Contents

Editorial and Announcements

Editorial (Actuarial Education Urbi et Orbi?) 163
The 3rd AFIR Colloquium 165
The XXIV ASTIN Colloquium in Cambridge 167
Obituary on Charles A. Hachemeister 183

Articles

K. K. Aase
Equilibrium in a Reinsurance Syndicate; Existence, Uniqueness and Characterization 185

Th. Mack
Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates 213

H. H. Panjer, S. Wang
On the Stability of Recursive Formulas 227

D. C. M. Dickson, H. R. Waters
Gamma Processes and Finite Time Survival Probabilities 259

Workshop

L. A. Roberts
Weighted Mortality Rates as Early Warning Signals for Insurance Companies 273

G. C. Taylor
The Incidence of Risk under Credit Insurance 287

B. Sundt
Improved Error Bounds for Bertram's Method 301

Ceuterick
EDITORIAL POLICY

*ASTIN BULLETIN* started in 1958 as a journal providing an outlet for actuarial studies in non-life insurance. Since then a well-established non-life methodology has resulted, which is also applicable to other fields of insurance. For that reason *ASTIN BULLETIN* has always published 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. Since the foundation of the AFIR section of IAA, i.e. since 1988, *ASTIN BULLETIN* has opened its editorial policy to include any papers dealing with financial risk.

*ASTIN BULLETIN* appears twice a year (May and November), each issue consisting of at least 80 pages.

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

MEMBERSHIP

ASTIN and AFIR are sections 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 of ASTIN receive *ASTIN BULLETIN* free of charge. As a service of ASTIN to the newly founded section AFIR of IAA, members of AFIR also receive *ASTIN BULLETIN* free of charge.

SUBSCRIPTION AND BACK ISSUES

*ASTIN BULLETIN* is published and printed for ASTIN by Ceuterick s.a., Brusselsestraat 153, B-3000 Leuven, Belgium.

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

The current subscription or back issue price per volume of 2 issues including postage is BEF 2 500.

Back issues up to issue 10 (= up to publication year 1979) are available for half of the current subscription price.

INDEX TO VOLUMES 1-20

The Cumulative Index to Volumes 1-20 is also published for ASTIN by Ceuterick at the above address and is available for the price of BEF 400.

Copyright © 1993 Ceuterick
EDITORIAL AND ANNOUNCEMENTS

ACTUARIAL EDUCATION URBI ET ORBI?

Actuarial science was born in the 17th century:
- Fermat and Pascal had successfully solved the 100 year old "problème des pistolets" and thus defined the rules for a solid calculus of probability.
- John Graunt had published the descriptive statistical analysis of demographic data in his "Observations made upon the Bills of Mortality".
- The Huyghens brothers had given a probabilistic interpretation of Graunt's tables.
- Edmond Halley had explicitly constructed a mortality table based on the yearly numbers of deaths observed in the city of Breslau.
- Jan de Witt had introduced the compound interest technique for the value of an annuity.

In the next century, all these elements were put together and became the fundamental pillars for the sound management of life insurance. Those who did so can rightly be called the first actuaries: James Dodson, Richard Price, William Morgan.

Although the title of actuary was used by the chief executive (and chief mathematician) of the Equitable Life Assurance Society from 1762 onwards, it seems that the title did not come into general use in British life insurance companies until the early 19th Century. The foundation of the Institute of Actuaries in 1848 and the Faculty of Actuaries in 1856 finally created the professional bodies which justified this title. In other countries the development was even slower, but one might say that at the turn of the 19th to the 20th century most European countries as well as the United States and Canada had founded their actuarial organization with the more or less similar purpose of allowing their members to use actuarial titles.

At present, it is general international practice to call each member of any national actuarial organization an "actuary". As long as the actuary only works in his "home market" one can live with this nomenclature. However, right now more and more national barriers—in particular for economic activities—disappear. For the actuary this means that his profession grows far beyond national frontiers.

Who is then such an "international" (in the full sense of the word) actuary? How does one become such an international actuary?

Typically, the modern actuary is an expert in only one of several fields (life insurance, pensions, general insurance...), but unless we want to split up the profession into several disjoint subgroups, it is essential that the actuary has a general understanding of all actuarial activities. This general understanding has also to become international, since the international actuary must understand what his colleagues brought up in other educational systems and working in different legal and cultural environments are doing.
This modern, internationally oriented, actuary needs a very sound education, which gives him a basis on which he can develop his actuarial skills, and also the ability and desire to be constantly improving those skills. It is counterproductive to aim at a uniform actuarial curriculum. These rather general goals of actuarial education are better attained by a diversity of educational systems.

It is a gross simplification, but one that we shall nevertheless make, to characterise the present systems of actuarial education in two categories: the professionally oriented systems, common in the Anglo-Saxon world, run by the professional actuarial associations; and the university oriented systems common in continental Europe, which emphasise more the mathematical foundations of actuarial science. Both systems have their defenders, who may feel that those educated in the other system are not "real actuaries", in the sense in which they understand the term.

The world needs both systems. If the professional practitioners lose touch with the ever-expanding mathematical foundation of their subject, they will find that they are no longer the technical experts their predecessors were. On the other hand, if actuaries devote themselves solely to the ivory towers of mathematical complexity, they will find that others start to provide the practical advice that insurance companies, pension funds and other financial organisations require.

The Groupe Consultatif (the liaison group of actuarial associations in the European Communities) has begun to compare in detail the educational systems within the European Communities. Even this is an immense task, made harder by different languages, different methods of teaching (lectures, text books, correspondence courses...) and the fact that those making the comparisons have already full-time jobs elsewhere.

The diversity of educational systems, which we believe should continue, needs, however, to be associated with explicit codes of professional conduct. The actuary must know the limitations of his (or her) own expertise, and must accept the consequences whenever he gives advice to others who may rely on it.

We do not need a streamlined international education system for actuaries, but we do need internationally compatible and enforceable ethical codes of conduct. Whether this is organised through the International Actuarial Association or through some other international actuarial organisation, which takes a professional rather than a scientific role, we hope that the diversity of actuarial education systems is recognised and preserved, while still allowing those actuaries, who wish to carry the responsibility, the opportunity to offer their expertise throughout the world.

HANS BÜHLMANN
ALOIS GISLER
HARRY REID
DAVID WILKIE
Rome, the Eternal City, was the scene for the 3rd AFIR International Colloquium, which took place there from 30 March to 2 April 1993. In such a location, the quandary of those who wondered whether to explore the classical antiquities, the churches of all ages, the innumerable art galleries, or just the space and peace of the Borghese Gardens nearby as an alternative to the Scientific Sessions, was understandable.

The hospitality and enthusiasm of our Italian hosts was overwhelming. A few Italians, indeed, were so enthusiastic to carry their visitors' luggage that they refrained from saying where they were carrying it to; but even that could be overlooked in the city of la dolce vita.

Each of the visitors will have taken home their own special memories. Many of the accompanying persons found the visit to Assisi most memorable. For me, the two special events were the flagwaving display at the Castello Odescalchi, where the closing dinner took place, and the concert in the Palazzo della Cancelleria, which Scots would be interested to know had been the home of Henry Stuart, Cardinal of York, the younger brother of Prince Charles Edward Stuart.

But the purpose of our visit to Rome was not to see the sights and be royally entertained, but to study the subject matter pertaining to "actuaries of the third kind". The previous colloquia, in Paris in 1990 and in Brighton in 1991, were organised differently from each other, and the Rome Colloquium in the Excelsior Hotel was yet again different. Authors were each given a few minutes to present their papers, and there was then general discussion of each group of papers. But this meant rather little discussion of each individual paper. The British, who perhaps enjoy questioning each other's work, are organising a single day seminar in London in November, at which the British papers presented to the Colloquium will be discussed again.

The three largest groups of participants were from Italy, France and the United Kingdom, and the most numerous papers came also from those countries; 17 from the United Kingdom, eight each from France and Italy. It is a generalisation, but I felt that one could characterise the British (and some American) papers as being written by practitioners who had problems for which they were seeking solutions. Those from Italy were contributed generally by university professors, who had theoretical solutions and were looking for practical problems to apply them to. Those from French authors seemed to reconcile the two, demonstrating sound mathematics applied to practical problems; does the fact that many of the French authors work for banks who are in the thick of market making, rather than insurance companies, have any bearing on this?

Besides discussion of the 59 papers, there were five invited lectures. Six had been planned, but one of the proposed speakers had just been appointed to a
high position in President Clinton's administration; this shows the Committee's good judgement in selecting the best. Fortunately Phelim Boyle and Hans Bühlmann and the three Italian speakers, Professors Pressacco, de Felice and Moriconi had not been elevated out of our sight.

The content of an hour's lecture on a mathematical subject is often best appreciated when the printed version is studied again at leisure later. The Italian organisers propose that the text of all five lectures will be published. In particular I shall find Professor Moriconi's work on "Analysing default-free bond markets by diffusion models", in which he separated out a yield curve for "real" interest rates, a yield curve for nominal interest rates, and a "yield curve" for prospective inflation, worthy of careful further study, particularly in connection with British index-linked stocks.

Phelim Boyle gave a most elegant and simple explanation of risk-neutral probabilities, which I hope will become widely copied. And Hans Bühlmann discussed, *inter alia*, the apparent paradox that, when investment returns are stochastic, the expected value of a unit at the end of one period is not equal to the present value of the expected return. In symbols: $E[1/(1+i)] \neq 1/E[1+i]$.

With certainty, the normal actuarial relationship that $v = 1/(1+i)$ holds. But to give a specific stochastic example, if $1+i$ is lognormally distributed, $E[1+i] = \exp(\mu + \sigma^2/2)$ whereas $E[1/(1+i)] = \exp(-\mu + \sigma^2/2)$. But if we set aside now the latter quantity, then at the end of one period our expectation is not unity but is $\exp(\sigma^2)$. Is it not therefore correct to set aside only $\exp(-\sigma^2/2)$?

It would be invidious to pick out a single paper as contributing the most to the sum of actuarial knowledge. All deserve careful study. Each student will find something that increases his or her understanding of the complex subject of investment mathematics and its actuarial implications, that stimulates him to further research, or that provides him with an answer, or a way to an answer, to a problem he has been seeking to solve. Colloquium papers are not just for those who attend. They remain as a record of the latest current ideas on the subject for others to study. The Transactions of the 3rd AFIR International Colloquium are worthy successors to the papers of the first two.

We look forward to the 4th International AFIR Colloquium in Orlando, 20-22 April 1994, which will no doubt differ from and be as valuable as its predecessors.

David Wilkie
The XXIV ASTIN Colloquium was held in St. John's College, Cambridge, with most participants staying in College and the colloquium sessions being held in the Palmerston Room of the recently constructed conference suite. The setting of the College, straddling the river Cam and looking out on to the green expanse of the backs, provided a beautiful backdrop for our discussions, formal and informal, and it seemed to be generally agreed that the lack of full hotel facilities was a small price to pay for the privilege of holding the colloquium in this unique environment.

The colloquium was attended by about 200 participants from 25 different countries and 61 accompanying persons.

The colloquium began with a reception in New Court Cloisters followed by a buffet in Hall and this set the standard of conviviality for the following four days as well as providing the first opportunity for informal discussions.

Scientific Programme

The scientific programme began with the opening ceremony and a formal introduction from John McCutcheon delivered in both English and French. He referred to the international nature of the profession, the increasing range of problems with which we are presented, and the wide variety of mathematical techniques required to solve them.

Invited Lectures

The first session consisted of an invited lecture given by Sir Brian Corby on the topic “Wider Horizons for Actuaries”. Summaries of this lecture, and of the invited lecture by Hans Bühlmann on the topic “Claims Reserves – Theory & Practice” are appended to these notes. The full text of Hans Bühlmann's lecture was subsequently circulated to participants and copies may be obtained from the conference department of the Institute of Actuaries.

Session A: Rating (Pure)

Ten papers were presented on this topic, and Stewart Coutts presented a summary of them, in which he classified them into three groups.

The first group of papers cover the analysis of portfolio segments. Coutts, Devitt & Shah develop a general insurance profit testing model equivalent to that used successfully in life assurance. They take the methodology further than most other authors in that they analyse information within a portfolio using GLIM. Dengsoe & Larsen consider rating the customer rather than the
product and conclude that the past claims experience of a customer for one line of business can be a guide to future experience not only in that line but other lines also. This has marketing as well as actuarial implications. Lemaire looks at the effect of introducing a high deductible as an alternative to a bonus-malus system. It is concluded that a high deductible increases the variability of claim payments but produces a fairer rating system for most policyholders. Orros & Pettengell present a simple interactive underwriting model for the premium rating of personal lines insurance business where the geographic area of the insured is an important insurance risk factor. The use of maps produced from a geographical information system is a key feature of the technique, which is intended as a practical premium rating tool for underwriters.

The second group of authors look at macro rating problems. Holm & Hoyland use a reserving package called ICRFS to evaluate the motor third party bodily injury risk in Norway. The importance of the technique is that instability in the payment year trend (inflation) can be identified and investigated. Krieter covers the use of the partnership clause in reinsurance treaties, considering the relationship between the cedent and the reinsurer on relatively short tail business and how profits and losses will emerge over time. He investigates how a fairer division of the profit or losses can be achieved. Slee considers the rating methodology for a class of business where no data is available. He suggests that the process should consist of five stages: hypothesis, measure, model, what if?, predict. The model must be dynamic in that it can be changed easily and must be readily communicated to the user.

The third group of papers are concerned with credibility theory and are in the main very theoretical. Dannenburg looks at the classical Bühlmann model and investigates the estimation of the credibility factor. The results are interesting in that the simulation models show that using the usual numerical procedures the credibility factor is under-estimated. Kling looks at the De Vylder model and shows that the Taylor expansion can be used iteratively to derive an approximation which under certain conditions converges to the best linear estimate. Schnieper addresses the problem that in standard credibility theory the structure parameters are estimated directly from the data, giving a posterior mean which is often biased. His approach is to treat the unknown parameters as random variables and to estimate simultaneously the a posteriori mean and the structure parameters.

Session B: Reinsurance

John Ryan presented a summary of the eleven papers on this topic, dividing them into four groups.

The first group covered Rating and Related Issues. Benktander develops a Stop Loss Rating Formula which is simple, elegant and practical. It is applicable for rating of a limited layer effectively as the difference between unlimited layers, although this is not generally considered to be an advisable method in practice. While stop loss has a limited market it does give the largest reduction in variance for a certain amount of risk premium. However,
unlimited stop-loss is not usually a practical proposition. Hurlimann derives a
distribution-free stop-loss premium formula for diatomic risks. He shows how
to guarantee a given return based on the assumption of additivity of stop-loss
reinsurance premiums and using means and standard deviations. The method
cannot be adapted to limited layers without more work. Kaas examines
relations between stop-loss premiums and variances. In the absence of reliable
data approximate methods like the translated Gamma-approximation become
viable. The paper offers much practical advice and useful techniques and
outlines some pitfalls. However, the variance adjustment works only for certain
adjustment points. Mirantis, Ryan and Salvatori propose a "marriage"
between the actuarial approach which essentially uses past loss data to derive
loss distributions to predict the future and an engineering approach which uses
fault or event trees in place of the decision trees in decision analysis. Emphasis
was made on the use of other disciplines apart from pure actuarial.

The second group on Pragmatic Problem Analysis comprised one paper, by Hirase which looks at the system for catastrophe reserves effectively used in
Japan, its problems and proposals for a new scheme. The system is of a kind
rarely seen in other countries. The question was raised as to how robust the
system would be to major catastrophes. In answer to a question on the tax
position it seems that there are some, but not complete, tax concessions.

The third group was on London Market and Other Market Issues. Craighead's contribution is an interesting paper giving background information on
the London market, with the intention of being practical. Devising a reinsurance
programme is a classic chicken and egg situation. The position now is that
there is no LMX-on-LMX market, any first tranche LMX being written for net
account, and high rates of premium are being charged for outwards reinsur-
ance. Sanders' paper on Catastrophe Excess of Loss Reinsurance contained
some interesting statistics. Current changes in the market are that there is no
more spiral, reinsurers are writing to aggregate exposures, rates have risen to
more accurately reflect risks, and the perception in the UK is that the major
risk is flood rather than windstorm. Today's scenario is that rates are
determined by 2 or 3 leaders, capacity has been severely reduced by large losses
and insolvencies, and direct rates are slower to react than reinsurance rates.
For reserving, Craighead curves are difficult to apply and there is no real
substitute for a full exposure analysis. The Hindley and Smith paper is on the
very important subject of financial reinsurance. Estimation must be made of
the value for money at the commencement of each year of account to ensure
equitable accounting treatment under the "clean sheet" principle, and equity
between names in Lloyds' syndicates should involve smoothing. The main
questions arising from this subject are whether financial reinsurance is worth-
while and whether Lloyds' syndicates should be allowed to purchase such
contracts. Buyers are often unaware of exactly what they are buying. A buyer
of financial reinsurance should reserve for all liabilities that might arise under a
contract. The FASB standards proposed in the USA are too rigid and could
produce nonsensical results, e.g. contracts specifically designed to overcome the
regulations. Financial reinsurance could be particularly useful when other
forms of reinsurance are unavailable, e.g. due to under-capacity in the market.

The fourth group of papers was on Mathematical Problem Analysis. The conclusion from DAHL's paper is that the share of a reinsurance treaty that optimises the expected utility equals the share that optimises the excess premium revenue over the zero-utility premium. This is an important result and is a useful formula subject to its being able to be evaluated. KREMER had developed a theory based on PML's. PML may be considered to be a property concept rather than a mathematical one and hence there is a question mark over the value of the results given the elaborate development. CENTENO investigated the effect of the retention limit on the risk reserve using the probability of ruin as a criterion. The topic is believed to be worthy of extensive analysis with more generalised assumptions. The conclusion, that the initial reserve is not in general an increasing function of the retention, having a minimum under fair assumptions, was variously described as "curious" and "amazing".

**Session C: Rating (Reserving)**

Eight papers were presented on this topic, and Philip Taylor presented an introduction and summary aimed at stimulating discussion. He classified them into two groups, according to whether they are concerned with traditional reserving models or more mathematical models.

The first group of papers re-examines traditional reserving models. SVENSDEN develops a reserving method which uses the life assurance technique of commutation columns. The reserve is calculated as a proportion of the total cost, the proportion depending only on the age of the claims. Changes in reserves may be analysed in terms of risk and interest components. AJNE looks at the conditions under which two sets of data may be combined without the results of a chain ladder calculation being distorted i.e. the chain ladder projections are additive. Necessary and sufficient conditions for this property are given, as are sufficient conditions for inequality. VERRALL considers the application of state space modelling to the chain ladder linear model in order to allow the development parameters to vary with accident year. The data is used in a recursive way so that there is an assumed relationship between the data in successive accident years. The paper by MACK, which was not circulated in advance, deals with the question of the stochastic model underlying the chain ladder. It is demonstrated that the underlying model is in fact a distribution free stochastic model rather than the loglinear approximation which has been used by several authors, and that these two models represent different philosophies of the claims process.

The second group of papers is concerned with research into new statistical methods of claims reserving. The first paper, by NORBERG, addresses the problem of drawing inferences from the data about the probabilistic laws governing claim size and claim development and using these as input to a model based on a marked Poisson process. Formulae for obtaining outstanding claims reserves from the model are derived. HESSELAGER also develops a model
based on the marked Poisson process, examining the various states such as IBNR, reported and settled. In order to compute outstanding claims reserves using this multi-state model, information on the frequency of transitions between states and the average payment per transition is required and must be derived from the data. Renshaw gives an overview of the potential of Generalised Linear Models (GLMs) for modelling the claims process in the presence of rating factors. Specific attention is focused on the variety of modelling distributions which can be used to model the claim frequency and claim severity components. Partrat examines the problem of modelling claim frequency when a claim may have several segments, such as Motor accidental damage and bodily injury, and illustrative examples are given. Poisson and mixed Poisson distributions are used in the modelling process.

Session D: Reporting

This Session was introduced by Colin Czapiewski. Whereas actuarial reporting is necessary and prevalent in life and non-life insurance it is only in life assurance that such reporting occurs widely in a statutory form. The system has worked well in life insurance and is clearly a lead that we as non-life actuaries can follow, showing that we too have something unique to offer. There has been considerable discussion in the UK as to whether there should be a statutory requirement for an Appointed Non-life Actuary but as yet no decision has been made. Non-life actuarial reporting can range in scope from an opinion on the quantum of technical claims reserves at year-end on business already written to responsibility for an insurance company being financially sound in its entirety, including whether new business rates are adequate.

Three distinct areas were considered in preparing subject matter for this Session:

— What the role of the actuary should be,
— To discover what progress other countries had made in actuarial reporting,
— The role of actuarial associations in different countries, especially as regards actuarial reporting.

Historically insurance supervisors have used simple methods to calculate minimum solvency requirements. More sophisticated methods have now been developed. RBC (Risk Based Capital), DST (Dynamic Solvency Testing), and simulation work carried out in the UK and Scandinavia have shown how actuaries are looking to improve their skills and to assist the supervisors. RBC methods, as used in the USA, involve considerable assumptions as to the methodology and to the parameters to be used. Some of the results from using these methods have been quite unexpected. Criticisms of RBC as used in the USA include:

— Risk factors initially calculated based on worst-case scenario rather than most likely,
— Small companies penalised by special additional loadings,
— Contracts where premiums are adjusted based on loss experience are not given due credit,
— Asset factors are those that were developed for life assurance,
— Insurers may be forced to perform to specifically meet RBC requirements
and this may not be in the best interests of policyholders.

Palmgren and Berg emphasise the benefit that many parties derive from
good actuarial reporting. They point to uniformity of standard reporting
including definitions and parameters used and concentrate on the experience,
exposure and economic aspects of actuarial reporting. The role of the actuary
will become far more important as insurance becomes more complex and
international.

Should actuaries from all countries get together to maintain higher and
consistent reporting standards?

The papers presented show the current situation in Eastern Europe, France
and the UK. Brief reports on actuarial reporting in Nigeria and Bulgaria (very
little!) were presented, together with the current situations in the USA and
Australia.

Varga describes the development of insurance in Hungary, including certain
aspects of the State Insurance Supervisory Authority and that part of the new
insurance law defining the role of the chief Actuary.

Daykin concentrates on the development of the actuarial profession in
Central and Eastern European countries. In the course of privatisation, the
insurance industry has faced severe problems. The role of the actuary in
helping to solve these problems is of great interest.

The paper by Bellando translates as State Control in France for Insurance.
The role of the Commissioner Controller is to constantly oversee insurance
companies and is essentially an actuarial role. The system appears to work
satisfactorily.

Scurfield describes the current UK situation and its history since the start
of non-life actuarial involvement about 25 years ago. Much has happened in
the UK over the last two or three years including statutory responsibility in
limited fields within Lloyd’s and the USA NAIIIO involvement. Guidance to
actuaries is provided by the Institute and Faculty. Statutory actuarial reporting
would provide a number of benefits.

In Canada the route has been towards DST. Some questions are whether
these methods should be approached by simulation or deterministically, has
Canada, got it right, or has the move to a regulatory role for Canadian
actuaries been too quick? The position in Canada is believed to be a possible
guide to the future for many countries. In general every insurance company
federally registered in Canada is required to have an appointed actuary. There
is no difference in the appointed actuary’s statutory responsibilities whether it
be for a life or non-life company. The new act imposes an obligation on the
Actuary to report on the year-end financial position and the regulators have
said they will request future financial condition reports for some non-life
companies (using DST). The Canadian Institute of Actuaries is pushing
aggressively to have regulators adopt discounted reserves simultaneously with
an explicit provision for adverse deviation.
This session gives an opportunity for actuaries to be involved in the most practical of matters, their communications with the outside world. As a profession we are on the verge of having a major input to the insurance industry world-wide. We must not miss this opportunity. We must bring added value to the insurance industry and be a credit to our profession.

Session E: Speaker’s Corner

Seven papers were presented in this session, including three which had not been circulated prior to the colloquium. As might be expected, these papers were diverse in their subject matter.

Gerber and Shiu show that the Esscher transform is an efficient technique for valuing derivative securities if the logarithms of the prices of the primitive securities are governed by certain stochastic processes with stationary and independent increments. Formulae are given for the pricing of European call options, and it is shown that the celebrated Black-Scholes option-pricing formula can be derived using this method. Klimkiewicz shows that Bühlmann’s classical credibility model is equivalent to a classical linear model and uses results relating to the estimators of variance components in linear models to show that the generally used estimator is optimal in the class of unbiased invariant quadratic estimators. Labie, Geerardyn and Goovaerts consider a modified Brownian motion process as a model for the surplus of an insurance portfolio. They take account of the effect of control activities forcing the surplus level upwards and the influence of dividend payments pushing the surplus downwards. The distribution of the surplus at time $t$ is derived under the assumption of no ruin. Parker presents two approaches to the modelling of interest randomness, namely modelling the force of interest accumulation function and the modelling of the force of interest directly. The expected value, standard deviation and coefficient of skewness of the value of immediate annuities are presented as illustrations. Boskov and Verrall give a method for premium rating by postcode area, based on spatial models in a Bayesian framework. A wide range of models within this class are suitable for use with insurance data, wherever there is a geographical area effect. Kreps and Steel describe a stochastic planning model for an insurance company, where the variables are connected by simple econometric equations whose form and parameters are generated by the data. The model gives surplus requirements as a function of both risk appetite and management scenarios. As a by-product, a stochastic model of loss development is generated. Petauton investigates the lower bounds for the distribution function of a positive random variable.

ASTIN General Assembly

G. Willem de Wit, Netherlands, who had been a member of the ASTIN Committee since 1985 and whose merits and work for ASTIN were acknowledged by the chairman, expressed his desire to resign. Bouke Posthuma,
Netherlands, was elected as his successor. Furthermore James MacGinnitie, USA, was welcomed as a new member and as one of the appointed IAA delegates on the ASTIN Committee.

It was announced that the next ASTIN Colloquium would take place in Cannes, France, on 11-15 September 1994. The main topics would be

- the financial stability of a non-life insurer;
- risk selection and the setting of premium rates;
- the analysis of major variations in loss experience.

Some discussion took place on the language difficulties which might be faced by those whose native language was not French, and ways of overcoming this would be investigated.

The following ASTIN Colloquium was expected to take place in Belgium in 1995, probably immediately following the Centenary International Congress to be held in Brussels.

The task force considering possible new formats for colloquia would continue its work, and would take account of suggestions made by delegates.

A prize had been instituted by the CAS to commemorate the work of Charles A. Hachemeister, with the intention of encouraging North American actuaries to become more familiar with ASTIN literature. It would be awarded to the colloquium or bulletin paper which was considered most beneficial to North American actuaries.

Social Programme

On Monday evening, participants and their accompanying persons were invited to an organ recital in St. John’s College Chapel, followed by dinner in the college dining hall.

On Tuesday there was a full day excursion which began with a walking tour of the Cambridge colleges. Coaches then arrived to take us to the Great Barn at Chilford Hall, a restored barn complete with vineyard. Lunch was served accompanied by the local wine, which was also available for purchase if desired. The afternoon offered a scenic ride through the Suffolk countryside to the medieval wool village of Lavenham, and the opportunity to explore the village at leisure before returning to Cambridge.

On Wednesday evening a drinks reception in St. John’s College preceded the gala dinner and dance in the Corn Exchange, which offered good food and wine, a convivial atmosphere, and music and dancing late into the night. John Martin, President of the Institute of Actuaries, gave a witty and entertaining after dinner speech.

Between or after the conference sessions, various delegates were spotted enjoying other traditional Cambridge pursuits, notably punting and croquet.

Sincere thanks are due to the organising committee and the staff of the Institute of Actuaries for their unstinting efforts to make the colloquium such a success both scientifically and socially.
Having recognised that even in July the British climate cannot be relied upon, the organisers thoughtfully arranged for each delegate to be equipped with an umbrella, courtesy of a well known firm of consulting actuaries. This forethought was appreciated during the walking tour of Cambridge, but the umbrellas also became quite a talking point, as will be seen from the following poem contributed by Bjorn Sundt:

What shall I do with my umbrella
When entering a plane
Or climbing out of a train?
Say, what shall I do with my umbrella
It's always on my mind
Where shall I find
A place to put my umbrella?

On registering they tell us
That we would need umbrellas.
They gave us blue and white ones,
But I'm not sure they're right ones.
I have some sort of feeling
That they are not appealing.
I'd rather call them ghastly.
They're long and Tillinghastly.
Although I might be wrong,
I'll sing my present song.

What shall I do with my umbrella?
It's so long, it's so bright.
I can't get that bloody thing out of my sight.
So what shall I do with my umbrella
For heaven’s sake
Why do they make
Such an enormous umbrella?

CAROLINE BARLOW, GRAHAM LYONS

APPENDIX A

"Wider Horizons for Actuaries" — Sir Brian Corby

In considering the possibility of expanding the areas of actuarial involvement, it is appropriate to reflect on how long it has taken to achieve the current level of involvement in non-life insurance. It is also appropriate to give a word of warning—the presentation of our results and ourselves is important, and it is vital that we should express ourselves in terms which are clearly understood, avoiding mathematical complexity. However interesting the theoretical probability distributions might be, it is necessary to address practical questions—how likely is an event to happen, what is the cost if it happens, and what other consequences will there be?
Actuarial science is unique in combining probability and compound interest, and this should be of wide applicability, since all human activity involves risk. Why therefore is there not wider actuarial involvement, and what opportunities exist for actuaries to widen their sphere of influence?

In the 6th annual lecture to the Geneva Association given in 1982, Professor Raymond Barr posed the following question: "We are living in a world where uncertainty is growing and risks are becoming ever more extensive. At the same time we note an upward trend in the demand for security. To what extent can the demand for security be met by the state without slowing down or hindering the adaptations made necessary and inevitable by the far reaching changes in the world?" This question concerns a choice for society, between security and progress.

At the same time, there is pressure for greater competition, more choice and a better deal for the consumer, but greater competition and choice implies greater risk, and there is thus a conflict between competition and risk aversion. There can be no absolute resolution of the conflict—it must be managed by the government, politicians and the public generally.

Against this background, a better understanding of risk—and of the fact that a risk free environment is impossible—is vital to society. Without this understanding, the process of protecting the consumer, particularly when the supply of goods or services goes wrong, can easily lead to those goods or services ceasing to be available.

It is interesting to consider people's perceptions of risk. The Royal Society study group on risk defined risk as “the probability that a particular adverse event occurs during a stated period of time or results from a particular challenge” where an adverse event is “an occurrence that is producing harm”. In relation to attitudes to risk, they concluded that “scientific determinism has replaced religion and magic as the way in which many ordinary people explain adversities and one consequence is that there is generally a search for causes and an attribution of responsibility or blame. These attributions lead to important differences in perceptions and there is urgent need for research in this area”. A further finding was that “At the more strategic levels, risk management is an essentially political process informed by technical estimates” and it was recommended that both the public and technical specialists should be involved in discussion and in the regulatory process with a view to achieving a more balanced approach to the existence of risk.

