_n]$. This is somewhat tedious and can be found in WITTING (1986), p.33–36.

3. THE LINEAR MARKOV PROPERTY IN SOME EXPERIENCE RATING MODELS

In the following models our starting point is always a real stochastic process $(Y_i)_{i\in N}$ with covariance structure given by $c_{i,k} = \mathrm{Cov}(Y_i, Y_k)(i, k \in N)$. Thereby Y_i may be interpreted as claim number or total of claims during the period i. At the end of period n the net premium P_{n+1} for the next period will be fixed by $P_{n+1} = \hat{E}[Y_{n+1} \mid Y_1, \quad, Y_n]$. Let $(X_i)_{i\in N}$ be a "linear cumulated transform" of the process $(Y_i)_{i\in N}$, i.e

$$X_i = \sum_{k=1}^{i} a_k Y_k$$

with appropriate coefficients $a_1, \ldots, a_i$ $(i \in N)$.
Let us further assume that the process $(X_i)_{i\in N}$ has the covariance structure given by

(12) $\quad \mathrm{Cov}(X_i, X_j) = g_i + f_i f_j$ for $i \leqslant j$ with $g_i > 0$ and $f_{i+1} - f_i \neq 0$.

With help of the multiplicative decomposition criterion (10) of Lemma 3 the following equivalence can easily be verified:

(13) $\quad (X_i)_{i\in N}$ is l.M. $\Leftrightarrow$ The quotient f_i/g_i is independent of i.

Indication of the proof of the " $\Rightarrow$ " part: From (10) it follows that the fractions

$$\frac{g_i + f_i f_j}{g_1 + f_1 f_j} \quad \text{and} \quad \frac{1 + (f_i/g_i) f_j}{1 + (f_1/g_1) f_j}$$

do not depend on j. This can only be true if f_i/g_i is independent of i because of our assumption in (12) that f_j depends on j.

3.1. *The Model of Jewell*

JEWELL (1975) considers the covariance structure given by $c_{i,k} = \delta_{ik}\gamma_k + \alpha_i\alpha_k$ with appropriate numbers $\gamma_k > 0$ and $\alpha_i \neq 0$ $(i, k \in N)$. This covariance structure is shown by Jewell to yield an explicit solution of the normal equations (2). Under which conditions has the transformed process $(X_i)_{i\in N}$ the linear Markov property? In Jewell's model we have

$$\mathrm{Cov}(X_i, X_j) = \sum_{m=1}^{i} a_m^2\gamma_m + \left(\sum_{m=1}^{i} a_m\alpha_m\right)\left(\sum_{k=1}^{j} a_k\alpha_k\right).$$

So (12) is fulfilled with

$$g_i = \sum_{m=1}^{i} a_m^2\gamma_m \text{ and } f_i = \sum_{m=1}^{i} a_m\alpha_m.$$

We conclude:

$$(X_i)_{i\in N} \text{ is l.M.} \underset{(13)}{\Leftrightarrow} \frac{\sum_{m=1}^{i} a_m\alpha_m}{\sum_{m=1}^{i} a_m^2\gamma_m} \text{ is independent of } i.$$

(14) $$\Leftrightarrow \frac{\alpha_m}{a_m\gamma_m} \text{ is independent of } m.$$

Therefore in Jewell's model the statistic $\Sigma(\alpha_k/\gamma_k)y_k$ is linear sufficient and the premium becomes

$$P_{n+1} = \hat{E}\left[Y_{n+1} \mid \sum_{k=1}^{n} \frac{\alpha_k}{\gamma_k} Y_k \right].$$

3.2. *The Classical Credibility Model*

It is obvious that the classical Buhlmann–Straub model (BUHLMANN and STRAUB, 1970) in the ordinary formulation is a special case of Jewell's model with known numbers α_{im} and γ_k. We choose a cumulative view onto the model. Considering one risk unit we denote by

- X_i, the cumulated total of claims up to the end of period i;
- Θ, a random risk parameter describing the unknown characteristics of the risk unit;
- p_i, a known cumulated measure of volume up to period i.

It is assumed that for given $[\Theta = \theta]$ the process $(X_i)_{i\in N}$ has independent increments, and $E[X_i \mid \Theta = \theta] = \mu(\theta)p_i$ with a measurable function $\mu(.)$ independent of i. From these assumptions it follows that:

$$\text{Cov}\,(X_i, X_j) = E[\text{Var}[X_i \mid \Theta]] + p_i p_j \text{Var}[\mu(\Theta)] \quad \text{for } i \leqslant j.$$

With (13) we conclude:

(15) $$(X_i)_{i\in N} \text{ is l.M.} \Leftrightarrow \frac{p_i}{E[\text{Var}[X_i \mid \Theta]]} \text{ is independent of } i.$$

In this case the l.M.-factor is (p_i).

It should be mentioned that (15) is fulfilled in the classical credibility model of Buhlmann and Straub.

3.3. *The Model of Shur*

SHUR (1972) considers the following model. The variables $Y_1, Y_2, \ldots$ have all the same expected value μ and the same variance σ^2, and the covariance structure is given by

(16) $$c_{i,k} = \rho^{|i-k|}\sigma^2 \text{ with } 0 \leqslant \rho \leqslant 1.$$

Hence the correlation between the total losses of two different periods decreases geometrically with the number of periods separating them. By inversion of the matrix

$$(c_{i,k})_{i,k=1}^{n} = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & . & \rho^{n-2} \\ \vdots & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & & \ldots & 1 \end{pmatrix} \sigma^2$$

and application of (3) one gets the credibility formula

$$P_{n+1} = \rho Y_n + (1-\rho)\mu. \tag{17}$$

Of course, this formula implies the linear Markov property of the process $(Y_i)_{i\in N}$. But this fact can already be detected directly from (16), because

$$c_{i,j} = \mathrm{Cov}(Y_i, Y_j) = \rho^{-i}\rho^j\sigma^2 \text{ for } i \leqslant j$$

fulfils the multiplicative decomposition criterion (10).

Following this way the inversion of the matrix $(c_{i,k})$ becomes unnecessary for the calculation of formula (17).

3.4. *A Model with Claim Numbers and Individual Claim Amounts*

Let us consider one risk unit with the following notations:

N_i is the cumulative number of claims up to the end of period i.

$Z^{(k)}$ is the amount of the kth individual claim. (It is assumed that these claims are numbered according to their order of occurrence.)

$X_i = \sum_{k=1}^{N_i} Z^{(k)}$ is the cumulative total of claims up to the end of period i.

Θ denotes a random risk parameter describing the unknown characteristics of the risk unit.

We make the following assumptions:

(A1) Given $[\Theta = \theta]$ the random variables $Z^{(1)}, Z^{(2)}, \ldots$ are i.i.d.

(A2) Given $[\Theta = \theta]$ the stochastic processes $(N_i)_{i\in N}$ and $(Z^{(k)})_{k\in N}$ are independent.

PROBLEM. Which are sufficient conditions such that the process $(X_i)_{i\in N}$ resp. the 2-dimensional process $((X_i, N_i)^T)_{i\in N}$ is l.M.? This would simplify the premium

$$P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_1, \ldots, X_n] \text{ resp.}$$

$$P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_1, \ldots, X_n, N_1, \ldots, N_n]$$

as usual, namely

$$P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_n] \text{ resp. } P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_n, N_n].$$

CONDITION I. Given $[\Theta = \theta]$ $(N_i)_{i\in N}$ is an inhomogeneous Poisson process, $\lambda_i g(\theta)$ being the Poisson parameter of N_i. Thereby $g(.)$ is a measurable function independent of i.

NOTE. It is not required in assumptions (A1)-(A2) and condition I that claim numbers and claim amounts are independent. They have only to be conditionally independent.

We get from assumptions (A1) and (A2) and Condition I: $(X_i)_{i\in N}$ is a process with conditional independent increments (given Θ), and

$$E[X_i \mid \Theta] = \lambda_i g(\Theta) E[Z^{(1)} \mid \Theta], \quad \mathrm{Var}[X_i \mid \Theta] = \lambda_i g(\Theta) E[(Z^{(1)})^2 \mid \Theta].$$

Therefore condition (15) is valid with $p_i = \lambda_i$ and the process $(X_i)_{i \in N}$ (and also the process $(N_i)_{i \in N}$) is l.M. with l.M.-factor (λ_i). It remains to prove the l.M.-property of the 2-dimensional process $((X_i, N_i)^T)_{i \in N}$. Because of Lemma 5 it suffices to check two 1-dimensional conditions (for fixed n):

(18) The sequence $X_1, \ldots, X_n, N_n$ is l.M.

(19) The sequence $N_1, \ldots, N_n, X_n$ is l.M.

It is true for $1 \leqslant i \leqslant n$ that:

$$\begin{aligned}
\mathrm{Cov}(X_i, N_n) &= \mathrm{Cov}(N_i, X_n) \\
&= E\left[\mathrm{Cov}\left(\sum_{k=1}^{N_i} Z^{(k)}, N_n \mid \Theta\right)\right] + \mathrm{Cov}\left(E\left[\sum_{k=1}^{N_i} Z^{(k)} \mid \Theta\right], E[N_n \mid \Theta]\right) \\
&\underset{(A1),(A2)}{=} E[E[Z^{(1)} \mid \Theta]\,\mathrm{Var}[N_i \mid \Theta]] + \lambda_i \lambda_n\, \mathrm{Cov}(E[Z^{(1)} \mid \Theta]\, g(\Theta), g(\Theta)) \\
&\underset{\text{cond I}}{=} \lambda_i \{E[E[Z^{(1)} \mid \Theta]\, g(\Theta)] + \lambda_n\, \mathrm{Cov}(E[Z^{(1)} \mid \Theta]\, g(\Theta), g(\Theta))\} \\
&= \lambda_i \times \text{term which is independent of } i.
\end{aligned}$$

Because each of the processes $(X_i)_{i \in N}$ and $(N_i)_{i \in N}$ has the l.M.-factor (λ_i) we get (18) and (19) by application of Lemma 3 (criterion (10)) and Lemma 4. The process $((X_i, N_i)^T)_{i \in N}$ is actually l.M.

Now we replace the Poisson assumption (condition I) by the hypothesis that the counting process $(N_i)_{i \in N}$ is l.M. and claim numbers and claim amounts are independent (level-2 assumption):

CONDITION II.

(A3) $(N_i)_{i \in N}$ is l.M. with l.M.-factor $(E[N_i])$.
(A4) $(N_i)_{i \in N}$ and Θ are independent.

REMARKS. (i) We have lost the convenient property that the increments of the process $(N_i)_{i \in N}$ resp. $(X_i)_{i \in N}$ are independent given Θ. Therefore it is not possible to apply the classical credibility model and condition (15) any longer.
(ii) Condition I implies (A3).
We need the further notation:

$$v_Z = E[\mathrm{Var}[Z^{(1)} \mid \Theta]] \quad \text{and} \quad w_Z = \mathrm{Var}[E[Z^{(1)} \mid \Theta]].$$

Then we get from the assumptions (A1)-(A4):

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= E[\mathrm{Cov}(X_i, X_j \mid \Theta, (N_k)_{k \in N})] \\
&\quad + \mathrm{Cov}(E[X_i \mid \Theta, (N_k)_{k \in N}], E[X_j \mid \Theta, (N_k)_{k \in N}])
\end{aligned}$$

$$= E[N_i \operatorname{Var}[Z^{(1)} \mid \Theta]] + \operatorname{Cov}(N_i E[Z^{(1)} \mid \Theta], N_j E[Z^{(1)} \mid \Theta])$$

(A1),(A2)

$$= E[N_i] v_Z + \operatorname{Cov}(N_i, N_j)\{w_Z + m_Z{}^2\} + E[N_i] E[N_j] w_Z$$

(A4)

$$= E[N_i] \times \text{factor which depends only on } j \text{ (and not on } i) \ (i \leqslant j).$$

(A3),(10)

Applying criterion (10) again we obtain that the process $(X_i)_{i \in N}$ is l.M. with l.M.-factor $(E[N_i])$. Furthermore it is true for $i \leqslant n$:

$$\operatorname{Cov}(X_i, N_n) = \operatorname{Cov}(N_i, X_n) = \operatorname{Cov}(N_i, E[X_n \mid \Theta, (N_k)_{k \in N}])$$

$$= \operatorname{Cov}(N_i, N_n E[Z^{(1)} \mid \Theta]) = \operatorname{Cov}(N_i, N_n) m_Z$$

(A1),(A2) (A4)

$$= E[N_i] \times \text{factor independent of } i.$$

Analogously with condition I the l.M.-property of the process $((X_i, N_i)^T)_{i \in N}$ follows.

4. THE LINEAR MARKOV PROPERTY IN SOME LOSS-RESERVING MODELS

The problem of estimating the ultimate loss reserve will not be presented with full rigour. Our only aim is to indicate the role of the l.M.-property in the most important loss-reserving models with credibility character.

The usual loss-reserving terminology is assumed to be known. Let Y_{ij} be the total of claims of accident year j which is reported during the development year i. Thereby we assume that each individual claim of accident year j is settled at its full amount immediately, i.e. there are no IBNER-claims resp. the IBNER-part is already contained in Y_{ij} as estimation.

The statistician considers each of the processes $(Y_{ij})_{i \in N}$ up to a certain time $n(j)$. For constituting the reserve he has to estimate the random variable

$$R_j = Y_{n(j)+1,j} + \ldots + Y_{\infty j}.$$

Because of the usual assumption of independent accident years it remains to evaluate

$$\hat{R}_j = \hat{E}[R_j \mid Y_{1j}, \ldots, Y_{n(j),j}] \text{ for each } j.$$

Modelling the development process $(Y_{ij})_{i \in N}$, different well-known experience rating models can be used.

4.1. *The Model of de Vylder*

DE VYLDER (1982) bases the development process on a special case of the (noncumulative) classical credibility model of Buhlmann–Straub. Therefore the covariance structure is contained in the model of Jewell. As described in Section

3.1 one may gain by linear transformation of the development process a process $(X_{ij})_{i\in N}$ which is I.M. That is, the reserve estimation becomes $\hat{R}_j = \hat{E}[R_j \mid X_{n(j),j}]$.

4.2. *The Model of Norberg*

NORBERG (1985) constructs a micro-model with claim numbers and individual claim amounts similar to the experience rating model in Section 3.4 with condition I. However, the distribution of the individual claim amounts may also depend on the reporting year. The resulting covariance structure of the development process becomes too complicated for calculating the theoretical credibility estimator up to an explicit formula. Therefore Norberg proposes numerical evaluation of the credibility estimator. In Norberg's model the cumulated claim number process is I.M. because of the Poisson assumption. This fact caused the present author to consider credibility estimators for the IBNR-claims in a distribution-free loss-reserving model where the Poisson assumption is replaced by the linear Markov property. This assumption is shown to be natural if the delay distribution does not depend on the hidden risk characteristics of the accident year (WITTING, 1986).

4.3. *The Model of Kramreiter and Straub*

Let us consider for fixed j the process $(X_{ij})_{i\in N}$ of the cumulative burning costs. KRAMREITER and STRAUB (1973) discuss the optimal unbiased homogeneous linear estimator of R_j with given statistical basis $X_{1j}, \ldots, X_{n(j),j}$ in a distribution-free model. "Optimal" means that the expected squared loss is minimized. This estimator exists and is uniquely determined. Kramreiter and Straub write the covariance structure in the form $\operatorname{Cov}(X_{ij}, X_{mj}) = c_{im}/p_j$, whereby p_j is a known volume measure of accident year j.

The most general covariance structure given by Kramreiter and Straub for which explicit calculation of the optimal homogeneous linear estimator remains possible is

$$c_{im} = c_i \prod_{k=i+1}^{m} \lambda_k \qquad \text{for } i \leqslant m,$$

where $(\lambda_i)_{i\in N}$ is a real sequence.

Because of criterion (9) in Lemma 3 this is exactly the linear Markov property of the process $(X_{ij})_{i\in N}$, which appears now as the actual assumption of the Kramreiter–Straub model.

General Remark on the Linear Markov Property

In the present paper we have only treated the case of a stochastic process ordered with respect to time. One may imagine the linear Markov property also with respect to other orders. An example for that is the recent paper of BUHLMANN and JEWELL (1986), who have used the linear Markov property for recursive calculation of the credibility estimator in a general hierarchical model.

REFERENCES

BUHLMANN, H and JEWELL, W S. (1986) Hierarchical credibility Preprint for the Summer School 1986 of the Association of Swiss Actuaries.