People's perceptions of risk often reflect subjectivity rather than an unbiased probabilistic assessment. This is illustrated by a Swedish report which commented that between 75% and 90% of car drivers consider themselves to be driving better than the average driver!

It is certain that failures will occur in the future, as they have in the past. It is important to recognise the dilemma between freedom and the desire for security, and failure should not necessarily cause us to call for changes in the system, which may nonetheless have worked well.

Perceptions of aggregation of risk must also be considered. A single disaster affecting a large number of lives is far more likely to lead to political action than a large number of smaller incidents.
There is a clear need for a better educated public and for the application of common sense. Politicians and the public must accept that zero risk is not attainable, and it is not possible to guarantee that adverse events "will not happen again".

It is important to recognise that there comes a point where the advantages of increased safety are not worth the restrictions which would be imposed. Risk is inseparable from a democratic society, since democracy implies choice which in turn implies risk. What is needed is the understanding and management of risk. The actuarial profession has a part to play in both of these aspects. It clearly has a role in relation to insurance, but its potential role is wider than that.

Insurance does not consist of selling security but rather of providing the means to manage uncertainty. Unfortunately, it appears that the limitations of insurance are not understood by the government or the public. One problem is that of moral hazard—the insurer and insured should be equally risk averse. Also, insurance may not be the solution where either the premium required would be too high to be socially acceptable or where the risk is very low but the cost if a loss occurs is very high. There is no counterpart to the "banker of last resort" for insurance, and this reinforces the view that certain risks are uninsurable.

The role of actuaries outside insurance will relate to the assessment of risk in other fields such as the evaluation of capital projects and, more significantly, the partial transfer of social security costs to the private sector. The private sector is likely to be keen to expand its business, but must be aware that there are limits to insurability. Enthusiasm must not lead to the acceptance of risks which cannot be insured against—particularly the uncertainties associated with inflation.

In conclusion, actuaries have a role to play in bringing about a better understanding of risk and contributing to a better informed society. At the same time, there is an opportunity to broaden the scope of the profession.

APPENDIX B

Lecture by Hans Bühlmann; Claims Reserves: Theory and Practice

Consider the origin of claims reserves. If an insurance claim occurs it entails a payment stream. The payment stream is the natural notion whereas claim amount and incurred claims are derived and somewhat artificial notions. From this it can be seen why reserving is a problem. It involves finding good methods for estimation of reserves. The functions used in non-life are comparable with the corresponding functions in life insurance.

The concept of a window is useful, the familiar triangle being the visible window with the variables outside the window requiring estimation. The actual paid and cumulative paid triangles are truly observed triangles, the incurred claims triangle itself being the result of estimation.

Considering the flow of claim payments for one accident year, what is required is to estimate the values outside the observed window, i.e. for future
years. This is only possible if we say what we mean by our estimate and its purpose. It is observed that only the sum of the future payments is needed, and that as this is a random variable, depending on the purpose of our prediction we may use different predictors, e.g. the mean, the median, a confidence interval, etc. What is being considered here is the conditional mean, i.e. given all information available to date.

The life actuary is much more interested in time points of payments using the axiom of "Time is Money". For him it is as equally important to calculate total future paid claims as the discounted value of future claims.

The following set of rules should be followed:

a) State clearly for what purpose the reserve calculation is needed.
b) State clearly how profits must be measured depending on the choice for the reserving methodology.
c) State clearly how the reserves should be financed.
d) Your managing director, having understood a), b), c), should choose between the undiscounted and discounted reserve options.

Ignoring the question of discounting, the essentials of any reserving methodology are:

— Describe the stochastic model underlying the calculations,
— Define the method of estimation,
— Discuss reliability of data and of results.

Concentrating on the first of these, using either the paid or incurred claim triangles, stochastic modelling is considered. Practically all such models can be written as regression models. The random variables $e^{(m)}_i$, the disturbances in the claims process for accident year $m$ at time $i$, are introduced, and used to define optimal solutions using expected square deviations. The models considered with their underlying stochastic models, are:

1. $AR_1$-models on cumulative claims only: Chain Ladder, London Chain Ladder and London Pivot,
2. $AR_1$-models including exposure: Simple Loss Ratio, Complementary Loss Ratio, Bornhuetter-Ferguson, Cape Cod,
3. Straightforward Extensions of 1: $AR_1$ replaced by $AR_p$, Explicit use of collateral knowledge,

We as actuaries should be particularly aware of the possible weakness of statistical methods, e.g.:

— It has been suggested that the effect of grouping of data is that any reasonable method applied to the two parts when a combined portfolio is split will yield a higher total reserve than the same method applied to the combined data. The results of some experiments to test this hypothesis have shown that it is not justified in practice.
— Many methods make implicit allowance for inflation. It is probably more reasonable to do this explicitly.
— It is impossible to make allowance in the reserving process for unforeseeable events.

Having covered how in theory claims reserves should be assessed, we then come to the practice. Reserving strategies can be used, for example, to either build hidden surplus or, alternatively, to show more favourable results than competitors. The gap between theory and practice is therefore between the actuary and the managing director.

Several principles learnt from life insurance are:

1. Ability to reproduce Claims Reserve valuations,
2. Normative Role of the Actuary,

The following proposals are put forward:

1. For the Actuary:
   — Accept the normative rôle we have to play in insurance,
   — Create a culture in claims reserving based on open communication.
2. For the Industry:
   — Create or extend a rating agency system on a world-wide basis,
   — Define non-government solvency standards, for solvency margins and reserve strength,
   — Extend the previous two proposals to all aspects of financial risk,
   — Support and encourage the rôle of the actuary as an independent moral and scientific force, on which one can build the long-term stability of insurance.

APPENDIX C

List of papers

Session A: Rating (Pure)

S.M. Coutts, E.R. Devitt & N. Shah
Financial Control in a General Insurance Operation

C. Dengsoe & C. Larsen
Customer-based Rating

J. Lemaire
High Deductibles Instead of Bonus-Malus: Can it Work?

G.C. Orros & C.T. Pettengell
An Interactive Underwriting Model for Premium Rating by Geographic Area
S. Holm & T. Hoyland
Rating the Risk of Bodily Injuries in Motor Insurance

F. Krieter
Rating the Partnership Clause

D. Slee
The Theory and Practice of Modern Rating Techniques

D. Dannenburg
Some Results on the Estimation of the Credibility Factor in the Classical Bühlmann Model

B. Kling
A Note on Iterative Non-Linear Regression in Credibility

R. Schnieper
On the Estimation of the Credibility Factor

Session B: Reinsurance

G. Benktander
A Stop Loss Rating Formula

W. Hürlimann
An Insurance Market Based Distribution free Stop-loss

R. Kaas
How to (and how not to) Compute Stop-loss Premiums in Practice

C. Miranthis, J. P. Ryan & L. Salvatori
Constructing a Reinsurance Programme for a Captive Insurance Company, Using Simulation Techniques

R. Hirase
Notes on Catastrophe Reserves on Non-life Insurance in Japan

D. H. Craighead
The Design of Catastrophe Reinsurance Programmes: Practical Problems

D. E. A. Sanders
When the Wind Blows - An Introduction to Catastrophe Excess of Loss Reinsurance

D. Hindley & A. Smith
Quantitative Analysis of Financial Reinsurance

P. Dahl
Expected Utility and Optimal Shares of Reinsurance Treaties

E. Kremer
A Probable Maximum Loss Theory

Maria de Lourdes Centeno
The Effect of the Retention Limit on the Risk Reference
Session C: Rating (Reserving)

O. A. Svendsen
A Deterministic Method for Claims Reserving in Non-life Insurance

B. Ajne
Additivity of Chain-ladder Projections Revisited

R. J. Verrall
Chain Ladder with Varying Run-off Evolutions

T. Mack
Which Stochastic Model is Underlying the Chain Ladder Method?

R. Norberg
Prediction of Outstanding Liabilities III: Parameter Estimation

O. Hesselager
A Markov Model for Loss Reserving

A. E. Renshaw
Modelling the Claims Process in the Presence of Covariates

C. Partrat
Compound Model for Two Dependent Kinds of Claim

Session D: Reporting

B. Palmgren & M. A. Berg
Experience, Exposure and Economics

T. Varga
The Development of Non-Life Actuarial Reporting in Hungary

C. D. Daykin
New Challenges for the General Insurance Actuary in Central and Eastern Europe

J. L. Bellando
Le Contrôle de l'Etat Exercé en France sur les Entreprises d'Assurance

H. H. Scurfield
Actuarial Reporting and Professional Standards – UK

Session E: Speakers Corner

H. U. Gerber & E. S. W. Shiu
Option Pricing by Esscher Transforms

T. K. Klimkiewicz
About Properties of Some Variance Estimators Used in Credibility Theory

E. Labie, J. Geerardyn & M. J. Goovaerts
Measuring the Effect of Control and Dividend on the Surplus
G. Parker
Two Stochastic Approaches for Discounting Actuarial Functions

M. Boskov & R. J. Verrall
Premium Rating by Geographic Area Using Spatial Models

R. E. Kreps & M. M. Steel
A Stochastic Planning Model for the Insurance Corporation of British Columbia

P. Petauton
Limites Inférieures de la Fonction de Répartition d'une Variable Aléatoire Positive
OBITUARY
Charles A. HACHEMEISTER

ASTIN lost one of its strongest contributors with the passing of Charles A. Hachemeister on September 9, after a year long fight with brain cancer.

Charlie was well known in the worldwide actuarial community. He was a fellow of the Casualty Actuarial Society, an Associate of the Society of Actuaries, and a member of the Swiss Actuarial Society. He served as a member of the ASTIN Committee from 1983 through 1988, on the Board of Casualty Actuarial Society from 1977 through 1979 and 1989 through 1991, and was an active contributor to Restin. As a Vice President at Prudential Reinsurance and then at F & G Re, he made many basic contributions to both the Actuarial Theory and Practice of Reinsurance. He authored numerous papers and presentations.

Because of his prominent roles in both the CAS and ASTIN, Charlie was always searching for ways to strengthen the relationship between the two organisations. He was instrumental in forming and was the first Chairman of the International Relations Committee of the CAS.

The CAS recently established an award for the ASTIN paper judged to be of most value to North American Actuaries. In recognition of Charlie's contributions, this has been named The Charles A. Hachemeister Award.

The international actuarial community, the Casualty Actuarial Society of America and ASTIN in particular have lost one of their great thinkers, a fascinating partner in any discussion, but above all a wonderful friend.

Jim Stanard
ARTICLES

EQUILIBRIUM IN A REINSURANCE SYNDICATE;
EXISTENCE, UNIQUENESS AND CHARACTERIZATION

BY KNUT K. AASE

Norwegian School of Economics and Business Administration
5035 Bergen-Sandviken, Norway

ABSTRACT

This paper attempts to give an overview of the pricing of risks in a pure exchange economy, where trade takes place at time zero and where uncertainty is revealed at time one. An economic equilibrium model under uncertainty is formulated, where conditions characterizing a Pareto optimal exchange equilibrium are derived. We present two sets of sufficient conditions for the existence of an equilibrium, and demonstrate how equilibria can be characterized through several examples. Uniqueness of equilibrium is also discussed. Special attention is given to the principal components that the premiums in a reinsurance market must depend upon. We also apply the general theory to the risk exchange problem between a policyholder and an insurer, and in particular we compute market premiums of the resulting optimal contracts.

It is emphasized throughout how the formulation of a competitive equilibrium, rather than merely a general risk exchange formulation, is of particular interest in deriving a well-defined and unique set of equilibrium premiums in an insurance market. The theory is put into a framework which is fruitful for extensions beyond the one-period case.

KEYWORDS

Reinsurance market; competitive equilibrium; uniqueness of premiums; Pareto optimality; risk exchange; private insurance; CAPM; risk tolerance; complete markets.

I. INTRODUCTION

The following model is interpreted as a reinsurance syndicate, in which I insurers trade among themselves. We take as given

(a) The preference of insurer \( i \in I = \{1, 2, \ldots, I\} \), \( \succ_i \), represented by expected utility \( E\{u_i(\cdot)\} \), where \( u'_i > 0 \) and \( u''_i < 0 \).

(b) The initial net reserve of insurer \( i \), represented by the random variable \( x_i \), \( i \in I \).

We assume that each \( x_i \in L^2(\Omega, \mathcal{F}, P) \) where \( (\Omega, \mathcal{F}, P) \) is the probability space on which all the \( x_i \)'s are defined, all the insurers agree on the probability
measure $P$ (homogeneous beliefs), and the events $\mathcal{F}$ on which $P$ is defined are generated by the I net reserves, which we sometimes call the initial portfolios, i.e., $\mathcal{F} = \sigma \{ x_1, x_2, \ldots, x_I \}$. This means that the uncertainty in the model is totally described by the initial portfolios $x_i$, $i \in I$. We briefly comment on the realism of homogeneous beliefs in reinsurance: The assumption about homogeneous beliefs appears reasonable for a reinsurance market, where trade is supposed to take place under conditions of umberrimae fidei, and no information is supposed to be hidden. Our pricing results for a reinsurance market are likely to influence premiums in the market for direct insurance as well. In the direct market the assumption about homogeneous beliefs seems more unrealistic. It is likely that the buyers of insurance have more information about the risk they try to cover, than the insurers. This asymmetric information gives rise to adverse selection. In addition, the buyers can directly or indirectly influence events so that the probability distributions of the risks are altered. This can happen since the insurer is usually unable to monitor the insured, and the phenomenon gives rise to moral hazard. Whereas the problem of moral hazard does not seem important in a reinsurance market, the problem of adverse selection may occur since the ceding company usually has more detailed information about the risks they underwrite than the reinsurers. It may of course be tempting for some direct insurer to sell some “bad risks” in the reinsurance market. In the long run this “practice” is not likely to pay off, since the reinsurance industry makes heavy use of a detailed rating system for insurance companies (i.e., Insurance Solvency International), and there exist penalties for such actions.

The competitive equilibrium (CE) that we shall demonstrate in this model, we claim to be of particular interest in insurance, where its importance has been partly overlooked. Insurers of today seem to be turning to finance markets and their models, often without the understanding of the most basic exchange economy that can be thought of, and which we think is of the utmost importance to general insurance markets; the syndicate described in this paper. The present model has also been a key motivation of much of the financial equilibrium theory which has dominated the literature of financial economics.

The usual formulation in insurance settings has been to derive the front of Pareto optimal (PO) risk exchanges, generally uncountable in number. This does not help to find unique premiums, as there will be one set of prices for each different Pareto optimal point. In order to find a well-defined set of premiums in this model, the budget constraints of the insurers must be employed. A well-posed model will then normally determine a unique set of equilibrium premiums subject to a normalization condition.

The paper is organized as follows: In Section 2 we present the economic model of uncertainty. Here we formulate one set of sufficient conditions, the Inada conditions, for the existence of a unique equilibrium, and we demonstrate some properties of this equilibrium. In Section 3 we demonstrate that the CE is Pareto optimal, and in Section 4 we present examples of how optimal sharing rules might look like, and what their market values are. Here we introduce a different set of sufficient conditions, called properness, for the
existence of an equilibrium, which turns out to be satisfied in the examples. We discuss when a syndicated market can restrict attention to proportional treaties, and when non-proportional treaties are needed. In the latter case, we argue that this has nothing to do with the market being “incomplete”, as has been suggested in the economic literature. At least this is a definition which we do not find fruitful. In Section 5 we demonstrate some properties of risk tolerances, and in Section 6 we risk adjust the probability measure in the present one-period framework. In Section 7 we present an insurance version of the capital asset pricing model, and in Section 8 we rewrite our results on portfolios to treat insurance premiums directly in this syndicate. In Section 9 we employ the results of Sections 2-4 to the general treatment of the risk exchange between a policyholder and an insurer, and in particular to the computation of market premiums of optimal contracts in such models. We end our exposition in Section 10 with some concluding remarks.

2. THE ECONOMIC MODEL OF UNCERTAINTY

In the market the I insurers exchange parts of their initial portfolios among themselves. As a result of these exchanges insurer $i$ obtains a final portfolio, represented by the random variable $y_i(x_1, x_2, \ldots, x_i)$. Market clearing requires that

$$\sum_{i \in I} y_i = \sum_{i \in I} x_i = x_M, \quad P\text{-a.s.,}$$

since the insurers only trade among themselves, where $x_M$ represents the “market portfolio”. If some allocation of risks $(y_1, y_2, \ldots, y_I)$ satisfy (2.1), it is called feasible. The premium functional we denote by $\pi(\cdot)$. In order to prevent arbitrage possibilities this must be a linear functional on $L^2(\Omega, \mathcal{F}, P)$.

As an illustration of this point, assume on the contrary that $\pi(y_1 + y_2) > \pi(y_1) + \pi(y_2)$ for two risks $y_1$ and $y_2$. Then one agent could insure the bundle $(y_1 + y_2)$ and reinsure separately $y_1$ and $y_2$. The cash flow at time zero equals $\{\pi(y_1 + y_2) - \pi(y_1) + \pi(y_2)\} > 0$. The cash flow at time one equals $-(y_1 + y_2) + y_1 + y_2 = 0$. This agent has no obligations at time one, so he has made a riskless profit at time zero. This is a money pump, or a “free lunch”, which is inconsistent with an economic equilibrium.

By the Riesz' representation theorem there exists some function $U' \in L^2(\Omega, \mathcal{F}, P)$ such that

$$\pi(x) = E\{xU'\}, \quad \forall x \in L^2(\Omega, \mathcal{F}, P).$$

Since $\mathcal{F} = \sigma\{x_1, x_2, \ldots, x_i\}$, it follows that $U' = U'(x_1, x_2, \ldots, x_i)$ is some Borel-measurable function of $(x_1, x_2, \ldots, x_i)$ (see e.g. TUCKER (1967), Th.1.1). (So far the prime on $U'$ is just a matter of notation. Later we show that under the present conditions $U'(\cdot)$ is also a derivative of some function as well.) The statement in (2.2) means, in economic terms, that the market is complete in the following sense (BORCH (1962)):
Definition 1:
A market model is complete if it assigns a unique value $\pi(x)$ to an arbitrary random variable $x \in L^2(\Omega, \mathcal{F}, P)$.

The economic theory of pricing of contingent claims started with Arrow's paper in (1953). Borch (1960-62-68) developed these ideas further, and below we present some of the elements of this theory. Consider the following problem:

\begin{equation}
\max_{y_i(x) \in L^1} E\{u_i\left(y_i(x_1, x_2, \ldots, x_I)\right)\}
\end{equation}

subject to the budget constraint

\begin{equation}
\pi(y_i) \leq \pi(x_i), \quad i \in I.
\end{equation}

In order to avoid bankruptcy problems we also assume that $y_i \geq 0$ a.s. An equilibrium is a collection $(\pi; y_1, y_2, \ldots, y_I)$ consisting of a price functional $\pi(.)$ and a feasible allocation $(y_1, y_2, \ldots, y_I)$ such that for each $i$, $y_i$ solves problem (2.3-4).

The market value of the portfolio cannot increase when exchanges are settled at market prices. The expected utility of the portfolio can however be increased by such exchanges, and this is the very purpose of reinsurance transactions.

The Lagrangian of this problem is

\begin{equation}
\mathcal{L}(y_i; \lambda_i) = E\{u_i\left(y_i(x_1, x_2, \ldots, x_I)\right) - \lambda_i(y_i - x_i)U'\}.
\end{equation}

For the purpose of the first result below, in addition to the assumptions (a) and (b) we now make the following three assumptions

(c) The derivatives $u_i'(x)$ satisfy

$$
\lim_{x \to -\infty} u_i'(x) = 0 \quad \text{and} \quad \lim_{x \to 0} u_i'(x) = +\infty.
$$

(d) The functions $x \to xu_i'(x)$ are all nondecreasing.

(e) The aggregate market portfolio $x_M \in [\delta, \Delta]$ almost surely for finite constants $\Delta > \delta > 0$.

The assumption $u_i'(0+) = +\infty$ guarantees that the constraint $x_M \geq 0$ will never be active, called the Inada condition. The condition $u_i'(x) \to 0$ as $x \to \infty$ can be thought of as a saturation effect. We now present a theorem giving sufficient conditions for the existence of a competitive equilibrium. Assumption (d) is sufficient for uniqueness. The theorem also characterizes the equilibrium.

Theorem 1:
Suppose assumptions (a)-(c) and (e) hold. Then there exists a CE characterized by

\begin{equation}
u_i'(y_i(x_1, x_2, \ldots, x_I)) = \lambda_i U'(x_1, x_2, \ldots, x_I), \quad i \in I, \ P\text{-a.s.},\end{equation}
where $\lambda_i$ are positive constants. If in addition (d) holds, then the CE is unique.

**Proof:** Existence of CE in infinite dimensional spaces under our conditions are shown by Duffie and Zame (1989). Uniqueness of an interior CE under the additional assumption (d) has been shown by Karatzas et al. (1988). As for the characterization (2.6), since the Bernoulli utility functions $u_i(\cdot)$ are concave, the program (2.3)-(2.4) is concave for each $i$. By the Saddle Point Theorem, if $(y_i, \lambda_i)$ is a saddle point of the Lagrangian for this program, $y_i(x)$ solves the given program for all $i$. Again because of concavity, the conditions (2.6) are the Euler equations of the maximization problem of the Lagrangian in $y(\cdot)$, which in the present situation are necessary and sufficient for the solution of this sub-problem, since the optimal solution happens to be interior by our conditions. Thus the equations (2.6) must hold.

**Remarks:**

— Uniqueness means relative to a normalization. In particular this means that if $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_I)$ and if $\lambda' = (\lambda'_1, \lambda'_2, \ldots, \lambda'_I)$ are two different vectors corresponding to a unique CE, then there exists a constant $c > 0$ such that $\lambda = c\lambda'$, $\pi(y; \lambda) = \pi(y; \lambda')$ and $y_i(x; \lambda) = y_i(x; \lambda')$, where $y_i(x; \lambda)$, $i = 1, 2, \ldots, I$, are the optimal sharing rules, or equilibrium allocations, corresponding to the vector $\lambda$, and $\pi(y; \lambda)$ stands for the corresponding pricing rule. Thus equilibrium premiums can be determined only up to a multiplicative constant, since there can always be a re-evaluation of currency; this is not going to affect, however, the way in which the insurers share the risks among themselves.

— The present economic interpretation of the function $U'(x)$ is that it represents the marginal utility of the market as a whole at the "portfolio point" $x$. Another common interpretation, especially in financial economics and in macroeconomics, is that $U'(x)$ represents the marginal utility function of some representative insurer, or even of some abstract central planner. A final interpretation is also possible; $U'(x)$ is the shadow price per unit of $P$-probability when $x(\omega) = x$ (we return to this in Section 3).

— In Section 4 we present a different set of sufficient conditions for the existence of an equilibrium. It turns out that these conditions are inconsistent with the Inada condition (c), but otherwise they appear to be less restrictive. Here we need the above conditions in order to secure an interior optimum.

Some immediate consequences of (2.6) are:

(i) $\frac{\partial U'(x_1, x_2, \ldots, x_I)}{\partial x_i} = \frac{1}{\sum_{r \in I} \lambda_r \frac{u_r''(y_r(x))}{u_r'(y_r(x))}}$, $i \in I$, P-a.s.
This follows from differentiating (2.6). Since the right-hand side above does not depend on \( i \), it follows directly that
\[
(2.7) \quad U'(x_1, x_2, \ldots, x_i) = U'(x_M), \quad P\text{-a.s.}
\]
so that only changes in the market portfolio \( x_M \) affects \( U' \). Similarly
\[
(i) \quad \frac{\partial y_i(x_1, x_2, \ldots, x_i)}{\partial x_j} = \frac{\lambda_i}{u''_i(y_i(x))} \sum_{r \in I} \frac{\lambda_r}{u''_r(y_r(x))}, \quad i, j \in I, \quad P\text{-a.s.}
\]
which again follows from (2.6). We notice that the right-hand side above does not depend upon \( j \). This means that the total derivative
\[
d y_i(x) = \frac{\partial y_i(x_1, x_2, \ldots, x_i)}{\partial x_j} \frac{d x_M}{d x_k}, \quad \text{for any } j, k \in I, \text{ and for all } i \in I, \quad P\text{-a.s.}
\]
so that only changes in the aggregate portfolio \( x_M \) affects the optimal final portfolios \( y_i \), i.e.
\[
(2.8) \quad y_i(x_1, x_2, \ldots, x_i) = y_i(x_M), \quad i \in I, \quad P\text{-a.s.}
\]
A consequence of this is that the syndicate members can hand in all their initial portfolios to a pool, and let the pool’s clerk distribute parts of \( x_M \) back to the syndicate’s members according to the optimal sharing rules \( y_i(x_M) \).

Here we remark that the consequences (2.7) and (2.8) could alternatively have been derived more directly from the Saddle Point Theorem.

3. PARETO OPTIMALITY

If feasible, an allocation \( y_1, y_2, \ldots, y_I \) is Pareto optimal if there is no feasible allocation \( z_1, z_2, \ldots, z_I \) such that
\[
E(u_i(z_i)) \geq E(u_i(y_i)) \quad \text{for all } i, \text{ with strict inequality for at least one } i.
\]
It is now easy to demonstrate that any competitive equilibrium allocation is Pareto optimal. In order to show this, let \( (U'(x_M), y_1, \ldots, y_I) \) denote the CE in Theorem 1, and suppose \( z \) is a feasible allocation which Pareto dominates \( y \). Then
\[
E(u_i(z_i)) > E(u_i(y_i)) \quad \text{for all } i \text{ with at least one strict inequality, say for insurer } j.
\]
Since
\[
E(u_j(z_j)) > E(u_j(y_j)),
\]
we know that \( \pi(z_j) > \pi(y_j) \). If for some \( i \) the quantity
\[
\delta = \pi(y_i) - \pi(z_i) > 0,
\]
we could let \( y_i^* = z_i + \delta U'(x_M)/\pi(U'(x_M)) \), from which
\[
\pi(y_i^*) > \pi(y_j).
\]
But then, since \( u_i \) is strictly increasing and \( U' > 0 \), we would have
\[
E u_i(y_i^*) > E u_i(y_j),
\]
which is impossible by the definition of an equilibrium. Thus \( \pi(z_i) \geq \pi(y_j) \) for all \( i \). Using this, we now have the contradiction
\[
\pi(\Sigma x_j) \geq \pi(\Sigma z_j) > \pi(\Sigma y_i) = \pi(\Sigma x_i),
\]
which proves the result.

When the optimal solution is interior, an alternative construction is the following: It is well known the Pareto optimal sharing rules are found by solving (see e.g. Borch (1960-62) or Wilson (1968))
\[
(3.1) \quad \max_{y_i(x_M) \in L^2} E \left\{ \sum_{i \in I} k_i u_i(y_i(x_M)) \right\}
\]
such that

$$\sum_{i \in I} y_i(x_M) = \sum_{i \in I} x_i = x_M, \quad P\text{-a.s.},$$

where $k_1, k_2, \ldots, k_I$ are arbitrary positive constants. In two dimensions there is a nice graphical illustration of this point. The Lagrangian of problem (3.1) is

$$(3.2) \quad \mathcal{J}(y_1, y_2, \ldots, y_I; \lambda(x_M)) = E \left\{ \sum_{i \in I} k_i u_i(y_i(x_M)) - \lambda(x_M) \sum_{i \in I} (y_i - x_i) \right\},$$

where the Lagrangian multiplier $\lambda(\cdot)$ is now a Borel-measurable function, so that $\lambda = \lambda(x_M)$ is an $\mathcal{F}$-measurable random variable.

**Theorem 2: (Borch's Theorem)**

Under our assumptions (a) and (b) on preferences, the Pareto optimal sharing rules $y_i(x_M)$ are characterized by

$$(3.3) \quad k_i u_i'(y_i(x_M)) = \lambda(x_M), \quad i \in I, \quad P\text{-a.s.},$$

where the $k_i$ are arbitrary positive constants.

**Proof:** We assume that $k_i$ can be chosen in such a way that the domains of the functions $k_i u_i(\cdot)$ have a nonvoid intersection. Then there exists at least one Pareto optimal treaty (see DU MOUCHEL (1968)). By the concavity of the Bernoulli utility functions $u_i(\cdot)$, our program is concave. If $(y_i, \lambda)$ is the saddle point of the Lagrangian in (3.2), $y_i$ solves the problem (3.1) since $\lambda(\cdot)$ is continuous (this latter property follows since any positive linear functional on $L^2$ is continuous). The saddle point must maximize the Lagrangian (3.2) in $y$, and this latter problem can be solved by the calculus of variations: Because of concavity of the $u_i$ for all $i$, a necessary and sufficient condition for this maximization is again given by the Euler equations. In this special case they are given by (3.3), since the derivatives of $y_i(x_M)$ with respect to $x_M$ are not entering the equations (3.2).

**Corollary 1:**

The competitive equilibrium of Theorem 1 is Pareto optimal.

**Proof:** By comparing (3.3) and (2.6-8), the result follows after simply identifying $U'(x_M)$ with $\lambda(x_M)$ and $k_i$ with $1/\lambda_i$. Alternatively, see the introduction to this section.
We may notice that the identification in the above proof also explains that the function $U'(\cdot)$ may be interpreted as a shadow price, which is exactly what $\lambda(\cdot)$ is. Finally we notice that the formulation (2.3) and (2.4) implies that the CE solutions $y_i(x_M), i \in I$, must satisfy individual rationality, since clearly the solution $y_i(x_M) = x_i$ is possible, where $x_i$ is square integrable and obviously satisfies (2.4).

4. existence and uniqueness of CE: examples

4.1. Introduction

Theorem 1 gives a set of sufficient conditions for a CE when premiums, as a result of a competitive equilibrium, are unique. Since equilibrium prices can be determined only up to a multiplicative constant, we should normally get unique premiums from the budget constraints after normalization. The family of solutions we get by varying the normalization constant will not affect the sharing rules, as will be demonstrated in the examples below. It turns out, however, that the conditions (c) and (d) for existence and uniqueness are far from necessary. After the examples we shall therefore present an alternative set of sufficient conditions for existence of CE. When the sharing rules are linear, it is possible to reach a Pareto optimum by an exchange of fractions of the initial portfolios. Linear sharing rules are optimal when the individual utility functions are members of the HARA class. In a reinsurance market this means that there should be no need for any other contract than the standard proportional reinsurance contract when this is true. Applied to a stock market, the assumption that the optimal sharing rules are linear implies that there should be no need for trading any other securities than ordinary shares (common stock). Non-proportional reinsurance and securities such as contingent claims and options both exist and are important, so we must conclude that the preferences of decision makers are at least so diverse that they cannot be represented HARA-utility functions only. For some reason many economists refer to a market in which it is impossible to reach a Pareto optimum through an exchange of proportions of the initial portfolio as an "incomplete market".

4.2. Illustrations

Example 1: Exponential utility.

Here

$$u'_i(x_i) = \exp\{-x_i/\alpha_i\}, \quad \alpha_i > 0, \quad i \in I.$$  

Notice that neither (c) nor (d) hold true here. Nevertheless we shall demonstrate both existence and uniqueness of an interior CE.