BUHLMANN, H and STRAUB, E. (1970) Glaubwurdigkeit fur Schadensatze *Bulletin of the Association of Swiss Actuaries* **70/1**, 111–133

DE VYLDER, F (1982) Estimation of IBNR-claims by credibility theory *Insurance Mathematics and Economics* **82/1**, 21–45.

DRYGAS, H (1983) Sufficiency and completeness in the general Gauss–Markov model *Sankhya* A **45**, 88–98

DRYGAS, H (1985) Linear sufficiency and some applications in multilinear estimation *Journal of Multivariate Analysis* **16**, 71–84.

FELLER, W. (1966) *An Introduction to Probability and its Applications*, Vol 2. Wiley, New York

JEWELL, W S (1975) Two classes of covariance matrices giving simple linear forecasts. *IIASA Research Memorandum* **75/17**

KRAMREITER, H and STRAUB, E. (1973) On the calculation of IBNR reserves II. *Bulletin of the Association of Swiss Actuaries* **73**, 177–190

NEUHAUS, W (1985) Choice of statistics in linear Bayes estimation. *Scandinavian Actuarial Journal* **85/1**, 1–26

NORBERG, R (1985) A contribution to modelling of IBNR claims. *Statistical Research Report of the University of Oslo* **85/4**

PAPOULIS, A (1965) *Probability, Random Variables and Stochastic Processes* McGraw-Hill, Tokyo

SHUR, W (1972) Discussion of Margolin's "Credibility of group insurance claim experience". *Transactions of the Society of Actuaries* **23**, 240–251

SUNDT, B (1979) Credibility and relative exchangeability *Scandinavian Actuarial Journal* **79**, 45–51

WITTING, T (1986) Uber Kredibilitatsschatzungen von reinen IBNR-Schaden Dissertation ETH Nr 8042

T. WITTING
ETH Zurich, Mathematikdepartement HG G 36.1, Ramistr. 101, CH-8092 Zurich.

WORKSHOP

THE SOLVENCY OF A GENERAL INSURANCE COMPANY IN TERMS OF EMERGING COSTS

BY C. D. DAYKIN, G. D. BERNSTEIN, S. M. COUTTS, E. R. F. DEVITT, G. B. HEY, D. I. W. REYNOLDS AND P. D. SMITH[1]

ABSTRACT

The authors challenge the traditional balance sheet concept of the solvency of a general insurance company and put forward an emerging costs concept, which enables the true nature of the assets and liabilities to be taken into account, including their essential variability. Simulation is suggested as a powerful tool for use in examining the financial strength of a company. A simulation model is then used to explore the resilience of a company's financial position to a variety of possible outcomes and to assess the probability that the assets will prove adequate to meet the liabilities with or without an assumption of continuing new business. This suggests the need for an appropriate asset margin assessed individually for each company. The implications for the management and supervision of general insurance companies are explored. The suggestion is made that the effectiveness of supervision based on the balance sheet and a crude solvency margin requirement is limited. More responsibility should be placed on an actuary or other suitably qualified professional individual to report on the overall financial strength of the company, both to management and to the supervisory authorities.

KEYWORDS

Solvency; financial strength; asset margin; emerging costs; simulation; professional report.

1. THE NATURE OF SOLVENCY

1.1. The financial position of a general insurance company is normally disclosed through annual accounts for shareholders and through returns to relevant supervisory authorities. Solvency is demonstrated by showing that the assets exceed the liabilities. To a large degree the bases are chosen by the company. For supervisory purposes it is not just a question of the assets exceeding the liabilities. The assets must normally exceed the liabilities by a specified margin.

1.2. In life assurance there is a report by the actuary on the valuation of the liabilities. By contrast the basis on which general insurance liabilities have been assessed is not usually stated. Furthermore, whereas in life assurance actuaries take account of the assets and effectively advise on the total financial strength of the company, there is no one with this role in a general insurance company. It is frequently the case that no specific account is taken of the suitability of the

[1] Solvency Working Party of the General Insurance Study Group of the Institute of Actuaries

assets to match the expected liabilities nor of the resilience of the balance sheet position disclosed to the inherent uncertainty in both assets and liabilities.

1.3. In principle, the balance sheet represents no more than the Directors' opinion about the financial position of the company. There is considerable uncertainty about the true amount of the liabilities and the realizable value of the assets. The auditors may place some restraints on how the Directors present the position but their role is largely confined to ensuring that what the Directors have done is reasonable.

1.4. There is in fact no single correct value that can be ascribed to either assets or liabilities. Different values may be appropriate according to one's perspective. Shareholders want a "true and fair" view, authorities want a cautious assessment of the position and tax authorities want as little as possible to be offset against taxable profits, to name but three interested parties. A balance sheet which shows a solvent position should reflect an expectation that the assets will be adequate, but it may, either deliberately or inadvertently, present a misleading picture. It certainly does not give any idea of the probability that the assets may prove to be inadequate to meet the liabilities.

1.5. In most countries a general insurance company is permitted by the supervisory authority to carry on writing business only if it has some specified excess of the value of the assets over the liabilities. This clearly increases the probability that the assets will prove sufficient to meet the liabilities, but most solvency margin requirements pay little or no attention to the differing degrees of uncertainty inherent in different types of business, nor do they distinguish adequately between the risks of running off the claims payments on the existing portfolio of business and the risks involved in continuing to write further business.

1.6. Reserving standards are frequently ill-defined or non-existent and do not require special provision to be made to cover the effects of changes in the value of assets on their adequacy to meet the liabilities. Problems may arise from some or all of the following:

adverse run-off of existing business;
poor underwriting experience;
failure to recover from reinsurers;
falls in asset values;
excessive expenses;
mismanagement, negligence or fraud.

A more extended description of the factors affecting solvency and a discussion of the interaction between solvency margin requirements and standards for technical provisions may be found in a paper by DAYKIN *et al.* (1984).

1.7. The object of a statutory solvency margin is two-fold. It reduces the probability that the assets will prove inadequate to meet the liabilities and it provides a buffer against further deterioration in a company's financial position which can occur in the period before its authorization to write new business can be withdrawn. The effect of a statutory minimum requirement is in practice also to set a somewhat higher formal standard in the market place.

1.8. A solvency margin is not required of other trading companies, but this can be said to reflect not only the nature of the business but also the extent of the insured's interest in the continued viability of the company. In many cases the insured can be exposed to quite serious liabilities in the event of the insurer failing to meet a claim. He cannot limit his liability in the way that he can with a trading company.

1.9. A company can carry on writing business only if the supervisory authority says that it meets the solvency requirements (cf. STEWART, 1971). The way in which they lay down requirements for this purpose will differ from the criteria which would be used by a Court in determining whether a company should be wound up. It is in fact relatively rare for insurance companies to be wound up by the Courts. It is more normal for the existing business to be run off to extinction or be transferred to another company. The latter procedure is more common in some countries than others.

2. BREAK UP OR GOING CONCERN?

2.1. The concern with safeguarding the position should the company cease trading is peculiarly the preserve of the supervisory authority. It arises because one of the main weapons available to the supervisory authority is the possibility of preventing a company from writing any further business. The supervisory authority will be subject to criticism if they stop a company from taking on any further business only when the position has been reached that the company cannot even meet its liabilities in respect of business already on its books. The consequence of this is that the supervisory authority will seek to close a company to new business when it can still be expected that the run-off of the existing liabilities will give rise to a surplus of assets, in other words the company is *de facto* solvent.

2.2. In order to achieve their objective of a "satisfactory" run-off, the supervisory authority is likely to take the view that outstanding claims provisions should be sufficient to enable all claims to be met with a reasonably high degree of probability. Failure to maintain an additional solvency margin over and above the outstanding claims provisions would not then imply that the company is unable to meet its existing liabilities, but that it does not have sufficient free resources to satisfy the supervisory authority that it should be permitted to continue writing business. Supervisory authorities using this approach are using what might be termed a "break-up" basis, i.e. it is assumed that no further business is written but existing business is run off to extinction.

2.3. In the EEC a two-stage solvency margin trigger has been adopted. The higher level is referred to as the required solvency margin and the lower is termed the guarantee fund. The origins of the EEC requirements have been described by DAYKIN (1984). If an insurer fails to maintain its required solvency margin it must provide to the supervisor a plan for the restoration of a sound financial position, which may include demonstration that on a properly drawn up business plan, and with realistic assumptions about profitability, the solvency margin will be restored within a reasonable short space of time. Only if the company fails to maintain the guarantee fund, set at one-third of the solvency margin, with a specified minimum in absolute terms, is immediate action to inject additional capital required in order to stave off withdrawal of authorization.

2.4. This approach is in practice not very different from one which assesses the company on a "going concern" basis. Theoretically, the main differences will be in relation to the provisions for outstanding claims, where a "going concern" basis might include less of a margin than a "break up" basis, and the provision for expenses, where a strict "break up" basis, would require a technical provision to be made to cover all the costs of running off the existing business. On a "going concern" basis these costs may be set against the continuing business of the company and it is probable that the past liabilities may be able to be run off for a lesser sum than on the break up basis. Yet another possible basis of assessment would be a "winding-up" basis, in which the assets are divided up and distributed on the basis of an estimate of the liabilities. This requires the assets to be realized at an early date, which in practice may have the effect of depressing market values. A summary of the main features of the different bases of assessment is given in Table 1.

2.5. A key objective of the management of an insurance company is to ensure that it does not have to cease trading. The accounts prepared for shareholders will reflect this by being prepared, as is the normal convention, on a going concern basis. Whilst continuing solvency is also a concern of the shareholders, in most cases this will be taken for granted, and the objective of the accounts should be to provide a true and fair view of the financial position of the company. For this purpose, technical provisions should not be overestimated or contain cautious margins and any adverse development of outstanding claims will emerge in due course and affect future profitability. In spite of the differences in the purposes for which the provisions are required, however, most companies adopt the same provisions for their accounts as they do for their statutory returns.

2.6. Whether seen from the viewpoint of the supervisory authority, from that of the shareholder or from that of an outside analyst, a common problem is the uncertainty as to the strength of the technical provisions. This might be helped

TABLE 1

COMPARISON OF ASSESSMENT BASES

	Assessment basis		
Assumption	Going concern	Break up	Winding-up
New business	Indefinite	None	None
Expenses	Claims settlement expenses only	All run-off expenses	All expenses of winding-up
Assets	Market or book values	Market value	Realization value
Liabilities	Best estimate	Cautious estimate	Best estimate of current value

by more clearly defined reserving standards and by more disclosure of the basis for the provisions, but there is still the fundamental weakness that the assets and the liabilities are not being valued on consistent bases and variability is not taken into account.
2.7. The division of a company's resources into technical provisions and "free" assets is not necessarily helpful from the point of view of establishing the true financial strength of the company. Most of the so-called surplus or "asset margin" (the excess of assets over liabilities) may in fact be needed to reduce the probability of being unable to meet the liabilities to an acceptably low level, particularly if the provisions are only "best estimates". To examine financial strength, all of the resources of the company need to be brought into consideration.

3 EMERGING COSTS

3.1. Although the current market value of the investments is increasingly coming to be disclosed in shareholders' accounts, at least in the UK, and is required by supervisory authorities for the statutory returns, its advantages are mainly in relation to its objectivity as a value to be placed on the company's investments, rather than in relation to its relevance to the ability of the company to meet its liabilities, even in the context of the break up basis.
3.2. The assets will not in practice have to be realized on a particular date and, in any case, by the time the accounts or returns have been prepared, the market value at the date to which those accounts relate is a matter of no more than historical interest. What is important is whether the proceeds of the assets, both capital and income, will prove sufficient to meet the liabilities as they emerge. This is true solvency.
3.3. The concept of projecting the emerging costs of the liabilities to which an enterprise is subject and placing them alongside the expected pattern of income is one which is familiar to actuaries in the life assurance and pensions contexts and is also fundamental to investment appraisal by economists in many other spheres of industry. However, little work seems to have been done on the application of the concept to general insurance companies.
3.4. There has been some theoretical consideration from the viewpoint of financial economics by KAHANE (1979) and KAHANE and BIGER (1977) which may not be widely known among actuaries. Actuarial concepts of looking at the company as a whole were applied to general insurance in a paper by BENJAMIN (1980) and the use of emerging costs was implicit in papers by RYAN (1980, 1984) on the use of simulation techniques in general insurance. COUTTS *et al.* (1984) set out more fully the fundamental concepts of the emerging costs of a general insurance company and a practical example was presented in a paper by DAYKIN and BERNSTEIN (1985) on run-off and asset risks.
3.5. The concept is a simple one. It involves analysing the inflows and outflows of actual cash in each successive year. The inflows may consist of some or all of

the following:

premium income;
interest and dividends on assets;
maturity proceeds of assets;
reinsurance recoveries in respect of claims.

The outflows may consist of the following:

claims settled or amounts paid on account;
reinsurance premiums;
expenses;
tax;
dividends.

3.6. The effect of the various items in each year will be either a net amount available for investment or a shortfall. In the latter case assets need to be sold to meet the deficit. So long as there are sufficient assets available to enable all the outflows to be met as they arise, the company is solvent in an absolute sense, whatever the balance sheet may have shown. If all the assets have been realized but net liabilities still remain, the situation is one of *de facto* insolvency.

3.7. An emerging costs analysis should be carried out on the totality of the assets and liabilities of the company. For this purpose the dividing line between technical provisions and asset margin is of no real importance although estimates of future claims payments are necessary. The uncertainties of general insurance are such that it will not generally be sufficient to use deterministic values for the liabilities and the assets. Some measures of variability need to be introduced. However, this should not be allowed to detract from the essential simplicity of the concept. It only means that some or all of the items listed above should be treated as random variables. To handle this the emerging costs can be examined using simulation.

3.8. A single simulation is one realization of a random process in which each of the required quantities is assigned a value. By examining a large number of simulations a picture can be obtained of the likely pattern of development resulting from the interaction of the various variables. Simulation permits the use of stochastic models for the investment processes and allows the uncertainty in the outstanding claims and in the profitability of new business to be taken into account. The approach has much in common with the ideas developed by the Finnish Solvency Working Party (PENTIKAINEN and RANTALA, 1982) and extended to cover run-off risk by PENTIKAINEN and RANTALA (1986), although they did not use a stochastic approach for the investments.

3.9. In practice the various elements may be modelled in a variety of different ways. For some purposes very complex models may be desirable; for others a simpler model may suffice, although any model which is going to give a reasonably realistic representation of the real world is bound to be fairly complex. The important principle is that the totality of the company's operations is being considered.

3.10. The procedure is very flexible. It might enable, for example, questions to be asked about the impact of alternative strategies for premium rating or investment and about the effect of possible adverse claims development or failure to recover from reinsurers. It provides a management tool and also seems to offer a way forward for more rational supervision. This would involve the submission to the authorities of a report on total financial strength by an actuary or other suitably qualified expert, as a supplement to minimum balance sheet requirements. The result would be a system much better able to take account of the true position of each company, having regard to the specific risks to which it is subject and the inherent uncertainties of both assets and liabilities.

4. THE SIMULATION MODEL

4.1. *General Structure*

4.1.1. In order to demonstrate the potential of the emerging costs approach we present here a model which provides a representation of the dynamics of a general insurance operation. In order to be reasonably realistic the model is quite complex but, however complicated the model, it is essential that the concepts should be capable of being put across in a straightforward way and the results must be capable of being presented in ways that can be directly related to management concerns such as corporate strategy and decision-making.

4.1.2. At its most basic, the model is a projection of cash flow, bringing together income from premiums and from assets and outgo in respect of expenses, tax, dividends and claims, determining the net balance for each year, investing or disinvesting as the case may be and proceeding similarly for as many years into the future as one wishes. It may be considered more fully in terms of three separate components:

liabilities arising from existing business;
future premiums and the liabilities resulting from the risks underwritten;
asset returns and asset value movements.

4.1.3. A mathematical formulation of the model is given in Appendices 1 and 2 and a description of the computer program in Appendix 3.

4.2. *Existing Liabilities*

4.2.1. The existing liabilities, as shown in the balance sheet, consist of estimates of outstanding claims, including IBNR, and unearned premium reserves (including any additional amount for unexpired risks). Unearned premiums can be dealt with along similar lines to new written premiums (see Section 4.3) since the uncertainty includes uncertainty about the adequacy of premium rates in relation to events which have not yet occurred.

4.2.2. As far as outstanding claims are concerned, there is uncertainty about the amounts of claims and about when they will be settled. The model needs to provide an adequate representation of this uncertainty. We make the simplifying assumption that the variability in rates of settlement can be subsumed into a variation in the amount of claims settled in each period.