Borch's Theorem gives

$$k_iu'_i(y_i(x_M)) = U'(x_M), \quad i \in I, \quad P\text{-a.s.},$$
which leads directly to

\[
U'(x_M) = \exp \left\{ \frac{K - x_M}{\sum_{i \in I} \alpha_i} \right\}, \quad \text{where} \quad K = \sum_{i \in I} \alpha_i \ln k_i.
\]

Notice how the marginal utility of the market depends upon the parameters \( \alpha_i \) from the individual preferences, and the positive constants \( k_i \). The latter can finally be determined from the budget constraints (2.4), which we return to below.

We also notice that the optimal sharing rules are given by

\[
y_i(x_M) = \frac{\alpha_i}{\sum_{j \in I} \alpha_j} x_M + \left( \alpha_i \ln k_i - \alpha_i \frac{K}{\sum_{j \in I} \alpha_j} \right), \quad i \in I, \quad \text{P-a.s.},
\]

which verifies that the sharing rules are linear in \( x_M \). The Arrow-Pratt measure of absolute risk aversion equals \( 1/\alpha_i \) for each insurer \( i \). Also the relative risk aversion is increasing in the net reserves for these insurers. The kind of treaty given in (4.2) seems common in reinsurance practice. Insurer \( i \) will hold a share \( \alpha_i/\Sigma \alpha_j \) of the total market, inversely proportional to his coefficient of absolute risk aversion. In order to compensate for the fact that the least risk-averse insurer will hold the larger proportion of the market, zero-sum side-payments, or fees, occur between the insurers. The last term in (4.2) represents these fees. The quotas are determined by the risk-aversion parameters only. Quota-share treaties with side-payments also occur when all the insurers have preferences represented by logarithmic utilities, quadratic utilities, as well as by power utility functions with the same exponent. For further details see LEMARIE (1990).

Let us for simplicity write (4.2) as

\[
y_i(x_M) = \frac{\alpha_i}{\sum_{j \in I} \alpha_j} x_M + \beta_i
\]

where

\[
\beta_i = \left( \alpha_i \ln k_i - \alpha_i \frac{K}{\sum_{j \in I} \alpha_j} \right).
\]

Employing the budget constraints (2.4), we determine these constants as follows

\[
\beta_i = \frac{E \left\{ x_i \exp \left\{ - \frac{x_M}{A} \right\} - \frac{\alpha_i}{A} x_M \exp \left\{ - x_M \frac{x_M}{A} \right\} \right\}}{E \left\{ \exp \left\{ - \frac{x_M}{A} \right\} \right\}}, \quad i = 1, 2, \ldots, I,
\]
where \( A = \Sigma \alpha_i \), so that the sharing rules are now uniquely determined. Moreover, the ray \( (k_1, k_2, \ldots, k_i) \) is unique modulo a multiplicative constant. Normalizing so that

\[
K = \sum_{i \in I} \alpha_i \ln k_i,
\]

we obtain the unique ray as follows

\[
k_i = \exp \left\{ \frac{\beta_i}{\alpha_i} \right\}^{\frac{K}{e^A}}, \quad i = 1, 2, \ldots, I.
\]

In the case where we have a riskless security in the economy in addition to the existing portfolios, it is natural to normalize so that \( E\{U'(x_M)\} = 1 \), in which case the normalization constant \( K \) is determined from

\[
e^{-\frac{K}{A}} = E\left\{ \exp \left\{ -\frac{x_M}{A} \right\} \right\},
\]

so that the unique vector of constants \( (k_1, k_2, \ldots, k_i) \) is given by

\[
k_i = \exp \left\{ \frac{\beta_i}{\alpha_i} \right\}^{\frac{K}{e^A}}, \quad i = 1, 2, \ldots, I.
\]

Finally the unique set of market premiums of the optimal portfolios \( y_i \) is given as

\[
\pi(y_i(x_M)) = \frac{\alpha_i}{A} \frac{E\{x_M \exp (-x_M/A)\}}{E\{\exp (-x_M/A)\}} + \beta_i = \frac{E\{x_i \exp (-x_M/A)\}}{E\{\exp (-x_M/A)\}}, \quad i \in I,
\]

i.e., by the Esscher premium principle of actuarial sciences (see BÖHLMANN (1980)).

We now present another example.

**Example 2: Logarithmic utility.**

Here

\[
u_i(x_i) = \ln (\beta_i + \alpha_i x_i), \quad \text{where} \quad (\beta_i + \alpha_i x_i) > 0 \quad \text{P-a.s.,} \quad \alpha_i > 0.
\]

The individual marginal utilities are given by

\[
u'_i(x_i) = \frac{\alpha_i}{\beta_i + \alpha_i x_i}, \quad i \in I, \quad \text{P-a.s.}
\]
and the absolute risk aversion and the relative risk aversion are both increasing with net reserve levels if $\beta_i > 0$. In this case neither condition (c) nor (d) hold. Borch's Theorem gives

\begin{equation}
\frac{k_i \alpha_i}{\beta_i + \alpha_i \gamma_i(x_M)} = U'(x_M), \quad i \in I, \quad \text{P-a.s.},
\end{equation}

which leads to

\begin{equation}
U'(x_M) = \frac{\sum_{i \in I} k_i}{\sum_{i \in I} \beta_i / \alpha_i + x_M}, \quad \text{P-a.s.}
\end{equation}

and the linear sharing rules

\begin{equation}
y_i(x_M) = \frac{k_i}{\sum_{j \in I} k_j} x_M + \left( \frac{k_i}{\sum_{j \in I} k_j} \sum_{j \in I} \frac{\beta_j}{\alpha_j} - \frac{\beta_i}{\alpha_i} \right), \quad i \in I, \quad \text{P-a.s.}
\end{equation}

Using the budget constraints (2.4) we obtain the unique ray $(k_1, k_2, \ldots, k_I)$ subject to $\sum k_i = k$, as

\[ k_i = k \mathbb{E} \left( \frac{x_i + \beta_i / \alpha_i}{A + x_M} \right), \quad i = 1, 2, \ldots, I, \]

where

\[ A = \sum_{j \in I} \frac{\beta_j}{\alpha_j}. \]

In the case when the normalization is $E\{U'(x_M)\} = 1$, then the constant $k$ equals

\[ k = \left( \mathbb{E} \left( \frac{1}{A + x_M} \right) \right)^{-1}, \]

so that the unique vector of positive constants $k_i$ is given by

\[ k_i = k \mathbb{E} \left( \frac{x_i + \beta_i / \alpha_i}{A + x_M} \right) \left( \mathbb{E} \left( \frac{1}{A + x_M} \right) \right)^{-1}, \quad i = 1, 2, \ldots, I, \]

Finally the market values of the optimal portfolios $y_i$ are given by

\[ \pi(y_i(x_M)) = k \mathbb{E} \frac{x_i}{A + x_M} \left( \mathbb{E} \left( \frac{1}{A + x_M} \right) \right)^{-1}, \quad i = 1, 2, \ldots, I, \]

which is a new "premium principle".

We now present an example where both the conditions (c) and (d) hold true.
Example 3: Power utility.

Here

$$u_i(x_i) = x_i^{\rho_i}, \quad \rho_i \in (0, 1), \quad i \in I.$$  

If condition (e) hold true, by Theorem 1 there exists a unique CE, which means that this model is complete by our Definition 1. Borch's Theorem gives

$$k_i \rho_i (y_i(x_M))^{\rho_i - 1} = U'(x_M), \quad i \in I,$$

and the optimal allocations are

$$y_i(x_M) = \left( \frac{1}{k_i \rho_i} \frac{U'(x_M)}{U'(x_M)^{\rho_i - 1}} \right)^{1/(\rho_i - 1)}, \quad i \in I,$$

so that from the market clearing condition we obtain the equation

$$\sum_{i \in I} \left( \frac{1}{k_i \rho_i} \frac{U'(x_M)}{U'(x_M)^{\rho_i - 1}} \right)^{1/(\rho_i - 1)} = x_M.$$

The normalization $E\{U'(x_M)\} = 1$ together with the budget constraints finally determine the constants $k_i$. As an illustration, consider the case where $I = 2$ and $\rho_1 = 1/2$, $\rho_2 = 3/4$. Here

$$y_1(x_M) = (k_1 \rho_1)^2 (U'(x_M))^{-2}, \quad y_2(x_M) = (k_2 \rho_2)^4 (U'(x_M))^{-4}.$$

Only the ratio between the two positive constants matter, so we can arbitrarily set $k_2 = 4/3$. The marginal utility of the market equals

$$U'(x_M) = \left( \frac{\sqrt{h + \sqrt{h + 4x_M}}}{2x_M} \right)^{1/2},$$

and the optimal allocations are

$$y_1(x_M) = \frac{1}{2} (\sqrt{h^2 + 4hx_M} - h), \quad y_2(x_M) = x_M + \frac{1}{2} (h - \sqrt{h^2 + 4hx_M}),$$

where

$$h = \left( \frac{\rho_1 k_1}{\rho_2 k_2} \right)^4.$$

The normalization $E\{U'(x_M)\} = 1$ finally gives the equation for the constant $h$ (or really $k_1$), in which case the unique CE is determined. It should be clear that this Pareto optimum can not be achieved by an exchange of proportional reinsurance contracts. Similarly, in a stock market this type of arrangement can not be reached by an exchange of common stock.
Let us finally check condition (e). Supposing that $x_M$ is uniformly distributed on the interval $[\delta, 1+\delta]$ where $\delta$ is some given parameter, we see that here is $\delta = 0$ allowed, since the integral $\int_0^\infty u^{-1/2} du$ converges. In this particular case (e) is too strong.

4.3. Uniform properness

We now turn to another set of sufficient conditions for the existence of an equilibrium.

Following MAS-COLELL (1986) and MAS-COLELL and ZAME (1991), we define an expected utility function $U(y) = E\{u(y)\}$ to be $x$-proper on $X = L^2_+ (\mathcal{F}, \mathcal{F}, P)$ (or uniformly proper) if there exists a scalar $e > 0$ such that for all $y$ in $X$, $|\alpha| > 0$ and $z$ in $X$, $U(y - \alpha x + z) \geq U(y)$ implies that $||z|| \geq \alpha e$. Here $||z|| = (E\{z^2\})^{1/2}$.

The interpretation is that the portfolio $x$ is desirable, in the sense that loss of an amount $\alpha x$ cannot be compensated for by an additional amount $\alpha z$ for any portfolio $z$, if $z$ is sufficiently "small". When preferences are convex, properness of $U$ at $y$ with respect to $x$ is equivalent to the existence of a premium functional $U'$ such that $\pi(x) = E\{z U'\} \geq \pi(y) = E\{y U'\}$ whenever $U(z) \geq U(y)$ and has the additional property that $\pi(x) > 0$. The portfolio $x$ in this definition is said to be extremely desirable for $U$. Thus, under risk aversion properness at $x_M$ is equivalent to the linear premium rule we know must exist, or individual properness at $x_M$ is equivalent to market supportability of $\pi$.

Now, it is known that properness of $U_i(y) = E\{u_i(y)\}$ at $x$ is equivalent to the assertion that the random variable $u'_i(x)$ satisfies $E\{(u'_i(x))^2\} < \infty$. A quasi-equilibrium is defined by the existence of a $U' \in X$, $U' \neq 0$, such that $\pi(x) = E\{x U'\} = \pi(y)$ and $\pi(v) \geq \pi(y)$ whenever $U_i(v) \geq U_i(y)$. A quasi-equilibrium is an equilibrium if $U_i(v) > U_i(y)$ implies that $\pi(v) > \pi(y)$ for all $i$. This latter property holds at a quasi-equilibrium if $\pi(x_i) > 0$ for all $i$. The following result is of interest in our model of an insurance market.

Thereom 3:

Suppose our conditions (a) and (b) hold and that there is any allocation $z \geq 0$ with $\Sigma z_i = x_M$ $P$-a.s. If $x_M > 0$ $P$-a.s. and $E\{(u'_i(z_i))^2\} < \infty$ for each $i$, then there exists a quasi-equilibrium.

— The proof of this theorem can be adapted from MAS-COLELL and ZAME (1991).

— Consider now the examples above, and suppose the theorem holds for $z = y$, the optimal allocation. The conditions for properness are then

$$E\{\exp (-2 y_i(x_M)/\alpha_i)\} \leq \infty, E\left\{ \frac{\alpha_i^2}{(\beta_i + \alpha_i y_i(x_M))^2} \right\} < \infty, E\{y_i(x_M)^{2(\alpha_i - 1)}\} < \infty.$$
for the Examples 1, 2 and 3 respectively. As an illustration, suppose that $x_M$ is exponentially distributed with probability density $f(x) = \lambda \exp\{-\lambda x\}$, $x \geq 0$, $f(x) = 0$, $x < 0$. The properness condition in Example 1 is then equivalent to $A = \Sigma x_i > 0$, which is indeed one of our assumptions. In Examples 2 and 3 the properness requirement does not seem to add any new restrictions to the ones that are already naturally present.

— One advantage with our conditions (a)-(e) is the interior solution they provide, which gives us the characterization (2.6). Theorem 3 does not rule out corner solutions.

— Uniform properness is incompatible with the condition $u_i'(0^+) = +\infty$.

— Uniform properness was used in a model of a reinsurance market in Aase (1990).

5. RISK TOLERANCE

Here we demonstrate a simple consequence of Borch's Theorem:

$$\frac{u_i''(y_i(x_M))}{u_i'(y_i(x_M))} \frac{y_i'(x_M)}{U'(x_M)} = \frac{U''(x_M)}{U'(x_M)}, \quad i \in I, \quad P\text{-a.s.,}$$

which follows from differentiating (2.6). Equation (5.1) can alternatively be written

$$y_i'(x_M) = \frac{1}{R_i(y_i(x_M))} \frac{1}{R(x_M)}, \quad i \in I, \quad P\text{-a.s.,}$$

where $R = -\frac{U''}{U'}$, and $R_i = -\frac{u_i''}{u_i'}$ stand for absolute risk aversion. Since $\Sigma_i y_i'(x_M) = 1$, we see that

$$\frac{1}{R(x_M)} = \sum_{i \in I} \frac{1}{R_i(y_i(x_M))}, \quad P\text{-a.s.}$$

The quantity $1/R$ is called the risk tolerance. The above result has been found by Borch (1985); see also Böhlmann (1980) for the special case of exponential utility functions. The result (5.3) says that in a Pareto optimum the risk tolerance of the market as a whole is equal to the sum of the risk tolerances of the participants. If one member is risk neutral, his risk tolerance will be infinite, and hence that of the market. This may be interpreted as saying that in a Pareto optimum all risk should be carried by the risk neutral participants. We can also easily derive the following

$$\frac{\partial y_i(x_M)}{\partial x_M} = \frac{R(x_M)}{R_i(y_i(x_M))}, \quad i \in I, \quad P\text{-a.s.}$$

If all the syndicate's members are strictly risk averse, then $R_i > 0$, and $R > 0$ follows from (5.3), so that $y_i'(x_M) > 0$ a.s. from (5.4). This means that as the
market portfolio increases, all the insurers increase their portfolios in a Pareto optimum.

6. RISK ADJUSTMENT OF THE PROBABILITY MEASURE

The premium functional $\pi$ can alternatively be represented by a risk adjusted probability measure as follows: Suppose there exists a riskless security $x_0$ in the economy, and assume without loss of generality that $x_0(\omega) = 1$ $P$-a.s. We can then normalize such that $E(U'(x_M)) = 1$, as we have suggested earlier. Suppose that $P[U'(x_M) > 0] = 1$. Define a new measure $P^*$ as follows:

\begin{equation}
P^*(A) = \int_A U'(x_M(\omega)) \, dP(\omega).
\end{equation}

Clearly $P^*(\Omega) = 1$ from our normalization assumption. Also it follows from integration theory that $P^*(\cdot)$ is countably additive, confirming that $P^*$ is a probability measure. Finally $P^*$ and $P$ are mutually absolutely continuous with respect to each other, meaning that if $P(B) = 0$ then $P^*(B) = 0$ and if $P^*(A) = 0$ then $P(A) = 0$ for any $A, B \in \mathcal{F}$. Using $P^*$ we can express the premium as follows

\begin{equation}
\pi(x) = E(U'(x_M)x) = \int _\Omega U'(x_M(\omega)) \, x(\omega) \, dP(\omega) = \int _\Omega x(\omega) \, dP^*(\omega) = E^*(x),
\end{equation}

where $E^*$ refers to the expectation operator under $P^*$. The interpretation is that the market premium can be computed using an altered probability measure $P^*$ corresponding to a world of market risk neutrality. We call $P^*$ the risk adjusted probability measure. Notice from (6.1) that the market's marginal utility $U'(x_M)$ corresponds to the Radon-Nikodym derivative of $P^*$ with respect to $P$, i.e.

\begin{equation}
U'(x_M) = \frac{dP^*}{dP}.
\end{equation}

This type of construction is of considerable importance in the time-continuous case (see e.g., AASE (1988-92-93).

Returning to the illustrations in Section 4.2, we now see that in general the Radon-Nikodym derivative depends on the preferences. This at least holds in equilibrium models. This fact should be contrasted with the literature on contingent claims analysis. In the arbitrage pricing theory, where the uncertainty is modeled by Ito-diffusions, this quantity is preference independent, which clearly does not hold when "jumps" can occur as in our model.
7. INSURANCE PREMIUMS

The foregoing has been formulated in terms of portfolios and market values of net reserves. To obtain market premiums of insurance contracts, we note that the net reserves of insurer $i$ consist of assets $a_i$, less of liabilities under the insurance contracts held by the insurer. Let the non-negative random variable $z_i(\omega)$ represent claim payments under the contracts if the state of the world becomes $\omega \in \Omega$, $i \in I$. Let the events be completely specified by $\mathcal{F} = \sigma(z_1, z_2, \ldots, z_I)$, so that the assets $a_i$ are riskless, and write

\begin{equation}
(7.1) \quad x_i = a_i - z_i, \quad i \in I.
\end{equation}

Now we have that

\begin{equation}
(7.2) \quad \pi(x_i) = a_i - \pi(z_i) = a_i - E\{U'(a_M - z_M)z_i\},
\end{equation}

where $a_M = \Sigma a_i$ and $z_M = \Sigma z_i$. We define the market disutility of claim payments by the function $V$, where

\begin{equation}
(7.3) \quad V(z_M) = U'(a_M - z_M).
\end{equation}

Clearly $V'(z_M) = -U''(a_M - z_M) > 0$ because of assumption (a) and (5.3). Formula (7.2) simply says that the market value of the insurer's portfolio is equal to his riskless assets less the market premium for insurance of the liabilities. This formula makes it easy to translate results expressed in terms of net reserve values into insurance premiums. Notice in particular that if for some portfolio $x_i$ the premium $\pi(x_i) < E(x_i)$, we get from (7.2) that the corresponding insurance premium satisfies $\pi(z_i) > E(z_i)$, so that the economic risk premium $\{\pi(z_i) - E(z_i)\}$ of this insurance contract is positive. After normalization, we find in general that the risk premium can be written as follows

\begin{equation}
(7.4) \quad \pi(z_i) - E(z_i) = \text{cov} \{z_i, V(z_M)\}, \quad i \in I.
\end{equation}

Since the marginal disutility of the market increases as the aggregate claims in the market increase, from (7.4) we may be tempted to believe that for claims which are positively correlated with $z_M$, the risk premium is positive, and for claims which are negatively correlated with $z_M$, the risk premium is negative. Both these cases make perfectly sense in a rational reinsurance market with risk averse insurers. However, there exist joint distributions for $z = (z_1, z_2, \ldots, z_I)$ under which this result may not hold true. Covariances are measures of linear statistical dependance, and can only be considered as a good measure of "stochastic association" under multinormality. In insurance an assumption of jointly normally distributed claims is usually not very realistic. Among other things can claims only take on non-negative values. We are therefore reluctant to use the nice results obtainable from an assumption of multinormality in insurance. Here we cite Harald Cramér (1930) who wrote: "...in many cases the approximation obtained by using the normal function is not sufficiently good to justify the conclusions that have been drawn in this way". In the last section of the paper we nevertheless briefly discuss multinormality.
8. RISK EXCHANGE BETWEEN A POLICYHOLDER AND AN INSURER

The problem of risk exchange between a buyer of insurance and an insurer has been extensively studied under varying conditions in insurance economics, and some of the contributions can also be found in the actuarial literature. By restricting attention to the buyers problem only, MÖSSIN (1968) showed that if the compensation \( c(x) = \alpha x \) is received by the policyholder if the damage amounts to \( x \), where \( 0 < \alpha < 1 \) is a constant, and if the premium paid is \( \alpha p \), then if \( p > E x \) it is never optimal to take full coverage. Borch later modified this, and considered instead a premium \( p = \alpha E x + c \), where \( c \geq 0 \) is some constant. He showed, simply using Jensen's inequality, that \( \alpha^* = 1 \) is optimal if it is rational for the risk-averse customer to buy insurance. The constant \( c \) he interpreted as administrative costs. ARROW (1974) used Borch's original risk-exchange model of (1960-62), and found that a policy with a deductible is optimal. His premium contains a fixed percentage loading, which has later been interpreted as a special example of a cost function by RAVIV (1979), who analyzed the problem for general cost functions, using the maximum principle.

Here we remark that a loading is perhaps more naturally associated with an economic risk premium. HOLMSTRÖM (1979) analyzed the problem under moral hazard, and showed that this gives rise to deductibles. Moral hazard is clearly a problem in this particular kind of risk exchange. ROTHSCILD and STIGLITZ (1976) considered the case with imperfect information, and demonstrated deductibles for low-risk individuals in a very simple model, and TOWNSEND (1979) established deductibles under a certain kind of non-observability, where there is a cost involved by verification of the true state. LANDSBERGER and MEILIJSON (1990), on the other hand, explained deductibles in insurance from another perspective, by the use of preferences derived from so called star-shaped utility functions. MOFFET (1979) used Borch's Theorem directly on the risk exchange problem that we discuss below.

In this section we want to demonstrate that the risk exchange model of this paper can be used to establish some simple, yet general results, still abstracting from the problems caused by asymmetric information and moral hazard. These results, we claim, constitute the natural benchmark from which refinements should be obtained. In particular we are interested in the form of the premium functional in this situation, derived from (2.2).

To this end we consider a policyholder with initial wealth \( w_1 \), utility function \( u_1 \) satisfying conditions (a) and (b). Against a premium \( p \) the insurer offers a policy that reimburses the policyholder an amount \( I(x) \) if a claim of amount \( x \) occurs. The insurer has initial wealth \( w_0 \), and his utility function we denote by \( u_0(\cdot) \) satisfying \( u_0'(\cdot) > 0, u_0''(\cdot) \leq 0 \). A natural constraint on the compensation function \( I(x) \) is

\[
0 \leq I(x) \leq x \quad \text{for all } x.
\]

Ignoring this constraint for the moment, a direct application of Borch's Theorem to the present sharing arrangement gives

\[
(8.1) \quad u_0'(w_0 + p - I(x)) = (k_1/k_0) u_1'(w_1 - p - x + I(x)) .
\]
Differentiating (8.1) with respect to $x$ leads to

$$
\frac{\partial I(x)}{\partial x} = \frac{(k_1/k_0) u''(w_1 - p - x + I(x))}{u_0''(w_0 + p - I(x)) + (k_1/k_0) u''(w_1 - p - x + I(x))}.
$$

Using (8.1), we get directly

$$
\frac{\partial I(x)}{\partial x} = \frac{u_0''(w_0 + p - I(x)) + u''(w_1 - p - x - I(x))}{u_0'(w_0 + p - I(x)) + u'(w_1 - p - x + I(x))},
$$

which can be written

$$
\frac{\partial I(x)}{\partial x} = \frac{R_1(w_1 - p - x + I(x))}{R_0(w_0 + p - I(x)) + R_1(w_1 - p - x + I(x))},
$$

where $R_1$ and $R_0$ again stand for the measures of absolute risk aversion. If both parties are risk averse, then from (8.3) we see that

$$
0 < I'(x) < 1 \quad \text{for all } x \geq 0.
$$

Letting $I(0) = 0$, the mean value theorem implies that

$$
0 < I(x) < x \quad \text{for all } x \geq 0.
$$

This means that the Pareto optimal sharing rule involves a positive amount of coinsurance, or full coverage is not Pareto optimal.

Notice that policies with a deductible can not be Pareto optimal. This follows since $I(x) = 0$, $x \leq d$, $I(x) = x - d$, $x > d$ has $I'(x) = 0$, $x \leq d$ and $I'(x) = 1$, $x > d$, both violating (8.4). This holds quite generally without using any constraints on the compensation function $I(x)$.

Referring to the literature cited above, policies with a deductible can only be Pareto optimal in models where one or more of the following are included; costs, moral hazard, asymmetric information, non-observability or alternative preferences (e.g., star-shaped utility).

**Example 1**: (Exponential utility).

Suppose $u_1(w_1) = 1 - \exp\{-aw_1\}$, $u_0(w_0) = 1 - \exp\{-bw_0\}$ for two positive constants $a$ and $b$. In this case $R_0 = b$ and $R_1 = a$, so the absolute risk aversions are constants and independent of wealth levels. It now follows directly from (8.3) that $I'(x) = a/(a+b)$, or $I(x) = ax/(a+b) + c$, where $c$ is an integration constant. If $I(0) = 0$, $c = 0$. In this case if $R_1 = a$ is large compared to $R_0 = b$, $I(x)$ is approximately equal to $x$, so that full coverage is then approximately Pareto optimal. In practice this seems reasonable, since the absolute risk aversion of the policyholder is usually large compared to that of the insurer.
From (8.3) we see that the conclusion of this example also holds quite generally, i.e., if \( R_1 \) is very large as compared to \( R_0 \) for all input values, then full coverage is approximately Pareto optimal. Quite generally, if we tried to solve the risk exchange problem in this section imposing the natural constraints on \( I(x) \), the application of the maximum principle would yield the same conclusions as above: The Pareto optimal deductible is zero in the absence of operating expenses (RAVIV (1979)).

Turning to the premium, the problem of determining \( p \) is usually overlooked or ignored in the above kind of analyses, where \( p \) is simply assumed to be given as “a positive number”.

Suppose we use the pricing principles of Section 2 of this paper, and apply them to the present “mini-market”. We would like to answer the question of how the resulting equilibrium-based premium \( p \) depends on the parameters of the problem. First we need to derive the shadow price. Using Borch’s Theorem we get

\[
(8.6) \quad k_1 u_1'(w_1 - p - x + I(x)) = U'(w_1 + w_0 - x)
\]
and

\[
(8.7) \quad k_0 u_0'(w_0 + p - I(x)) = U'(w_1 + w_0 - x).
\]

The budget constraints of the two parties are

\[
(8.8) \quad \pi \{ y_0(x_M) \} = \pi \{ x_0 \}
\]
and

\[
(8.9) \quad \pi \{ y_1(x_M) \} = \pi \{ x_1 \}.
\]

Here \( y_0(x_M) = w_0 + p - I(x) \), \( y_1(x_M) = w_1 - p - x + I(x) \), \( x_M = w_1 + w_0 - x = w - x \), \( x_0 = w_0 \) and \( x_1 = w_1 - x \). Using (8.8) we have \( E \{ (w_0 + p - I(x)) U'(x_M) \} = E \{ w_0 U'(x_M) \} \), and since

\[
(8.10) \quad p = \pi \{ I(x) \} = E \{ I(x) U'(x_M) \},
\]
we obtain that the by now familiar normalization \( E \{ U'(x_M) \} = 1 \) must hold.

Consider the following example:

**Example 2**: (Exponential utility, continued).

Using the results of the above example and of Example 1 in Section 4, we get the following: The shadow price equals \( U'(w - x) = \exp \{(K - w + x)/A \} \) where \( K = (\ln k_1)/a + (\ln k_0)/b \) and \( A = 1/a + 1/b \). From the normalization we find the constant \( K = w - A \ln \{ E[\exp(x/A)] \} \). Furthermore, from Example 1 we get, in the case where \( I(0) = 0 \), that the market premium \( p \) is given by

\[
(8.11) \quad p = E \{ I(x) U'(x_M) \} = \frac{a}{a + b} \cdot \frac{E \{ x \exp(x/A) \}}{E \{ \exp(x/A) \}}.
\]

As an illustration, suppose that \( x \) is exponentially distributed with parameter \( \lambda \), so that \( \text{Ex} = 1/\lambda \). Then the simple formula \( p = a/[\lambda (a + b) - ab] \) obtains,
where the parameter \( \lambda > 1/A \). Notice that \( p \) increases with \( Ex \) and with \( a \). If \( \lambda < a \), then \( p \) increases with \( b \), with the opposite result if \( \lambda > a \). Notice that an increase in \( R_1 = a \) has here two effects, both working in the same direction: First the absolute risk aversion of the policyholder increases, and second the coverage increases; so we would expect a large premium \( p \) in both cases. An increase in \( R_0 = b \) implies on the one hand less coverage, but on the other hand the insurer becomes more averse towards risk. These two mutually competing facts explain the more complex comparative statics for \( b \).

The risk neutral case can be studied by letting \( b \to 0 \). Then \( p \to Ex \) follows from (8.11) by the monotone convergence theorem, i.e., we obtain the usual “actuarial fair” premium in the limit. Alternatively we could try the characterization in Section 2 directly with \( u_0(w) = b + cw \), \( c > 0 \), a constant. It is then straightforward to show that the shadow price \( U'(x_M) \equiv 1 \), again leading to premium given by the “principle of equivalence” above. (Formally the latter derivation is not valid when \( u''_0(w) = 0 \) for all \( w \).) ○

Examples 1 and 2 are somewhat specialized in that the absolute risk aversion is independent of wealth. In general we should also expect the premium to depend on the aggregate level of wealth \( w \) in the market. This is indeed of importance in actual markets where insurance contracts are traded at market prices. Consider the following example:

**Example 3:** (Power utility).

Suppose \( u_0(w) = u_1(w) = w^p \) where \( p \in (0, 1) \). In this case the shadow price is

\[
U'(x_M) = \frac{(w - x)^{p-1}}{\left( \frac{1}{(pk_0)^{1/(p-1)}} + \frac{1}{(pk_1)^{1/(p-1)}} \right)},
\]

which becomes, after the standard normalization

\[
U'(x_M) = \frac{(w - x)^{p-1}}{E\{(w - x)^{p-1}\}}.
\]

The Pareto optimal sharing rule satisfies

\[
\frac{\partial I(x)}{\partial x} = \frac{w_0 + p - I(x)}{w - x},
\]

depending, as we see, on the premium \( p \). Solving this differential equation under the condition \( I(0) = 0 \) gives

\[
I(x) = \frac{w_0 + p}{w} x,
\]

i.e., full coverage is only Pareto optimal if \( p = w_1 \). The present problem is
well-posed for the above utility functions only if \( x \leq \min (w_0, w_l) \) \( P \)-a.s. Since \( p < w_l \) must generally hold, coinsurance results. The premium \( p \) must satisfy

\[
p = E\{xU'(x_M)(w_0 + p)/w\},
\]

which leads to

\[
p = \left( \frac{w}{E\{xU'(x_M)\}} - 1 \right)^{-1} w_0.
\]

Notice how the premium \( p \) in general depends on the wealth level \( w \). It is seen that unless the wealth of the customer is too large, i.e., when \( w_l < E\{xU'(x_M)\} \), the premium decreases as \( w_0 \) increases as well as when \( w_l \) increases, whereas the premium increases as a function of \( w_0 \) when \( w_l > E\{xU'(x_M)\} \). In general the premium is a decreasing function of \( w \). This is in accordance with the general observation that the premiums tend to decrease as the "capacity" (= \( w \)) in the market increases and vice versa.