4.2.3. The first stage is to estimate the expected claims payments in each successive year for each year of origin. In order to do this, fixed settlement patterns have to be specified in constant money terms. The model permits different run-off patterns to be assumed for different types of business. Inflation then has to be allowed for. Future inflation is generated by a stochastic model and this is combined with the expected settlements in constant money terms to give the expected development of claim amounts. The inflation model is an integral part of the models used for the assets (see Paragraph 4.4.4).

4.2.4. The variability of claim amounts payable in each period can be dealt with in a variety of ways. In an earlier paper describing the application of a similar model to a run-off situation, DAYKIN and BERNSTEIN (1985) proposed that the actual outstanding claims settled in each year in respect of each year of origin should be varied. They assumed that each separate entry in the run-off triangle was distributed about the mean estimate of claims settled at that particular duration for that year of origin in accordance with a log normal distribution. This was attractive as a means of simulating the interaction between different years of origin and different classes of business, but it resulted in a somewhat lengthy simulation process.

4.2.5. In order to simplify the model and allow account to be taken of different sizes of company, the model presented here uses an aggregate approach, whereby the amount that is varied is the total amount of claims settled in a particular period, for all years of origin combined. This aggregate figure is assumed to vary according to a normal distribution with a standard deviation of the type:

$$aX + b\sqrt{X}$$

where X is the mean estimate of total claim payments in the year and a and b are suitably chosen constants. We understand that a similar formula is used by the Finnish supervisory authority for their statutory minimum solvency margin (see Appendix 5 for discussion of this formula which can be considered to be an approximation to the formula derived by BUCHANAN and TAYLOR, 1986).

4.2.6. The amounts payable in future years in respect of risks arising from future written premiums and from unearned premium reserves are included with the amounts payable in respect of existing liabilities before applying the overall variability formula. The extent of the assumed variability can be adjusted by varying the constants a and b in the formula above. For a standard basis we have assumed that they take the values 0.15 and 75 respectively, with claims amounts being expressed in £ sterling. The amounts payable in successive years are assumed to vary independently of each other. The variability is intended to cover not only stochastic variability of claim amounts, but also uncertainty about the expected run-off model in constant money terms. Uncertainty about future inflation is dealt with separately.

4.2.7. Two typical run-off patterns have been assumed, characterized as short and long-tailed. Details are given in Appendix 3. In order to place a value on the technical provisions which would be established at the outset in respect of the outstanding claims, it has been assumed that inflation would be allowed for at 5% a year and that the resulting outstanding claims would not be discounted. (For further discussion on the interaction between the reserving basis and the solvency margin, see Paragraph 5.2.)

4.2.8 In practice an actual outstanding claims portfolio could be used as the basis for the input to the model in respect of existing liabilities. It would need to be expressed as an expected run-off in constant money terms. For illustrative purposes, however, we have assumed that the outstanding claims have been generated in a similar way to the liabilities in respect of future written premiums, by specifying a rate of real premium growth and claim ratios. For the purpose of generating the outstanding claims at the base date no variability was assumed in the historic claim ratios, in contrast to the process described in Section 4.3. In conjunction with the specified run-off patterns and the inflation model, the liabilities generated in this way give rise to estimates of outstanding claims payable in each future year in respect of each past year of origin.

4 3. *Future Written Premiums*

4.3.1. Future premiums are generated from an assumed initial premium level and an assumed real annual growth rate. The effects of inflation are then built in explicitly. Although the existing portfolio of business is generated by assuming a past pattern of premium growth, as described in Paragraph 4.2.8 above, a different growth rate assumption may be made for the future. The proportions of written premiums which are assumed to relate to different types of business can be specified. The written premiums are taken to be net of commission and initial expenses.

4.3.2. For each year for which additional premiums are assumed to be written, a ratio of claims to premiums net of commission and expenses is generated for each type of business. The ratio is assumed to be normally distributed with mean and standard deviation to be specified. The resulting ratio is applied to the assumed net written premiums to produce an initial estimate of total claims in respect of that business, without any allowance for future inflation or for discounting. This ratio is such that a value of 100% implies break-even if future investment income exactly balances inflation. The assumed proportions of claims settled in each future year are then applied to obtain uninflated estimates of expected claims payments. Future inflation, as generated by the model described below (Paragraph 4.4.5), is incorporated when the expected claim payments in terms of constant money have been aggregated with the corresponding estimates in respect of the existing liabilities. The combined estimates are then varied as described in Paragraph 4.2.5 above.

4.3.3. Since the claim ratios generated are ratios of claims to written premiums net of commission and expenses, no explicit allowance needs to be made for these items of outgo. Expenses of claims settlement are assumed to be included in the costs of settled claims.

4.3.4. This relatively simple approach has been used as a practical expedient in view of the complexity of the underlying risk process. An alternative approach, described by BEARD *et al.* (1984) and developed in the Report to the Finnish Solvency Working Party (PENTIKAINEN and RANTALA, 1982), would be to treat the basic claims process as a Poisson process and then build on a series of "structure variables" to take account of:

trends of claims frequency;
long term variations in premium rate adequacy;
year to year fluctuations in mean claims frequency.

Further assumptions then have to be made about the claims size distribution.

4.3.5. Whilst it is clearly possible to specify a model which takes explicit account of each of these, the added complexity can only be justified if the parameters of the model can be satisfactorily determined. We have not as yet been able to assemble data in a suitable form for calibrating such a model. The problem of calibration still arises with the simpler model, but it is intuitively more accessible and enables judgement to be applied in the area which is probably of the greatest importance, i.e. changes in the relationship between premium levels prevailing in the market and the underlying risk premium. This is the factor described as "long-term cycles" by PENTIKAINEN and RANTALA (1982).

4.3.6. Although the adequacy of premium rates does exhibit the characteristics of a business cycle, experience seems to show that the variation does not have a regular periodicity or a constant amplitude. A considerable degree of judgement is needed to decide where in the "cycle" the industry finds itself at any particular moment. Our model allows for the user to give explicit consideration to this and requires the mean claim ratio for the next couple of years to be estimated. If the model were to be used to examine the effects of future written premiums over a longer period than 2 years, further consideration would need to be given to modelling this component. The assumption of a normal distribution of claim ratios about the mean is a not unreasonable approximation, bearing in mind the large numbers of claims involved.

4.4. *Asset Variability*

4.4.1. The variability inherent in the asset portfolio of a company depends on the nature and distribution of the assets. The realizable value of many assets will vary from day to day as market conditions change. In our model, the initial distribution of the assets by category has to be specified and the various components of the asset distribution are then analysed separately, simulating the income generated and the capital value of each type of asset for each future year. Rules need to be specified for investment and disinvestment.

4.4.2. Three different types of asset are assumed: cash, irredeemable Government securities and ordinary shares. A more realistic model would replace investment in irredeemable Government securities with short, medium and long-dated securities. In practice, however, short-dated securities behave somewhat like cash and long-dated securities like irredeemables, so the model can be regarded as a tolerable proxy and gets round the problem of the reducing life of dated stocks over time. Allowance has not been made for a proportion of the assets being effectively non-interest-bearing (e.g. agents' balances) but this could easily be done.

4.4.3. The development of the various components of the asset distribution has been represented by a series of interrelated stochastic processes, suggested by WILKIE (1984, 1986) which generate future scenarios for the values of different types of asset and the income from them. Although Wilkie's models were not originally intended to be used for relatively short-term simulations such as those with which we are concerned, we have adopted them as a readily available and coherent model of asset movements and inflation. Further work is needed on suitable asset models and the sensitivity of the results to the particular models used. Our results show that this is a most important aspect of the whole simulation model.

4.4.4. The models are described in detail in Appendix 2. In addition to the models of asset returns and asset values, the Wilkie models include a model for inflation and this has been used where it is needed in the simulation of the liabilities.

4.4.5. The initial asset mix is based on assets covering the technical provisions and assets representing the asset margin. Different distributions may be specified for each. A variety of different investment and disinvestment strategies may be applied to the total funds.

4.5. *Results of the Simulations*

4.5.1. The number of potential combinations of variables is vast, even allowing each variable to take only three or four different values. We have limited our considerations by adopting a standard set of parameters and normally varying only one parameter at a time.

4.5.2. The simulation process involves sampling scenarios from an infinite set and the results are necessarily subject to statistical error. For any particular case which is of interest more simulations can be carried out in order to improve the accuracy of the estimate. In order to illustrate the results on a large number of scenarios, we have limited our considerations to 1000 simulations for each. For each parameter combination the same 1000 sets of random numbers are used, so that the comparisons are not significantly affected by any bias in the particular sets of random numbers chosen.

4.5.3. Figure 1 shows, for illustrative purposes, the results of 100 simulations, assuming no new business. This demonstrates the general shape of the results, which is common to all the scenarios, although the variability differs greatly. The

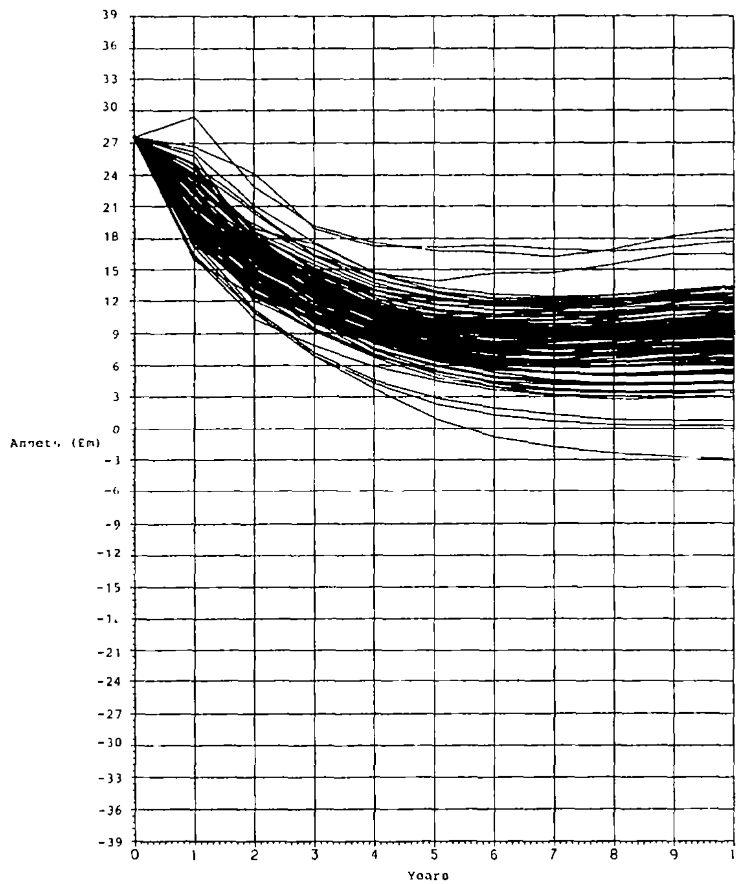


FIGURE 1 Run-off of assets assuming no new business (100 simulations)

graph shows the assets of the model company year by year throughout the run-off of the business.

4.5.4. When a line goes below the x-axis, this implies that all the assets have been exhausted on that particular simulation. If that should occur before the end of the run-off, true insolvency has occurred. In describing the results of the simulations, insolvency is used in this sense, without regard to the way in which the financial position of the company might be presented in the accounts or statutory returns at the base date or at any later date.

4.5.5. We thus define:

an insolvency occurs when the assets run out before all the liabilities have been met (on an emerging costs basis).

In the simulations a realization which runs into insolvency is allowed to continue

TABLE 2
DISTRIBUTION OF ASSETS AT END OF RUN-OFF FROM 1000 SIMULATIONS ON STANDARD BASIS

Remaining assets[1] as % of net written premiums[2]	Number of cases: Pure run-off	Number of cases: 2 years' new business
Less than 0	8	50
0–20	31	34
20–40	67	62
40–60	128	65
60–80	124	81
80–100	136	78
100–120	128	93
120–140	104	91
140–160	69	84
160 and over	205	362
Mean	112	144

1 Deflated to the date of assessment using the retail prices index
2 Premiums net of commission and initial expenses

by borrowing (at the rate of interest on cash plus a margin of 3%); this permits one to see how insolvent it becomes.

4.5.6. On the standard basis, insolvency in this sense occurred in 8 cases out of 1000 with no new business and 50 cases with 2 years' new business. The distribution of assets at the end of the run-off, deflated to the date of assessment using the retail prices index, is summarized in Table 2. The written premiums in question are those in the year before the base date. It should be recalled that the written premiums are net of commission and expenses. Results expressed as percentages of net written premiums can be rated down (say, by applying a factor of 75% or 80%, depending on the type of business) to obtain comparable results in terms of gross written premiums. The mean level of remaining assets for the 1000 simulations was 112% of net written premiums with no new business and 144% with 2 years' new business, with standard deviation of 70% and 109% respectively of net written premiums.

4.5.7. Full details of the assumptions underlying the standard basis are given in Appendix 4. However, we will return to the results after commenting on the application of the model.

4.6. *Application of the Model*

4.6.1. A simulation model of the insurance company, based on the emerging costs concept, provides a powerful and flexible tool for examining the dynamics of an insurer's operation, for exploring the effects of uncertainty and for developing the financial aspects of corporate strategy within a logical framework. This should be of value both to management and to the supervisory authorities.

Crucial to this process would be the presence of a suitably qualified actuary or other expert within the company, or acting as consultant, who could develop a suitable model and apply the necessary judgement to the use of the model in the circumstances of the particular company. The responsible expert would report to management on the financial strength of the company, taking all relevant factors into account.

4.6.2. The simulation approach would also enable the actuary to advise management on the potential effects of different new business and investment strategies, the risks involved and the return on capital which might be expected if additional capital is injected to enable a particular strategy to be adopted.

4.6.3. A report on the financial strength of the company could accompany the statutory returns to the supervisory authorities. The actuary would be answerable to the supervisor on the details of this report. One could envisage this leading to an informed dialogue between the supervisory authority and the company under scrutiny on the nature of the proposed corporate strategy, whether in relation to investment policy, growth or premium levels. The supervisor could then ask for an assessment of the effect of alternative strategies and seek agreement with the company on appropriate changes to its strategy as a condition for being permitted to continue writing business.

5. SOLVENCY CONSIDERATIONS

5.1. The results of the simulations can be presented in terms of numbers of insolvencies out of a given number of simulations. This is an estimate of the probability of ruin. Each result derives from an assumption about the excess of assets over technical provisions (the "asset margin") and a specified basis for calculating the latter. Given a basis for the technical provisions, the process can be used to derive the required initial asset margin in order to achieve a specified probability of ruin in a particular case.

5.2. The required asset margin will clearly differ according to differing definitions of the technical provisions. Table 3 illustrates this point. The table shows the technical provisions on the standard basis described above and the technical provisons on alternative bases as to inflation and discounting, but for the same set of outstanding claims. The table shows what asset margins would be necessary, expressed both as a percentage of technical provisions and as a percentage of net written premiums, in order to achieve the same degree of overall security as the technical provisions on the standard basis. Technical provisions on the standard basis are calculated assuming 5% inflation and no discounting. Thus if the reserves do not allow for any future inflation, or have been discounted using a rate of interest equal to the assumed rate of inflation, an asset margin of 21% of net written premiums or 9% of technical provisions would be needed to produce the same level of total assets as the technical provisions alone on the standard basis. The figures in this table underline the arbitrary nature of a statutory solvency requirement unless standards of technical provisions can be adequately specified.

TABLE 3
TECHNICAL PROVISIONS AND ASSET MARGINS

Reserving basis (net inflation assumed) %	Technical provisions[1]	Asset margin to achieve same security as standard: % of net written premiums[2]	% of technical provisions
−5	20450	39	19
0	22264	21	9
5	24364	0	0
10	26805	−24	−9
15	29656	−53	−18

1 Based on 40% long tail business and 60% short tail
2 Premiums net of commission and initial expenses

5.3. First we give some results for a pure run-off, i.e. with no future premiums assumed to be written. The outstanding claims and unexpired risks are allowed to emerge and the adequacy of the total assets (technical provisions and asset margin) is examined. Table 4 shows the number of insolvencies and the mean assets remaining at the end of the run-off and the standard deviation of the assets remaining on some alternative bases. Table 5 gives a similar set of results with the inclusion of 2 further years' written premiums. Appendix 4 gives details of all the assumptions and a full set of results.

5.4. Tables 6 and 7 show the asset margins required to achieve a probability of ruin of 1 in 100 for each of the combinations of assumptions in Tables 4 and 5 respectively, assuming that the technical provisions are established on the standard basis of 5% inflation and no discounting. The asset margins are shown in terms of both net written premiums in the year before the base date and as a percentage of technical provisions at the base date. The results can be expressed in terms of net written premiums even for the pure run-off case, since these are the premiums in the year before the base date when premiums are assumed to cease. As described in Paragraph 4.2.8, we have in fact generated the outstanding claims from past premiums. The difference between Tables 6 and 7 provides a measure of the additional capital needed in order to go on writing business for two more years.