In the limiting case where \( p \to 1 \), \( U'(x_M) \to 1 \) \( a.s. \) and \( p \to w_0 Ex/(w-Ex) \) by the dominated convergence theorem. In the limit, approaching risk neutrality, the optimal compensation scheme is

\[
I(x) = \frac{w_0 x}{(w-Ex)},
\]

costing its actuarial fair value \( E(I(x)) = w_0 Ex/(w-Ex) \).

As an illustration, suppose that \( x \) is uniformly distributed on \( (0, w_l) \), where \( w_0 > w_l \). The premium is then

\[
p = \frac{w_0 [w (w^p - w_0^p)/\rho - (w^{p+1} - w_0^{p+1})/(\rho+1)]}{(w^{p+1} - w_0^{p+1})/(\rho+1)},
\]

which depends on the aggregate wealth \( w \) of the two parties, their attitude towards risk as measured by \( \rho \) and the reserves \( w_0 \) of the insurer. As \( \rho \to 1 \), this expression is seen to converge to

\[
Pl = \frac{w_l w_0}{2(w-w_l/2)},
\]

which is exactly \( E[I_1(x)] \), where \( I_1(x) = (w_0 + p_1)x/w \) is the optimal sharing rule for this particular premium \( p_1 \).

9. AN INSURANCE VERSION OF THE CAPITAL ASSET PRICING MODEL

We now discuss the case when \( x = (x_1, x_2, \ldots, x_t) \) is jointly multinormally distributed. As noted before, this case has limited applicability in insurance economics. However some of the results in this model remain true even if the assumption of normality is dropped. The first problem we encounter is to find a set of sufficient conditions for the existence of a competitive equilibrium. Our earlier theorems can not be directly applied here, since \( x \) can take on negative values with positive probability. Nielsen (1990) has a set of sufficient conditions for the existence of equilibrium in a CAPM-model in financial economics. In his model the investor has a utility function \( U(\mu, \sigma) \) which is a function of the mean and the standard deviation of the total portfolio return. Mean-variance behavior is consistent with expected utility maximization with general utility functions if the returns follow the distributions described by Chamberlain (1983) and Owen and Rabinowitch (1983), which include the multinormal distribution. In the present model the optimal allocation may not
be a linear function of \( x_M \), in which case \( y \) is not necessarily multinormally distributed. Expected utility maximization with general utility functions can not, in our model of an insurance market, in general be represented by mean-variance utility functions \( U(\mu, \sigma) \), unless the utility functions happen to be members of the HARA-family with the same cautiousness. In this latter case Nielsen's sufficient conditions are possibly appropriate in our model.

For the moment supposing an interior equilibrium exists, its characterization is then straightforward. Under our normalization assumption \( E\{U'(x_M)\} = 1 \), the premium functional can be written

\[
\pi(x_i) = E(x_i) + \text{cov}(x_i, U'(x_M)) \quad \text{for all the } x_i.
\]  

Here \( (\pi(x_i) - E(x_i)) \) is the economic risk premium of \( x_i \). From the assumption of multinormality it follows that

\[
\text{cov}(x_i, U'(x_M)) = E\{U''(x_M)\} \text{cov}(x_i, x_M) \quad \text{for all the } x_i.
\]  

Since (9.2) holds for each of the initial portfolios, clearly

\[
\pi(x_i) = E(x_i) + E\{U''(x_M)\} \text{cov}(x_i, x_M) \quad \text{for all the } x_i.
\]  

By summation over \( i \) we obtain

\[
\pi(x_M) = E(x_M) + E\{U''(x_M)\} \text{var}(x_M)
\]

from the linearity of the pricing functional \( \pi(\cdot) \) and from standard properties of the expectation and the covariance operators. Rearranging, we finally have the insurance version of the capital asset pricing model as follows:

\[
(9.3) \quad \pi(x_i) - E(x_i) = \frac{\text{cov}(x_i, x_M)}{\text{var}(x_M)} (\pi(x_M) - E(x_M)), \quad \text{for all } i,
\]

The risk premium of any of the initial portfolios can be written as the risk premium of the market portfolio multiplied by the normalized covariance term, the portfolio's beta in the market.

The result (9.2) is often referred to as Stein's lemma. The first derivation in the economics literature seems to be due to Rubinstein (1973). Using a Taylor series expansion, he assumed that the function \( U' \) possesses derivatives of all orders and that these functions can be integrated. Below we give a simple proof, where \( U' \) need not even be one time differentiable for (9.3) to result by the above procedure.

**Lemma 1:**

Suppose \((X, Y)\) is jointly bivariate normally distributed. Then

\[
(\text{a}) \quad \text{cov}(X, g(Y)) = \frac{\text{cov}(Y, g(Y))}{\text{var}Y} \text{cov}(X, Y).
\]
Suppose $g'(\cdot)$ exists for all real numbers and that $E|g'(Y)| < \infty$. Then

$$(b) \quad E(g'(Y)) = \frac{\text{cov}(Y, g(Y))}{\text{var} Y}.$$ 

**Proof:** From the assumption of binormality it follows that $E(X|Y) = \alpha + \beta Y$, where

$$\beta = \frac{\text{cov}(X, Y)}{\text{var} Y} \quad \text{and} \quad \alpha = EX - \beta EY.$$ 

Also $\text{cov}(X, g(Y)) = E\{E[Xg(Y)|Y]\} - EXg(Y) = \alpha g(Y) + \beta E(Yg(Y)) - EEXg(Y) = \beta[E(Yg(Y)) - EYEg(Y)]$, proving (a). As for (b), by integration by parts, using the assumption that the expectation of $g'(Y)$ exists, we find

$$Eg'(Y) = - \int_{-\infty}^{\infty} g'(y) f_Y(y) dy = \frac{1}{\text{var} Y} E\{g(Y)(Y - EY)\}$$

where $f_Y(y)$ is the normal probability density function for $Y$. This proves (b). \(\diamondsuit\)

Note that (9.3) follows from (a) only. Thus the assumptions that $U'$ is one time differentiable and that the expectation of $U'(x_M)$ exists, are really not needed in the above step. If $U''$ exists for all reals together with its expectation, then (a) and (b) imply (9.2). For an extension of this result, proved by entirely different methods, see Wei and Lee (1988).

Note that we have used the equilibrium-result in Section 2 that $U'(x) = U'(x_M)$.

We may also find the sign of the risk premium of any Pareto optimal, linear sharing rule $y_i(x_M)$. In this case we find

$$(9.4) \quad \pi(y_i(x_M)) - E(y_i(x_M)) = E\left( \frac{\partial y_i(x_M)}{\partial x} \right)(\pi(x_M) - E(x_M)), \text{ for all } i.$$ 

By (5.4) we notice that this beta is positive. The risk premium of all the portfolios have then the same sign as the risk premium of the market portfolio, which in this case is negative. This result corresponds to "investors hold efficient portfolios in capital market equilibrium" in the theory of capital markets, whereas the fact that $E\{y_i'(x_M)|y_i\} > 0$ corresponds to "efficient portfolios have positive betas". Notice that the negative risk premiums here only mean that the insurers require a positive expected return on their reinsurance exchanges, since this expected return simply equals $[E(y_i) - \pi(y_i)]/\pi(x_i) > 0$.

Returning to (9.3) suppose that one of the initial portfolios, $x_1$ say, has a negative correlation with the market. The market finds this portfolio so valuable that it accepts a negative expected return on $x_1$ in equilibrium.
As for the characterization (9.3), there might exist other joint probability distributions giving the same separation result (see for example Ross (1976) and Chamberlain (1983)). A different line of attack has traditionally been to impose further conditions on preferences. For example, if the marginal utility of the market $U'$ is linear, then this separation follows as well.

Apparently this result seems to require no assumptions regarding the joint probability distribution of $x$. However, linear marginal market utility is usually a consequence of quadratic utility functions representing the preferences of the individual insurers, which means that the probabilities of falling beyond the satiation points should equal zero in order for condition (a) to remain valid. Otherwise the preferences are not monotonic, and risky investments are inferior compared to the riskless. Thus conditions must then indirectly be imposed on the joint probability distribution of $x$ as well. For example is the multinormal distribution not acceptable in this situation. In such cases the conditions of Theorem 1 may be met, and the characterization (9.3) be valid. We should add, however, that one obvious advantage of imposing distributional assumptions on $x$ rather than assumptions directly on preferences (if these can at all be avoided), is that the former can be empirically tested using statistical methods, whereas the latter are much harder to verify/refute from available data.

The classical version of the one-period CAPM in a capital market was developed by Sharpe, Lintner and Mossin. The classical one-period CAPM has also been developed without the assumption of a riskless asset by Black (1972). In a multiperiod setting Merton (1972) has developed an intertemporal CAPM, where the prices of the risky assets are assumed to follow Ito-diffusions. In a dynamic, intertemporal reinsurance context, where the claims processes are represented by random, marked point processes, an insurance version of an intertemporal CAPM can be found in Aase (1993).

10. SUMMARY

From the above analysis we observe that the premiums in a reinsurance market typically must depend on:

(i) The stochastic properties of the risk itself.
(ii) The stochastic relationship between the particular risk $z$ and claims in the market as a whole, described by the covariance between $V(z_M)$ and $z$.
(iii) The attitude towards risk in the market as a whole, represented by $V = U'$.
(iv) The total assets of all the insurers in the market, represented by $a_M$.

A realistic theory of insurance premiums must of course take all these four elements into account. This is however rarely done in actuarial risk theory. Several books have been written on insurance premium principles, some even recent, where only the first of these four elements are covered.

Some obvious weaknesses of the above model are the following. There is in reality no time dimension in these models; trade is supposed to take place only at one point in time, and the world more or less ends at the next time point. In
models of (re)insurance markets the risks may be more realistically represented by random, marked point processes. A model where trade can take place at any time point $t$ in an interval $[0, T]$ is given in Aase (1992-93). There it is shown that the market's attitude towards risk can be separated into two components; one related to frequency risk and the other related to claim size risk, given that an accident has occurred. In order to fully understand these results, however, it appears to be essential to have the above model in mind. This is so since the present derivation basically tells us what happens at each time point of jump of the vector $x$ of the stochastic process representing the exogenously given portfolios in the reinsurance market. For example is our interpretation of the market marginal utility crucial also in the dynamic case. Therefore the one-period analysis can be viewed as a necessary preparation in order to proceed to more realistic, but at the same time more complicated and mathematically challenging models of equilibrium premium formation in a dynamic exchange economy under uncertainty.

ACKNOWLEDGEMENT

I would like to thank Lars Tyge Nielsen for stimulating discussions on some topics of this paper, and also an anonymous referee for discovering several errors and misprints.

REFERENCES


Knut K. Aase

*Norwegian School of Economics and Business Administration*

5035 Bergen-Sandviken, Norway.
A distribution-free formula for the standard error of chain ladder reserve estimates is derived and compared to the results of some parametric methods using a numerical example.

Keywords
Claims reserving; chain ladder; standard error.

1. Introduction
The chain ladder method is probably the most popular method for estimating IBNR claims reserves. The main reason for this is its simplicity and the fact that it is distribution-free, i.e. that it seems to work with almost no assumptions. On the other hand, it is well-known that chain ladder reserve estimates for the most recent accident years are very sensitive to variations in the data observed. Moreover, in recent years many other claims reserving procedures have been proposed and the results of all these procedures vary widely and also differ more or less from the chain ladder result. Therefore it would be very helpful to know the standard error of the chain ladder reserve estimates as a measure of the uncertainty contained in the data and in order to see whether the difference between the results of the chain ladder method and any other method is significant or not.

Up to now only a few papers on claims reserving have been published which also consider the calculation of the standard error of the reserve estimate: In the papers by Taylor/Ashe 1983, Zehnwirth 1985, Renshaw 1989, Christofides 1990, Verrall 1990, Verrall 1991 essentially the same method for the calculation of the standard error is used, namely a least squares regression approach which (with the exception of Taylor/Ashe) is applied to the logarithms of the incremental claims amounts (i.e. assuming a lognormal distribution). Slightly different approaches have been proposed by Wright (1990, via a generalized linear model and the method of scoring) and Mack (1991, using a gamma distribution and maximum likelihood estimation). All methods cited require a rather high amount of programming in order to calculate the many covariances between the parameter estimators.
In the present paper, a very simple formula for the standard error of chain ladder reserve estimates is developed. The decisive step towards this formula was made by SCHNIEPER (1991). In order to calculate the rate for a casualty excess of loss cover he used a mixture of the Bornhuetter-Ferguson technique and the chain ladder method. Within this model he developed an approximation to the standard error of the estimated premium rate using a Taylor series approximation.

The present paper adapts Schnieper's idea to the claims reserving situation and contains the following additional points:

1. The model is specialized for the pure chain ladder case. This makes things easier and also makes it possible to replace the Taylor series approximation with a more exact procedure.
2. An estimate of the process variance is additionally included in the standard error of the reserve estimate. This is necessary here because the claims reserve is a random variable and not a parameter like the net premium (= expected value).
3. Schnieper intuitively claimed that the chain ladder development factors were "not strongly correlated". We prove that they are in fact uncorrelated and that therefore the reserve estimate is unbiased.
4. Besides the standard error for each accident year, a formula for the standard error of the overall reserve estimator is given, too, which takes the correlations between the estimates for the individual accident years into account.

Finally, two numerical examples are given and the results are compared to the results obtained by the aforementioned methods of Taylor/Ashe, Zehnwirth, Renshaw/Christofides, Verrall and Mack.

2. NOTATIONS AND BASIC RESULTS

Let $C_{ik}$ denote the accumulated total claims amount of accident year $i$, $1 \leq i \leq I$, either paid or incurred up to development year $k$, $1 \leq k \leq I$. We consider $C_{ik}$ a random variable of which we have an observation if $i+k \leq I+1$ (run-off triangle). The aim is to estimate the ultimate claims amount $C_u$ and the outstanding claims reserve

$$R_i = C_{il} - C_{i,i+1-i}$$

for accident year $i = 2, \ldots, I$.

The basic chain ladder assumption is that there are development factors $f_1, \ldots, f_{I-1} > 0$ with

$$E(C_{i,k+1}|C_{i1}, \ldots, C_{ik}) = C_{ik}f_k, \quad 1 \leq i \leq I, \quad 1 \leq k \leq I-1.$$  

The chain ladder method consists of estimating the $f_k$ by

$$\hat{f}_k = \sum_{j=1}^{i-k} C_{j,k+1} / \sum_{j=1}^{i-k} C_{jk}, \quad 1 \leq k \leq I-1,$$
and the ultimate claims amount $C_{ii}$ by

$$\hat{C}_{ii} = C_{i,t+1-i} \cdot \hat{f}_{t+1-i} \cdots \hat{f}_{t-1},$$

or equivalently the reserve $R_i$ by

$$\hat{R}_i = C_{i,t+1-i} (\hat{f}_{t+1-i} \cdots \hat{f}_{t-1} - 1).$$

Because the chain ladder algorithm does not take into account any dependencies between accident years, we can additionally assume that the variables $C_{ik}$ of different accident years, i.e.

$$\{C_{ii}, \ldots, C_{ii}\}, \{C_{jj}, \ldots, C_{jj}\}, i \neq j,$$

are independent.

This must be regarded as a further implicit assumption of the chain ladder method. In practice, the independence of the accident years can be distorted by certain calendar year effects like major changes in claims handling or in case reserving.

The following theorem makes it clear that (1) and (2) are indeed the implicit assumptions of the chain ladder method.

**Theorem 1:** Let $D = \{C_{ik}|i+k \leq I+1\}$ be the set of all data observed so far. Under the assumptions (1) and (2) we have

$$E(C_{ii}|D) = C_{i,t+1-i} \cdot \hat{f}_{t+1-i} \cdots \hat{f}_{t-1}.$$ 

**Proof:** We use the abbreviation

$$E_i(X) = E(X|C_{ii}, \ldots, C_{i,t+1-i}).$$

Then (2) and repeated application of (1) yield

$$E(C_{ii}|D) = E_i(C_{ii})$$

$$= E_i(E(C_{ii}|C_{ii}, \ldots, C_{i,t+1-i}))$$

$$= E_i(C_{i,t+1-i} \cdot \hat{f}_{t-1})$$

$$= E_i(C_{i,t+1-i} \cdot \hat{f}_{t-1})$$

$$= etc.$$ 

$$= E_i(C_{i,t+1-i} \cdot \hat{f}_{t+1-i} \cdots \hat{f}_{t-1})$$

$$= C_{i,t+1-i} \cdot \hat{f}_{t+1-i} \cdots \hat{f}_{t-1}. \square$$

This theorem shows that the estimator $\hat{C}_{ii}$ has the same form as $E(C_{ii}|D)$ which is the best forecast of $C_{ii}$ based on the observations $D$. The next theorem shows that estimating $\hat{f}_{t+1-i} \cdots \hat{f}_{t-1}$ by $\hat{f}_{t+1-i} \cdots \hat{f}_{t-1}$ is indeed a reasonable procedure.

**Theorem 2:** Under the assumptions (1) and (2) the estimators $\hat{f}_k$, $1 \leq k \leq I-1$, are unbiased and uncorrelated.
Proof: Let

\[ B_k = \{ C_{ij} | j \leq k, i+j \leq I+1 \}, \; 1 \leq k \leq I. \]

Then (2) and (1) yield

\[ E(C_{i,k+1}|B_k) = E(C_{i,k+1}|C_{iI}, \ldots, C_{ik}) = C_{ik}f_k. \]

We therefore have

\[ E(\hat{f}_k|B_k) = \sum_{j=1}^{I-k} E(C_{j,k+1}|B_k) \sum_{j=1}^{I-k} C_{jk} = f_k, \]

which immediately gives the unbiasedness

\[ E(\hat{f}_k) = E(E(\hat{f}_k|B_k)) = f_k, \; 1 \leq k \leq I-1, \]

of the parameter estimates. Also, the \( \hat{f}_k \) are uncorrelated because for \( j < k \)

\[ E(\hat{f}_j \hat{f}_k) = E(E(\hat{f}_j \hat{f}_k|B_k)) \\
= E(\hat{f}_j E(\hat{f}_k|B_k)) \\
= E(\hat{f}_j) f_k \\
= E(\hat{f}_j) E(\hat{f}_k). \]

The uncorrelatedness of the \( \hat{f}_k \)'s is surprising because \( \hat{f}_{k-1} \) and \( \hat{f}_k \) depend on the same data \( C_{1k} + \ldots + C_{i-k,k} \). The foregoing proof of the uncorrelatedness easily extends to arbitrary products of pairwise different \( \hat{f}_k \), i.e. we have

\[ E(\hat{f}_{i+1-f_1-i} \ldots \hat{f}_{i-1}) = f_{i+1-f_1-i} \ldots f_{i-1}, \]

which shows that \( \hat{C}_{il} = C_{i,i+1-i} \hat{f}_{i+1-i} \ldots \hat{f}_{i-1} \) is an unbiased estimator of \( E(C_{il}|D) = C_{i,i+1-i} \hat{f}_{i+1-i} \ldots \hat{f}_{i-1} \). In the same way, the reserve estimator \( \hat{R}_i = \hat{C}_{il} - C_{i,i+1-i} \) is an unbiased estimator of the true reserve \( R_i = C_{il} - C_{i,i+1-i} \).

3. Calculation of Mean Squared Error and Standard Error

The mean squared error \( mse(\hat{C}_{il}) \) of the estimator \( \hat{C}_{il} \) of \( C_{il} \) is defined to be

\[ mse(\hat{C}_{il}) = E((\hat{C}_{il} - C_{il})^2|D) \]

where \( D = \{ C_{ik} | i+k \leq I+1 \} \) is the set of all data observed so far. Note that we are not using the unconditional mean squared error \( E((\hat{C}_{il} - C_{il})^2) = E(E((\hat{C}_{il} - C_{il})^2|D)) \) as this averages over all possible data \( D \) from the underlying distribution. Instead, in practise, we are more interested in the conditional mean squared error of the particular estimated amount \( \hat{C}_{il} \) based on the specific data set \( D \) observed and therefore have to use \( E((\hat{C}_{il} - C_{il})^2|D) \) which just gives us the average deviation between \( \hat{C}_{il} \) and \( C_{il} \) due to future randomness only.

First, we see that

\[ mse(\hat{R}_i) = E((\hat{R}_i - R_i)^2|D) = E((\hat{C}_{il} - C_{il})^2|D) = mse(\hat{C}_{il}). \]
Next, because of the general rule $E(X-a)^2 = \text{Var}(X) + (E(X) - a)^2$ we have

$$mse(\hat{C}_{il}) = \text{Var} (C_{il}|D) + (E(C_{il}|D) - \hat{C}_{il})^2$$

which shows that the mean squared error is the sum of the stochastic error (process variance) and of the estimation error.

In order to further calculate the $mse$ we need a formula for the variance of $C_{ik}$. From the fact that $\hat{f}_k$ is the $C_{ik}$-weighted mean of the individual development factors $C_{i,k+1}/C_{ik}, \, 1 \leq i \leq I-k$, we can induce that $\text{Var} (C_{i,k+1}/C_{ik}|C_{i1}, \ldots, C_{ik})$ should be inversely proportional to $C_{ik}$, or equivalently

(3) $\text{Var} (C_{i,k+1}|C_{i1}, \ldots, C_{ik}) = C_{ik} \sigma_k^2, \, 1 \leq i \leq I, \, 1 \leq k \leq I-1,$

with unknown parameters $\sigma_k^2, \, 1 \leq k \leq I-1$. This is the variance assumption which is implicitly underlining the chain ladder method.

Later on, we will need an estimator for $\sigma_k^2$. Similarly as for $\hat{f}_k$ it can be shown that

$$\hat{\sigma}_k^2 = \frac{1}{I-k-1} \sum_{i=1}^{I-k} C_{ik} \left( \frac{C_{i,k+1}}{C_{ik}} - \hat{f}_k \right)^2, \, 1 \leq k \leq I-2.$$ 

is an unbiased estimator of $\sigma_k^2, \, 1 \leq k \leq I-2$. We still lack an estimator for $\sigma_{I-1}$. If $\hat{f}_{I-1} = 1$ and if the claims development is believed to be finished after $I-1$ years, we can put $\hat{\sigma}_{I-1} = 0$. If not, we extrapolate the usually exponentially decreasing series $\hat{\sigma}_1, \ldots, \hat{\sigma}_{I-3}, \hat{\sigma}_{I-2}$ by one additional member, for instance by loglinear regression or more simply by requiring that

$$\hat{\sigma}_{I-3}/\hat{\sigma}_{I-2} = \hat{\sigma}_{I-2}/\hat{\sigma}_{I-1}$$

holds at least as long as $\hat{\sigma}_{I-3} > \hat{\sigma}_{I-2}$. This last possibility leads to

$$\hat{\sigma}_{I-1}^2 = \min (\hat{\sigma}_{I-2}^2/\hat{\sigma}_{I-3}^2, \min (\hat{\sigma}_{I-3}^2, \hat{\sigma}_{I-2}^2))$$

which has been used in the examples.

Now, we are able to state and prove the main result

**Theorem 3:** Under the assumptions (1), (2) and (3) the mean squared error $mse(\hat{R}_l)$ can be estimated by

$$\overline{mse}(\hat{R}_l) = \hat{C}_{il} \sum_{k=I+1-l}^{I-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left( \frac{1}{\hat{C}_{ik}} + \frac{1}{\sum_{j=1}^{I-k} C_{jk}} \right)$$

where $\hat{C}_{ik} = C_{i,I+1-i}\hat{f}_{I+1-i} \cdots \hat{f}_{k-I}, \, k > I+1-i$, are the estimated values of the future $C_{ik}$ and $\hat{C}_{i,I+1-i} = C_{i,I+1-i}$. 

Proof: We use the abbreviations

\[ E_i(X) = E(X|C_{il}, \ldots, C_{i, t+1-i}) \]
\[ \text{Var}_i(X) = \text{Var}(X|C_{il}, \ldots, C_{i, t+1-i}) . \]

We start from

\[ \text{mse}(\hat{R}_i) = \text{Var}(C_{il}|D) + (E(C_{il}|D) - \hat{C}_{il})^2 . \]

Repeated application of the basic chain ladder assumption (1) and of the above variance assumption (3) yields for the first term of \( \text{mse}(\hat{R}_i) \)

\[
\begin{align*}
\text{Var}(C_{il}|D) &= \text{Var}_i(C_{il}) \\
&= E_i(\text{Var}(C_{il}|C_{il}, \ldots, C_{i, t-1})) + \\
&\quad + \text{Var}_i(E(C_{il}|C_{il}, \ldots, C_{i, t-1})) \\
&= E_i(C_{i, t-1}) \sigma^2_{t-1} + \text{Var}_i(C_{i, t-1})f^2_{t-1} \\
&= E_i(C_{i, t-2})f^2_{t-2} \sigma^2_{t-2} + E_i(C_{i, t-2}) \sigma^2_{t-2} f^2_{t-1} + \\
&\quad + \text{Var}_i(C_{i, t-2})f^2_{t-2} f^2_{t-1} \\
&= \text{etc.} \\
&= C_{i, t+1-i} \sum_{k=t+1-i}^{t-1} f_{t+1-i} \cdots f_{k-1} \frac{\sigma^2_k}{C^2_{ik}} f^2_{k+1} \cdots f^2_{t-1}
\end{align*}
\]

because of \( \text{Var}_i(C_{i, t+1-i}) = 0 \).

Due to Theorem 1 we obtain for the second term of \( \text{mse}(\hat{R}_i) \)

\[ (E(C_{il}|D) - \hat{C}_{il})^2 = C^2_{i, t+1-i}(f_{t+1-i} \cdots f_{t-1} - \hat{f}_{t+1-i} \cdots \hat{f}_{t-1})^2 . \]

In practice, we must find estimators for these two terms of \( \text{mse}(\hat{R}_i) \). For the first term this will be done by replacing the unknown parameters \( f_k \) and \( \sigma_k^2 \) with their estimators \( \hat{f}_k \) and \( \hat{\sigma}^2_k \), i.e. we estimate \( \text{Var}(C_{il}|D) \) by

\[
\begin{align*}
C_{i, t+1-i} \left( \sum_{k=t+1-i}^{t-1} \hat{f}_{t+1-i} \cdots \hat{f}_{k-1} \frac{\sigma^2_k}{C^2_{ik}} \hat{f}^2_{k+1} \cdots \hat{f}^2_{t-1} \right) \\
= \hat{C}^2_{il} \sum_{k=t+1-i}^{t-1} \frac{\hat{\sigma}^2_k}{\hat{C}_{ik}} \hat{f}^2_k
\end{align*}
\]

where we have used the notation \( \hat{C}_{ik} \) introduced in the theorem.

But in the second term (*) of \( \text{mse}(\hat{R}_i) \) we can not simply replace \( f_k \) with \( \hat{f}_k \) because this would yield 0. We therefore use a different approach. We can write

\[
F = f_{t+1-i} \cdots f_{t-1} - \hat{f}_{t+1-i} \cdots \hat{f}_{t-1} \\
= S_{t+1-i} + \cdots + S_{t-1}
\]

with

\[
S_k = \hat{f}_{t+1-i} \cdots \hat{f}_{k-1} (f_k - \hat{f}_k) f_{k+1} \cdots f_{t-1}
\]
and therefore

\[ F^2 = (S_{I+1-i} + \ldots + S_{I-1})^2 \]

\[ = \sum_{k=I+1-i}^{I-1} S_k^2 + 2 \sum_{j<k} S_j S_k. \]

Now we replace \( S_k^2 \) with \( E(S_k^2|B_k) \) and \( S_j S_k \), \( j < k \), with \( E(S_j S_k|B_k) \). This means that we approximate \( S_k^2 \) and \( S_j S_k \) by averaging over as little data as possible such that as many values \( C_{ik} \) as possible from the observed data are kept fixed. Because of \( E((f_k - \hat{f}_k)^2|B_k) = 0 \) (see the proof of Theorem 2) we obtain \( E(S_j S_k|B_k) = 0 \) for \( j < k \). Because of

\[ E((f_k - \hat{f}_k)^2|B_k) = \text{Var}(\hat{f}_k|B_k) \]

\[ = \sum_{j=1}^{I-k} \text{Var}(C_{j,k+1}|B_k) \left( \sum_{j=1}^{I-k} C_{jk} \right)^2 \]

we obtain

\[ E(S_k^2|B_k) = \hat{f}_{I+1-i}^2 \cdot \ldots \cdot \hat{f}_{I-1}^2 \sigma_k^2 \sigma_{k+1}^2 \cdot \ldots \cdot \sigma_{I-1}^2 \left( \sum_{i=1}^{I-k} C_{ik} \right). \]

Taken together, we replace \( F^2 = (\Sigma S_k)^2 \) with \( \Sigma_k E(S_k^2|B_k) \) and because all terms of this sum are positive we now can replace all unknown parameters \( f_k \), \( \sigma_k^2 \) with their unbiased estimators \( \hat{f}_k \), \( \sigma_k^2 \). Altogether, we estimate \( F^2 = (\hat{f}_{I+1-i} \cdot \ldots \cdot \hat{f}_{I-1} \cdot \hat{f}_{I+1-i} \cdot \ldots \cdot \hat{f}_{I-1})^2 \) by

\[ \hat{f}_{I+1-i}^2 \cdot \ldots \cdot \hat{f}_{I-1}^2 \sqrt{\hat{f}_{k+1}^2 \cdot \ldots \cdot \hat{f}_{I-1}^2} \left( \sum_{i=1}^{I-k} C_{ik} \right) \]

\[ = \hat{f}_{I+1-i}^2 \cdot \ldots \cdot \hat{f}_{I-1}^2 \sum_{k=I+1-i}^{I-1} \frac{\sigma_k^2}{\hat{f}_k^2} \left( \sum_{j=1}^{I-k} C_{jk} \right). \]

This finally leads to the estimator stated in the theorem.

The square root s.e. \( (\hat{R}_i) \) of an estimator of the mean squared error is defined to be the standard error of \( \hat{R}_i \).

Often the standard error of the overall reserve estimate \( \hat{R} = \hat{R}_2 + \ldots + \hat{R}_i \) is of interest, too. In this case we cannot simply add together the values of \( (\text{s.e.}(\hat{R}_i))^2 \), \( 2 \leq i \leq I \), because they are correlated via the common estimators \( \hat{f}_k \) and \( \sigma_k^2 \). We therefore proceed as before and obtain:
Corollary: With the assumptions and notations of Theorem 3 the mean squared error of the overall reserve estimate \( \hat{R} = \hat{R}_2 + ... + \hat{R}_t \) can be estimated by

\[
\widehat{mse}(\hat{R}) = \sum_{i=2}^{t} \left\{ (\text{s.e.} \ (\hat{R}_i))^2 + \hat{C}_{il} \left( \sum_{j=i+1}^{l} \hat{C}_{jl} \right) \sum_{k=i+1-i}^{l-1} \frac{2 \hat{\sigma}_k^2 \hat{\mu}_k^2}{\sum_{a=1}^{l-k} C_{nk}} \right\}
\]

Proof: We have

\[
mse \left( \sum_{i=2}^{t} \hat{R}_i \right) = E \left( \left( \sum_{i=2}^{t} \hat{R}_i - \sum_{i=2}^{t} R_i \right)^2 \right) = E \left( \left( \sum_{i=2}^{t} \hat{C}_{il} - \sum_{i=2}^{t} C_{il} \right)^2 \right) = \text{Var} \left( \sum_{i=2}^{t} C_{il|D} \right) + E \left( \sum_{i=2}^{t} C_{il|D} \right) - \sum_{i=2}^{t} \hat{C}_{il} \right)^2.
\]

The independence of the accident years yields

\[
\text{Var} \left( \sum_{i=2}^{t} C_{il|D} \right) = \sum_{i=2}^{t} \text{Var} \ (C_{il|D}),
\]

whose summands have already been calculated in the proof of Theorem 3. Furthermore

\[
\left( E \left( \sum_{i=2}^{t} C_{il|D} \right) - \sum_{i=2}^{t} \hat{C}_{il} \right)^2 = \left( \sum_{i=2}^{t} (E(C_{il|D}) - \hat{C}_{il}) \right)^2 = \sum_{i,j} (E(C_{il|D}) - \hat{C}_{il}) \cdot (E(C_{ij|D}) - \hat{C}_{ij}) \]

\[
= \sum_{i,j} C_{i, l+1-i} C_{j, l+1-j} F_i F_j
\]

with

\[
F_i = f_{i+1-i} \cdots f_{i-1} - \hat{f}_{i+1-i} \cdots \hat{f}_{i-1}.
\]

Observing

\[
mse (\hat{R}) = Var \ (C_{il|D}) + (C_{i, l+1-i} F_i)^2
\]

(cf. (*) in the proof of theorem 3) we see that

\[
mse \left( \sum_{i=2}^{t} \hat{R}_i \right) = \sum_{i=2}^{t} mse (\hat{R}_i) + \sum_{2 \leq i < j \leq t} 2 \cdot C_{i, l+1-i} C_{j, l+1-j} F_i F_j.
\]
An analogous procedure as for $F^2$ in the above proof yields for $F_i F_j$, $i < j$, the estimator

$$
\hat{f}_{i+1-j} \cdots \hat{f}_{i-j} \hat{f}_{i+1-k} \cdots \hat{f}_{i-k} \hat{f}_{k+1-k} \cdots \hat{f}_{i-1-k} \sum_{i=1}^{l} C_{nk}.
$$

This completes the proof.

4. EXAMPLES

In the first example we use the TAYLOR/ASHE (1983) data, which were also used by VERRALL (1990, 1991).

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>RUN-OFF TRIANGLE (ACCUMULATED FIGURES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>$C_{n1}$</td>
</tr>
<tr>
<td>1</td>
<td>357848</td>
</tr>
<tr>
<td>2</td>
<td>352118</td>
</tr>
<tr>
<td>3</td>
<td>290507</td>
</tr>
<tr>
<td>4</td>
<td>310608</td>
</tr>
<tr>
<td>5</td>
<td>443160</td>
</tr>
<tr>
<td>6</td>
<td>396132</td>
</tr>
<tr>
<td>7</td>
<td>440832</td>
</tr>
<tr>
<td>8</td>
<td>359480</td>
</tr>
<tr>
<td>9</td>
<td>376686</td>
</tr>
<tr>
<td>10</td>
<td>344014</td>
</tr>
</tbody>
</table>

This yields the following parameter estimates ($k = 1, \ldots, 9$):

\[ \hat{f}_k: 3.49, 1.75, 1.46, 1.174, 1.104, 1.086, 1.054, 1.077, 1.018 \]
\[ \hat{\sigma}_k/1000: 160, 37.7, 42.0, 15.2, 13.7, 8.19, 0.447, 1.15, 0.477 \]

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>ESTIMATED RESERVES $\hat{R}$, IN 1000 S</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 2$</td>
<td>95</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>470</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>710</td>
</tr>
<tr>
<td>$i = 5$</td>
<td>985</td>
</tr>
<tr>
<td>$i = 6$</td>
<td>1419</td>
</tr>
<tr>
<td>$i = 7$</td>
<td>2178</td>
</tr>
<tr>
<td>$i = 8$</td>
<td>3920</td>
</tr>
<tr>
<td>$i = 9$</td>
<td>4279</td>
</tr>
<tr>
<td>$i = 10$</td>
<td>4626</td>
</tr>
</tbody>
</table>

overall | 18681     | 16652     | 19512      | 19124      | 18085      | 22301      |
The results for 'Taylor/Ashe' and 'Verrall 1991' have been taken from these papers. Taylor/Ashe produced much lower standard errors than the other methods. This is due to the fact that their reserve estimates employed only 6 parameters (as compared to 19 of the other methods) and that they additionally used the information on the numbers of claims finalized.

Renshaw and Christofides describe the same loglinear regression method which is also identical to Verrall's (1990) Bayesian approach without any prior information. Therefore the results for 'Renshaw/Christofides' have been taken from VERRALL (1990), Table 2.

The results for 'Zehnwirth' have been obtained by using his ICRFS software package version 6.1 employing one of his fixed parameter development factor models which he calls 'chain ladder model'. We have used it without any further adjustment. It should be remarked that this is not what Zehnwirth intends, as his software package is a modelling framework and any initial model should be further adjusted interactively with the help of the indications and plots given by the program. Without any further adjustment this 'chain ladder model' is identical to the Renshaw/Christofides model, i.e. it is a loglinearized approximation of the usual chain ladder model. The fact that it leads to slightly lower results is attributable to using a different estimator for the model variance.

The results for 'Mack 1991' have been obtained according to a previous paper (MACK (1991)) of the author but additionally an estimate of the process variance has been included, as this is the case with all the other methods.

The estimated reserves of all methods except 'Taylor/Ashe' differ by less than 20% and are therefore according to Table 3 within one standard error. For the chain ladder method neither the reserve estimates nor the standard errors are systematically higher or lower than for the other methods (except 'Taylor/Ashe'). The reason for the comparatively high chain ladder standard
error of 80% for accident year 2 is the fact that the reserve $\hat{R}_2$ itself is very low in comparison to the other reserves $\hat{R}_3, \ldots, \hat{R}_{10}$: If we look at the sequence $\hat{R}_{10}, \hat{R}_{9}, \ldots, \hat{R}_4, \hat{R}_3$ we see that $\hat{R}_2$ is always greater than $\hat{R}_1/2$ but $\hat{R}_2$ is smaller than $\hat{R}_3/4$. This fact is very well reflected by the high standard error of 80%.

A closer look at the Taylor/Ashe data shows that the individual development factors \( C_{i,k+1}/C_{ik}, 1 \leq i \leq I-k \), do not fluctuate much around their mean value $f_k$ so that the whole triangle can be considered as relatively regular. Therefore Taylor/Ashe were able to dispense with taking logarithms and thus avoided the problem of transforming back the result into the original data space. We therefore give a second example, which is less regular and where the claims amounts of the most recent accident years are much lower than in the previous years. These data (mortgage guarantee business) were compiled from a competition presented by SANDERS (1990).

### Table 4

Run-off triangle (accumulated figures)

<table>
<thead>
<tr>
<th>i</th>
<th>$C_{n1}$</th>
<th>$C_{n2}$</th>
<th>$C_{n3}$</th>
<th>$C_{n4}$</th>
<th>$C_{n5}$</th>
<th>$C_{n6}$</th>
<th>$C_{n7}$</th>
<th>$C_{n8}$</th>
<th>$C_{n9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58046</td>
<td>127970</td>
<td>476599</td>
<td>1027692</td>
<td>1360489</td>
<td>1647310</td>
<td>1819179</td>
<td>1906852</td>
<td>1950105</td>
</tr>
<tr>
<td>2</td>
<td>24492</td>
<td>141767</td>
<td>724288</td>
<td>2142656</td>
<td>2961978</td>
<td>3683940</td>
<td>4048898</td>
<td>4115760</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>32848</td>
<td>274682</td>
<td>1522637</td>
<td>3203427</td>
<td>4445927</td>
<td>5158781</td>
<td>5342335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>21439</td>
<td>529828</td>
<td>2900301</td>
<td>4999019</td>
<td>6460112</td>
<td>6853904</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>40379</td>
<td>763394</td>
<td>2920745</td>
<td>4989572</td>
<td>5648563</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>90748</td>
<td>951994</td>
<td>4210640</td>
<td>5866482</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>62096</td>
<td>868480</td>
<td>1954797</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>24983</td>
<td>284441</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>13121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter estimates \((k = 1, \ldots, 8)\): 

$\hat{f}_k$: 11.1, 4.09, 1.71, 1.28, 1.14, 1.069, 1.026, 1.023 

$\hat{\sigma}_k^2/1000$: 1787, 977, 194, 42.8, 27.0, 5.57, 1.26, 0.285

### Table 5

Estimated reserves $\hat{R}_i$ in 1000s

<table>
<thead>
<tr>
<th>i</th>
<th>Chain ladder</th>
<th>Renshaw Christofides</th>
<th>Zehnwirth</th>
<th>Mack 1991</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>93</td>
<td>91</td>
<td>87</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>265</td>
<td>275</td>
<td>262</td>
<td>199</td>
</tr>
<tr>
<td>4</td>
<td>834</td>
<td>818</td>
<td>778</td>
<td>682</td>
</tr>
<tr>
<td>5</td>
<td>1568</td>
<td>1979</td>
<td>1884</td>
<td>1639</td>
</tr>
<tr>
<td>6</td>
<td>3696</td>
<td>5497</td>
<td>5231</td>
<td>4420</td>
</tr>
<tr>
<td>7</td>
<td>3487</td>
<td>6650</td>
<td>6328</td>
<td>5378</td>
</tr>
<tr>
<td>8</td>
<td>2956</td>
<td>4331</td>
<td>4122</td>
<td>3143</td>
</tr>
<tr>
<td>9</td>
<td>1647</td>
<td>2339</td>
<td>2226</td>
<td>1555</td>
</tr>
<tr>
<td>overall</td>
<td>14547</td>
<td>21980</td>
<td>20919</td>
<td>17078</td>
</tr>
</tbody>
</table>


TABLE 6

<table>
<thead>
<tr>
<th></th>
<th>Chain ladder</th>
<th>Renshaw Christofides</th>
<th>Zehnwirth</th>
<th>Mack 1991</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 2$</td>
<td>65%</td>
<td>90%</td>
<td>80%</td>
<td>60%</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>53%</td>
<td>60%</td>
<td>53%</td>
<td>41%</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>38%</td>
<td>51%</td>
<td>45%</td>
<td>37%</td>
</tr>
<tr>
<td>$i = 5$</td>
<td>38%</td>
<td>48%</td>
<td>42%</td>
<td>35%</td>
</tr>
<tr>
<td>$i = 6$</td>
<td>28%</td>
<td>46%</td>
<td>41%</td>
<td>33%</td>
</tr>
<tr>
<td>$i = 7$</td>
<td>37%</td>
<td>47%</td>
<td>42%</td>
<td>34%</td>
</tr>
<tr>
<td>$i = 8$</td>
<td>61%</td>
<td>50%</td>
<td>47%</td>
<td>36%</td>
</tr>
<tr>
<td>$i = 9$</td>
<td>133%</td>
<td>66%</td>
<td>64%</td>
<td>47%</td>
</tr>
<tr>
<td>overall</td>
<td>26%</td>
<td>24%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Here all results have been calculated by the author. In comparison with the standard errors of the first example, the chain ladder standard errors now reflect very well the generally higher uncertainty of this second triangle and especially the uncertainty of the last two accident years where the relative standard errors are very high because the reserve estimates are comparatively low. The most extreme deviation between the reserve estimates of the different methods is for accident year 7 where the ‘Renshaw/Christofides’ reserve exceeds the chain ladder reserve by 2.5 standard errors.

Altogether, if the impressions of these two examples can be taken as typical, we can conclude that the standard errors are of about the same size for the chain ladder as with the other methods, although they do not show such a smooth pattern as these because the other methods use only one $\sigma^2$ parameter as compared to $I-1$ of chain ladder. But this could also be achieved for the chain ladder method by smoothing out the $\sigma^2$’s by means of an exponential function $\exp(a-bk)$.

Finally, we must bear in mind that these standard errors can only reflect the estimation error and the statistical error, but not the specification error, i.e. the fact that the model chosen can be wrong or that the future development may not be in accordance with past experience.

ACKNOWLEDGEMENT

I am indebted to Alois Gisler for pointing out the correct definition of the mean squared error and some further useful remarks.

REFERENCES


THOMAS MACK
Münchener Rückversicherungs-Gesellschaft, Königinstrasse 107, D-80791 München.
ON THE STABILITY OF RECURSIVE FORMULAS

BY HARRY H. PANJER AND SHAUN WANG

University of Waterloo, Ontario, Canada

ABSTRACT

Based on recurrence equation theory and relative error (rather than absolute error) analysis, the concept and criterion for the stability of a recurrence equation are clarified. A family of recursions, called congruent recursions, is proved to be strongly stable in evaluating its non-negative solutions. A type of strongly unstable recursion is identified. The recursive formula discussed by Panjer (1981) is proved to be strongly stable in evaluating the compound Poisson and the compound Negative Binomial (including Geometric) distributions. For the compound Binomial distribution, the recursion is shown to be unstable. A simple method to cope with this instability is proposed. Many other recursions are reviewed. Illustrative numerical examples are given.

KEYWORDS

Recursive formula; compound distribution; probability of ruin; dominant solution; subordinate solution; congruent recursion; index of error propagation; stable; strongly stable; strongly unstable; relative error analysis; empirical inflation factor.

1. INTRODUCTION

Compound distributions are used extensively in modeling the total claims for insurance portfolios. Consider the family of claim frequency distributions satisfying the recursion:

\[ p_n = p_{n-1} \left( a + \frac{b}{n} \right), \quad n = 1, 2, 3, \ldots \]

where \( p_n \) denotes the probability that exactly \( n \) claims occur in a fixed time interval such as one year and \( p_0 \) is an initial value. If the claim severity has a probability function (p.f.) \( f(x) \), \( x > 0 \), the total claims has a compound distribution with a p.f.:

\[ g(x) = \sum_{n=0}^{\infty} p_n f^{*n}(x), \quad x \geq 0. \]
Panjer [12] has shown that, if the claim severity distribution is defined on the positive integers with a p.f. \( f(x) \), \( x > 0 \), the compound distribution in (2) can be evaluated recursively as:

\[
(3) \quad g(x) = \sum_{j=1}^{x} \left( a + b \frac{j}{x} \right) f(j) g(x-j), \quad x = 1, 2, 3, \ldots
\]

\[
(4) \quad g(0) = p_0.
\]

This recursive formula is very useful for computer programming and significantly reduces the computing time comparing with the brute-force method directly using formula (2).

As with any algorithm, round-off errors are inevitable since computers only represent a finite number of digits. Practical observations show that algorithm (3) works well in evaluating compound distributions. However, in the actuarial literature, there are also some comments which diverge from the above observations and make the picture somewhat fuzzy. There is an obvious need for a clearer picture of the stability of recursive computation.

To convey some impression that round-off errors are not necessarily small, we start with a numerical example.

**Example 1:** In a compound Poisson model, the claim frequency has a Poisson distribution with mean \( \lambda = 10 \), the claim severity has a two points distribution:

\[
f(1) = .95, \quad f(2) = .05.
\]

By directly applying recursion (3) in the usual *forward* direction:

\[
(5) \quad g(x) = \frac{\lambda}{x} \left[ f(1) g(x-1) + 2 f(2) g(x-2) \right],
\]

\[
(6) \quad = \frac{10}{x} \left[ .95 g(x-1) + .1 g(x-2) \right],
\]

with initial values

\[
(7) \quad g(-1) = 0, \quad g(0) = \exp(-\lambda) = \exp(-10),
\]

one can obtain the compound distribution easily.

Values at \( x = 9 \) and \( x = 10 \) are

\[
(8) \quad g(9) = .1140989798, \quad g(10) = .1183785348.
\]

Equation (6) can be used in the *backward* direction as:

\[
(9) \quad g(x-2) = g(x) - 9.5 g(x-1).
\]

With \( g(10) \) and \( g(9) \) as starting points, we obtained the surprising results in Table 1 when 6 digits of floating points are used. One can see that round-off errors blow up rapidly!
ON THE STABILITY OF RECURSIVE FORMULAS

TABLE I
AN EXAMPLE USING ALGORITHM (3) IN THE BACKWARD DIRECTION

<table>
<thead>
<tr>
<th>points</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.099850</td>
</tr>
<tr>
<td>7</td>
<td>.078315</td>
</tr>
<tr>
<td>6</td>
<td>.054807</td>
</tr>
<tr>
<td>5</td>
<td>.027538</td>
</tr>
<tr>
<td>4</td>
<td>.067231</td>
</tr>
<tr>
<td>3</td>
<td>-.501005</td>
</tr>
<tr>
<td>2</td>
<td>5.02847</td>
</tr>
<tr>
<td>1</td>
<td>-49.2735</td>
</tr>
<tr>
<td>0</td>
<td>478.155</td>
</tr>
</tbody>
</table>

The catastrophic instability in the backward direction can indicate strong stability in the forward direction. The well-known Miller's algorithm (see [16], p. 153) is based on this principle. Thus, the stability of a recursion depends on the direction in which it is used. In this paper, unless otherwise stated, the direction of recursive evaluation is the forward direction.

2. RELATIVE ERROR VS ABSOLUTE ERROR

GOOVAERTS and DE VYLDER [9] (p. 57) have discussed the propagation of absolute errors of the recursion (3). Based on their analysis about the inflation of absolute errors, they concluded that the recursion (3) seems to be unstable.

There is nothing wrong in their error analysis, but the conclusion they drew is inappropriate because the absolute error has little bearing on the behavior of errors relative to the required solution. We want to stress one basic point in standard numerical analysis: “as a measure of accuracy, the absolute error may be misleading and the relative error more meaningful” – BURDEN and FAIRES [1] (p. 13). The criterion for the stability of an algorithm should be relative error, rather than absolute error.

Example 2: For a Poisson distribution with a large mean $\lambda$, say $\lambda = 1000$, assume ideal computing which gives exact solutions using the recursion:

$$p_n = \frac{\lambda}{n} p_{n-1}, \quad n \geq 1.$$  

(9)

Thus, in the above ideal computing process there is no error propagation. Rounding errors only occur when the computer outputs the exact solution. Only a finite number $r$ ($r$ can be any desired number) digits can be represented in the output. In this way, both the first point, $p_0$, and the mean point, $p_{1000}$, are obtained. When $r = 10$, one has

$$p_0 = .5075958897 \times 10^{-434}, \quad \text{and} \quad p_{1000} = 0.1261461134,$$
with absolute errors of about
\[ 10^{-444}, \quad \text{and} \quad 10^{-12}, \]
respectively.

For any value of \( r \), the absolute error is inflated \( 10^{432} \) times when the recursive evaluation moves from \( p_0 \) to \( p_{1000} \). Obviously one cannot conclude that the algorithm (9) is unstable.

On the other hand, one can see that the algorithm (9) is stable by observing a constant relative error in the evaluation process (the relative errors for \( p_0 \) and \( p_{1000} \) are about the same at \( 10^{-7} \)).

To conclude this section, we cite Oliver’s ([11], p. 324) argument about the criterion of stabilities of recursions:

"If we should wish to determine the number of significant figures in the computed values, then the absolute stability of the relation is quite irrelevant; what matters is the behavior of the propagated errors relative, not to unity, but to the required solution."

3. LINEAR RECURSIONS OF FINITE ORDER

Consider the linear homogeneous recurrence equation in the forward direction

\[ g(x) = \sum_{j=1}^{m} A_j(x) g(x-j), \quad x > k, \quad A_m(x) \neq 0, \]

where \( m \) is called the order of the recurrence equation. The point \( k \) is the starting point of the recursion and \( g(k-m+1), \ldots, g(k) \) are the initial values.

For any given initial conditions

\[ \{ g(j) = \alpha_j; \ j = k-m+1, \ldots, k\}; \ (\alpha_{k-m+1}, \ldots, \alpha_k) = \alpha, \]

the linear recurrence equation (10) has one and only one solution, \( g_{\alpha,k}(x) \). Any solution of (10) can be represented by its initial values. Also, the solution \( g_{\alpha,k}(x) \) linearly depends upon the initial vector \( \alpha \):

\[ g_{c_1 \alpha + c_2 \beta, k}(x) = c_1 g_{\alpha,k}(x) + c_2 g_{\beta,k}(x) \]

The homogeneous linear recurrence equation (10) possesses a linearly independent set of solutions \( \{ g^{(h)}(x), 1 \leq h \leq m \} \), called a fundamental set, and any solution of (10) can be expressed as a linear combination of these functions.

**Definition 1:** A solution \( g(x) \) of equation (10) is called a dominant solution, if for any solution \( h(x) \) of equation (10) there exists a constant \( C > 0 \), such that

\[ |g(x)| \geq C|h(x)|, \quad x > K \quad \text{for some} \quad K \geq k. \]
A solution $h(x)$ of equation (10) is called a **subordinate** solution, if there exists a solution $g(x)$ of equation (10) such that

$$\lim_{x \to \infty} \left| \frac{g(x)}{h(x)} \right| = \infty;$$

in this case, we say that $g(x)$ dominates $h(x)$.

It should be noted that some solutions may be neither dominant nor subordinate. However, for most recurrence equations that are encountered in practical applications, their coefficients $A_j(x)$ satisfy some regularity conditions and there exists a fundamental set $\{g^{(h)}(x), 1 \leq h \leq m\}$ such that

- $g^{(1)}(x)$ is a dominant solution and free from zero for $x$ sufficiently large;

- $\lim_{x \to \infty} g^{(1)}(x)/g^{(h)}(x) = \infty, \text{ for } 2 \leq h \leq m$.

(See CASH [2], p. 2; WIMP [23], p. 19 and p. 272-9).

**Remarks:**

For positive arithmetic severities with finite support, by a simple rescaling, one can assume that $f(x)$ is defined on positive integers with finite support $\{x_1, x_2, \ldots, x_r\}$ such that

$$1 \leq x_1 < x_2 < \ldots < x_r < \infty,$$

$$\gcd(x_1, x_2, \ldots, x_r) = 1,$$

where $\gcd$ stands for greatest common divisor. In this case, formula (3) becomes a special case of (10) with $m = x_r$ and $k = 0$:

$$g(x) = \sum_{j=1}^{m} \left( a + b \frac{j}{x} \right) f(j) g(x-j)$$

with initial values:

$$\{g(x) = 0; x = -m+1, \ldots, -1\}; g(0) = p_0 > 0.$$

**4. RELATIVE STABILITY THEORY**

For the general linear recurrence equation (10), Oliver\(^1\) [11] proposed a theory of relative stability. Oliver's relative stability theory is presented with modifications and refinement.

\(^1\) J. Oliver, wrote his Ph.D. dissertation partly on the relative stability theory of linear recurrence algorithms under J.C.P. Miller at Cambridge.
4.1. Concepts and definitions

Definition 2: The desired solution of recursion (10) is a special solution to be computed, which can be represented by the initial values

\[ g(j) = \alpha_j; \ j = k-m+1, \ldots, k; \quad (\alpha_{k-m+1}, \ldots, \alpha_k) = \alpha. \]

We denote this desired solution as \( g_{\alpha, k}(x) \).

Notation: We use \( \varepsilon \) to denote absolute errors and \( \eta \) to denote relative errors.

Two possible ways to generate round-off errors are: (i) rounding, and (ii) chopping. Most computers use rounding; however, some computers do use chopping.

As indicated in Example 2, when the desired solution is a rapidly varying solution, the absolute round-off errors also vary rapidly. However, Oliver [11] (p. 326-7) pointed out that, for a rapidly varying solution, floating point arithmetic would be used. If floating point arithmetic is used then the actual relative round-off errors \( \eta \), are fairly evenly distributed within a small range

\[ \{ [-\eta, \eta] \text{ if rounding is used,} \]
\[ \{ [-\eta, 0] \text{ if chopping is used.} \]

If \( r \) digits are assigned by a user to the computer, \( r+1 \) digits would be actually used by the computer to leave some room for rounding or chopping. Then every real number in the floating-point range of the computer can be represented with a relative error bounded by

\[ \tilde{\eta} = \begin{cases} 0.5 \times 10^{-r} & \text{if rounding is used,} \\ 10^{-r} & \text{if chopping is used.} \end{cases} \]

(See Dahlquist and Bjorck [4], p. 45).

To symbolize this fact, we give the following definition.

Definition 3: The basis relative error generator \( \tilde{\eta}_{\text{gen}} \) is a random variable uniformly distributed on

\[ \{ [-\tilde{\eta}, \tilde{\eta}] \text{ if rounding is used,} \]
\[ \{ [-\tilde{\eta}, 0] \text{ if chopping is used.} \]

During the recursive evaluation by computers, each of the initial values \( g(j); \ j = k-m+1, \ldots, k \) has only initial round-off error. After the starting point \( k \), there are two sources of errors in each step of the evaluation of \( g(x) \):

\(^2\) To be consistent, 'the number of digits' will refer to the number of digits assigned to the computer.
(i) the propagation of earlier errors, and (ii) the newly generated round-off error when the computer outputs its ‘exact’ result assuming that all inputs are exact. We assume that the newly generated round-off errors are independent and identically distributed random variable $\eta_{\text{gen}}$. Obviously, for any newly generated error, it will be propagated in the same way as the true ‘value’ and thus satisfies the recursion (10).

**Definition 4:** The relative error for the initial value $g(j) = \alpha_j$ is $\eta_j$ (a value of $\eta_{\text{gen}}$). The propagation of the initial value errors is a solution $\varepsilon_k(x)$ of (10) which satisfies the initial condition:

$$
(22) \quad \varepsilon_k(j) = \eta_j \alpha_j, \quad j = k - m + 1, \ldots, k.
$$

We shall adopt the following convention: if one of the initial values $\alpha_j$ is zero, then the actual value used will be correct. This is equivalent to assuming that the computer can represent zero exactly, i.e. all bits set to zero. For example, in the initial conditions (18) of the recursion (17), the first $m - 1$ initial values are zero, and in actual computing they are used as zero without error. OLIVER [11] (p. 330) also supports this convention.

**Definition 5:** The (newly generated) round-off relative error at point $\tau$ ($\tau > k$) is $\eta_{\tau}$ (a value of $\eta_{\text{gen}}$). The propagation of the round-off error at $\tau$ is a solution $\varepsilon_{\tau}(x)$ which satisfies the initial condition at $\tau$:

$$
(23) \quad \{\varepsilon_{\tau}(\tau - m + j) = 0; \ j = 1, \ldots, m - 1\}; \ \varepsilon_{\tau}(\tau) = \eta_\tau g_{\text{gen}}(\tau).
$$

### 4.2. The Basic Error Propagation

Consider the first order homogeneous linear recursion:

$$
(24) \quad g(x) = cg(x - 1), \quad x \geq 1.
$$

For recursion (24), it is easy to see that the propagated value of any generated error remains constant relative to the solution $g(x)$:

$$
(25) \quad \frac{\varepsilon_i(x)}{g(x)} = \eta_i, \quad i = 0, 1, 2, \ldots
$$

An upper bound for the accumulated relative error is

$$
(26) \quad \frac{\Sigma_{i=0}^x |\varepsilon_i(x)|}{|g(x)|} \leq \frac{\Sigma_{i=0}^x |\varepsilon_i(x)|}{|g(x)|} \leq (x + 1) \eta.
$$

Note that at worst the accumulated relative error increases linearly with the number of points that have been evaluated. “This is an acceptable form of error accumulation, since if floating point arithmetic is used then doubling the range of evaluation corresponds to the loss of a single binary digit (in terms of error bounds rather than actual errors).” – OLIVER [11] (p. 325).
We define the basic error propagation for which (26) holds, i.e. relative error bound grows linearly with a slope no greater than 1, and we judge the acceptability of error behavior in the general case by comparing it with the above basic error propagation.

4.3. Index of error propagation

**Definition 6**: The range of interest for recursion (10) is the interval \([k, R]\) over which the values of \(g(x)\) are to be computed.

**Definition 7**: The index of error propagation for the recursion (10) evaluating the desired solution \(g_{\bar{z}, k}(x)\) over the range \([k, R]\) is defined by

\[
I(k, R) := \sup_{x \in [k, R]} \left\{ \frac{1}{(x-k+1)\bar{h}} \left| \sum_{i=k}^{x} e_i(x) \right| \right\}
\]

In evaluating the desired solution,

1. if \(I(k, R)\) is bounded, we say that the recursion (10) is **stable** over the range \([k, R]\).
2. if \(I(k, R) \leq 1\), we say that the recursion (10) is **strongly stable** over the range \([k, R]\).
3. if \(I(k, R) = \infty\), we say that the recursion (10) is **unstable** over the range \([k, R]\).

In other words, a recursive evaluation is stable if the round-off error grows linearly, and being strongly stable if the linear slope is bounded by 1; a recursive evaluation is unstable if the round-off error grows more rapidly than linear; for example, exponentially.

**Theorem 1**: The linear recursion (10) is stable for evaluating its dominant solutions, and unstable for evaluating its subordinate solutions.

This result can be found in **WIMP** [23] (p. 10) and **CASH** [2] (p. 3). Here we just give an intuitive interpretation.

Let \(g^{(h)}(x), (h = 1, 2, \ldots, m)\), be a fundamental set of (10) such that \(g^{(1)}(x)\) is a dominant solution and

\[
\lim_{x \to \infty} \frac{g^{(h)}(x)}{g^{(1)}(x)} = 0, \quad \text{for} \quad h = 2, \ldots, m.
\]

The solution \(g_{\bar{z}, k}(x)\) to be computed can be written as a linear combination of this fundamental set:

\[
g_{\bar{z}, k}(x) = d_1 g^{(1)}(x) + \ldots + d_m g^{(m)}(x),
\]
ON THE STABILITY OF RECURSIVE FORMULAS

where

\[ d_1 \begin{cases} = 0 & \text{if } g_{\alpha,k}(x) \text{ is subordinate} \\ \neq 0 & \text{if } g_{\alpha,k}(x) \text{ is dominant} \end{cases} \]

On the other hand, the round-off error propagation \( \varepsilon_{r}(x) \), as a disturbance solution, can be written as a linear combination of the fundamental set:

\( \varepsilon_{r}(x) = c_1 g^{(1)}(x) + \ldots + c_m g^{(m)}(x) \),

where even though \( c_1 \) is small, but with probability 1 that \( c_1 \neq 0 \). Since \( c_1 \neq 0 \), one has

\[ \lim_{x \to \infty} \frac{\varepsilon_{r}(x)}{g_{\alpha,k}(x)} = \begin{cases} \infty & \text{if } g_{\alpha,k}(x) \text{ is subordinate} \\ \frac{c_1}{d_1} & \text{if } g_{\alpha,k}(x) \text{ is dominant} \end{cases} \]

where \( \frac{c_1}{d_1} \) can be made arbitrarily small by using sufficient number of digits.