5.5. It is clear that the results obtained depend critically on the models used and the parameters assumed. More work is needed on a number of different aspects. However, the results presented do appear consistent and sensible and variations in relation to changing parameter values conform with general reasoning.

5.6. It is difficult from these results to draw conclusions about an appropriate level of a minimum statutory solvency margin. In fact we have avoided using the term solvency margin in this section because of its special significance in statutory terms and have referred to the necessary margin as the asset margin. Our asset margins relate to particular assumptions about the basis for the technical provisions and provide a defined degree of security in relation to specified scenarios

TABLE 4

SUMMARY OF RESULTS FOR PURE RUN-OFF OF BUSINESS (WITH 1000 SIMULATIONS)

Assumptions	No of insolvencies	Mean assets remaining[1] (%)	Standard deviation of assets remaining[1] (%)
Standard basis	8	112	70
1 Net written premiums[2]			
(a) £1m a year	20	113	75
(b) £10m a year (s)	8	112	70
(c) £100m a year	6	112	69
2 Proportion of long-tailed business			
(a) 20% of net written premiums[2]	3	94	55
(b) 40% of net written premiums[2] (s)	8	112	70
(c) 60% of net written premiums[2]	13	130	85
3 Initial asset distribution: Cash / Gilts / Equities			
(a) TP + AM / — / —	3	95	53
(b) — / TP + AM / —	20	120	98
(c) — / — / TP + AM	49	136	115
(d) $\frac{1}{2}$TP / $\frac{1}{2}$TP / AM (s)	8	112	70
4 Initial asset margin.			
(a) 0% of net written premiums[2]	134	52	55
(b) 20% of net written premiums[2]	36	83	62
(c) 40% of net written premiums[2] (s)	8	112	70
(d) 60% of net written premiums[2]	2	147	80
(e) 80% of net written premiums[2]	0	180	90
5 Asset selling rules			
(a) Equities, gilts; cash	9	102	66
(b) Cash; gilts; equities	7	123	79
(c) In proportion to holdings (s)	8	112	70
(d) Sell best performer first	14	108	70

1 Deflated to the date of assessment using the retail prices index and expressed as a percentage of net written premiums[2] in the year before the date of assessment (see Appendix 3 6 8)

2 Premiums net of commission and expenses

(s) indicates the assumption made for the standard basis

on the basis of our model. A statutory solvency margin, in the sense in which it is usually used, provides a general level of security, independent of the particular circumstances of the company, against all possible future scenarios, including the effect of unquantifiable risks such as fraud, mismanagement and the failure of reinsurers.

5.7. A starting point for consideration of an appropriate level of statutory solvency margin might be to look at the asset margin for a company with a fairly standard distribution of business, a moderate growth rate and investment entirely in cash. In our view the resulting margin ought to be in two parts:

a percentage of the technical provisions at the assessment date;
a percentage of written premiums.

The former represents the margin required in respect of the run-off risks and the latter the margin required in respect of writing up to two years' further new

TABLE 5

SUMMARY OF RESULTS WITH 2 FURTHER YEARS' BUSINESS (WITH 1000 SIMULATIONS)

Assumptions	No of insolvencies	Mean assets remaining[1] (%)	Standard deviation of assets remaining[1] (%)
Standard basis	50	144	109
1 Net written premiums[2]			
(a) £1m a year	61	144	117
(b) £10m a year (s)	50	144	109
(c) £100m a year	43	144	107
2 Proportion of long-tailed business			
(a) 20% of net written premiums[2]	48	120	91
(b) 40% of net written premiums[2] (s)	50	144	109
(c) 60% of net written premiums[2]	52	168	130
3 Real growth rate (past and future)			
(a) −20% a year	53	171	134
(b) No growth (s)	50	144	109
(c) +50% a year	83	144	121
4 Mean claim ratio[3] (short-tailed)			
(a) 80% of net written premiums[2]	7	187	117
(b) 100% of net written premiums[2] (s)	50	144	109
(c) 125% of net written premiums[2]	165	90	103
5 Variability of claim ratio (short-tailed)			
(a) Standard deviation 5% NWP[2]	49	144	109
(b) Standard deviation 10% NWP[2] (s)	50	144	109
(c) Standard deviation 15% NWP[2]	48	144	111
6 Mean claim ratio[3] (long-tailed)			
(a) 80% of net written premiums[2]	17	163	106
(b) 100% of net written premiums[2] (s)	50	144	109
(c) 125% of net written premiums[2]	105	120	114
7 Variability of claim ratio (long-tailed).			
(a) Standard deviation 10% NWP[2]	50	144	109
(b) Standard deviation 15% NWP[2] (s)	50	144	109
(c) Standard deviation 20% NWP[2]	49	144	110
8 Initial asset distribution: Cash / Gilts / Equities			
(a) Cash TP + AM; Gilts —; Equities —	46	118	86
(b) Cash —; Gilts TP + AM; Equities —	79	155	152
(c) Cash —; Gilts —; Equities TP + AM	86	181	172
(d) Cash ½TP; Gilts ½TP; Equities AM (s)	50	144	109
9 Initial asset margin			
(a) 0% of net written premiums[2]	196	74	98
(b) 40% of net written premiums[2] (s)	50	144	109
(c) 80% of net written premiums[2]	11	216	133

1 Deflated to the date of assessment using the retail prices index and expressed as a percentage of net written premiums[2] in the year before the date of assessment (see Appendix 3 6 8)

2 Premiums net of commission and expenses

3 Ratio of claims (including claims settlement expenses), without allowance for future inflation or for discounting, to premiums net of commission and expenses (see Paragraph 4 3 2).

(s) indicates the assumption made for the standard basis

TABLE 6

ASSET MARGINS REQUIRED TO ACHIEVE 1/100 PROBABILITY OF RUIN — NO FUTURE NEW BUSINESS

Assumptions				Asset margin as % of NWP[1]	Asset margin as % of technical provisions
Standard basis				40	15
1	Net written premiums [1]				
	(a) £1m a year			55	25
	(b) £10m a year (s)			40	15
	(c) £100m a year			35	15
2	Proportion of long-tailed business·				
	(a) 20% of net written premiums[1]			30	15
	(b) 40% of net written premiums[1] (s)			40	15
	(c) 60% of net written premiums[1]			45	15
3	Initial asset distribution				
	Cash	Gilts	Equities		
	(a) TP + AM	—	—	30	10
	(b) —	TP + AM	—	60	25
	(c) —	—	TP + AM	80	35
	(d) $\frac{1}{2}$TP	$\frac{1}{2}$TP	AM (s)	40	15
4	Asset selling rules				
	(a) Equities; gilts, cash			40	15
	(b) Cash, gilts, equities			35	15
	(c) In proportion to holdings (s)			40	15
	(d) Sell best performer first			55	25

1 Premiums net of commission and expenses

(s) indicates the assumption made for the standard basis

business. To the new business margin might be added a contingency loading to cover other unquantifiable risks.

5.8. This would provide a basic safety net for an average company, assuming that technical provisions were at least up to the standard envisaged. Statutory reserving standards might be necessary to achieve this, since it has to be recognized that a solvency margin requirement based on technical provisions has a similar weakness to one based on written premiums. If the provisions are understated the requirement is reduced, whereas it should in fact be higher.

5.9. Alongside such a basic solvency requirement would be a requirement for a report by an actuary or other expert on the overall financial strength of the company. This would transcend the arbitrary dividing line between technical provisions and solvency margin and would take specific account of the nature of the business written by the company, the proportions of different types of business, the assets held, and all other relevant factors, including the nature of and the security of the reinsurance programme.

5.10. If a requirement for an actuarial report is not introduced, then further consideration would need to be given to whether the solvency margin requirement should include components relating to the assets held and the reinsurance recoveries expected. Regard should also be had to the nature of the outstanding claims portfolio and the type of business being written. However, such a solution would be far from ideal.

TABLE 7

ASSET MARGINS REQUIRED TO ACHIEVE 1/100 PROBABILITY OF RUIN — TWO YEARS' NEW BUSINESS

Assumptions	Asset margin as % of NWP[1]	Excess asset margin as compared to run-off (as % of NWP[1])
Standard basis	90	50
1 Net written premiums[1]		
(a) £1 a year	100	45
(b) £10m a year (s)	90	50
(c) £100m a year	80	45
2 Proportion of long-tailed business		
(a) 20% of net written premiums[1]	80	50
(b) 40% of net written premiums[1] (s)	90	50
(c) 60% of net written premiums[1]	95	50
3 Real growth rate (past and future).		
(a) −20% a year	100	35
(b) No growth (s)	90	50
(c) +50% a year	115	85
4 Mean claim ratio[2] (short-tailed)		
(a) 80% of net written premiums[1]	30	0
(b) 100% of net written premiums[1] (s)	90	50
(c) 125% of net written premiums[1]	125	75
5 Variability of claim ratio (short-tailed)		
(a) Standard deviation 5% NWP[1]	80	45
(b) Standard deviation 10% NWP[1] (s)	90	50
(c) Standard deviation 15% NWP[1]	85	45
6 Mean claim ratio[2] (long-tailed)		
(a) 80% of net written premiums[1]	50	20
(b) 100% of net written premiums[1] (s)	90	50
(c) 125% of net written premiums[1]	115	60
7 Variability of claim ratio (long-tailed):		
(a) Standard deviation 10% NWP[1]	85	50
(b) Standard deviation 15% NWP[1] (s)	90	50
(c) Standard deviation 20% NWP[1]	90	50
8 Initial asset distribution: Cash / Gilts / Equities		
(a) TP + AM / — / —	85	55
(b) — / TP + AM / —	110	50
(c) — / — / TP + AM	135	55
(d) ½TP / ½TP / AM (s)	90	50

1 Premiums net of commission and expenses

2 Ratio of claims (including claims settlement expenses), without allowance for future inflation or for discounting, to premiums net of commission and expenses (see Paragraph 4 3 2)

(s) indicates the assumption made for the standard basis

6. REINSURANCE

6.1. Reinsurance business accepted may be regarded as another class of business, which is often particularly volatile and unpredictable. Appropriate reserving levels for casualty reinsurance business are likely to present particular problems, since it can take many years for the liabilities (including IBNR) to develop fully. Solvency margins certainly ought to have regard to this uncertainty. In principle there seems no reason why the simulation approach should not also provide some insights in this area of an insurers' portfolio.

6.2. Much more difficult to handle in the context of the assessment of financial strength is the security of reinsurance cessions. Many insurers are critically dependent on their ability to recover from reinsurers, since the size of the risks they write is such as to bankrupt or cripple them if they had to bear the liability alone. One safeguard against reinsurance failure is to spread reinsurance cessions widely, so that there is not any great dependence on particular reinsurers. However, this does not remove the need to look carefully at the security of individual reinsurers chosen for the programme.

6.3. From the reserving point of view, a decision has to be made on the extent to which reinsurance recoveries can be relied on. Extreme caution might point towards reserving for the full gross liability but this is not a practical commercial possibility in most cases. Clearly recoveries from reinsurance companies already known to be in trouble should be ignored or heavily discounted, but it is more difficult to know what should be done when there are no specific known problems. In accounting terms it may be difficult to set up a provision against an unseen and unquantifiable possibility of reinsurance failure. On the other hand the accountancy concept of prudence would preclude taking credit in advance for receipts which are uncertain, so it would be possible to justify taking only partial credit for reinsurance recoveries, depending on an assessment of the viability of the reinsurers.

6.4. The issue is of particular importance in considering the overall financial strength of the company. This would be one aspect which the actuary would need to cover in his report. Different approaches may be acceptable in different circumstances but simulation does seem to offer a promising way forward. Further work is clearly needed in this area to develop ways of modelling reinsurance recoveries. It has been assumed in our model that all claims are net of reinsurance. This may be good enough for many companies, with relatively little dependence on reinsurance. However, it will be far from adequate for other companies for which the possibility of failure to recover from reinsurers is a significant one and the potential impact disastrous in solvency terms. Some tentative ideas of a possible way of tackling this are set out in Appendix 6.

6.5. A detailed examination of the reinsurance programme can hardly be practicable for the supervisory authorities and here again it seems that an actuary's report would help. No general solvency requirement can be a substitute for this. The practice adopted for the EEC solvency margins of reducing the solvency margin requirement calculated on the basis of gross written premiums to allow for reinsurance based on actual recoveries in the past three years, but with a maximum reduction of 50%, is a very rough and ready solution and does not have any regard to the actual dependence on reinsurers for future recoveries. With non-proportional reinsurance the premium can be very small in relation to the potential liability, so no simple percentage of premium is likely to make sense as a solvency margin. A percentage of anticipated recoveries from reinsurers would have a stronger rationale, but it would be difficult to find a logical basis for any particular percentage.

7. CONCLUSIONS

7.1. We have outlined the weaknesses in the traditional balance sheet concept for describing the true financial strength of a general insurance company. Assets and liabilities should not be treated as independent aspects and much more attention needs to be focused on the uncertainties and on the company's resilience in the face of such uncertainties. Appropriate techniques have been developed by actuaries for dealing with these problems in the life and pensions areas and similar principles can be used to begin to tackle the general insurance problem. The parallels are drawn out in a paper by COUTTS and DEVITT (1986).

7.2. However, there are also differences, arising mainly from the greater volatility of claim amounts in general insurance. The problem of variability can be explored by means of simulation. A simulation model of a general insurance company provides a powerful tool for analysing the impact of all types of uncertainty and assessing the true financial strength of the company.

7.3. A solvency margin requirement expressed in terms of a simple percentage of written premiums (or in terms of a percentage of technical provisions, which might be more appropriate to cover the run-off risk) cannot have proper regard to the risk to which each company is subject, whether as regards the assets or liabilities. It must, therefore, be seen as a general underlying safety net, providing a margin against the effects not only of stochastic variations but also of mismanagement, fraud or simply error, and permitting the statutory authority to operate a satisfactory control system.

7.4. Despite our strong belief that the solvency margin should relate to the various risks affecting the financial position of an insurance company, we acknowledge that there will be interest in the use of our model to provide a rationale for a minimum statutory solvency margin. Time has so far prevented us from carrying out sufficient simulations to explore the full implications of the assumptions made and, in particular, the response of the Wilkie model to changes in parameters. We have also only shown the results for a probability of ruin of 1% and this level is of course crucial to the resulting asset margins.

7.5. Nevertheless at this level of security Table 6 shows that, for a moderate sized company, writing £100m of net premiums but otherwise on our standard basis, the margin necessary to cover the run-off risks would be 15% of technical provisons, assuming that the provisions for outstanding claims are set up on an undiscounted basis with allowance for inflation at 5% (the mean value used in the Wilkie model). The margin might be reduced to 10% if all the investments are assumed to be held in cash. Such a margin may be over-stringent as a minimum for larger companies but Table 6 indicates that it should be higher for small companies. A similar standard to a 10% margin would be obtained for the mix of business considered here by setting up provisions for outstanding claims allowing for inflation at 10% with no discounting.

7.6. Care has to be taken in interpreting the extra margin implied to be necessary to allow for the risks contingent on writing new business for two years. The margins to cover the run-off risk have been expressed for this purpose in terms of net written premium and these margins (for the respective sets of assumptions) have then been subtracted from the margins obtained assuming two further years in business. It could be argued that if the risks of new business and run-off are to be provided for independently, then the model should be run with no past business in order to assess the appropriate margin for new business risks. We have not done this as we do not believe that the two issues are independent, there being interactions in regard to both assets and the variability of the run-off of claims. Assuming that the margins expressed as a percentage of net written premiums are additive, Table 7 indicates a margin of a 50% premium net of commission and expenses for a £100m company, otherwise on our standard basis apart from investment being entirely in cash. This might be equivalent to 35–40% of actual gross written premiums.

7.7. Such a solvency margin requirement appears rather high and it is worth considering briefly some of the major factors which give rise to it. A significant part arises from the effect of simulated future inflation and the possibility that returns on cash will not be adequate to compensate for it. This suggests that the risks might be reduced with greater use of index-linked stocks.

7.8. Much of it also arises from the assumption on the standard basis of a mean claim ratio of 100% of net written premiums. As described in Paragraph 4.3.2, this implies break-even if future investment income exactly balances inflation. Thus the assumption is that business is written on a basis where the only profit on an expected value basis is to the extent that a positive real rate of return can be obtained. This might be perceived as too stringent for a minimum solvency margin requirement, although it is not unrealistic in current conditions. The requirement could be reduced by about 1% of actual written premiums for every percentage point by which the expected claim ratios are reduced below 100%.

7.9. Any general solvency requirement will have its limitations. Apart from the points mentioned in Paragraph 7.3, there is also the problem of relating the requirement to written premiums or to technical provisions, which may themselves be more adequate for some companies than for others. The adequacy of the technical provisions is of particular importance (cf. Paragraph 5.2), since they determine what assets are apparently available as a margin. There is, therefore, a need for consistent standards to be applied in setting technical provisions, suggesting that there would be considerable advantages in requiring the provisions to be established on the basis of advice from an actuary or other claims reserving expert, acting within the framework of an appropriate professional standard. However, it has to be acknowledged that there is always likely to be some uncertainty about the strength of technical provisions.

7.10. We have also argued that a crude minimum solvency margin requirement cannot adequately have regard to the true level of risk for a particular company.

The supervisory authority is not well-placed to assess each company's risk situation in detail on an individual basis and the answer would seem to be to rely on an appointed actuary or other similarly qualified person within the company (or acting as a consultant to the company). The actuary would be responsible for reporting both to management and to the supervisory authority on the financial strength of the company, taking all relevant factors into account. A summary of the actuary's report could appear in the statutory returns, with full details being available to the supervisory authority on request. The supervisory authority would be able to question the actuary on the effects of alternative assumptions and could then discuss with management an appropriate strategy for reducing the risk profile to an acceptable level.

7.11. The actuary would need to use simulation techniques in performing his duties. There is plenty of scope for developing appropriate simulation models for this task and one such model is presented here as an example of what can be done. Apart from providing a framework for analysing the existing position of the company, such models could be powerful tools for answering a wide variey of "what if?" questions, such as:

what changes do there need to be to premium rates to make a particular line of business worth writing?
is the investment strategy too risky with the present asset margin?
what additional capital would be needed to pursue a particular strategy?
will the strategy give a reasonable expected return on the additional capital?

ACKNOWLEDGEMENTS

8.1. The Working Party wishes to acknowledge a debt of gratitude to the members of the Finnish Solvency Working Party for starting them off along the lines of this paper and, in particular, to Professor Teivo Pentikainen for his advice and encouragement. Our thanks are also due to many UK actuaries who commented on the work at different stages.

APPENDIX 1
DESCRIPTION OF SIMULATION MODEL OF GENERAL INSURANCE COMPANY

A1.1. In standard risk theory the year to year transition formula is of the form:

$$\Delta U = B + I - X - C - T$$

where

ΔU is the change in the solvency margin U;
B is the earned premium income, including safety and expense loadings;
I is the net income from investments;
X is incurred claims;
C is the cost of administration, reinsurance etc.;
T is dividends, tax, etc.

By implication, incurred claims includes changes to estimates of outstanding claims generated in previous years and included in the technical provisions at the start of the year in question. This formulation is also deficient in that changes in the values of investments are ignored.

A1.2. *General Formula*

More generally, we define:

$$\begin{aligned}\Delta A(j) &= A(j) - A(j-1)\\ &= \sum_k A_k(j) - \sum_k A_k(j-1)\\ &= \sum_k A_k(j-1)[\{1 + y_k(j-1)\}\{1 + g_k(j-1)\} - 1]\\ &\quad + \left\{B(j) - C(j) - T(j) - \sum_{i \leqslant j} x(i;j)\right\}\end{aligned}$$

where

$A(j)$ is the total value of the assets at the end of year j; $A_k(j)$ is the total value of component k of the asset portfolio at the end of year j (in our model $k = 1$ for cash, 2 for irredeemable government securities, 3 for ordinary shares);

$y_k(j)$ is the yield on asset component k at the end of the year j. In particular, in our model:

$y_1(j) = c(j) - 0.01$
$y_2(j) = c(j)$
$y_3(j) = y(j)$
where $c(j)$ is the yield on 2.5% Consols;

$y(j)$ is the dividend yield on the Financial Times Actuaries All-Share Index;

$g_k(j)$ is the proportionate change in capital values between the end of years j and $(j+1)$. In particular, in our model:

$$g_1(j) = 0$$

$$g_2(j) = \frac{c(j)}{c(j+1)}$$

$$g_3(j) = \frac{d(j+1)y(j)}{d(j)y(j+1)}$$

where $d(j)$ is an index of share dividends (= dividend yield × price index) corresponding to $y(j)$;

$B(j)$ is the *written* premium income in year j including safety and expense loadings;

$C(j)$ is the cost in year j of administration, commission, reinsurance etc.;

$T(j)$ is the amount paid out in dividends and tax in year j;

$X(i;j)$ is the amount settled in year j in respect of claims arising in year i.

We now define $B'(j)(=B(j)-C(j))$ as the written premiums in year j net of commission and all expenses other than claims settlement expenses and $X(i;j)$ as including claims settlement expenses.

A1.3. *Asset and Inflation Models*

The asset components $A_k(j)$ can be defined in a variety of ways relative to the total $\sum_k A_k(j)$. For example, if investment or disinvestment is proportional to the value of assets brought forward to the end of the year from the previous year-end,

$$A_k(j)=\frac{A(j)}{A(j-1)}A_k(j-1).$$

If proportions $p_k(\sum_k p_k=1)$ are specified such that p_k of any new investment is invested in component k:

$$A_k(j)=A_k(j-1)\{1+y_k(j-1)\}\{1+g_k(j-1)\}$$
$$+p_k\left\{B(j)-C(j)-T(j)-\sum_{i\leqslant j}X(i;j)\right\}.$$

We also define $q(j)$ as the retail price index at the end of year j and $r(j)$ as the price growth in year j:

$$r(j)=\frac{q(j)}{q(j-1)}-1.$$

The variables $q(j)$, $d(j)$, $y(j)$ and $c(j)$ are defined by an interrelated set of autoregressive models, described in detail in Appendix 2.

A1.4. *Tax and Dividends Model*

The dividends and tax term is expressed in terms of the investment income and an input parameter t, representing the proportion of investment income absorbed by tax and dividends paid to shareholders, by the following:

$$T(j)=t\sum_k A_k(j-1)y_k(j-1).$$

A1.5. *Model of Claims Generation Process*

We define written premiums in the year prior to the date of assessment (taken as the time $j = 0$) as $B(0)$ and the rate of growth of written premiums before and after that date as e_1 and e_2. Then:

$$\begin{aligned} B'(j) &= B'(0)(1+e_1)^j \quad (j<0) \\ &= B'(0)(1+e_2)^j \quad (j \geqslant 0) \end{aligned}$$

and

$$B_k'(j) = f_k B'(j) \text{ for } k = 1, 2, 3$$

where f_k is the proportion of written premiums in respect of type of business k ($k = 1$ for short-tailed, 2 for long-tailed and 3 for very long-tailed).

Claims are assumed to be generated from written premiums by means of a variable claims ratio and specified proportions settled in each year of the run-off. Thus the estimated payment in year j in respect of premiums written in year i is given by:

$$X(i; j) = \sum_k s^k(j) R_k(i) B_k'(i) \prod_{l=i+1}^{j} \{1 + r(l)\}$$

where

> $R_k(i)$ is the uninflated, undiscounted claims ratio in year i assumed to be normally distributed with mean R_k and standard deviation σ_r^k. For $i \leqslant 0$, $R_k(i) = R_k$.
>
> $s^k(j)$ is the proportion of uninflated, undiscounted claims from type of business k that are assumed to be settled in development year j.

A1.6. *Model of Claims Settlement*

Claims settled in each year of development are aggregated from all the separate years of origin, whether before or after the date of assessment. The total amount of claims settled in year j, $X(j)$, is assumed to be normally distributed with mean $\bar{X}(j)$ and standard deviaton $a\bar{X}(j) + b\sqrt{\bar{X}(j)}$ where a and b are specified constants and $\bar{X}(j)$ is defined as:

$$\bar{X}(j) = \sum_{i \leqslant j} \sum_k s^k(j-i) R_k(i) B_k'(i) \prod_{l=i+1}^{j} \{1 + r(l)\}.$$

A1.7. *Technical Reserves*

The technical reserves, TR(0), at the date of assessment are calculated from the estimates of claims to be settled in future years arising from premiums earned prior to the date in question. They allow for inflation at a specified rate, r, and discounting at a specified rate, d. They can be expressed as follows:

$$\mathrm{TR}(0) = \tfrac{1}{2} \sum_k B_k'(0) R_k \sum_{j=0}^{20} s^k(j)(1+r)^j(1+d)^{-j}$$

$$+ \sum_k R_k \sum_{i=-20}^{-1} B_k'(i) \sum_{j=-i}^{20} s^k(j)(1+r)^j(1+d)^{-j} + \tfrac{1}{2} B_k'(0).$$

The initial solvency margin, SM(0), is defined as a function of written premiums in the year before the date of assessment:

$$\mathrm{SM}(0) = \alpha B'(0).$$

The initial assets are thus given by:

$$A(0) = \mathrm{TR}(0) + \mathrm{SM}(0).$$

APPENDIX 2

DESCRIPTION OF STOCHASTIC MODELS USED FOR ASSETS AND INFLATION

A2.1. The investment and inflation models used are those proposed by WILKIE (1984, 1986). A summary of the specification of the model is given below. The variables used are:

- $q(t)$ The UK retail prices index.
- $d(t)$ An index of share dividends.
- $y(t)$ The dividend yield on these same share indices, that is, the dividend index at the specified date divided by the share price index at that date.
- $c(t)$ The yield on 2.5% Consols (irredeemable), which is taken as a measure of the general level of fixed interest yields in the market.

A2.2. The model used for $q(t)$ is:

$$\nabla \ln\{q(t)\} = \mu_q + \alpha_q[\nabla \ln\{q(t-1)\} - \mu_q] + \sigma_q z_q(t)$$

where the backwards difference operator ∇ is defined by

$$\nabla x(t) = x(t) - x(t-1)$$

and $z_q(t)$ is a sequence of independent identically distributed unit normal variates. The values adopted for the parameters are:

$$\mu_q = 0.05, \ \alpha_q = 0.6, \ \sigma_q = 0.05.$$

A2.3. The model for $y(t)$ is:

$$\ln\{y(t)\} = \omega_y \nabla \ln\{q(t)\} + y_n(t)$$

where

$$y_n(t) = \ln(\mu_y) + \alpha_y[y_n(t-1) - \ln(\mu_y)] + \sigma_y z_y(t)$$

and $z_y(t)$ is a sequence of independent identically distributed unit normal variates. The values adopted for the parameters are:

$$\mu_y = 0.04, \ \alpha_y = 0.6, \ \omega_y = 1.35, \ \sigma_y = 0.175.$$

A2.4. The model for $d(t)$ is:

$$\nabla \ln\{d(t)\} = \omega_d\left(\frac{\delta_d}{1-(1-\delta_d)\bar{B}}\right)\nabla \ln\{q(t)\} + \alpha_d \ \nabla \ln\{q(t)\}$$
$$+ \beta_d\sigma_y z_y(t-1) + \sigma_d z_d(t) + \gamma_d \ \sigma_d z_d(t-1)$$

where the backwards step operator $\bar{B}$ is defined by

$$\bar{B}x(t) = x(t-1)$$

and hence

$$\bar{B}^n x(t) = x(t-n)$$

and $z_d(t)$ is a sequence of independent identically distributed unit normal variates.

The term in parentheses above involving δ_d represents an infinite series of lag effects, with exponentially declining coefficients:

δ_d,
$\delta_d(1-\delta_d)$,
$\delta_d(1-\delta_d)^2$, etc.

The sum of these coefficients is unity, so this part of the formula represents the lagged effect of inflation, with unit gain. This means that if retail prices rise by 1% this term will also, eventually, rise by 1%. We can alternatively describe it as the "carried forward" effect of inflation $m(t)$, where

$$m(t) = \delta_d \ \nabla \ln\{q(t)\} + (1-\delta_d)m(t-1),$$

from which we see that the amount that enters the dividend model each year is δ_d times the current inflation rate, plus $(1-\delta_d)$ times the amount brought forward from the previous year, and that this total is then carried forward to the next year. The values adopted for the parameters are:

$$\omega_d = 0.8, \ \delta_d = 0.2, \ \alpha_d = 0.2, \ \beta_d = -0.2,$$
$$\gamma_d = 0.375, \ \sigma_d = 0.075.$$

A2.5. The model for $c(t)$ is:

$$c(t) = \omega_c\left(\frac{\delta_c}{1-(1-\delta_c)\bar{B}}\right) \ \nabla \ln\{q(t)\} + n(t),$$

where

$$\ln\{n(t)\} = \ln(\mu_c) + (\alpha_c\bar{B} + \beta_c\bar{B}^2 + \gamma_c\bar{B}^3)[\ln\{n(t)\} - \ln(\mu_c)] + \phi_c\sigma_y z_y(t) + \sigma_c z_c(t),$$

where $z_c(t)$ is a sequence of independent identically distributed unit normal variates.

The term in parentheses in δ_c has a similar form to the δ_d term in the dividend model, though the parameter value is different. It represents the current value of expected future inflation as an exponentially weighted moving average of past rates of inflation. The values adopted for the parameters are:

$$\omega_c = 1.0,\ \delta_c = 0.045,\ \mu_c = 0.035,\ \alpha_c = 1.20,$$
$$\beta_c = -0.48,\ \gamma_c = 0.20,\ \phi_c = 0.06,\ \sigma_c = 0.14.$$

A2.6. Interested readers are referred to WILKIE (1986) for interpretation of what the model implies and how it can be used. A fuller description of the derivation of the model is given in WILKIE (1984).

A2.7. There is no specific provision in Professor Wilkie's model for cash as an investment. We have assumed that the return on cash for any year is the Consols yield at the start of the year less one percentage point.

APPENDIX 3
THE SIMULATION PROGRAM

A3.1.1. In order to simulate the run-off of an insurance company it is necessary to make decisions in regard to many parameters. Furthermore it is not easy, *ab initio*, to select sensible values for many of them. The program is written in such a way as to allow for a range of values of each of the parameters. As there are at least 20 parameters or values that may vary, and several may have up to 8 or 9 values, it is impossible to provide for every possible combination of values of the parameters. Not only would the program take too long to run but the volume of output would be too great to comprehend.

A3.1.2. The program is written, therefore, to allow each of the parameters to vary in turn over its whole range, whilst the others are kept constant at a "normal" or standard level. It also permits an analysis by two parameters at a time, for every possible combination of the various levels of those two parameters.

A3.1.3. For each parameter combination the same 1000 sets of random numbers are used, so that the comparisons are not significantly affected by any bias in the particular sets of random numbers chosen.

The Basis of the Simulations

A3.2.1. The program works from a series of written premiums, going back sufficiently far into the past to include every year for which claims are still to be run off. Provision is made for three alternative bases for the future.

1. A wind-up — an assumed return of the unearned premium reserve (UPR) as the policyholders claim on the liquidator for the unearned part of their premiums.
2. A run-off — the UPR is translated into a pattern of future claims payments and included with payments in respect of the outstanding claims and IBNR.

3. A continuing business — the future period of writing premiums can be selected and after that there is a run-off as in 2 above.

A3.2.2. It is necessary to generate claim ratios for each type of business and for provision to be made for the claims ratios to vary stochastically. The classes of business are characterized by the length of run-off period and settlement pattern and the proportions of business written in each category of tail are set by three parameters.

A3.2.3. The investment model is that given by WILKIE (1986). The investment mix may be varied according to the nature of the business and the initial investment mix is specified separately for the technical provisions and asset margin. The rules for selling and buying investments may be selected. Buying is likely to occur where there is a continuing business and written premiums are growing but it can also arise in the later years of a run-off where the income from the assets is large, particularly in the case of larger initial asset margins.

A3.2.4. The volume of written premiums may be allowed to grow or diminish over the years since this affects the ratio of outstanding claims to the latest year's written premiums and also the relative importance of income and outgo in respect of future business where this is assumed.

A3.2.5. Corporation tax is payable by a general insurance company in the UK on its profits, which include capital gains as well as income and exclude any allowance for indexation of the purchase price of securities. However, such "income" is not subject to tax if it is used to pay claims and expenses, and it seems likely that with a company that is in any danger of becoming insolvent there will be past losses carried forward, as well as future claims outgo, that will probably absorb most, if not all, of the income. This will mean that the effective rate of tax on interest will be very low. Provision is made for notional rates of tax for the first five years, at rates well below the current rates of corporation tax. The "tax" is assumed also to include the payment of dividends to shareholders. This will result in an overstatement of the outcome in scenarios where the company remains solvent, but this is not the main feature of the results with which we are concerned. The tax treatment in the model could clearly be made more sophisticated.