Therefore, a recursive evaluation by (10) is stable if the desired solution \( g_{\alpha,k}(x) \) is dominant; and is unstable if the desired solution \( g_{\alpha,k}(x) \) is subordinate.

Also, we can see that, regardless of our desired solution, the computation always generate a dominant solution \( g_{\alpha,k}(x) + \varepsilon_{r}(x) \).

Example 3: Consider the following linear recursion:

\[ g(x) = g(x-1) - \frac{3}{16} g(x-2), \quad x > 2. \]

Equation (32) has a fundamental set of solutions

\[ g^{(1)}(x) = (.75)^x, \quad g^{(2)}(x) = (.25)^x. \]

Where \( g^{(1)}(x) \) is a dominant solution, and \( g^{(2)}(x) \) is a subordinate solution. A combination \( c_1 g^{(1)}(x) + c_2 g^{(2)}(x) \) is a dominant solution if and only if \( c_1 \neq 0 \). Also, a solution \( g(x) \) is a dominant solution if and only if

\[ \lim_{x \to \infty} \frac{g(x)}{g(x-1)} = .75. \]

(1). Evaluate the desired solution \( g^{(1)}(x) \) by recursion (32) with initial values:

\[ g^{(1)}(1) = .75, \quad g^{(1)}(2) = .75^2. \]
The computed results for some selected points are listed in Table 2 (5 digits are used in the evaluation).

**TABLE 2**
EVALUATION OF THE DOMINANT SOLUTION \( g^{(1)}(x) \)

<table>
<thead>
<tr>
<th>x</th>
<th>calculated ( g^{(1)}(x) )</th>
<th>relative error</th>
<th>x</th>
<th>calculated ( g^{(1)}(x) )</th>
<th>relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.23730</td>
<td>-.000019753</td>
<td>75</td>
<td>.42622 \times 10^{-9}</td>
<td>.000089971</td>
</tr>
<tr>
<td>10</td>
<td>.056318</td>
<td>-.000079649</td>
<td>100</td>
<td>.32074 \times 10^{-12}</td>
<td>.00061680</td>
</tr>
<tr>
<td>20</td>
<td>.0031710</td>
<td>-.000066832</td>
<td>200</td>
<td>.10291 \times 10^{-24}</td>
<td>.00047191</td>
</tr>
<tr>
<td>30</td>
<td>.00017858</td>
<td>-.000011704</td>
<td>300</td>
<td>.33020 \times 10^{-37}</td>
<td>.00091698</td>
</tr>
<tr>
<td>40</td>
<td>.000010057</td>
<td>-.000041251</td>
<td>400</td>
<td>.10592 \times 10^{-49}</td>
<td>.0010888</td>
</tr>
<tr>
<td>50</td>
<td>.56639 \times 10^{-6}</td>
<td>.00012068</td>
<td>500</td>
<td>.33970 \times 10^{-62}</td>
<td>.0010686</td>
</tr>
</tbody>
</table>

From Table 2, one can observe that the relative error grows very slowly. The accumulated error at \( x \) is bounded by \( (x-1) \bar{\eta} \), (i.e. the evaluation of \( g^{(1)}(x) \) is stable).

(II). Evaluate the desired solution \( g^{(2)}(x) \) by recursion (32) with initial values:

\[
 g^{(2)}(1) = .25, \quad g^{(2)}(2) = .25^2.
\]

The computed results for some selected points are listed in Table 3 (5 digits are used in the evaluation).

**TABLE 3**
EVALUATION OF THE SUBORDINATE SOLUTION \( g^{(2)}(x) \)

<table>
<thead>
<tr>
<th>x</th>
<th>calculated ( g^{(2)}(x) )</th>
<th>relative error</th>
<th>( g^{(2)}(x)/g^{(2)}(x-1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.015625</td>
<td>0</td>
<td>.25000</td>
</tr>
<tr>
<td>5</td>
<td>.0009763</td>
<td>-.00026880</td>
<td>.24995</td>
</tr>
<tr>
<td>10</td>
<td>.8781 \times 10^{-6}</td>
<td>-.079245</td>
<td>.23643</td>
</tr>
<tr>
<td>20</td>
<td>-.42568 \times 10^{-8}</td>
<td>-.4681.4</td>
<td>.75032</td>
</tr>
<tr>
<td>30</td>
<td>-.23977 \times 10^{-9}</td>
<td>-.27644 \times 10^{5}</td>
<td>.75001</td>
</tr>
<tr>
<td>50</td>
<td>-.76036 \times 10^{-12}</td>
<td>-.96387 \times 10^{18}</td>
<td>.75000</td>
</tr>
</tbody>
</table>

From Table 3, one can observe that the round-off errors blow up rapidly. The recursive evaluation is very unstable. By checking the ratio \( g^{(2)}(x)/g^{(2)}(x-1) \) of the computed results, one can see that the computed solution (eventually) follows a pattern of a dominant solution.

**Remarks:**

1. In general, every (non-trivial) linear recursion is stable for some solution and unstable for other solutions. Thus it is meaningless to merely talk about the stability of a recursion without mentioning the desired solution.
However, for simplicity, when we talk about the stability of a recursion without specifying which solution it refers to, we assume that the desired solution is implicitly known.

2. It is the rate of growth of the desired solution with respect to other solutions of the recursive equation that determines whether or not the recursive computation is successful. In terms of initial value representations, the family of subordinate solutions form an \( m-1 \) dimensional surface in the \( m \) dimensional space of all solutions of (10). As a result of round-off errors and higher-order round-off errors, the disturbance solution can be in any direction in the space of all solutions of (10). Therefore, in general, no matter whether the desired solution is dominant or not, the computed result follows a pattern of a dominant solution. When the desired solution is a subordinate solution, round-off errors will blow up and make the recursive evaluation ineffective.

We have clarified the stability concept of linear recursions. In the next section, we shall give a family of recursions whose non-negative solutions are dominant solutions.

5. CONGRUENT RECURSIONS OF FINITE ORDER AND THEIR DOMINANT SOLUTIONS

**Definition 8:** A linear recurrence equation of the form:

\[
g(x) = \sum_{j=1}^{m} B_j(x) f(j) g(x-j),
\]

with the following restrictions:

- \( f(x) \) is non-negative with finite support on \( \{x_1, x_2, \ldots, x_r\} \) which satisfies (15) and (16). Note that \( f(x) \) does not have to be a probability function
- \( B_j(x), j = 1, 2, \ldots, m, \) are strictly positive functions of \( x > 0 \)

is called a congruent recursion of finite order \( m \).

In this section, we are going to give the dominant solutions of congruent recursions.

We first discuss a set of solutions \( g^{(h)}(x) \) of (34) with starting point \( k \geq 0 \) and initial values

\[
g^{(h)}(k-m+j) = \delta_{h,j} = \begin{cases} 1 & \text{if } j = h, \\ 0 & \text{if } j \neq h, \end{cases} \quad 1 \leq h, j \leq m.
\]

**Proposition 1:** For a positive number \( n \), if it is a linear combination of \( x_1, x_2, \ldots, x_r \) with coefficients in \( \mathbb{Z}^0 = \{0, 1, 2, \ldots\} \), then

\[
g^{(h)}(k+h+n) > 0.
\]
Proof: Since

\[ g^{(h)}(k - m + h) = 1, \quad B_j(\cdot)f(\cdot) \geq 0, \]

from equation (34), we have

\[ g^{(h)}(k + h) \geq B_m(k + h)f(\cdot) > 0. \]

Now from point \(k + h\), apply the recursion (34) again:

\[ g^{(h)}(k + h + x_i) \geq B_{x_i}(\cdot)f(x_i)g^{(h)}(k + h) > 0 \]

\[ \ldots \]

By induction, for any \(n\) which is a linear combination of \(x_1, x_2, \ldots, x_r\) with coefficients in \(\mathbb{Z}^0\), we have

\[ g^{(h)}(k + h + n) > 0. \]

Lemma 1: Let \(x_1, x_2, \ldots, x_r \in \mathbb{Z}\) with \(x_i\) not all zero. The following statements are equivalent:

- \(x_1z_1 + x_2z_2 + \ldots + x_rz_r = 1\) has a solution in integers \(z_i\);
- \(gcd(x_1, x_2, \ldots, x_r) = 1\).

Proof: See Flath ([7], p. 13).

Proposition 2: Let \(x_1, x_2, \ldots, x_r\) be positive integers with

\[ gcd(x_1, x_2, \ldots, x_r) = 1. \]

There exists a constant \(N_0\), such that for any integer \(k > N_0\), \(k\) can be expressed as a linear combination of \(x_1, x_2, \ldots, x_r\) with coefficients in \(\mathbb{Z}^0 = \{0, 1, 2, \ldots\}\).

Proof: From Lemma 1, there exist \(z_1, z_2, \ldots, z_r \in \mathbb{Z}\) such that

\[ x_1z_1 + x_2z_2 + \ldots + x_rz_r = 1. \]

Let

\[ N_0 = x_1(|z_1|x_1 + |z_2|x_2 + \ldots + |z_r|x_r). \]

For any \(n > N_0\), apply the division algorithm to positive integers \(n - N_0\) and \(x_1\), we obtain

\[ n = N_0 + ux_1 + v, \quad u \geq 0, \quad 0 \leq v < x_1. \]

By replacing \(v\) with

\[ v(x_1z_1 + x_2z_2 + \ldots + x_rz_r), \]
we obtain
\[ n = N_0 + v(x_1 z_1 + x_2 z_2 + \ldots + x_r z_r) + u x_1, \]
which is a linear combination of \( x_1, x_2, \ldots, x_r \) with coefficients in \( \mathbb{Z}^0 \). \( \square \)

**Theorem 2:** For the congruent recursion (34), the solution \( g^{(h)}(x) \) is non-negative and free from zero when \( x \) gets large:
\[ g^{(h)}(x) \geq 0, \text{ for } x \geq k; \quad g^{(h)}(x) > 0, \text{ for } x > N_0; \]
where \( N_0 \) is some constant.

**Proof:** It is an immediate application of Propositions 1 and 2. \( \square \)

Now we are ready to generalize our results to the solutions of equation (34) with an arbitrary non-negative initial vector \( \vec{\alpha} \) with at least one positive element:
\[ g_{\vec{\alpha}, k}(x) = \sum_{j=1}^{m} \alpha_{k-m+j} x^j, \quad (j = 1, 2, \ldots, m); \quad \sum_{j=1}^{m} \alpha_{k-m+j} > 0. \]

**Theorem 3:** For the congruent recursion (34) with initial conditions (38), the solution \( g_{\vec{\alpha}, k}(x) \) is non-negative and free from zero when \( x \) gets large:
\[ g_{\vec{\alpha}, k}(x) \geq 0, \text{ for } x \geq k; \quad g_{\vec{\alpha}, k}(x) > 0, \text{ for } x > N_0 \]
where \( N_0 \) is some constant.

We say that \( \vec{\alpha} \geq \vec{\beta} \) if an only if \( \alpha_{k-m+j} \geq \beta_{k-m+j} \) for \( j = 1, 2, \ldots, m \).

**Theorem 4 (Comparison):** For the congruent recursion (34), if \( \vec{\alpha} \geq \vec{\beta} \), then
\[ g_{\vec{\alpha}, k}(x) \geq g_{\vec{\beta}, k}(x), \text{ for } x \geq k. \]

**Proof:** Since \( \vec{\alpha} \geq \vec{\beta} \), we have \( \vec{\alpha} - \vec{\beta} \geq 0 \) and from equation (34) we have
\[ g_{\vec{\alpha}, k}(x) - g_{\vec{\beta}, k}(x) = g_{\vec{\alpha} - \vec{\beta}, k}(x) \geq 0. \] \( \square \)

**Theorem 5:** For the congruent recursion (34) with initial conditions (38), the solution \( g_{\vec{\alpha}, k}(x) \) is a dominant solution.

**Proof:** From Theorem 3, we can move the starting point from \( k \) to a new point \( K \) such that
\[ \vec{y} = (y_{k-m+1}, \ldots, y_K) = \{g_{\vec{\alpha}, k}(K-m+1), \ldots, g_{\vec{\alpha}, k}(K)\} \]
has all its \( m \) components strictly positive.

Obviously \( g_{\vec{\alpha}, k}(x) \) and \( g_{\vec{y}, k}(x) \) are the same for \( x > K \).
Let $h(x)$ be any solution of the congruent recursion (34), and

$$
\beta = (\beta_{K-m+1}, \ldots, \beta_K) = \{h(K-m+1), \ldots, h(K)\}.
$$

Since a finite number of values are always bounded, there is a positive constant $\xi$ ($0 < \xi < \infty$) such that

$$
\xi \cdot |\beta_{K-m+j}| \geq |\beta_{K-m+j}|, \quad j = 1, 2, \ldots, m.
$$

Therefore

$$
g_{\bar{z}, k}(x) = g_{\bar{z}, k}(x) \geq \xi^{-1}|h(x)|, \quad \text{for } x > K.
$$

6. NON-HOMOGENEOUS RECURSIONS OF INFINITE ORDER

Now we extend our discussions to a general family of non-homogeneous recursions of infinite order.

**Definition 9:** A recurrence equation of the form

$$
g(x) = \sum_{j=1}^{x} A_j(x) g(x-j) + H(x), \quad x > k \geq 0,
$$

is called a **non-homogeneous** recursion of **infinite** order.

**Definition 10:** The recurrence equation of infinite order

$$
g(x) = \sum_{j=1}^{x} A_j(x) g(x-j), \quad x > k \geq 0,
$$

is called the **homogeneous counterpart** of recursion (42).

The homogeneous counterpart (43) is also a special case of (42) with $H(x) = 0$. When $H(x) = 0$, the homogeneous counterpart of equation (43) is itself.

For an example, when claim severity has a infinite support, the recursion (3) is homogeneous recurrence equation of infinite order.

**Definition 11:** The desired solution is a special solution of the non-homogeneous recursion (42) to be computed, which can be represented by the initial values:

$$
g(j) = \alpha_j, \quad j = 0, 1, \ldots, k; \quad (\alpha_0, \alpha_1, \ldots, \alpha_k) = \bar{\alpha}.
$$

We denote this desired solution as $g_{\bar{z}, k}(x)$.

In the above definition, without loss of generality, we assumed that the initial points are $\{0, 1, \ldots, k\}$. If initial points are $\{r, r+1, \ldots, r+k\}$, one can always introduce a new variable $x' = x-r$ and get a new equation in terms of $x'$. Of course, the stabilities for these two recursions are equivalent. Note that, by a transformation $x' = x-(k-m+1)$ the recursion (10) of finite order is a special case of (42) with $H(x) = 0$ and $A_j(x) = 0$ for $j > m$. 


Since both the desired solution \( \hat{g}_{a,k}(x) \) and the computed values \( \tilde{g}_{a,k}(x) \) satisfy the non-homogeneous recursion (42), the accumulated absolute error \( \hat{g}_{a,k}(x) - \tilde{g}_{a,k}(x) \) satisfies the homogeneous counterpart (43).

**Definition 12:** The relative error for the initial value \( g(j) = \alpha_j \) is \( \eta_j \). The propagation of initial value errors is a solution \( \varepsilon_k(x) \) of the homogeneous counterpart (43) with the initial condition

\[
(45) \quad \varepsilon_k(j) = \eta_j \alpha_j, \quad j = 0, 1, \ldots, k.
\]

**Definition 13:** The (newly generated) round-off relative error at point \( \tau (\tau > k) \) is \( \eta_\tau \). The propagation of the round-off error at \( \tau \) is a solution \( \varepsilon_\tau(x) \) of the homogeneous counterpart (43) with the initial condition

\[
(46) \quad \varepsilon_\tau(j) = 0, \quad j = 0, 1, \ldots, \tau - 1; \quad \varepsilon_\tau(\tau) = \eta_\tau \hat{g}_{a,k}(\tau).
\]

Other definitions (e.g. index of error propagation and strongly stable, etc.) can be similarly defined as in the finite homogeneous case.

**Definition 14:** A non-homogeneous congruent recursion of infinite order is defined by:

\[
(47) \quad g(x) = \sum_{j=1}^{x} B_j(x) g(x-j) + H(x), \quad x > k,
\]

with \( B_j(x) \geq 0 \), and \( H(x) \geq 0 \).

The dominance ranking between the desired solution and the error solution determines whether the recursive evaluation is successful or not.

Unlike its homogeneous counterpart, a non-homogeneous first order recursion is not necessarily stable. This is because that, for a non-homogeneous recursion, the desired solution and the error solution satisfy two different equations.

**Example 4:** Consider the first order forward recursion:

\[
(48) \quad g(x) = g(x-1) - 0.5^x, \quad x \geq 1,
\]

with an initial value \( g(0) = 1 \). The desired solution is \( g(x) = 0.5^x \). A fundamental set of the homogeneous counterpart is given by \( g^{(1)}(x) = 1 \). Since \( g^{(1)}(x) \) dominates \( g(x) \), the recursive evaluation is unstable in evaluating \( g(x) \). This instability can be easily verified on a computer. If 5 digits are used, the computed results for the points after \( x = 40 \) become a constant \( 0.79228 \times 10^{-7} \), which again follows a pattern of a dominant solution.

Similarly, we have a comparison theorem.
Theorem 6: Let \( g_{\bar{a}, k}(x) \) and \( g_{\bar{b}, k}(x) \) be two solutions of the non-homogeneous congruent recursion (47), \( e_{\bar{b}, k}(x) \) be a solution of the homogeneous counterpart of (47). If \( \bar{a} \geq \bar{b} \), then

\[
g_{\bar{a}, k}(x) \geq g_{\bar{b}, k}(x) \geq e_{\bar{b}, k}(x), \quad \text{for} \quad x > k.
\]

From the above theorem, or by mathematical deduction, for non-negative initial vector \( \bar{a} \), the solution \( g_{\bar{a}, k}(x) \) of (47) is non-negative.

Theorem 7 (Strongly Stable): A non-homogeneous congruent recursion of infinite order (47) is strongly stable in evaluating \( g_{\bar{a}, k}(x) \) provided that \( \bar{a} \) is non-negative.

Proof: After the initial points, any vanishing of \( g_{\bar{a}, k}(x) \) results solely from zeros in the initial values and does not depend on previous non-zero \( g_{\bar{a}, k}(x) \) values. There is no error in this case.

We need only to be concerned with positive values of \( g_{\bar{a}, k}(x) \).

For the propagation \( e_k(x) \) of initial value errors, since

\[
|e_k(j)| \leq \bar{\eta} g(j) = \bar{\eta} \bar{z}_j, \quad j = 0, 1, \ldots, k,
\]

from Theorem 6, we have

(49) \[
\frac{|e_k(x)|}{g_{\bar{a}, k}(x)} \leq \bar{\eta}, \quad x > k.
\]

For the propagation \( e_{\tau}(x) \) of the newly generated round-off error at point \( \tau \), since

\[
e_{\tau}(j) = 0, \quad (j = 0, 1, \ldots, \tau - 1); \quad |e_{\tau}(\tau)| \leq \bar{\eta} g_{\bar{a}, k}(\tau),
\]

we have

(50) \[
\frac{|e_{\tau}(x)|}{g_{\bar{a}, k}(x)} \leq \bar{\eta}, \quad x > k.
\]

Therefore,

\[
\frac{1}{(x-k+1)\bar{\eta}} \left| \sum_{i=k}^{x} e_i(x) \right| \leq 1, \quad x > k.
\]

The strongly stable condition (27) holds. \( \square \)

In the proof, the inequalities (49) and (50) can be very loose. Thus, 1 is only a gross upper bound for \( I(k, \infty) \). It can be much less than 1 in actual error propagation. Another important factor is the offset of positive and negative relative errors when rounding is used by the computer.
Theorem 8: In evaluating non-negative solutions of the congruent recursion (47), if rounding is used by the computer, the accumulated relative error at point \( x \) is a random variable \( \tilde{u}(x) \) with values in
\[
[-(x-k+1)\tilde{\eta}, (x-k+1)\tilde{\eta}].
\]
\( \tilde{u}(x) \) has a mean of zero and variance \( (x-k+1)\tilde{\eta}^2/3 \).

Proof: It is a direct result from a sum of \( x-k+1 \) i.i.d. random variables which are uniformly distributed on \( [-\tilde{\eta}, \tilde{\eta}] \).

From Theorem 8, even though the upper bound for accumulated relative errors at \( x \) grows linearly with \( x \), the standard deviation is only a constant multiple of \( \sqrt{x-k+1} \). A 99% confidence interval for \( \tilde{u}(x) \) is approximately
\[
[-1.5 \sqrt{x-k+1} \tilde{\eta}, 1.5 \sqrt{x-k+1} \tilde{\eta}].
\]

7. FORWARD DIRECTION VS BACKWARD DIRECTION

The earlier discussions can also be easily extended to recursions in the backward direction. For simplicity, we only discuss recursions of finite order.

Definition 15: A recurrence equation of the form
\[
g(y) = \sum_{j=1}^{m} A_j(y) g(y+j) + H(y), \quad y < k,
\]
with
\[
g(j) = \alpha_j, \quad (j = k, \ldots, k+m-1), \quad \vec{\alpha} = (\alpha_k, \ldots, \alpha_{k+m-1})
\]
is called a non-homogeneous recursion in the backward direction with starting point \( k \) and initial vector \( \vec{\alpha} \). We denote this solution as \( g_{\vec{\alpha}, k}(y) \).

Definition 16: A non-homogeneous congruent recursion in the backward direction is defined by:
\[
g(y) = \sum_{j=1}^{m} B_j(y) g(y+j) + H(y), \quad y < k,
\]
with \( B_j(y) \geq 0 \), and \( H(y) \geq 0 \).

Similarly, we have a strongly stable theorem.

Theorem 9: The non-homogeneous congruent recursion (55) is strongly stable in evaluating its non-negative solution \( g_{\vec{\alpha}, k}(y) \) in the backward direction.

When a congruent recursion in the forward direction is rewritten as a recursion in the backward direction, it is no longer a congruent recursion in the backward direction. Thus, 'congruent is direction dependent!'

The links between the two directions are important.
For a first order homogeneous recursion, since there is no dominance ranking among the solutions, the recursion is strongly stable in both directions.

For a second order homogeneous recursion, there are only two solutions in a fundamental set. If the recursion is unstable in one direction, which means the undesired error solution grows unboundedly with respect to the desired solution, then this undesired error solution will decrease rapidly in the reverse direction, and thus the recursion is stable in the reverse direction.

For a recursion of order \( m \geq 2 \), its solutions are ranked by their dominance relationship. There may be solutions which are subordinate in both directions; for these solutions, the recursion is unstable in both directions. Nevertheless, if the desired solution dominates all other solutions (in a fundamental set) in one direction, then the same desired solution will be dominated by other solutions in the reverse direction. Thus, if a recursion is stable in one direction, it is unstable in the reverse direction. In general, the more stable a recursion is in one direction, the more unstable when it is used in the reverse direction.

**Definition 17:** Assuming that \( m \geq 2 \), a recursion is called **strongly unstable** in one direction for a desired solution if it is strongly stable in the reverse direction for the same desired solution.

The next two theorems follow directly from this definition.

**Theorem 10:** A recurrence equation (\( m \geq 2 \))

\[
g(x) = B_m(x) g(x-m) - \sum_{j=1}^{m-1} B_j(x) g(x-j) - H(x), \quad x > k,
\]

with \( B_j(x) \geq 0 \) and \( H(x) \geq 0 \) is strongly unstable in the forward direction in evaluating its non-negative solutions.

**Theorem 11:** A recurrence equation (\( m \geq 2 \))

\[
g(y) = B_m(y) g(y+m) - \sum_{j=1}^{m-1} B_j(y) g(y+j) - H(y), \quad y < k,
\]

with \( B_j(y) \geq 0 \) and \( H(y) \geq 0 \) is strongly unstable in the backward direction in evaluating its non-negative solutions.

**Example 5:** Reconsider Example 1. From Theorem 7, the forward recursion (6) is strongly stable in evaluating its non-negative solutions. From Theorem 11, the backward recursion (8) is strongly unstable in evaluating its non-negative solutions.
Example 6: Consider the recurrence equations for modified Bessel functions (see Press, et al. [16], p. 192):

\begin{align}
I_{n+1}(x) &= -(2n/x) I_n(x) + I_{n-1}(x), \\
K_{n+1}(x) &= +(2n/x) K_n(x) + K_{n-1}(x).
\end{align}

Since $I_n(x)$ and $K_n(x)$ are non-negative solutions for $x \geq 0$, the recursion (58) is strongly unstable in the forward direction, and the recursion (59) is strongly stable in the forward direction.

8. EMPIRICAL INFLATION FACTOR

In this section, based on the signs of the coefficients $A_j(x)$ and the term $H(x)$ in (42), we investigate the growth of the relative errors in each step of the recursive evaluation.

Lemma 2: Let $a$ and $b$ be two positive real values, with their estimates $\hat{a}$ and $\hat{b}$ having relative errors $\eta_1, \eta_2$, respectively. Then, $\hat{a} + \hat{b}$ as an estimate of $a + b$, has a relative error

\begin{equation}
\frac{a}{a+b} \eta_1 + \frac{b}{a+b} \eta_2
\end{equation}

which is bounded by $[-\eta, \eta]$ where

\begin{equation}
\eta = \max (|\eta_1|, |\eta_2|).
\end{equation}

As a special case, if $\eta_2 = 0 (\hat{b}$ is exact), then, $\hat{a} + \hat{b}$, as an estimate of $a + b$, has a relative error which is less than $\eta_1$. We say that the relative error is damped.

Lemma 3: Let $a$ and $b$ be two positive real values, with their estimates $\hat{a}$ and $\hat{b}$ having relative errors $\eta_1, \eta_2$, respectively. Then, $\hat{a}\hat{b}$ as an estimate of $ab$, has a relative error $\eta_1 + \eta_2$, provided that $\eta_i$ is small relative to 1 ($\eta_i \ll 1, i = 1, 2$). As a special case, $\hat{a}\hat{b}$, as an estimate of $ab$, has a relative error $\eta_1$.

Lemma 4: Let $a$ and $b$ be two positive real values, with their estimates $\hat{a}$ and $\hat{b}$ having relative errors of any value in the range $(-\bar{\eta}, \bar{\eta})$. Then, $\hat{a} - \hat{b}$, as an estimate of $a - b$, can have a relative error of any value in the range $(-\gamma\bar{\eta}, \gamma\bar{\eta})$, where

\begin{equation}
\gamma = \frac{a+b}{|a-b|},
\end{equation}

is called the error inflation factor.

In Lemma 4, one can see that, when $a \approx b$, $\gamma$ can be infinitely large, which causes extraordinary unstable result. This should be avoided in any computing schemes.
Consider the non-homogeneous recursion (42) of infinite order. The value \( g(x) \) at point \( x \) depends upon all previous values \( g(x-j), j = 1, \ldots, x \). In each step of recursive evaluation, there are \( x+1 \) terms involved:

\[ H(x) \quad \text{and} \quad A_j(x) \, g(x-j), \quad (j = 1, \ldots, x). \]

Some of them may be positive, and some may be negative. To indicate clearly the sign of each term, we re-write the equation (42) into the following form:

\[
(63) \quad g(x) = \sum_{j=1}^{x} s_j(x) \, B_j(x) \, g(x-j) + H^+(x) - H^-(x),
\]

such that

\[
(64) \quad B_j(x) \, g(x-j) = |A_j(x) \, g(x-j)| \geq 0,
\]

and

\[
(65) \quad s_j(x) = \begin{cases} 
1 & \text{if } A_j(x) \, g(x-j) > 0, \\
0 & \text{if } A_j(x) \, g(x-j) = 0, \\
-1 & \text{if } A_j(x) \, g(x-j) < 0,
\end{cases}
\]

and

\[
(66) \quad H^+(x) = \frac{|H(x)| + H(x)}{2}, \quad H^-(x) = \frac{|H(x)| - H(x)}{2}.
\]

**Definition 18:** Associated with the computed solution \( g(x) \), we define a positive part \( g_+(x) \) and a negative part \( g_-(x) \) at each point \( x \) such that

\[
(67) \quad g_+(x) = \sum_{j=1}^{x} B_j(x) \, g(x-j) + H^+(x) \geq 0
\]

\[
(68) \quad g_-(x) = \sum_{j=-1}^{x} B_j(x) \, g(x-j) + H^-(x) \geq 0
\]

\[
(69) \quad g(x) = g_+(x) - g_-(x)
\]

**Definition 19:** An empirical inflation factor at \( x \) is defined by

\[
(70) \quad \hat{\gamma}(x) = \frac{g_+(x) + g_-(x)}{|g_+(x) - g_-(x)|}, \quad \text{if } g_+(x) \neq g_-(x);
\]

and

\[
(71) \quad \hat{\gamma}(x) = \infty, \quad \text{if } g_+(x) = g_-(x).
\]
Definition 20: If \( k \) is the starting point of the recursion (63), then we define an empirical accumulated relative error bound recursively:

\[
\hat{u}(x) = \hat{u}(x-1) \hat{v}(x) + \bar{\eta}, \quad x > k,
\]

with initial value \( \hat{u}(k) = \bar{\eta} \).

Theorem 12: Assuming that \( r \) digits are used. For the computed value \( g(x) \) by recursion (42), an empirical upper bound of the relative error is given by

\[
\hat{u}(x) \leq \begin{cases} 
(x-k+1) \prod_{i=k}^{x} \hat{v}(i) \times 0.5 \times 10^{-r} & \text{if rounding is used}, \\
(x-k+1) \prod_{i=k}^{x} \hat{v}(i) \times 10^{-r} & \text{if chopping is used}.
\end{cases}
\]

Proof: It can be easily verified by mathematical induction.

Definition 21: We say that the number of significance digits in the computed value \( g(x) \) is \( v(x) \) if the relative error is less than \( 10^{-v(x)} \).

One can empirically estimate the number of significant digits \( v(x) \) in the computed value \( g(x) \) by the following inequality:

\[
v(x) \geq \hat{v}(x) = \left\lfloor -\log_{10} \hat{u}(x) \right\rfloor = \left\lfloor -\frac{\ln \hat{u}(x)}{\ln 10} \right\rfloor
\]

where \( \ln x \) denotes the natural logarithm of \( x \), \( \lfloor x \rfloor \) denote the largest integer which is no greater than \( x \). For example, \( \lfloor 2.317 \rfloor = 2 \), and \( \lfloor -2.317 \rfloor = -3 \).