The Investment Model

A3.3. The Wilkie model has been used, notwithstanding the author's warning that it was not developed for short-term forecasting. Wilkie's own view is that its use in these simulations can be justified. We have examined the output from the model over several hundred simulations of 30 years and have satisfied ourselves that the variations in the values do not appear unreasonable in the light of experience over recent years. However, this includes the possibility of a collapse in the market such as occurred in 1974 and it might be thought that such a collapse would require special dispensations allow the majority of insurers to continue to write business. Care should, therefore, be exercised in interpreting the results in so far as they depend upon the impact of temporary abrupt falls in market values.

Future Statutory Solvency

A3.4.1. For a continuing company it is necessary to examine the financial position at the end of each year, if not more often. Accounts and returns have to be presented and a simulation of the future development of the company for management purposes would need to have regard to how the position might appear in presentational terms at each future reporting date.

A3.4.2. For a company that is already being run-off or to test what would happen in such circumstances, the reporting constraint is less relevant and our aim has been to look at "true" solvency, rather than the position as constrained by reporting conventions. The model simply looks at the adequacy of the assets to meet the liabilities as they are simulated to arise during the run-off. It does not check the solvency position as it might be reported to shareholders or to the supervisory authority at points during the run-off. Such a factor could be introduced if a procedure for deciding on appropriate bases for the technical provisions in future years were to be defined.

The Choice of Parameters and Their Values

A3.5. Every parameter is allowed to have at most 9 values, but need not be given more than 1. The parameters are numbered 1 to 13 and their levels 1 to 9. The value for level 5 is the standard and a value must be inserted for this parameter in every case, even if it is not included in the list of parameters to be analysed, since the program requires a value to be assigned for every parameter. A detailed list of the parameters and the factors underlying their choice is given below:

1. *Written premiums.* The values used are £1,000,000, £10,000,000 and £100,000,000 a year. For a larger amount of business the purely stochastic variation would be negligible in comparison with other variability, so that the results would be unlikely to differ significantly from those for £100,000,000 a year. The written premiums are taken as being net of initial expenses and commission.
2. *Claim ratio — very long.* This is the claim ratio for future business of a very long-tailed nature, i.e. with a run-off period of 20 years. So far no such business has been included in the simulations. Claim ratios are assumed to include the expenses of claim settlement with actual claim costs but they are related to written premiums net of commission and expenses (cf. Paragraph 4.3.2).
3. *Standard deviation — very long.* This is the standard deviation of the above claim ratio.
4. *Claim ratio — long* This is the claim ratio for future business of a long-

tailed nature, i.e. with a run-off period of 10 years. We have used values of 80%, 100%, 125% and 150%.

5. *Standard deviation — long.* We have used 5%, 10%, 15%, 20% and 25% of written premiums.

6. *Claim ratio — short.* This is the claim ratio for future business of a short-tailed nature. We have used 80%, 100%, 125% and 150% here also.

7. *Standard deviation — short.* We have used 5%, 10%, 15% and 20% of written premiums.

8. *Growth rates.* Separate real growth rates may be assumed before and after the assessment date. Rates varying from 0.8 to 1.5 have been used and the effect of zero real growth up to the date of assessment and positive or negative growth thereafter, and vice versa, have been examined. Inflation is automatically allowed for in the program so that the growth assumptions relate to growth in real terms.

9. *Proportions of business.* These are the proportions of written premiums represented by very long-tailed business, long-tailed business and short-tailed business. Only the first two are given: the program calculates the short-tailed and checks that it is not negative.

10. *Asset mix — solvency margin.* The proportions of equities and gilts are given separately. The proportion of "cash" is calculated and checked to see that it is not negative.

11. *Asset mix — technical provisions.* As above (10).

12. *Asset margin.* This is expressed as a percentage of the net written premiums in the last year before the date of assessment. This margin is allowed to range from nil to 120%. The normal value has been taken as 40%. Different reserving strength, arising from the assumptions made in calculating the outstanding claims, allowing for inflation and for discounting, can be studied by looking at different asset margins (cf. Paragraph 5.2). On the standard basis the technical provisions are established using 5% inflation and no discounting.

13. *Selling rules.* There are 8 alternative rules, namely:

- (*a*) Sell equities until they are exhausted, then gilts and finally cash.
- (*b*) Equities, cash, gilts.
- (*c*) Gilts, equities, cash.
- (*d*) Gilts, cash, equities.
- (*e*) Cash, gilts, equities.
- (*f*) Cash, equities, gilts.
- (*g*) Sell rateably (i.e. in proportion to the current value of holdings).
- (*h*) Sell each year whatever has performed best since the start of the run-off.

Investment (where there is a surplus of income over outgo) is always done rateably. It is also necessary to specifiy:

1. The number of future years. This is limited to the range 1 to 10 but the parameter can take the values 0 or -1 meaning that we are assuming no new business written and that we have either a run-off (0) or a wind-up (-1).

2. The number of simulations.
3. The number of parameters to be analysed, that is 1 or 2.
4. The existence of the very long-tailed class. This was included as an option to avoid having very long loops which are not needed where there is no such class. It was merely a program-writing device.

The Program Plan

A3.6. The program has been written to permit it to be run on FORTRAN IV (otherwise known as FORTRAN 66). In particular we have avoided the use of negative values in arrays. For this purpose we have assumed that the past is represented by years 1 to 20 and the future by years 21 to 46. Whilst this means that some arrays have to be larger than they would otherwise need to be, the simplification is worthwhile. The program is divided into sections:
1. *Initialization.* This sets out the values of the parameters, dimensions the arrays and sets some initial values. The values of the parameters could be inserted by lead cards if preferred. This section also includes the values of: number of future years, number of parameters, very long-tailed option and the number of simulations. This section also contains some data manipulation and checking to avoid time-consuming operations later in the program.
2. *The random number generator.* This generates the necessary number of random normal variates and stores them in an array for use by the later stages of the program. This ensures that the same numbers are used for every variation within a single simulation. They are recalculated for each further simulation. The random number generator of the machine has been used to generate uniform random variates in the range 0 to 2. After subtracting 1 these are used in Marsaglia's polar method to generate the corresponding random normal variates. This method requires pairs of uniform randoms and produces normal variates if, and only if, the sum of the squares of the two variates is less than 1. The program counts the number of useful pairs and stops when it has enough to fill the array.

We have tested this process and found that a distribution of 3 million variates was very closely normal, using 9-figure tables of the normal integral for the test. This however does not test that they come in a random order and we have further tested them to count the number of cases where there is a run of 1 increase or 2 increases and so on up to 7 increases. It is not difficult to calculate theoretically the expected number of such runs both upwards and downwards and their expected size. The results are within expected limits. The methods will be described in detail in a paper to be written by two of the authors of this report, together with notes on the times taken to make the calculations. These seem to vary considerably from one method to another. It is perhaps worth mentioning that what we require are representative sequences rather than purely random ones. Kendall and Babington Smith noted in 1938 that a sequence of $10^{10^{10}}$ random numbers is almost certain to contain a sequence of a million zeros (or, for that matter any other sequence you care to specify). This might be a random sequence but it is not very useful in practice.
3. *Investment values.* The program now calculates the investment values for up

to 26 future years, depending on the particular run-off period involved. The values are of:

1. The retail price index.
2. Equity dividends.
3. Equity yield.
4. Gilt yield.
5. Equity price.
6. Gilt price.
7. Cash yield.
8. Borrowing rate.
9. A net income multiplier (see below).
10. An equity price ratio.
11. A gilt price ratio.
12. A mean retail price index.

The reason for a "cash" yield is that gilts are assumed to relate to medium or long term, whilst cash is either cash on deposit or very short term gilts. It is assumed that the cash yield is 1% below that of gilts and that when cash becomes negative and we have to borrow, it is at a rate 2% higher than the gilt rate. The gilt and equity yields have a minimum of 0.5%. The equity price ratio is the square root of the ratio of the equity price at the end of the relative year to its value at the start of the year. Its purpose is to revalue equities from the year-end value, on which the income is based, to the mid-year value at which it is assumed that sales take place or purchases are made. After the mid-year transactions the remaining values of gilts and equities are updated to the year-end by a further multiplication by the equity (or gilt) price ratio. Although interest is calculated on the values at the start of the year, allowance is made for the loss of income on selling during the year by multiplying the net outgo by a factor of 1 plus half the average yearly yield on the investments. Whilst this assumes that the values of all three classes are equal, the effect of differences is likely to be too small to be of any consequence in practice.

4. *Best investment.* The next section is really a continuation of section 3 in that it calculates which of the three classes of investment has performed best since the start of year 21 and stores this information for use later in the program.

5. *Outstanding claims.* The program now calculates the outstanding claims at the end of year 20. For each earlier year the program calculates the claims according to the mean claim ratios and then, using the run-off rates shown in Table A3.1, calculates the amounts, in constant money terms, which it expects to pay out in each future year.

These are stored in an array by year of expected payment and the total is accumulated, allowing for 5% future inflation, in a variable TOTOS which is the total provision for claim amounts outstanding at the end of year 20. By using these run-offs we have automatically taken into account the IBNR claims. If we have a wind-up situation then TOTOS is the technical provision. If we are considering a run-off or a continuing business then we must add 50% of the written

TABLE A3 1

From Abbott, *et al*, 1981

Duration from year of origin	Proportion of claims settled (%) Short-tail	Long-tail
0	61 2	5 6
1	24 1	25 3
2	5 2	18 7
3	3 7	13 2
4	2 7	10.4
5	2 2	7 9
6	0 9	6 4
7	—	4 6
8	—	3 8
9	—	3 0
10	—	1 1
	100 0	100 0

premium for year 20 into TOTOS as the unearned premium reserve. This figure for technical provisions, together with the asset margin obtained from the product of the assumed asset margin percentage and the written premiums for year 20, enable us to calculate the initial amounts of each type of asset using the specified proportions. This is the initial investment portfolio.

6. *Future premiums.* We next add into the arrays of future payments the expected contribution to claims outgo arising from future written premiums and from the unearned premium reserve for the last year, to give the expected claims outgo in constant money terms.

7. *Emerging costs.* The program now has the information to enable it to calculate the expected payments in each future year. The claims outgo is adjusted for inflation according to the Wilkie model and is allowed to vary stochastically. We assume a normal distribution and a formula of $0.15X = 75\sqrt{X}$ as the standard deviation for the total claim outgo in any year. The square root factor is dominant for the smaller amounts and the smaller companies but for the larger companies the stochastic variation is negligible and it is only realistic to assume some sort of overall secular variation (see Appendix 5).

We take the values of assets at the beginning of the year and calculate the income on each type of asset, reducing the total income for the year by the tax factor where appropriate. We then have the outgo, adjusted to allow for inflation and stochastic variation, less the income and less any written premiums for a continuing business. As mentioned in A3.6.3 we adjust for the loss of part of the year's investment income as a result of net selling during the year (or vice versa in a net buying situation). Investment or disinvestment is assumed to take place at mid-year values. If there is net investment, it is assumed to be made proportionately to the existing values of the three classes of investment. Where there is net outgo, the specified selling rule is applied.

8. *Final assets.* This process continues until the last year's claims outgo has been paid. The final assets are in the currency of the final year as a result of the application of the investment model which revalues the assets, combined with the models

for income and outgo in each year which allow implicitly for future inflation. In order to bring the final asset value into the currency of the start of the run-off, it is divided by the ratio of the retail price index in the final year to that at the date of assessment. The result is then expressed as a percentage of the written premiums in year 20 (the year before the date of assessment). These values from the 1000 simulations are grouped into ranges and output as a distribution, together with their mean and standard deviation.

APPENDIX 4
RESULTS OF SIMULATION

Full details of the results of 1000 simulations on a variety of different bases are set out in Tables A4.1 to A4.4. Tables A4.1 and A4.2 show summary distributions of the simulations by the assets remaining at the end, as well as the number of insolvencies and the mean and standard deviation of the distributions. Results are also given for a few additional variants not tabulated in Tables 4 and 5. Tables A4.3 and A4.4 also include a number of additional variants and Table A4.4 shows the additional asset margin required in the case of 2 years' new business as compared to the pure run-off with the same assumptions (in so far as these are applicable).

The tables show the standard basis at the top and also in each of the groups of alternative assumptions (marked (s)). The variants examine the effect of varying the one assumption referred to, whilst leaving all the other assumptions the same as in the standard basis.

The assumptions underlying the standard basis are as follows:

Net written premiums[1]	£10m a year
Proportion of long-tailed business	40% of net written premiums
Past growth	In line with inflation
Future growth	In line with inflation
Mean claim ratio[2] (short-tailed)	100% of net written premiums[1]
Standard deviation of CR[2] (short-tailed)	10% of net written premiums[1]
Mean claim ratio[2] (long-tailed)	100% of net written premiums[1]
Standard deviation of CR[2] (long-tailed)	15% of net written premiums[1]
Initial asset distribution	Technical provisions: 50% cash; 50% gilts Asset margin: 100% equities
Asset selling rule	Proportionate to holdings
Asset margin (for Tables A4.1 and A4.2)	40% of net written premiums

1 Premiums net of commission and expenses
2 Ratio of claims (including claims settlement expenses), without allowance for future inflation or for discounting, to premiums net of commission and expenses (see Paragraph 4 3 2)

TABLE A4 1

SUMMARY OF RESULTS FOR PURE RUN-OFF OF BUSINESS (WITH 1000 SIMULATIONS)

Assumptions	No of insolvencies	No of simulations with remaining assets[1] of. 0%–40%	40%–80%	80%–120%	120%–160%	Over 160%	Mean assets remaining[1] %	Standard deviation of assets remaining[1] %
Standard basis	8	98	252	264	173	205	112	70
1 Net written premiums.[2]								
(a) £1m a year	20	119	227	250	171	213	113	75
(b) £10m a year (s)	8	98	252	264	173	205	112	70
(c) £100m a year	6	93	258	275	164	204	112	69
2 Proportion of long-tailed business·								
(a) 20% of net written premiums[2]	3	126	321	293	153	104	94	55
(b) 40% of net written premiums[2]	8	98	252	264	173	205	112	70
(c) 60% of net written premiums[2]	13	88	197	230	184	288	130	85
3 Initial asset distribution: Cash / Gilts / Equities								
(a) TP + AM / — / —	3	128	303	295	162	109	95	53
(b) — / TP + AM / —	20	159	220	208	142	251	120	98
(c) — / — / TP + AM	49	124	182	181	149	315	136	115
(d) $\frac{1}{2}$TP / $\frac{1}{2}$TP / AM (s)	8	98	252	264	173	205	112	70
(e) $\frac{1}{2}$TP + $\frac{1}{2}$AM / $\frac{1}{2}$TP + $\frac{1}{2}$AM / —	14	127	260	255	150	194	107	72
(f) $\frac{1}{2}$TP + $\frac{1}{2}$AM / — / $\frac{1}{2}$TP + $\frac{1}{2}$AM	12	85	222	278	184	219	118	72
(g) TP / — / AM	3	93	283	312	176	133	102	54

TABLE A4 1 (*Continued*)

Assumptions	No of insolvencies	No. of simulations with remaining assets[1] of					Mean assets remaining[1] %	Standard deviation of assets remaining[1] %
		0%–40%	40%–80%	80%–120%	120%–160%	Over 160%		
Standard basis	8	98	252	264	173	205	112	70
4. Initial asset margin								
(a) 0% of net written premiums[2]	134	336	292	138	58	44	52	55
(b) 20% of net written premiums[2]	36	210	312	222	117	102	83	62
(c) 40% of net written premiums[2] (s)	8	98	252	264	173	205	112	70
(d) 60% of net written premiums[2]	2	34	149	244	212	360	147	80
(e) 80% of net written premiums[2]	0	12	74	184	211	519	180	90
(f) 100% of net written premiums[2]	0	4	34	117	186	659	212	100
5 Asset selling rules:								
(a) Equities, gilts, cash	9	133	285	262	152	159	102	66
(b) Equities, cash, gilts	16	142	233	241	143	225	113	81
(c) Gilts, equities, cash	5	123	246	285	181	160	105	61
(d) Gilts, cash, equities	3	67	249	314	201	166	111	58
(e) Cash; gilts, equities	11	134	212	237	168	238	120	83
(f) Cash, equities, gilts	7	78	245	244	185	241	123	79
(g) In proportion to holdings (s)	8	98	252	264	173	205	112	70
(h) Sell best performer first	14	131	239	268	158	190	108	70