Example 7: Reconsider the backward recursion (8) in Example 1. Now we calculate the estimated \( \hat{v}(x) \) and compare it with the actual \( v(x) \) at each point. The results are listed in Table 4.

<table>
<thead>
<tr>
<th>( x )</th>
<th>computed ( g(x) )</th>
<th>( \hat{v}(x) )</th>
<th>( \hat{y}(x) )</th>
<th>exact ( g(x) )</th>
<th>actual ( v(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.099850</td>
<td>4</td>
<td>22.71</td>
<td>.0998450</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>.078315</td>
<td>3</td>
<td>25.22</td>
<td>.0783629</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>.054807</td>
<td>2</td>
<td>28.15</td>
<td>.0543124</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>.027538</td>
<td>0</td>
<td>38.81</td>
<td>.0325723</td>
<td>0</td>
</tr>
</tbody>
</table>

The catastrophic instability of the backward recursion (8) can be seen from the large inflation factors \( \hat{y}(x) \) in Table 4.
Remarks:

If all the terms are of the same sign, (i.e. either \( g_+ (x) = 0 \) or \( g_- (x) = 0 \) for all \( x \geq k \)), then \( \tilde{\mu}(x) = (x - k + 1) \eta \) and

\[
\nu(x) \geq \begin{cases} 
  r + [\log_{10} 2 - \log_{10} (x-k+1)] & \text{if rounding is used,} \\
  r + [-\log_{10} (x-k+1)] & \text{if chopping is used.}
\end{cases}
\]

Our earlier results about the non-negative solutions of congruent recursions are ‘recovered’.

One should interpret the inflation factors with care. For an example, in evaluating the dominant solution \( g^{(0)}(x) \) in Example 3, the inflation factors are a constant \( \hat{\gamma} = 1.6667 \), but error inflations seldom occur and the evaluation is stable.

9. APPLICATIONS

Note that the recursion (3) is a special case of (47) with \( H(x) = 0 \) and starting point \( k = 0 \). The initial value \( g(0) \) is positive and the desired compound distribution is non-negative. If the claim frequency is in the family of Poisson, Negative Binomial or Geometric distributions, we have, from PANJER [12],

\[
B_j(x) = a + b \frac{j}{x} > 0, \quad j = 1, \ldots, x.
\]

As an immediate application of Theorem 7, the recursion (3) is strongly stable in evaluating compound Poisson, compound Negative Binomial and compound Geometric distributions.

In using recursion (3) to evaluate compound Poisson, compound Negative Binomial and compound Geometric distributions, the accumulated relative error bound grows linearly with a slope no greater than 1. If the evaluation starts at point \( x = 0 \) and \( r \) digits are used, a guaranteed number of significance digits in the computed \( g(x) \) can be estimated by the following simple inequality:

\[
\nu(x) \geq \begin{cases} 
  r + [\log_{10} 2 - \log_{10} (x+1)] & \text{if rounding is used,} \\
  r + [-\log_{10} (x+1)] & \text{if chopping is used.}
\end{cases}
\]

If rounding is used by the computer, with a probability of 99%,

\[
\nu(x) \geq r + \left[ \log_{10} \frac{4}{3} - \frac{1}{2} \log_{10} (x+1) \right].
\]

For example, if both claim frequency and claim size have a mean 1000, one wishes to get an accuracy with relative errors less than \( 10^{-7} \) over the interested range \([0, 10^7]\). One can achieve this accuracy by using 14 digits. Also, with (at least) 99% confidence, one can achieve this accuracy by using only 11 digits. This strongly stable property has practical significance in applications of
If one increases the number of points by a factor of 100 in the discretization of severity distribution, simply adding 2 digits can keep the same level of accuracy.

As an application of Lemma 2 and Lemma 3, the effect of round-off coefficients can be considered. For any finite number of positive values, their summation has the same level of relative error, and their product has a relative error bound which is the summation of individual relative error bounds. For any non-negative solution of the recursion (47), if the relative round-off errors of $B_j(x)$ and $f(j)$ are i.i.d. random variable $\eta_{gen}$, then the index of relative error propagation enlarges only by a constant multiple of 3. One additional digit is sufficient to protect the solution from round-off errors in the coefficients.

The condition (76) does not hold for the family of compound Binomial distributions. Compound Binomial distributions share a special feature that it has only finite support when claim size has finite support. Since the desired solution eventually becomes zero in the forward direction, it cannot be a dominant solution. From Theorem 1, recursion (3) is unstable in evaluating compound Binomial distributions. This instability can be encountered at the right tail of the compound distribution in the forward direction. A special treatment for compound Binomial distributions is given in the next section.

10. THE CASE OF COMPOUND BINOMIAL

In this section, we investigate in more detail about the instability of compound Binomial distribution. Based on some special features of compound Binomial distribution, a simple method to cope with this instability is given.

Consider the case that the claim frequency has a Binomial distribution:

\begin{equation}
    p_n = \frac{N!}{n!(N-n)!} \theta^n (1-\theta)^{N-n}, \quad 0 \leq n \leq N.
\end{equation}

Then, in recursion (3),

\begin{equation}
    a = -\frac{\theta}{1-\theta}, \quad b = (N+1)\frac{\theta}{1-\theta}.
\end{equation}

Example 8 \footnote{All the numerical examples in this paper are done on Maple V [6], on which one can freely assign the number of digits. Rounding is used by Maple V.}: Consider compound Binomial distribution with parameters

\begin{align*}
    \theta &= .95, \quad N = 100,
\end{align*}

and with claim severity distribution as in Table 5.

In order to investigate how unstable the recursion (3) is in evaluating this compound Binomial distribution, we use 200 digits in the calculation. The
empirical inflation factors are calculated along with the recursive evaluation. The results for some selected points are listed in Table 6.

From Table 6, one can see that the error inflation factor remains flat at 1 when \( x \leq 100 \), and accelerates after \( x > 100 \). The accelerating growth in the error inflation factors indicates that the recursive evaluation becomes more and more unstable when it proceeds to the right half of the compound Binomial distribution. Even 200 digits can not protect the desired solution from the disturbance of rounding-off errors! In the computed values of \( g(x) \), we obtained the following absurd results:

\[
g(898) = -1.9502 \times 10^{-93}, \quad \text{and} \quad g(1000) = -5.9052 \times 10^{-70}.
\]

The computed \( g(x) \) becomes negative at \( x = 898 \), which tells us that the empirical estimates \( \hat{\nu}(x) \) and \( \hat{\nu}(x) \) after point \( x = 898 \) are no longer reliable.

### 10.1. A combined usage of two directions

This method involves two recursions: (i) the forward direction, and (ii) the reverse recursion in the backward direction staring at the end point \( mN \).

When the claim severity has a finite support \( \{x_1, x_2, \ldots, x_r\} \), recursion (3) can be written into a recursion (17) of finite order \( m = x_r \). The recursion (17) can be easily turned into a backward recursion:

\[
(81) \quad g(y) = \frac{1}{P(y)} \left\{ g(y+m) - \sum_{j=1}^{m-1} \left( a + b \frac{m-j}{y+m} \right) f(m-j) g(y+j) \right\}
\]

where

\[
(82) \quad P(y) = \left( a + b \frac{m}{y+m} \right) f(m).
\]
For compound Binomial distributions, the boundary condition at the end point $mN$ is known and can be used as an initial value for the backward recursion:

\[(83) \quad g(mN) = \theta^N f(m)^N, \quad g(mN+j) = 0, \quad \text{for } j = 1, 2, \ldots \]

**Theorem 13:** In evaluating compound Binomial distributions, we have the following results:

1. The forward recursion (3) is locally strongly stable over the range $[0, N+1]$.
2. The forward recursion (17) is locally strongly stable over the range $[0, N+1]$, and becomes strongly unstable when it proceeds to the range $[mN-N-1, mN]$.
3. The backward recursion (81) is locally strongly stable over the range $[mN-(N+1), mN]$, and becomes strongly unstable when it retreats to the range $[0, N+1]$.
4. As a special case, when $m = 2$, a combination of recursions in both directions gives a locally strongly stable evaluation over the interested range $[0, mN]$.

**Proof:** When $0 < x \leq N+1$, we have

\[a + b \frac{j}{x} \geq 0, \quad j = 1, 2, \ldots \]

Therefore, the coefficients of the forward recursion (3) are all non-negative over the range $[0, N+1]$.

When $mN-(N+1) \leq y \leq Nm$, we have

\[P(y) > 0, \quad \text{and} \quad -\left(a + b \frac{m-j}{y+m}\right) \geq 0, \quad j = 1, \ldots, m-1.\]

Thus the coefficients of the backward recursion (81) are all non-negative over the range $[mN-(N+1), mN]$.

From earlier results, the theorem is proved.

One can see the connection between compound Binomial and compound Poisson distributions. When $N \to \infty$ and $\theta = \lambda/N \to 0$, the limiting distribution of compound Binomial is nothing but a compound Poisson distribution, which is strongly stable over the range $[0, \infty)$.

**Example 9:** Reconsider the compound Binomial distribution discussed in Example 8. We use both the forward recursion (3) and the backward recursion (81) to evaluate the compound Binomial distribution. This time, instead of using 200 digits, we use only 20 digits in the evaluation. The results are displayed in Table 7.
Table 7

Evaluate compound binomial recursively in both directions

<table>
<thead>
<tr>
<th>x</th>
<th>$g(x)$ (forward)</th>
<th>$g(x)$ (backward)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$7.888609052 \times 10^{-130}$</td>
<td>$1.025580868 \times 10^{11}$</td>
</tr>
<tr>
<td>1</td>
<td>$2.248253579 \times 10^{-129}$</td>
<td>$-0.885719952 \times 10^{10}$</td>
</tr>
<tr>
<td>...</td>
<td>$1.254727678 \times 10^{-32}$</td>
<td>$-0.570650733 \times 10^{-6}$</td>
</tr>
<tr>
<td>305</td>
<td>$2.472423462 \times 10^{-2}$</td>
<td>$0.2472423462 \times 10^{-2}$</td>
</tr>
<tr>
<td>306</td>
<td>$2.694072242 \times 10^{-2}$</td>
<td>$0.2694072242 \times 10^{-2}$</td>
</tr>
<tr>
<td>378</td>
<td>$0.8779196867 \times 10^{-2}$</td>
<td>$0.8779196867 \times 10^{-2}$</td>
</tr>
<tr>
<td>379</td>
<td>$0.8381164919 \times 10^{-2}$</td>
<td>$0.8381164919 \times 10^{-2}$</td>
</tr>
<tr>
<td>...</td>
<td>$0.2300721278 \times 10^{25}$</td>
<td>$0.109963604 \times 10^{-20}$</td>
</tr>
<tr>
<td>999</td>
<td>$-0.2066050091 \times 10^{11}$</td>
<td>$0.3684354379 \times 10^{-166}$</td>
</tr>
<tr>
<td>1000</td>
<td>$0.3516174897 \times 10^{11}$</td>
<td>$0.3684354379 \times 10^{-162}$</td>
</tr>
</tbody>
</table>

From Table 7, one can observe that the computed results by recursions in both directions meet each other over the middle range $[305, 379]$ in their first 10 non-zero digits. If we use the results of forward recursion for points before 379, and the results of backward recursion for points after 305, then we have confidence in that there are at least 10 significant digits in the combined results.

Note that in (79), $a = \frac{1}{1 - \theta}$, thus $a \to \infty$ as $\theta \to 1$.

In this numerical example, $\theta = 0.95$, which gives a large negative value $a = -19$ and thus causes rapid round-off error blow-ups. The effectiveness of both forward and backward recursions are compared in Table 8, for different values of $\theta$, and in Table 9, for different values of $N$.

Table 8

Evaluate compound binomial with severity $A$ and $N = 100$ in two directions
(20 digits are used to ensure 10 significance digits in the computed result)

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>forward range</th>
<th>forward mass</th>
<th>backward range</th>
<th>backward mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0 → 965</td>
<td>$1 - 6.7415 \times 10^{-136}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0 → 926</td>
<td>$1 - 9.0647 \times 10^{-123}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0 → 889</td>
<td>$1 - 2.4936 \times 10^{-117}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0 → 841</td>
<td>$1 - 7.4181 \times 10^{-110}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0 → 801</td>
<td>$1 - 4.9640 \times 10^{-10}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0 → 772</td>
<td>$1 - 7.8696 \times 10^{-9}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0 → 747</td>
<td>$1 - 3.6413 \times 10^{-7}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0 → 732</td>
<td>$1 - 3.9255 \times 10^{-6}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0 → 692</td>
<td>$1 - 4.4574 \times 10^{-4}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0 → 624</td>
<td>$1 - 3.3322 \times 10^{-3}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0 → 523</td>
<td>$1 - 4.4942 \times 10^{-3}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>0 → 379</td>
<td>$1 - 8.7773 \times 10^{-3}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td>0 → 214</td>
<td>$1 - 1.5973 \times 10^{-3}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ON THE STABILITY OF RECURSIVE FORMULAS

TABLE 9

Evaluate compound Binomial with Severity \( A \) and \( \theta = .5 \) in two directions
(20 digits are used to ensure 10 significance digits in the computed result)

<table>
<thead>
<tr>
<th>( N )</th>
<th>Forward range</th>
<th>Forward mass</th>
<th>Backward range</th>
<th>Backward mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0 \rightarrow 100</td>
<td>1</td>
<td>0 \rightarrow 100</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>0 \rightarrow 192</td>
<td>1 - .20096 \times 10^{-31}</td>
<td>0 \rightarrow 200</td>
<td>.99630</td>
</tr>
<tr>
<td>50</td>
<td>0 \rightarrow 422</td>
<td>1 - .14619 \times 10^{-31}</td>
<td>0 \rightarrow 500</td>
<td>.004359</td>
</tr>
<tr>
<td>100</td>
<td>0 \rightarrow 747</td>
<td>1 - .36413 \times 10^{-74}</td>
<td>0 \rightarrow 1000</td>
<td>.43928 \times 10^{-13}</td>
</tr>
<tr>
<td>200</td>
<td>0 \rightarrow 1337</td>
<td>1 - .81694 \times 10^{-111}</td>
<td>0 \rightarrow 500</td>
<td>.67602 \times 10^{-46}</td>
</tr>
<tr>
<td>500</td>
<td>0 \rightarrow 2996</td>
<td>1 - .52880 \times 10^{-207}</td>
<td>0 \rightarrow 10000</td>
<td>.31541 \times 10^{-165}</td>
</tr>
<tr>
<td>1000</td>
<td>0 \rightarrow 5763</td>
<td>1 - .87859 \times 10^{-372}</td>
<td>0 \rightarrow 10000</td>
<td>.87859 \times 10^{-372}</td>
</tr>
</tbody>
</table>

Remarks:

1. In terms of probability mass (not number of points) covered by the valid range in which the accuracy meets a specified level, the effectiveness of the forward direction increases when \( N \) increases, and increases when \( \theta \) decreases. This can be seen from Table 8 and Table 9, which is also consistent with the result in Theorem 13. However, the forward direction can be very unstable when \( \theta \) gets close to 1 or the claim distribution is highly negative skewed. In such cases, the backward recursion can play a major part in evaluating the compound distribution.

2. From Table 8, we can see that, when \( \theta \leq .5 \), the backward direction can give accurate results for more than one third of the points over the whole range; however, their total probability mass is very small. Thus, when \( \theta \leq .5 \), the actual usefulness of the backward direction can be used to check the accuracy of the forward direction.

3. In most insurance applications, \( \theta \leq .5 \) and \( N \) is large and the claim size distribution \( f(x) \) is positively skewed, if additional digits are used in the evaluation, one should not be bothered by seeing negative probabilities in the extreme far right tail, since almost all of the compound distribution except the very extreme right tail has been evaluated with desired accuracy. A check of accuracy can be done by a recursive evaluation in the backward direction. If two directions do not meet over the middle range, increasing the number of digits in the evaluation can make them so.

4. As mentioned by CHAN ([3], p. 175) and SHIU ([19], p. 181), the famous J. C. P. Miller formula has been used to evaluate the power of polynomials and the N-fold convolution of arithmetic distributions. Essentially, J. C. P. Miller formula is a variant of recursion (3) for the compound Binomial case. Assume that a discrete distribution \( f(x) \) is defined on integers \( \{x_0, x_1, \ldots, x_r\} \). With a transformation \( x' = x - N x_0 \), the N-fold convolution of \( f(x) \) is equivalent to a compound Binomial with \( \theta = 1 - f(x_0) \). In such situations, the Binomial parameter \( \theta \) can be very close to 1, or \( f(x) \)
itself can have a high negative skewness, which may cause difficulties when using the recursion in the forward direction. This instability can be easily handled by using two recursive evaluations in both directions.

11. REVIEW OF OTHER RECURSIONS

11.1. The generalized \((a, b)\) class

Sundt and Jewell [2] extended recursion (3) to a larger family of claim frequencies

\[
\frac{p_n}{p_{n-1}} = a + \frac{b}{n}, \quad n = r + 1, \ldots
\]

The compound distribution for this family of claim frequency satisfies:

\[
g(x) = \sum_{j=1}^{x} \left( a + b \frac{j}{x} \right) f(j) \, g(x-j) + \sum_{i=1}^{r} \left( p_i - \left( a + \frac{b}{i} \right) p_{i-1} \right) f^{*i}(x).
\]

Among the generalized claim frequencies (84), the class with \(r = 1\) is of special interest and is given a name \((a, b)\) class. In the \((a, b)\) class, \(p_0\) can be any value in the interval \([0, 1]\). The family of frequencies in (1) given by Panjer [12] is a subclass of the \((a, b)\) class with \(r = 0\) and is called the \((a, b, 0)\) subclass. The family of frequencies in the \((a, b)\) class with \(r = 1\) and \(p_0 = 0\), is called the \((a, b, 1)\) subclass.

As counterparts of the \((a, b, 0)\) subclass, truncated Poisson, truncated Negative Binomial, truncated Geometric and truncated Binomial are members in the \((a, b, 1)\) subclass. Another member in the \((a, b, 1)\) subclass is the logarithmic distribution. Sundt and Jewell [20] and Willmot [21] completed the enumeration of members in the \((a, b, 1)\) subclass by adding in the extended truncated Negative Binomial (ETNB) distribution.

For members in the \((a, b, 1)\) subclass, we can modify the probabilities at zero arbitrarily. We name the members of the \((a, b)\) class as: zero-modified Poisson, zero-modified Negative Binomial, zero-modified Geometric, zero-modified Binomial, zero-modified extended Negative Binomial, and log-zero distribution.

For claim frequencies in the \((a, b)\) class, the non-homogeneous recursion (85) becomes

\[
g(x) = \sum_{j=1}^{x} \left( a + b \frac{j}{x} \right) f(j) \, g(x-j) + (p_1 - (a + b) p_0) f(x).
\]

If we decompose the recursion (86) into the following form:

\[
g(x) = \sum_{j=1}^{x-1} \left( a + b \frac{j}{x} \right) f(j) \, g(x-j) + p_1 f(x) + (a + b) (g(0) - p_0) f(x),
\]

Sundt and Jewell [2] extended recursion (3) to a larger family of claim frequencies
and utilize the initial condition $g(0) = p_0$, recursion (86) can be reduced to:

\begin{equation}
  \sum_{j=1}^{x-1} \left( a + \frac{j}{x} \right) f(j) g(x-j) + p_1 f(x),
\end{equation}

with two initial values

\begin{equation}
  g(0) = p_0, \quad g(1) = p_1 f(1).
\end{equation}

From Theorem 7, one can easily see that the recursion (88) is strongly stable in evaluating compound zero-modified Poisson, compound zero-modified Negative Binomial, compound zero-modified Geometric and compound log-zero distributions.

The recursion (88) is unstable in evaluating compound zero-modified Binomial. The method developed for compound Binomial in the last section can be applied to this case without any difficulty.

For the compound zero-modified extended Negative Binomial distribution, we have

\begin{equation}
  0 < a < 1, \quad b = (r-1)a, \quad -1 < r < 0.
\end{equation}

For positive claim severities with a finite support \( \{x_1, \ldots, x_r = m\} \), we have

\begin{equation}
  a + \frac{j}{x} > 0, \quad j = 1, \ldots, m, \quad \text{for } x > (1 + |r|)m.
\end{equation}

Therefore, for compound zero-modified extended Negative Binomial distribution, the recursion (86) is stable. Also, once recursive evaluation has reached at a point \( k > (1 + |r|)m \), the recursive evaluation for future points are strongly stable.

In all the previous recursions for aggregate claims, it was assumed that claims were positive valued. For non-negative claim severities including zero claims, Panjer and Willmot [15] proposed a simple method, by which the spike at zero can be easily removed and the previous recursions for positive claims can be used.

11.2. Compound Poisson \((a, b)\) (CPAB) class

Willmot and Panjer [22] discussed various contagious counting distributions which involve a sequential usage recursion (3). For example, a compound Poisson Inverse Gaussian (P-IG) distribution can be evaluated by a two-stage usage of recursion (3): (i) a compound ETNB over the claim severity distribution; (ii) a compound Poisson with the compound ETNB distribution obtained in the first stage as its severity distribution.

If each recursive evaluation is stable, their combined usage is also stable. Islam and Consul ([10], p. 93) commented that the use of CPAB frequency model may cause serious numerical instabilities. Clearly their comment was wrong.
11.3. Improved recursions aren’t improved

When the claim frequency has a Poisson distribution, and the claim severity has a special pattern of piecewise constant or piecewise linear, DE PRIL [17] gave some simplified recursions in terms of numbers of calculations required. However, since both positive and negative signs evenly appeared in the coefficients of the recursions, they are unstable and thus not really improved.

11.4. Probability of ultimate ruin

PANJER [13] proposed a method of direct evaluation of the probability of ruin. Since the desired probability is a compound Geometric distribution, the recursive evaluation is strongly stable.

GOOVAERTS and DE VYLDER [9] proposed a different approach to approximate the probability of ruin. The upper bounds are evaluated by a recurrence equation:

\[ \hat{\Psi}_u(x) = \frac{1}{1 + \theta} \left\{ K(x) - \sum_{i=1}^{x} \Delta K((i-1)h) \hat{\Psi}_u((x-i)h) \right\}, \quad x = 1, 2, \ldots \]  

The lower bounds are evaluated by a recurrence equation:

\[ \hat{\Psi}_l(x) = \frac{1}{1 + \theta + \Delta K(0)} \left\{ K(x) - \sum_{i=1}^{x-1} \Delta K(ih) \hat{\Psi}_u((x-i)h) \right\}, \quad x = 1, 2, \ldots \]

Since

\[ K(s) = \int_{s}^{\infty} \frac{1 - F(y)}{p_1} \, dy, \quad s \geq 0, \]

we have

\[ -\Delta K(ih) = \int_{ih}^{(i+1)h} \frac{1 - F(y)}{p_1} \, dy > 0. \]

Therefore, the recursions (92) and (93) are indeed strongly stable in evaluating the desired ruin probability.

RAMSAY [18] recently commented that his numerical result did not agree with that of GOOVAERTS and DE VYLDER [9] and was unable to explain the difference ([18], p. 58). Now it becomes clear that, the instability that RAMSAY [18] discussed about was not from inherent rounding error accumulations by using recursions (92) and (93), but from the unstable evaluation of the coefficients \( \Delta K(ih) = K((i+1)h) - K(ih) \) by subtracting two nearly equal numbers. Also, the inaccuracy in the numerical results of GOOVAERTS and DE VYLDER [9] can be explained by the slow convergence (as proved by Ramsay) of the approximation scheme of Goovaerts and De Vylder, and not because of the instability of the recursions.
11.5. Probability of finite time ruin

In their paper [5], Dickson and Waters suggested a method of recursive evaluation of finite time ruin probabilities. DICKSON and WATERS [5] (p. 211) commented that they experienced some numerical instabilities when using a combination of two recursions. One (see (4.2) of DICKSON and WATERS [5], p. 208) is now known as strongly stable; the other recursion (see (3.2) of DICKSON and WATERS [5], p. 206) involves many differencing terms. It can be verified that, (3.2) of DICKSON and WATERS is unstable in evaluating the desired probabilities.

The basic ideas and results in this paper can be extended and applied to other recursions (not necessarily in actuarial field). For unstable recursions, alternative methods of evaluation merit further research.

REFERENCES


Harry H. Panjer and Shaun Wang

Faculty of Mathematics, Institute of Insurance & Pension Research, University of Waterloo, Waterloo Ontario, N2L 3G1, Canada.
In this paper we derive formulae for finite time survival probabilities when the aggregate claims process is a Gamma process. We illustrate how a compound Poisson process can be approximated by a Gamma process and by a process defined as a translated Gamma process. We also show how survival probabilities for a compound Poisson process can be approximated by those for a Gamma process or a translated Gamma process.

**Keywords**

Gamma process; finite time; survival probability.

1. **Introduction**

The Gamma process was introduced into the actuarial literature by Dufresne, Gerber and Shiu [1991]. They defined a Gamma process as a limit of compound Poisson processes and proposed it as a model for the aggregate claims process. They discussed many of the properties of this process and, in particular, showed how to calculate the probability of ultimate ruin for such a process. This paper takes this study a stage further and is concerned mainly with the probability of ruin/survival in finite time.

In the next section we derive (very simple) formulae for the probability of survival in finite time for a Gamma process. In §3 we show how to approximate a compound Poisson process by a Gamma process and investigate numerically how well the probability of survival in finite time for the former process is approximated by that for the latter process. The advantage in approximating a compound Poisson process in this way is that, as we show in §2, the probability of survival in finite time for a Gamma process is particularly easy to calculate. In §4 we introduce what we term a translated Gamma process and we carry out an investigation similar to that in §3.

2. **Finite time survival for the Gamma process**

The main result in this section gives formulae for the probability of survival in finite time for a Gamma process. These formulae are derived from standard formulae for a compound Poisson process. Before deriving these results we show how to define/construct a (standardised) Gamma process as a limit of
compound Poisson processes. This construction follows precisely the method outlined by Dufresne et al. [1991, §§2 and 3]. However, we provide more details of the construction since some of these details are important for the proof of the main result.

For $x > 0$ define the function $Q(x)$ as follows:

$$Q(x) = \int_{x}^{\infty} \frac{\exp\{-y\}}{y} \, dy$$

(This is precisely the same as the function $Q(x)$ in Dufresne et al. [1991, § 3] with their parameters $a$ and $b$ both taken to be 1.)

Now let $\{S(t; 1)\}_{t>0}$ be a compound Poisson process such that:

(a) the number of claims occurring in $(0, t]$ has a Poisson distribution with mean $tQ(1)$, and,
(b) individual claims amounts have a distribution function $P(y; 1)$, where:

$$P(y; 1) = 0 \quad \text{for } y < 1$$

$$= \frac{Q(1) - Q(y)}{Q(1)} \quad \text{for } y \geq 1.$$  

For $n = 2, 3, \ldots$, define the process $\{S(t; n)\}_{t>0}$ as follows:

$$(2.1) \quad S(t; n) = S(t; n-1) + X_n(t)$$

where $\{X_n(t)\}_{t>0}$ is a compound Poisson process, independent of $\{S(t; n-1)\}_{t>0}$, such that:

(c) the number of claims occurring in $(0, t]$ has a Poisson distribution with mean $tQ(1/n) - Q(1/(n-1))$, and,
(d) individual claim amounts have a distribution function $\Pi(y; n)$, where:

$$\Pi(y; n) = 0 \quad \text{for } y < 1/n$$

$$= \frac{Q(1/n) - Q(y)}{Q(1/n) - Q(1/(n-1))} \quad \text{for } 1/n \leq y < 1/(n-1)$$

$$= 1 \quad \text{for } y \geq 1/(n-1)$$

Then it is easy to show that for $n = 1, 2, 3, \ldots$ $\{S(t; n)\}_{t>0}$ is a compound Poisson process such that:

(e) the number of claims occurring in $(0, t]$ has a Poisson distribution with mean $tQ(1/n)$, and,
(f) individual claims amounts have a distribution function $P(y; n)$, where:

$$P(y; n) = 0 \quad \text{for } y < 1/n$$

$$= \frac{Q(1/n) - Q(y)}{Q(1/n)} \quad \text{for } y \geq 1/n$$

(Note that $S(t; n)$ and $P(y; n)$ in this paper were denoted $S(t; x)$ and $P(y; x)$, respectively, with $x = 1/n$ by Dufresne et al. [1991].)
The standardised Gamma process is the process \( \{S_{SG}(t)\}_{t > 0} \) defined by:

\[
S_{SG}(t) = \lim_{n \to \infty} S(t; n) \quad \text{for } t > 0
\]

It is important to note that this limit exists surely, rather than just almost surely, for each \( t > 0 \), since, from (2.1), \( S(t; n) \) is monotonic non-decreasing as \( n \to \infty \).

Since \( S(t; n) \) converges surely to \( S_{SG}(t) \), it also converges in distribution. However, Dufresne et al. [1991, § 3] show that \( S(t; n) \) converges in distribution to a random variable with a Gamma \((t, 1)\) distribution. Hence, \( S_{SG}(t) \) has a Gamma \((t, 1)\) distribution. Finally in this construction/definition, for any \( \alpha > 0 \) and \( \beta > 0 \), define a new stochastic process \( \{S_G(t)\}_{t > 0} \) as follows:

\[
S_G(t) = \beta^{-1} S_{SG}(\alpha t)
\]

We will refer to \( \{S_G(t)\}_{t > 0} \) as a Gamma \((\alpha, \beta)\) process, so that the standardised Gamma process, \( \{S_{SG}(t)\}_{t > 0} \), is a Gamma \((1, 1)\) process. Note that the random variable \( S_G(t) \) has a Gamma \((\alpha t, \beta)\) distribution. (Note also that we are parameterising the Gamma distribution so that \( S_G(t) \) has mean \( \alpha t/\beta \).)

The remainder of this section will be concerned with the standardised Gamma process. We are regarding this as a model for the aggregate claims process for a risk, so that \( S_{SG}(t) \) represents the aggregate claims generated by this risk in the period \((0, t]\). We assume that premium income is received continuously at constant rate \( c \) per unit time for this risk. We assume that

\[
c > E[S_{SG}(1)] = 1
\]

We denote by \( \delta_{SG}(U, t) \) the probability of survival, i.e. non-ruin, up to time \( t \) for this process given initial surplus \( U \geq 0 \), so that:

\[
\delta_{SG}(U, t) = \Pr(S_{SG}(\tau) \leq U + ct \quad \text{for all} \quad \tau, 0 < \tau \leq t)
\]

The main result of this section is the following:

Result:

\[
(2.3) \quad \delta_{SG}(0, t) = F_{SG}(ct, t) - \frac{1}{c} F_{SG}(ct, t+1)
\]

\[
(2.4) \quad \delta_{SG}(U, t) = F_{SG}(U+ct, t) - c \int_0^t f_{SG}(U+cs, s) F_{SG}(c(t-s), t-s) ds + \sum_{n=1}^\infty c \int_0^t f_{SG}(U+cs, s) F_{SG}(c(t-s), t-s+1) ds
\]
where \( f_{SG}(x, t) \) and \( F_{SG}(x, t) \) are the density function and distribution function, respectively, of a Gamma \((t, 1)\) random variable.