1 Deflated to the date of assessment and expressed as a percentage of net written premiums[2] in the year before the date of assessment (see Appendix 3 6 8)
2 Premiums net of commission and expenses
(s) indicates the assumption made for the standard basis

TABLE A4 2

SUMMARY OF RESULTS WITH 2 FURTHER YEARS' BUSINESS (WITH 1000 SIMULATIONS)

Assumptions	No of insolvencies	No of simulations with remaining assets[1] of:					Mean assets remaining[1] %	Standard deviation of assets remaining[1] %
		0%–40%	40%–80%	80%–120%	120%–160%	Over 160%		
Standard basis	50	96	146	171	175	362	144	109
1. Net written premiums [2]								
(a) £1m a year	61	106	144	155	161	373	144	117
(b) £10m a year (s)	50	96	146	171	175	362	144	109
(c) £100m a year	43	96	144	181	172	364	144	107
2 Proportion of long-tailed business:								
(a) 10% of net written premiums[2]	52	136	217	220	164	211	108	83
(b) 20% of net written premiums[2]	48	116	190	212	159	275	120	91
(c) 40% of net written premiums[2] (s)	50	96	146	171	175	362	144	109
(d) 60% of net written premiums[2]	52	83	120	141	159	445	168	130
(e) 80% of net written premiums[2]	53	77	99	117	126	528	191	151
(f) 90% of net written premiums[2]	58	71	93	106	118	554	203	162
3 Future real growth rate (in constant money terms).								
(a) −20% a year (real past growth −20% p a)	53	78	118	143	144	464	171	134
(b) No growth (no real past growth) (s)	50	96	146	171	175	362	144	109
(c) +50% a year (real past growth +50% p a)	83	94	139	149	155	380	144	121
(d) +30% a year (no real past growth)	66	93	131	144	151	415	155	127
(e) +50% a year (no real past growth)	86	80	122	130	137	445	164	141
4 Mean claim ratio[3] (short-tailed)								
(a) 80% of net written premiums[2]	7	47	105	129	177	535	187	117
(b) 100% of net written premiums[2] (s)	50	96	146	171	175	362	144	109
(c) 125% of net written premiums[2]	165	167	177	168	120	203	90	103
(d) 150% of net written premiums[2]	380	158	183	110	67	102	35	102

TABLE A4 2 (*Continued*)

SUMMARY OF RESULTS WITH 2 FURTHER YEARS' BUSINESS (WITH 1000 SIMULATIONS)

Assumptions	No of insolvencies	No of simulations with remaining assets[1] of: 0%–40%	40%–80%	80%–120%	120%–160%	Over 160%	Mean assets remaining[1] %	Standard deviation of assets remaining[1] %
Standard basis	50	96	146	171	175	362	144	109
5. Variability of claim ratio (short-tailed)								
(a) Standard deviation 5% NWP[2]	49	97	142	178	170	364	144	109
(b) Standard deviation 10% NWP[2] (s)	50	96	146	171	175	362	144	109
(c) Standard deviation 15% NWP[2]	48	96	148	168	173	367	144	111
(d) Standard deviation 20% NWP[2]	52	94	144	170	172	368	145	112
6 Mean claim ratio[3] (long-tailed)								
(a) 80% of net written premiums[2]	17	62	136	161	195	429	163	106
(b) 100% of net written premiums[2] (s)	50	96	146	171	175	362	144	109
(c) 125% of net written premiums[2]	105	139	155	160	138	303	120	114
(d) 150% of net written premiums[2]	195	139	164	143	119	240	97	120
7 Variability of claim ratio (long-tailed)								
(a) Standard deviation 5% NWP[2]	49	84	159	172	171	365	144	109
(b) Standard deviation 10% NWP[2]	50	89	156	172	168	365	144	109
(c) Standard deviation 15% NWP[2] (s)	50	96	146	171	175	362	144	109
(d) Standard deviation 20% NWP[2]	49	101	137	176	176	361	144	110
(e) Standard deviation 25% NWP[2]	50	104	135	174	172	365	144	111
8 Initial asset distribution: Cash / Gilts / Equities								
(a) TP + AM / — / —	46	130	186	204	173	261	118	86
(b) — / TP + AM / —	79	121	141	147	124	388	155	152
(c) — / — / TP + AM	86	91	112	131	113	467	181	172
(d) $\frac{1}{2}$TP / $\frac{1}{2}$TP / AM (s)	50	96	146	171	175	362	144	109
(e) $\frac{1}{2}$TP + $\frac{1}{2}$AM / $\frac{1}{2}$TP + $\frac{1}{2}$AM / —	61	120	162	159	163	335	136	113
(f) $\frac{1}{2}$TP + $\frac{1}{2}$AM / — / $\frac{1}{2}$TP + $\frac{1}{2}$AM	43	77	141	174	159	406	152	112
(g) TP / — / AM	41	97	166	208	182	306	128	87

9 Initial asset margin								
(a) 0% of net written premiums[2]	196	204	197	154	97	151	74	98
(b) 20% of net written premiums[2]	103	148	197	167	140	244	109	105
(c) 40% of net written premiums[2] (s)	50	96	146	171	175	362	144	109
(d) 60% of net written premiums[2]	24	54	115	151	158	497	180	123
(e) 80% of net written premiums[2]	11	29	75	127	140	619	216	133
(f) 100% of net written premiums[2]	3	18	45	88	128	718	252	144
10 Asset selling rules								
(a) Equities, gilts, cash	56	120	161	175	150	338	136	113
(b) Equities, cash, gilts	63	103	156	169	138	371	144	120
(c) Gilts; equities; cash	51	106	154	165	181	343	137	103
(d) Gilts, cash, equities	35	87	141	192	197	348	142	99
(e) Cash, gilts, equities	41	76	143	173	173	394	154	116
(f) Cash; equities, gilts	52	90	144	153	168	393	151	119
(g) In proportion to holding (s)	50	96	146	171	175	362	144	109
(h) Sell best performer first	58	101	154	164	159	364	142	114

1 Deflated to the date of assessment and expressed as a percentage of net written premiums[2] in the year before the date of assessment (see Appendix 3 6 8)
2 Premiums net of commission and expenses
3 Ratio of claims (including claims settlement expenses), without allowance for future inflation or for discounting, to premiums net of commission and expenses (see Paragraph 4.3.2)
(s) indicates the assumption made for the standard basis

TABLE A4 3

ASSET MARGINS REQUIRED TO ACHIEVE 1/100 PROBABILITY OF RUIN — NO FUTURE NEW BUSINESS

Assumptions				Asset margin as % of NWP[1]	Asset margin as % of technical provisions
Standard basis				40	15
1 Net written premiums[1]					
(a) £1m a year				55	25
(b) £10m a year (s)				40	15
(c) £100m a year				35	15
2 Proportion of long-tailed business					
(a) 20% of net written premiums[1]				30	15
(b) 40% of net written premiums[1] (s)				40	15
(c) 60% of net written premiums[1]				45	15
3. Initial asset distribution					
	Cash	Gilts	Equities		
(a)	TP + AM	—	—	30	10
(b)	—	TP + AM	—	60	25
(c)	—	—	TP + AM	80	35
(d)	$\frac{1}{2}$TP	$\frac{1}{2}$TP	AM (s)	40	15
(e)	$\frac{1}{2}$TP + $\frac{1}{2}$AM	$\frac{1}{2}$TP + $\frac{1}{2}$AM	—	50	20
(f)	$\frac{1}{2}$TP + $\frac{1}{2}$AM	—	$\frac{1}{2}$TP + $\frac{1}{2}$AM	45	20
(g)	TP	—	AM	30	10
4. Asset selling rules:					
(a) Equities; gilts; cash				40	15
(b) Equities, cash; gilts				50	20
(c) Gilts, equities; cash				35	15
(d) Gilts, cash, equities				30	10
(e) Cash; gilts, equities				35	15
(f) Cash, equities; gilts				45	20
(g) In proportion to holdings (s)				40	15
(h) Sell best performer first				55	25

1. Premiums net of commission and expenses
(s) indicates the assumption made for the standard basis

TABLE A4 4

ASSET MARGINS REQUIRED TO ACHIEVE 1/100 PROBABILITY OF RUIN — TWO YEARS NEW BUSINESS

Assumptions	Asset margin as % of NWP[1]	Excess asset margin as compared to pure run-off (as % of NWP[1])
Standard basis	90	50
1 Net written premiums[1]		
(a) £1m a year	100	45
(b) £10m a year (s)	90	50
(c) £100m a year	80	45
2 Proportion of long-tailed business		
(a) 10% of net written premiums[1]	75	45
(b) 20% of net written premiums[1]	80	50
(c) 40% of net written premiums[1] (s)	90	50
(d) 60% of net written premiums[1]	95	50
(e) 80% of net written premiums[1]	100	40
(f) 90% of net written premiums[1]	105	40

TABLE A4 4 (*Continued*)

ASSET MARGINS REQUIRED TO ACHIEVE 1/100 PROBABILITY OF RUIN — TWO YEARS' NEW BUSINESS

Assumptions				Asset margin as % of NWP[1]	Excess asset margin as compared to pure run-off (as % of NWP[1])
Standard basis				90	50
3 Future growth rate (in constant money terms)					
(a) −20% a year (real past growth −20% p.a.)				100	35
(b) No growth (no real past growth) (s)				90	50
(c) +50% a year (real past growth +50% p.a.)				115	85
(d) +30% a year (no real past growth)				100	65
(e) +50% a year (no real past growth)				120	85
4 Mean claim ratio (short-tailed)					
(a) 80% of net written premiums[1]				30	0
(b) 100% of net written premiums[1] (s)				90	50
(c) 125% of net written premiums[1]				125	75
(d) 150% of net written premiums[1]				180	115
5 Variability of claim ratio (short-tailed)					
(a) Standard deviation 5% NWP[1]				80	45
(b) Standard deviation 10% NWP[1] (s)				90	50
(c) Standard deviation 15% NWP[1]				85	45
(d) Standard deviation 20% NWP[1]				90	50
6 Mean claim ratio[2] (long-tailed)					
(a) 80% of net written premiums[1]				50	20
(b) 100% of net written premiums[1] (s)				90	50
(c) 125% of net written premiums[1]				115	60
(d) 150% of net written premiums[1]				150	85
7 Variability of claim ratio (long-tailed)					
(a) Standard deviation 5% NWP[1]				80	45
(b) Standard deviation 10% NWP[1]				85	50
(c) Standard deviation 15% NWP[1] (s)				90	50
(d) Standard deviation 20% NWP[1]				90	50
(e) Standard deviation 25% NWP[1]				90	50
8 Initial asset distribution					
	Cash	Gilts	Equities		
(a)	TP + AM	—	—	85	55
(b)	—	TP + AM	—	110	50
(c)	—	—	TP + AM	135	55
(d)	$\frac{1}{2}$TP	$\frac{1}{2}$TP	AM (s)	90	50
(e)	$\frac{1}{2}$TP + $\frac{1}{2}$AM	$\frac{1}{2}$TP + $\frac{1}{2}$AM	—	90	55
(f)	$\frac{1}{2}$TP + $\frac{1}{2}$AM	—	$\frac{1}{2}$TP + $\frac{1}{2}$AM	90	45
(g)	TP	—	AM	75	45
9 Asset selling rules					
(a) Equities, gilts, cash				95	55
(b) Equities; cash; gilts				95	45
(c) Gilts; equities; cash				90	55
(d) Gilts, cash, equities				70	40
(e) Cash, gilts, equities				85	50
(f) Cash, equities, gilts				95	50
(g) In proportion to holdings (s)				90	50
(h) Sell best performer first				95	40

1 Premiums net of commission and expenses

2 Ratio of claims (including claims settlement expenses), without allowance for future inflation or for discounting, to premiums net of commission and expenses (see Paragraph 4 3 2)

(s) indicates the assumption made for the standard basis

APPENDIX 5
VARIABILITY OF CLAIMS OUTGO

A5.1. In DAYKIN and BERNSTEIN (1985) it was assumed that the amount of the payments made in each development year for each year of origin varied log-normally. This meant that a payment amount that was to be varied stochastically was multiplied by $\exp(RS+M)$ where R is a random normal variate, S the standard deviation and M the mean. In order that the overall mean should be correct the value of M has to be equal to minus half the square of the standard deviation. This formula is suitable for a single payment, but in most cases the payment amounts considered were the totals of several or many individual amounts. Furthermore different values would need to be adopted for funds of different sizes if account was to be taken of the fact that variation is not the same for a small fund as for a large one.

A5.2. This was cumbersome and not entirely satisfactory, so an alternative approach was sought. The formula should reflect the number of payments involved and, if possible, the ratio of the standard deviation to the mean (the coefficient of variation). Consideration was given to the estimation of the numbers of claims (or claim payments) in each year's totals. We were unable to obtain any figures from actual portfolios but information from returns to the supervisory authority and from other sources suggested that for short-tailed business an average payment rising from £500 in the year of occurrence by multiples of 2 to £16,000 in the last year of development was not unreasonable. For long-tailed business the average payments rose over 10 years from £800 to £15,000.

A5.3. We assumed that coefficients of variation were in the range of 2 to 10, increasing at later durations as fewer, larger claims are settled. We were then able to estimate both the numbers of claims and their average amounts for different mixes of business by year of development. For this purpose it was assumed that claims were identical with payments, and whilst this is clearly not the case, it is not thought that it would make much difference if we were able to make more detailed assumptions. These calculations suggested that the formula for standard deviation should be a multiple of the square root of the number of claims, or its deemed equivalent, the total amount of payment. For convenience we used the amount of money, even though inflation would involve a change in the multiplier over time.

A5.4. It must be realized that precision was out of the question since we could not take into account all the possible variations in the make-up of a portfolio. It was also necessary to have regard to the fact that the bulk of the outstanding claims are paid in the first two or three years of run-off and relate primarily to the latest two or three years' business. Calculations showed that out of total outstandings of £1 million about one-half was paid in the first year and a quarter in the next year. By year 7 the payments were under £20,000, so that variation in these later years was less significant in the overall context. What is more, for many insurers the later payments, if they turn out to be large, may well be recoverable from reinsurers and so not form part of the problem for net run-off patterns. It simply moves the problem to another area. Further consideration would need to be given to the variability of the tail in the case of a company with a lot of long-tailed business and relatively high retentions.

A5.5. Experiment suggested that a multiplier of about 50 to 100 times the square root of the amount (in pounds sterling in 1986) was of the right order of magnitude. However, it was clear that whilst this gave a reasonable amount of variation for the smaller insurer it was wholly inadequate for a large one. In present conditions most of the variation for the larger fund arises from secular change and this is more likely to be proportional to the actual amount to be paid than to its square root. The problem is to choose a multiplier to give a realistic variation. Experience over recent years suggests that it must be at least 0.1, to give a variation of 20% in 95% of all cases. We finally adopted the formula

$$\mathrm{SD} = aX + b\sqrt{X}$$

using values of 0.15 for a and 75 for b.

A5.6. This formula is similar to one which we understand was introduced by the Finnish supervisory authority in 1952. Whilst we are well aware of the approximations and assumptions involved in its derivation, we think it is adequate for the purpose, although it can be considered as simply one of a class of possible formulae. It also greatly simplifies the calculations. As indicated above, the earlier paper calculated the outgo for each future year for each year of occurrence and for each length of tail separately and applied the stochastic factor to each such amount. The main effect of this was to reduce the overall variation compared with applying the same formula to the total and this effect can be achieved by adjusting the overall level of the variation. It was decided, therefore, to calculate the total outgo in each year, including that from future business where appropriate, and apply the variability factor to the total.

A5.7. It is interesting to compare the values produced by the formula with those from the exponential basis. The comparisons, with values of R corresponding to the 5%, 25% and 50% points, are shown in Table A5.1. The correspondence

TABLE A5 1

STOCHASTIC MULTIPLIER $(1 + RS)$ FOR DIFFERENT VALUES OF R AND STANDARD DEVIATION (S)

	Random normal variate (R)				
	−1 96	−0 675	0	0 675	1 96
log-normal					
$S = 0\ 3$	0 53	0 78	0 96	1 17	1.72
$S = 0\ 5$	0 33	0 63	0 88	1.24	2 35
square root formula ($S = 0\ 15X + 75\sqrt{X}$)					
$X = 100{,}000$	0 24	0 74	1 0	1 26	1.76
$X = 1{,}000{,}000$	0.56	0.85	1 0	1 15	1.44
$X = 10{,}000{,}000$	0 66	0 88	1 0	1.12	1 34

between the two formulae, coupled with the size of variation by insurer, suggests that the new formula is in line with the old but more realistic in its relation to the actual amounts of payments.

APPENDIX 6

POSSIBLE APPROACH TO SIMULATING REINSURANCE RECOVERIES

A6.1. It is not possible to simulate reinsurance recoveries in our model in any very precise way, firstly because it is too complicated and secondly because the model simulates claims only in aggregate. It would in principle be possible to think in terms of a specified number of reinsurers, each bearing a share of the anticipated reinsurance recoveries, and find a way to model the failure of reinsurers. Rather easier, and probably no less realistic, would be to go directly to the proportion recovered. One way of approaching the problem is set out below.

A6.2. Reinsurers would be allocated to say, three categories — strong, average and weak. For any class of business the proportion of reinsurance recoveries anticipated from each of the three categories of reinsurer would be input as data. The model would then be to apply a process, defined separately for each category, to determine the proportion not recovered in respect of any particular year's estimated gross claim payments. There remains, of course, the problem of estimating gross claims payments and simulating their out-turn, so that there would be considerable practical problems in implementing an approach of this sort.

A6.3. The probability of recovery would be related to the gross claims out-turn. This could be done by taking the estimate of gross claims paid in the year in question to be the mean estimate of claims paid, based on proportions expected to be settled in the year, the rate of inflation assumed in setting the technical provisions and, in the case of claims arising from future business, the mean claim ratio. There would then be a set of formulae, one for each category of reinsurer, to define the proportion of gross claims paid in the year which is assumed not to be recovered, based on the ratio of gross claims out-turn to estimated gross claims for the year. For year j we might, for example, define the proportion not recovered, $Y(j)$, by:

Weak

$$Y(j) = \frac{k(j)}{200} \qquad 0 < k(j) < 200$$

Average

$$Y(j) = \frac{k(j) - 50}{500} \qquad 50 < k(j) < 550$$

Strong

$$Y(j) = \frac{k(j) - 100}{800} \qquad 100 < k(j) < 900$$

where

$$1 + \frac{k(j)}{100} = \frac{X(j)}{\sum_{i \leqslant j} \hat{X}(i:j)}$$

$X(j)$ = actual total gross claims settled in year j

and

$\sum_{i \leqslant j} \hat{X}(i:j)$ = expected gross claims settlement in year j in respect of year of origin i on basis of mean claims ratio, assumed settlement pattern and expected inflation.

In terms of the notation of Appendix 1:

$$\hat{X}(i;j) = \sum_k s^k(j-1)R_k B_k'(i)(1+r)^{j-i}.$$

A6.4. The formulae can obviously be adapted to reflect one's ideas of a plausible model for reinsurance recoveries. The general principle of these illustrative formulae is that one would expect higher proportions not recovered for weaker reinsurers and that, above a certain threshold, higher claims relative to the expected level of claims imply a higher proportion not recovered. These formulae do not attempt to distinguish between high claims as a result of high initial loss ratios, high inflation and adverse development. In principle one could also develop some form of cumulative trigger so that failure to recover increased with a series of high claims payments rather than simply on the basis of a single year.

A6.5. Consideration would also need to be given to whether to apply the formulae to all classes together or each class separately. Possibly the most realistic would be to apply it to the total claims on those classes of business where significant amounts are reinsured.

A6.6. The simple approach suggested here may not be sufficiently realistic for some companies for whom reinsurance recovery is a major issue. Further development of these ideas is clearly needed. However, it is suggested that it may be possible to obtain a useful indication of the role of reinsurance in a particular case by the use of straightforward models.

REFERENCES

ABBOTT W. M., CLARKE, T. G and TREEN, W. R (1981) Some financial aspects of a general insurance company *Journal of the Institute of Actuaries* **108**, 119.

BEARD, R E , PENTIKAINEN, T and PESONEN, E (1984) *Risk Theory* (3rd edn) Chapman and Hall, London

BENJAMIN, S (1980) Solvency and profitability in insurance *Transactions of the 21st International Congress of Actuaries* **1**, 33

BUCHANAN, R A and TAYLOR, G C (1986) The management of solvency, Presented to the International Conference on Insurance Solvency, Philadelphia, USA

COUTTS, S. M. and DEVITT, E. R F. (1986) The assessment of the financial strength of insurance companies — a generalised cash-flow model Presented to the International Conference on Insurance Solvency, Philadelphia, USA.

COUTTS, S. M., DEVITT, E R F and ROSS, G A. F (1984) A probabilistic approach to assessing the financial strength of a general insurance company *Transactions of the 22nd Internal Congress of Actuaries* **3**, 129.

DAYKIN, C D (1984) The development of concepts of adequacy and solvency in non-life insurance in the EEC. *Transactions of the 22nd International Congress of Actuaries* **3**, 299

DAYKIN, C. D and BERNSTEIN, G D (1985) A simulation model to examine questions of solvency in the light of asset and run-off risks Presented to ASTIN Colloquium in Biarritz, October 1985

DAYKIN, C D, DEVITT, E R., KHAN, M R and MCCAUGHAN, J P. (1984) The solvency of general insurance companies *Journal of the Institute of Actuaries* **111**, 279.

KAHANE, Y. (1979) Solidity, leverage and the regulation of insurance companies *The Geneva Papers on Risk and Insurance* No 14.

KAHANE, Y and BIGER, N (1977) Balance sheet optimisation in inflationary circumstances — the case of non-life insurance companies *The Journal of Insurance Issues and Practices* **1**, No 2

PENTIKAINEN, T and RANTALA, J (1982) *Solvency of Insurers and Equalization Reserves* Helsinki.

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

RYAN, J P (1980) An application of model office techniques to solvency testing for a non-life office. *Transactions of the 21st International Congress of Actuaries* **1**, 403

RYAN, J. P (1984) Application of simulation techniques to solvency testing for a non-life office *Transactions of the 22nd International Congress of Actuaries* **3**, 269.

STEWART, C M. (1971) The assessment of solvency *Astin Bulletin* **6**, 79.

WILKIE, A. D (1984) Steps towards a comprehensive stochastic investment model. Occasional Actuarial Research Discussion Paper No 36, Institute of Actuaries.

WILKIE, A D (1986) A stochastic investment model for actuarial use *Transactions of the Faculty of Actuaries* **39**, 341

CHRIS DAYKIN
Government Actuary's Department, 22 Kingsway, London WC2B 6LE, UK

BOOK REVIEWS

H. U. GERBER (1986). *Lebensversicherungsmathematik.* Springer Verlag, Berlin etc.; Vereingung Schweizerischer Versicherungs-mathematiker, Zurich. XIII, 125 pages, DM 98.00.

In the past decades there has been — and to a certain extent still is — a gap between practitioners on the one hand and researchers in the field of actuarial mathematics on the other, at least in Germany. While many mathematicians at the universities have been inclined almost to ignore actuarial applications, people from insurance companies have had the impression that stimulating and innovative new ideas for their business were not to be expected from scientists working in the ivory-tower of a university. These facts need to be kept in mind when a new book on life insurance mathematics has to be assessed, written in German by a leading expert in the fields of actuarial mathematics and risk theory.

It is true that for the techniques of life insurance elementary deterministic models based on the calculation of interest suffice and will still suffice for the foreseeable future, at least in the mass business. Nevertheless, these traditional models are unsatisfactory because they do not take into account the random character of insurance processes on the one side nor do they take advantage of data-processing and modern computers on the other.

This is the starting point of the present monograph: mortality tables are replaced by stochastic models based on a random variable T denoting the residual life time of a person, and tabulations of commutation functions and the like are supplanted by algorithms, especially recursions. Compared with this, the topics of the individual chapters are fairly conventional.

Chapter 1. Calculation of Interest. Indispensable preliminaries, presented concisely and elegantly.

Chapter 2. The Future Life Time of a Person Aged x. The stochastic model — the life time variable and distribution — and related notation.

Chapter 3. Capital Assurances. Discussion of the cash value of a capital. Since this is a random variable by definition, not only the net single premium — the expected value — but also the variance of the cash value is of interest.

Chapter 4. Life Annuities. Calculation of cash values and net single premiums.

Chapter 5. Net Premiums. Derivation of well-known formulae for various types of insurance.

Chapter 6. Net Level Premium Reserves. Again, standard results as well as results appearing in the stochastic model only, are proved, e.g. Hattendorf's theorem.

Chapter 7. Several Causes of Decrement. Inclusion of an additional random variable describing the cause of decrement.

Chapter 8. Joint Life Assurance. Here the advantages of the stochastic model are particularly obvious, and analogies with reliability theory are made (joint life status/series structure, last-survivor status/parallel structure).
Chapter 9. Aggregate Claims of a Collective. Risk theoretic considerations, especially the development and numerical treatment of the distribution function of the aggregate claims.
Chapter 10. Inclusion of Expenses. Incorporation of the third base of calculation.
Chapter 11. Estimation of Probabilities of Death. Classical methods and procedures from mathematical statistics.
Appendix A Commutation Values
Appendix B. Simple Interest.

The present monograph thus has as many chapters as the first volume of Saxer's standard work. Regarding the contents, the amount of overlap is about sixty per cent. The book is written clearly, precisely and elegantly. As in his pioneering book on risk theory, the author succeeds brilliantly in bridging the gap between intuition and rigour.

Compared with this, there are only a few minor points to be criticized. First of all, the use of stochastic models appears to be a bit half-hearted now and then, especially so with respect to their connections with reliability theory. Symptomatically, in the Foreword a probability space (Ω, A, P) is mentioned in passing, whereas in the text the symbol *Pr*, which is never defined explicitly, is used whenever probabilities are represented — even 'probabilities' of the type $Pr(t < T < t + dt)$.

Naturally, the practical needs of an actuary over and above the technical and mathematical aspects, e.g. statement of accounts, are not met by the present book. However, practitioners from *life* assurance might be interested by the material presented in Chapter 9 under the topic of reinsurance.

These objections, however, cannot detract from the substantial merits of Gerber's excellent book for which success both with practitioners and theorists can be predicted without any risk whatsoever.

W.-R. HEILMANN

BJØRN SUNDT (1984). *An Introduction to Non-Life Insurance Mathematics.* Veroffentlichungen des Instituts fur Versicherungswissenschaft der Universitat Mannheim, Vol. 28, Verlag Versicherungswirtschaft, Karlsruhe. 168 pages, DM 24.00.

In his foreword to the book the editor writes: "Textbooks in Non-Life Insurance Mathematics are rare. So it is a pleasure for me that Dr. Sundt was willing to write down his lectures given at Mannheim during the summer of 1983." A practitioner might be deterred by these introductory sentences, since lectures for students of mathematics in Germany usually are filled with abstract theories and complicated proofs. Dr. Sundt's book, however, is well suited as an introduction for practitioners into methods of modern risk theory. Abstract theories are avoided whenever possible (e.g. credibility estimators are derived without any Hilbert space theory), and complicated proofs are substituted by informal deductions (e.g. Edgeworth expansions are introduced without mentioning Cramér's condition on the underlying characteristic function). Abstract models are motivated and explained using realistic actuarial problems, thus demonstrating that the mathematics presented in the book is in fact applicable. This can best be seen in the chapters on experience rating containing a nice introduction to credibility theory. On the other hand, rigorous proofs are given whenever they are informative, short and easy. So the practitioner will be enabled and encouraged to build his own model when models presented in the book do not fit his actuarial problem.

The main chapters of the book are credibility theory, bonus systems, the risk process, the accumulated claim distribution, claims reserves, and utility theory. These subjects include the most interesting and promising subjects of recent research in risk theory. In the chapter on credibility theory, the simple standard model as well as the Bühlmann–Straub model and the Hachemeister regression model are presented. The bonus malus chapter is concerned with the computation of the premium for each bonus class when the transition rules are fixed. In the chapter on the risk process, the claim number process is modelled by a general non-homogeneous counting process, while for the accumulated claims process, a homogeneous compound Poisson process is used. Ruin probabilities and the adjustment coefficient are introduced, and the optimal reinsurance results of Waters are presented. The chapter on the accumulated claim distribution includes stop loss inequalities, recursive algorithms of Panjer, Edgeworth expansions and normal power approximations, and solvency control problems solved using normal power approximations. In the claims reserves chapter we find the indispensable chain ladder method as well as Taylor's separation method and Straub's burning cost model. The chapter on utility theory is concerned with inequalities between zero utility premiums and between expected utilities with different compensation functions, and the definition of Pareto-optimality.

The whole book is clearly written and easy to understand. Perhaps due to personal taste, I do not believe in normal power approximations, I prefer approximations by compound Poisson distributions. Sundt's statement "...in practical applications the NP-approximation seems to perform very well" can be criticized. Approximations are needed only in those cases in which the exact computation is impossible, e.g. for large portfolios. The performance of an approximation can, however, be checked only if one can compare the exact values and the approximations, or if theoretical error bounds are available. For the normal power approximation, no theoretical error bounds exist (not even for the simple case of identically distributed risks). Theoretical error bounds exist for the approximation with compound Poisson distributions. Nevertheless, normal power approximations are frequently used in practice.

This book by Dr. Sundt can be recommended as an introductory textbook into modern risk theory for students as well as practitioners. Risk theory is a young and rapidly growing discipline with possible applications in life and non-life insurance, and in non actuarial branches. I think that Dr. Sundt's book will pave the way for the future application of new methods in risk theory like hierarchical credibility models or the estimation of accumulated claims distributions.

In his preface Dr. Sundt reports on the way his book was created. The preface is written in a very modest manner. There is not enough space to include the first paragraph of the preface here but in brief it states that the book was created by pure chance. I am pleased that this rare event actually happened, and all readers of the book will be pleased, too.

CHRISTIAN HIPP

GUIDELINES TO AUTHORS

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

   Hans Buhlmann,
Mathematik, ETH-Zentrum,
CH-8092 Zurich, Switzerland

   D Harry Reid,
Eagle Star Insurance Company Ltd,
1 Threadneedle Street, London EC2R 8BE, England

   or to the Co-Editor

   Alois Gisler,
Winterthur, Schweizerische Versicherungs-Gesellschaft,
General Guisan-Strasse 40,
8401 Winterthur, Switzerland

   Submission of a paper is held to imply that it contains original unpublished work and is not being submitted for publication elsewhere
   Receipt of the paper will be confirmed and followed by a refereeing process, which will take about three months

2. Manuscripts should be typewritten on one side of the paper, double-spaced with wide margins Typists should check a recent issue of ASTIN BULLETIN for style

3. Papers should be written in English or in French

4. The first page of each paper should start with the title, the name(s) of the author(s), and an abstract of the paper as well as some major keywords An institutional affiliation can be placed between the name(s) of the author(s) and the abstract

5. Footnotes should be avoided as far as possible

6. The author should aim at having as little 'white' on the page as possible. To avoid breaking the text alignment, one could use the solidus instead of built-up fractions, transposed row vectors instead of column vectors, etc Free-standing formulae should be well separated from the text

7. Upon acceptance of a paper, any figures should be drawn in black ink on white paper in a form suitable for photographic reproduction with lettering of uniform size and sufficiently large to be legible when reduced to the final size

8. References should be arranged alphabetically, and for the same author chronologically Use a, b, c, etc to separate publications of the same author in the same year For journal references give author(s), year, title, journal (in italics, cf point 10), volume (in boldface, cf point 10), and pages For book references give author(s), year, title (in italics), publisher, and city

   Examples

   BARLOW, R E and PROSCHAN, F (1975) *Mathematical Theory of Reliability and Life Testing* Holt, Rinehart, and Winston, New York.
   JEWELL, W. S. (1975a) Model variations in credibility theory In *Credibility Theory and Applications* (ed P M KAHN), pp 193-244, Academic Press, New York
   JEWELL, W S (1975b) Regularity conditions for exact credibility *ASTIN Bulletin* **8**, 336-341

   References in the text are given by the author's name followed by the year of publication (and possibly a letter) in parentheses

9. The address of at least one of the authors should be typed following the references

Continued overleaf

Guidelines to Authors *continued from inside back cover*

10. Italics (boldface) should be indicated by single (wavy) underlining Mathematical symbols will automatically be set in italics, and need not be underlined unless there is a possibility of misinterpretation
Information helping to avoid misinterpretation may be listed on a separate sheet entitled 'special instructions to the printer' (Example of such an instruction Greek letters are indicated with green and script letters with brown underlining, using double underlining for capitals and single underlining for lower case)

11. Authors will receive from the publisher two sets of page proofs together with the manuscript One corrected set of proofs plus the manuscript should be returned to the publisher within one week Authors may be charged for alterations to the original manuscript

12. Authors will receive 50 offprints free of charge Additional offprints may be ordered when returning corrected proofs A scale of charges will be enclosed when the proofs are sent out