**Proof:** For \( n = 1, 2, 3, \ldots \) and \( U \geq 0 \) define:

\[
\delta(U, t; n) = \Pr(S(\tau; n) \leq U + ct \text{ for all } \tau, 0 < \tau \leq t)
\]

The first step in the proof of this result is to show that:

\[
(2.5) \quad \lim_{n \to \infty} \delta(U, t; n) = \delta(U, t)
\]

To see this, note from the construction of the processes \( \{\{S(t; n)\}_{t>0}\} \) and from (2.2) that for any sample path \( \omega \) and any \( \tau \in (0, t] \):

\[
\ldots S(\tau; n-1)(\omega) \leq S(\tau; n)(\omega) \leq \ldots \leq S_{SG}(\tau)(\omega)
\]

and:

\[
(2.6) \quad \lim_{n \to \infty} S(\tau; n)(\omega) = S_{SG}(\tau)(\omega)
\]

Hence:

\[
\lim_{n \to \infty} \delta(U, t; n) \geq \delta(U, t)
\]

and the limit on the left does exist. If this limit is strictly greater than \( \delta(U, t) \), then there must exist some sample path \( \omega \) and some \( \tau \in (0, t] \) such that for all \( n \):

\[
S(\tau; n)(\omega) \leq U + ct < S_{SG}(\tau)(\omega)
\]

which contradicts (2.6). This proves (2.5).

Since \( \{\{S(t; n)\}_{t>0}\} \) is a compound Poisson process, we have:

\[
(2.7) \quad \delta(0, t; n) = \frac{1}{ct} \int_0^{ct} F(y, t; n) \, dy
\]

where \( F(y, t; n) \) is the distribution function of \( S(t; n) \). See SEAL [1978b, Ch. 4]. The convergence in distribution of the processes \( \{\{S(t; n)\}_{t>0}\} \) and the fact that \( F_{SG}(y, t) \) is everywhere continuous show that for all \( y \) and all \( t \):

\[
F_{SG}(y, t) = \lim_{n \to \infty} F(y, t; n)
\]

Applications of this result, of (2.5) and of the Bounded Convergence Theorem to (2.7) show that:

\[
(2.8) \quad \delta_{SG}(0, t) = \frac{1}{ct} \int_0^{ct} F_{SG}(y, t) \, dy
\]
Since $S_{SG}(t)$ has a Gamma $(t, 1)$ distribution, we can rewrite (2.8) as follows:

\[
\delta_{SG}(0, t) = \frac{1}{ct} \int_0^{ct} \int_0^x y^{t-1} e^{-y} \frac{dy}{\Gamma(t)}
\]

\[
= \frac{1}{ct} \int_0^{ct} \int_y^{ct} y^{t-1} e^{-y} \frac{dy}{\Gamma(t)}
\]

\[
= \int_0^{ct} y^{t-1} e^{-y} \frac{dy}{\Gamma(t)} - \frac{\Gamma(t+1)}{ct \Gamma(t)} \int_0^{ct} y^{t} e^{-y} \frac{dy}{\Gamma(t+1)}
\]

\[
= F_{SG}(ct, t) - \frac{1}{c} F_{SG}(ct, t+1)
\]

This proves (2.3). Formula (2.4) can be derived as follows. Using the familiar general reasoning argument (see Seal [1974, p. 126]), we have:

\[
\delta_{SG}(U, t) = F_{SG}(U+ct, t) - c \int_0^{t} \delta_{SG}(0, t-s) f_{SG}(U+cs, s) \, ds
\]

Formula (2.4) is obtained by substituting (2.3) into this last expression.

Table 1 shows values of $\delta_{SG}(U, t)$ for the standardised Gamma process for various combinations of $U$ and $t$. The premium has been taken to be 1.1 per unit time, so that the premium loading factor is 10%.

**TABLE 1**

VALUES OF $\delta_{SG}(U, t)$ WHEN $c = 1.1$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$U = 0$</th>
<th>$U = 1$</th>
<th>$U = 5$</th>
<th>$U = 10$</th>
<th>$U = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.39352</td>
<td>0.82211</td>
<td>0.99706</td>
<td>0.99998</td>
<td>1.00000</td>
</tr>
<tr>
<td>2</td>
<td>0.30244</td>
<td>0.71818</td>
<td>0.99108</td>
<td>0.99991</td>
<td>1.00000</td>
</tr>
<tr>
<td>3</td>
<td>0.25906</td>
<td>0.65070</td>
<td>0.98304</td>
<td>0.99975</td>
<td>1.00000</td>
</tr>
<tr>
<td>4</td>
<td>0.23254</td>
<td>0.60271</td>
<td>0.97378</td>
<td>0.99947</td>
<td>1.00000</td>
</tr>
<tr>
<td>5</td>
<td>0.21423</td>
<td>0.56647</td>
<td>0.96393</td>
<td>0.99907</td>
<td>1.00000</td>
</tr>
<tr>
<td>6</td>
<td>0.20064</td>
<td>0.53790</td>
<td>0.95386</td>
<td>0.99852</td>
<td>1.00000</td>
</tr>
<tr>
<td>7</td>
<td>0.19005</td>
<td>0.51467</td>
<td>0.94384</td>
<td>0.99784</td>
<td>1.00000</td>
</tr>
<tr>
<td>8</td>
<td>0.18151</td>
<td>0.49532</td>
<td>0.93402</td>
<td>0.99703</td>
<td>1.00000</td>
</tr>
<tr>
<td>9</td>
<td>0.17445</td>
<td>0.47890</td>
<td>0.92449</td>
<td>0.99608</td>
<td>1.00000</td>
</tr>
<tr>
<td>10</td>
<td>0.16848</td>
<td>0.46475</td>
<td>0.91531</td>
<td>0.99503</td>
<td>1.00000</td>
</tr>
<tr>
<td>20</td>
<td>0.13662</td>
<td>0.38506</td>
<td>0.84296</td>
<td>0.98058</td>
<td>0.99999</td>
</tr>
<tr>
<td>30</td>
<td>0.12306</td>
<td>0.34924</td>
<td>0.79653</td>
<td>0.96428</td>
<td>0.99996</td>
</tr>
<tr>
<td>40</td>
<td>0.11534</td>
<td>0.32837</td>
<td>0.76483</td>
<td>0.94935</td>
<td>0.99987</td>
</tr>
<tr>
<td>50</td>
<td>0.11030</td>
<td>0.31460</td>
<td>0.74194</td>
<td>0.93644</td>
<td>0.99971</td>
</tr>
<tr>
<td>100</td>
<td>0.09919</td>
<td>0.28380</td>
<td>0.68487</td>
<td>0.89569</td>
<td>0.99801</td>
</tr>
<tr>
<td>200</td>
<td>0.09351</td>
<td>0.26783</td>
<td>0.65187</td>
<td>0.86585</td>
<td>0.99409</td>
</tr>
<tr>
<td>300</td>
<td>0.09195</td>
<td>0.26345</td>
<td>0.64234</td>
<td>0.85617</td>
<td>0.99176</td>
</tr>
<tr>
<td>400</td>
<td>0.09138</td>
<td>0.26182</td>
<td>0.63782</td>
<td>0.85233</td>
<td>0.99055</td>
</tr>
<tr>
<td>500</td>
<td>0.09113</td>
<td>0.26112</td>
<td>0.63717</td>
<td>0.85063</td>
<td>0.98993</td>
</tr>
<tr>
<td>600</td>
<td>0.09102</td>
<td>0.26080</td>
<td>0.63645</td>
<td>0.84983</td>
<td>0.98961</td>
</tr>
<tr>
<td>700</td>
<td>0.09097</td>
<td>0.26065</td>
<td>0.63610</td>
<td>0.84944</td>
<td>0.98945</td>
</tr>
<tr>
<td>800</td>
<td>0.09094</td>
<td>0.26057</td>
<td>0.63592</td>
<td>0.84924</td>
<td>0.98936</td>
</tr>
<tr>
<td>900</td>
<td>0.09093</td>
<td>0.26053</td>
<td>0.63582</td>
<td>0.84913</td>
<td>0.98931</td>
</tr>
<tr>
<td>1000</td>
<td>0.09092</td>
<td>0.26051</td>
<td>0.63578</td>
<td>0.84908</td>
<td>0.98928</td>
</tr>
</tbody>
</table>
3. THE GAMMA PROCESS APPROXIMATION TO A COMPOUND POISSON PROCESS

In this section we demonstrate how finite time survival probabilities for a compound Poisson process can be approximated by those for a Gamma process.

Let \( \{S(t)\}_{t>0} \) be a compound Poisson process with Poisson parameter \( \lambda \). Let \( P(x) \) denote the individual claim amount distribution and let \( p_k \) denote the \( k \)-th moment about zero of this distribution. We shall approximate this compound Poisson process by a Gamma \((\alpha, \beta)\) process \( \{S_\alpha(t)\}_{t>0} \). We find the parameters \( \alpha \) and \( \beta \) of the Gamma process by matching the first two moments of the two processes (assuming that these moments exist). For each value of \( t>0 \) we set

\[
E[S(t)] = \lambda t p_1 = \alpha t / \beta = E[S_\alpha(t)]
\]

\[
V[S(t)] = \lambda t p_2 = \alpha t / \beta^2 = V[S_\alpha(t)]
\]

which gives

\[
\beta = p_1 / p_2 \quad \text{and} \quad \alpha = \lambda p_1^2 / p_2
\]

Note that the parameters \( \alpha \) and \( \beta \) are independent of \( t \).

The surplus process associated with the compound Poisson process is \( \{U(t)\}_{t>0} \) where \( U(t) = U + ct - S(t) \). \( U \) is the initial surplus and \( c \) is the premium income per unit time. In our numerical examples we write \( c = (1+\theta)\lambda p_1 \) where \( \theta \) is the premium loading factor. The finite time survival probability for this process is \( \delta(U, t) \) defined by

\[
\delta(U, t) = \Pr(S(t) < U + ct \quad \text{for all} \quad 0 < t \leq t)
\]

We will approximate this probability by \( \delta_\alpha(U, t) \) defined by

\[
\delta_\alpha(U, t) = \Pr(S_\alpha(t) < U + ct \quad \text{for all} \quad 0 < t \leq t)
\]

where the parameters \( \alpha \) and \( \beta \) of \( \{S_\alpha(t)\}_{t>0} \) are given by (3.1). We calculate \( \delta_\alpha(U, t) \) using the standardised Gamma process and the identity

\[
\delta_\alpha(U, t) = \Pr(S_\alpha(t) < U + (1+\theta)\alpha t / \beta \quad \text{for all} \quad 0 < t \leq t)
\]

\[
= \Pr(S_\alpha(t) < U + (1+\theta)\alpha t / \beta \quad \text{for all} \quad 0 < t \leq t)
\]

\[
= \Delta S_\alpha(0) \quad \text{for all} \quad 0 < t \leq t)
\]

In our numerical examples we set \( \lambda = 1 \) and use two distributions for individual claims amounts:

**Distribution 1**: \( P(x) \) is the Gamma \((1/3, 1/3)\) distribution. For this distribution, \( p_1 = 1 \) and \( p_2 = 4 \), so that \( \alpha = \beta = 0.25 \) in the approximating Gamma process. Thus we approximate \( \delta(U, t) \) by \( \delta_\alpha(U/4, t/4) \).

**Distribution 2**: \( P(x) \) is the Pareto \((3, 2)\) distribution. For this distribution, \( p_1 = 1 \) and \( p_2 = 4 \), so that \( \alpha = \beta = 0.25 \) in the approximating Gamma process.

The parameters of these distributions have been chosen so that the same Gamma process approximates each compound Poisson process. Although the
first two moments are the same for each of Distributions 1 and 2, \( p_3 \) does not exist for Distribution 2, but does exist for Distribution 1. Hence we would expect that the Gamma process approximation would be better for the compound Poisson process with Distribution 1 rather than Distribution 2.

In Figures 1 to 5 the crosses denote the ratios \( \delta(U, t) : \delta_{SG}(\beta U, \alpha t) \) (shown as percentages) for selected values of \( U \) and \( t \) for the surplus process associated with Distribution 1 when \( \theta = 0.1 \). A logarithmic scale for time has been used in these graphs. The values of \( \delta(U, t) \) have been calculated (approximately) using the methods described by Dickson and Waters [1991]. We note that for each of these values of \( U \) the approximate values are fairly close to the true values. They are within 4% of the true values for all values of \( U \) shown, and within 1% for virtually all combinations of \( U \) and \( t \), except when \( U = 4 \). We conclude that the approximation to \( \delta(U, t) \) is reasonable in this case.

**Figure 1.** Ratios of exact to approximate values of \( \delta(0, t) \) for Gamma individual claims.

**Figure 2.** Ratios of exact to approximate values of \( \delta(4, t) \) for Gamma individual claims.
Figure 3. Ratios of exact to approximate values of $\delta(20, t)$ for Gamma individual claims.

Figure 4. Ratios of exact to approximate values of $\delta(40, t)$ for Gamma individual claims.

Figure 5. Ratios of exact to approximate values of $\delta(100, t)$ for Gamma individual claims.
The crosses in Figures 6 to 10 denote the corresponding ratios for the same combinations of $U$ and $t$ for the surplus process associated with Distribution 2, again with $\theta = 0.1$ and with a logarithmic scale for time. The approximations are generally worse for this distribution, although the approximate values are within 1% of the true values when $U = 40$ and when $U = 100$.

We have used the same Gamma process to approximate to two different compound Poisson processes. It is therefore no surprise that the approximation is better for one of the compound Poisson processes. In the following section we present a method which allows us to approximate survival probabilities for compound Poisson processes with identical first two moments by different Gamma processes.

Figure 6. Ratios of exact to approximate values of $\delta(0, t)$ for Pareto individual claims.

Figure 7. Ratios of exact to approximate values of $\delta(4, t)$ for Pareto individual claims.
FIGURE 8. Ratios of exact to approximate values of $\delta(20, t)$ for Pareto individual claims.

FIGURE 9. Ratios of exact to approximate values of $\delta(40, t)$ for Pareto individual claims.

FIGURE 10. Ratios of exact to approximate values of $\delta(100, t)$ for Pareto individual claims.
4. The translated Gamma process approximation to a compound Poisson process

As in the previous section, \( \{S(t)\}_{t>0} \) is a compound Poisson process with Poisson parameter \( \lambda \); \( P_k \) denotes the \( k \)-th moment about zero of the individual claim amount distribution. We shall approximate \( \{S(t)\}_{t>0} \) by what we term a translated Gamma process \( \{S_{\theta \lambda}(t)\}_{t>0} \).

For all \( t > 0 \) we define

\[
S_{\theta \lambda}(t) = S_{\lambda}(t) + \theta \lambda t
\]

where \( \{S_{\lambda}(t)\}_{t>0} \) is a Gamma \((\theta \lambda, \lambda)\) process and \( \theta \) is some constant (which may be positive or negative). The parameters \( \alpha, \beta \) and \( \theta \) of the process \( \{S_{\theta \lambda}(t)\}_{t>0} \) are chosen such that for all \( t > 0 \)

\[
E[S(t)] = \theta \lambda tp_1 = \alpha \beta + \theta \lambda t = E[S_{\theta \lambda}(t)]
\]

\[
V[S(t)] = \lambda tp_2 = \alpha \beta^2 = V[S_{\theta \lambda}(t)]
\]

\[
Sk[S(t)] = \lambda tp_3/(\lambda tp_2)^{3/2} = 2/(\alpha t)^{1/2} = Sk[S_{\theta \lambda}(t)]
\]

i.e. we are matching the mean, variance and coefficient of skewness of \( S(t) \) and \( S_{\theta \lambda}(t) \) for all \( t \) (again assuming that these quantities exist). These identities give the parameter values as

\[
\alpha = 4 \lambda p_3/p_3^2 \quad \beta = 2 p_2/p_3 \quad \theta = (p_1 - 2 p_2^2/p_3)
\]

As in the previous approximation, the parameters are all independent of \( t \). We now approximate

\[
\delta(U, t) = \Pr (S(t) \leq U + \theta \lambda t) \quad \text{for all} \quad \tau, 0 < \tau \leq t
\]

by

\[
\delta_{\theta \lambda}(U, t) = \Pr (S_{\theta \lambda}(t) \leq U + \theta \lambda t) \quad \text{for all} \quad \tau, 0 < \tau \leq t
\]

\[
= \Pr (S_{\lambda}(t) + \theta \lambda t \leq U + \theta \lambda t) \quad \text{for all} \quad \tau, 0 < \tau \leq t
\]

\[
= \Pr (S_{\lambda}(t) \leq U + (\theta - \theta \lambda) \tau) \quad \text{for all} \quad \tau, 0 < \tau \leq t
\]

Note that

\[
c - \theta \lambda t = (1 + \theta) \lambda p_1 - \lambda (p_1 - 2 p_2^2/p_3)
\]

\[
= \theta \lambda p_1 + 2 \lambda p_2^2/p_3 = \theta \lambda p_1 + \alpha/\beta
\]

Thus \( c - \theta \lambda t > E[S_{\lambda}(1)] \) (where \( S_{\lambda}(1) \) has a Gamma \((\alpha, \beta)\) distribution) regardless of the values of \( k \). Since \( \lambda p_1 = \alpha/\beta + k \), \( \delta_{\theta \lambda}(U, t) \) represents the finite time survival probability when the aggregate claims process is a Gamma \((\alpha, \beta)\) process and when the premium loading factor is \( \hat{\theta} = \theta (1 + k \beta / \alpha) \). Hence we can approximate \( \delta(U, t) \) by \( \delta_{\theta \lambda}(U, t) \) using a premium loading factor of \( \hat{\theta} \).

We illustrate the approximation method for two distributions for individual claim amounts:

**Distribution 1:** \( P(x) \) is the Gamma \((1/3, 1/3)\) distribution. The first three moments of this distribution are \( p_1 = 1 \), \( p_2 = 4 \) and \( p_3 = 28 \). Let \( \lambda = 1 \) and let \( \theta = 0.1 \), giving \( \alpha = 16/49 \), \( \beta = 2/7 \) and \( k = -1/7 \). Then \( \alpha/\beta = 8/7 \) and \( c - k = 87/70 \), so that \( \hat{\theta} = 7/80 \). We approximate \( \delta(U, t) \) by \( \delta_{\theta \lambda}(U, t) \) using a premium loading factor of \( 7/80 \).
In Figures 1 to 5 the circles denote the ratios $\delta(U, t) : \delta_{SG}(\beta U, \alpha t)$ for the same values of $U$ and $t$ as before for the surplus process associated with Distribution 1. Again a logarithmic scale has been used for time. The approximate values are very close to the exact ones except when $U = 0$ (within 0.3% when $U = 4$, and within 0.06% when $U = 20, 40$ and 100). It is no surprise that the translated Gamma process gives superior approximations to those in the previous section when $U > 0$ since the approximating process matches one further feature of the given process. The reason why the approximation is poor when $U = 0$ is discussed later in this section.

**Distribution 3:** $P(x)$ is the Pareto $(4, 3)$ distribution. (Note that the third moment of Distribution 2 does not exist and so we cannot use the translated Gamma approximation in this case.) For this distribution $p_1 = 1$, $p_2 = 3$ and $p_3 = 27$. Again let $\lambda = 1$ and let $\theta = 0.1$ giving $\alpha = 4/27$, $\beta = 2/9$ and $k = 1/3$. Then $\alpha/\beta = 2/3$ and $c-k = 23/30$, so that $\theta = 0.15$. Hence we approximate $\delta(U, t)$ by $\delta_{SG}(2U/9, 4t/27)$ using a premium loading factor of 0.15. In Figures 6 to 10 the circles denote the ratios $\delta(U, t) : \delta_{SG}(\beta U, \alpha t)$ for the same values of $U$ and $t$ as before for the surplus process associated with Distribution 3. The approximate values are fairly close to the exact ones when $U > 0$ (within 6% when $U = 4$, and within about 1% when $U = 20, 40$ and 100) but the approximations are not as close as for Distribution 1. Note that the ratios are not directly comparable with those for Distribution 2 as the parameter values for the distributions are different. Nevertheless, when $U > 0$ this approximation represents an improvement over the method described in the preceding section.

Since the first three moments of the approximating translated gamma process match those of the compound Poisson process for all values of $t$, we would expect $\delta_{RG}(U, t)$ to be a reasonable approximation to $\delta(U, t)$. This is indeed the case when $U > 0$. However, when $U = 0$, the approximations are poor and are much worse than those in the previous section. We can see why this is so by writing down formulae for the exact and approximate survival probabilities. The formula for the exact survival probability is

$$\delta(0, t) = \frac{1}{ct} \int_0^{ct} F(x, t) \, dx$$

where $F(x, t) = \Pr(S(t) \leq x)$, and $S(t)$ has a compound Poisson distribution. The approximation in Section 3 is

$$\delta_G(U, t) = \frac{1}{ct} \int_0^{ct} F_G(x, t) \, dx$$

where $F_G(x, t) = \Pr(S_G(t) \leq x)$, and $S_G(t)$ has a Gamma distribution whose first two moments match those of the compound Poisson distribution. It is not surprising that (4.2) is a good approximation to (4.1). Not only are the formulae for $\delta(0, t)$ and $\delta_G(0, t)$ of the same form, but $F_G(x, t)$ may be
regarded as an intuitively reasonable approximation to $F(x, t)$ since these two distributions have the same first two moments.

In this section $\delta(0, t)$ is approximated by

$$
\delta_{TG}(0, t) = \frac{1}{(c-k)t} \int_0^{(c-k)t} F_G(x, t) \, dx
$$

Comparing (4.1) and (4.3), we see that there are two differences: the factor $c$ in (4.1) is replaced by $c-k$ in (4.3), and $F_G(x, t)$ does not have the same first two moments as $F(x, t)$. It is not surprising that (4.3) is not as good an approximation to (4.1) as is (4.2). In particular, in the limiting case when $t \to \infty$ we can measure the difference between (4.1) and (4.3) since

$$
\lim_{t \to \infty} \delta(0, t) = 1/(1+\theta) \quad \text{and} \quad \lim_{t \to \infty} \delta_{TG}(0, t) = 1/(1+\hat{\theta})
$$

(See Dufresne et al. [1991].)

5. CONCLUDING REMARKS

The computer time required to produce approximate values for $\delta(U, t)$ using the methods of the previous two sections is substantially less than that required for the algorithms described by Dickson and Waters [1991]. Although these algorithms produce very accurate values, the amount of computer time required for large values of $U$ and $t$ can be considerable. The examples in Section 4 show that the approximations to $\delta(U, t)$ are good for large values of $U$ and $t$. We conclude that the approximation method of the previous section can be used to produce fast and fairly reliable estimates of $\delta(U, t)$ for such combinations of $U$ and $t$.

The underlying idea in Sections 3 and 4 has been to approximate a compound Poisson process $\{S(t)\}_{t>0}$ by a Gamma process $\{S_G(t)\}_{t>0}$ and by a translated Gamma process $\{S_{TG}(t)\}_{t>0}$, respectively. In each case the approximation is the result of matching an appropriate number of moments. The probability of survival for the compound Poisson process, $\delta(U, t)$, is then approximated by the corresponding probability for the Gamma process or the translated Gamma process. However, there is an alternative and related way of approximating this probability. We have the following formulae:

$$
\delta(0, t) = \frac{1}{ct} \int_0^{ct} F(x, t) \, dx
$$

$$
\delta(U, t) = F(U+ct, t) - c \int_0^t \delta(0, t-s) f(U+cs, s) \, ds
$$

(Formula (5.1) is, of course, the same as formula (4.1).)

We can now approximate $\delta(0, t)$ and $\delta(U, t)$ by approximating $F(\cdot, t)$ and $f(\cdot, s)$ by Gamma distributions or translated Gamma distributions with the
same first two or three moments as the original distributions. If we approximate $F(\cdot, t)$ and $f(\cdot, s)$ by Gamma distributions, it can be checked that the result is the same as that achieved by approximating the original compound Poisson process by a Gamma process, as in Section 3. However, replacing $F(\cdot, t)$ and $f(\cdot, s)$ in (5.1) and (5.2) by translated Gamma distributions with the same first three moments, an idea originally discussed by Seal [1978a], is not the same as approximating the original process by a translated Gamma process as described in Section 4. To see this, note that Seal’s method leads to the following approximation for $\delta(0, t)$:

$$\delta(0, t) \approx \frac{1}{ct} \int_0^{ct} F_{\gamma}(x, t) \, dx$$

where $F_{\gamma}(x, t)$ is a translated Gamma distribution with the same first three moments as $F(x, t)$. Formula (5.3) is clearly not the same as formula (4.3). An advantage of Seal’s method as compared to the method of Section 4 is that (5.3) is a better approximation to (4.1) than is (4.3). A disadvantage of Seal’s method is that it leads to a slightly more complicated (approximate) formula for $\delta(0, t)$ and hence for $\delta(U, t)$.

**REFERENCES**


Seal, H.L. (1974) The numerical calculation of $U(w, t)$, the probability of non-ruin in an interval $(0, t)$. *Scandinavian Actuarial Journal*, 121–139.


---

**DAVID C.M. DICKSON**  
*Centre for Actuarial Studies, University of Melbourne, Parkville, Victoria 3052, Australia*

**HOWARD R. WATERS**  
*Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, Scotland*
WORKSHOP

WEIGHTED MORTALITY RATES AS EARLY WARNING SIGNALS FOR INSURANCE COMPANIES

BY LEIGH A. ROBERTS

Institute of Statistics and Operations Research, Victoria University, Wellington, New Zealand

ABSTRACT

Weighted mortality rates are commonly used in actuarial work, but the inter-relationship between the weights used and the underlying mortality rates seems not to have been widely investigated.

Calculation of the ratio of weighted mortality rates to conventional mortality rates provides a simple means for an insurance company to track changes in the underlying mortality of its portfolio over time, and acts as an early warning system for possible deterioration of underwriting results. Asymptotic distributions are found for this ratio, and for the mortality rates themselves. It is suggested that insurance companies commence to gather data for the calculation of this ratio for the insurance sector as a whole, for the main annuity and assurance classes.

KEYWORDS

Asymptotic distribution; decremental rates; early warning signals; insurance; ratio of weighted to unweighted mortality rates; sums at risk; underlying mortality trends; weighted mortality rates.

1. INTRODUCTION

Conventional mortality rates are calculated as a ratio of the number of deaths to the number of life years lived which gave rise to those deaths, and measure the probability of dying within the next year at a given age. A variant on this theme in common use in actuarial work is the use of weighted mortality rates, attaching weights to each death and unit of exposure to risk: in life insurance these weights are typically sums assured or numbers of policies. For monetary weights, for instance, the numerator is expressed as dollars or pounds which have 'died', the denominator in terms of those sums which were exposed to this risk, measured in dollar or pound years.

Weighting the rate estimates in this way is a natural thing to do, in that what matters ultimately to an insurance company is the monetary amounts requiring to be paid out. Multiplying total sums at risk in an age interval, for example,
by the central weighted mortality rate yields an estimate of total payments to
that group over the next year, provided that the weights remain unaltered. This
point notwithstanding, the use of weighted rates is clearly somewhat dangerous
if the underlying mortality rates for a given age vary within the portfolio, and
the inter-relationship between the weights and the mortality rates is not
understood, or changes unexpectedly over time.

The statistical heterogeneity caused by confounding mortality effects and
financial factors may pose insurance companies a serious problem, because,
while mortality changes relatively slowly over time, the inter-relationship
between sums at risk and mortality may change suddenly and unpredictably.
The 1956 Finance Act in the UK, to take an extreme example, radically
changed the pattern of annuity purchases in the UK overnight: the annuitants' expe-
enience gathered by the insurance companies had thereafter to be separated
into pre-1957 and post-1956 subgroups (CMIR1 (1973, p. 29)). Less extreme
examples occur every year with changes to the financial environment, such as
adjustments to tax rates, altering the balance of the different socio-economic
groups purchasing various financial instruments, and changing the relative sizes
of transactions as well. The danger is all the more acute in that such changes will
tend to occur uniformly over ages, exacerbating the financial consequences.

When numbers of policies are used as weights, on the other hand, the link
between the weights and the underlying mortality may not be subject to such
capricious political changes; nevertheless, the relationship is not at all well
understood, and the effect of duplicate policies, especially at advanced ages,
can be very marked (e.g., CMIR2 (1976, p. 69); see also CMIR6 (1983,
p. 45)).

Given the relatively crude understanding of the inter-relationship between
weights used and the underlying mortality rates, the apparent lack of effort
expended to investigate this connection is surprising, to say the least. It is not
that the literature has nothing to offer on weighted averages of mortality rates.
On the contrary, directly standardised demographic rates, of fertility or
mortality etc., assume precisely the form of a weighted average; and ratios of
weighted average mortality rates, such as the Comparative Mortality Factor
(CMF) and the Standardised Mortality Ratio (SMR), are of central impor-
tance in demography (Cox (1976, p. 298)). These two ratios, moreover, possess
direct analogues in economics in the Laspeyres and Paasche price indices,
which in turn have an extensive literature (Kitagawa (1964, p. 302)).

Nor is there a paucity of study on the dependence of mortality rates on many
other factors, such as impairments. Both the North American Society of
Actuaries and the UK actuarial bodies, amongst many other organisations,
publish regular reports on impaired lives' mortality, and there is a voluminous
literature. Haberman (1988), for example, deals with weighted averages of
$q/q'$, where $q$ is the mortality rate of the impaired lives' group, and $q'$ that of
the population at large; and both the CMF and the SMR can also be recast as
weighted averages of this ratio (Cox (1976, p. 298)).

The point is, however, that the emphasis in most mortality studies is on the
comparison between two sets of mortality rates, be they for two regions, or for
impaired and normal lives, etc.; the initial difficulty arising is that of finding appropriate weights to use in taking the averages of the rates. Here we are interested instead in the effects of using different possible weights applied to the one set of underlying mortality rates. The ratio $R$ defined below is the analogue of the Area Comparability Factor, a factor designed to correct for the varying population profiles in different regions when calculating inter-censal mortality behaviour (Cox (1976, p. 305)).

Nevertheless, the literature is not entirely silent on the dependence of mortality on the size of the insurance policy. In accord with the fact that the Society of Actuaries uses sums insured as weights for its mortality investigations, while the UK actuarial bodies generally use numbers of policies, most of this work originates from North America.

An early writer treating the behaviour of weighted rates such as those considered in this paper is Cody (1941), who finds expressions for the mean and variance of initial weighted mortality rates, and derives the ratio of the variances of the weighted and unweighted rates. More recently, Klugman (1981) has contrasted the mean square errors of weighted and unweighted initial mortality rates, setting off the larger variance of the weighted rate against the bias implicit in using the unweighted rate. These theoretical developments notwithstanding, there is very little insurance data to which to apply the theory. Virtually the only publicly available data is gathered by the Society of Actuaries in its quinquennial survey of Mortality on Policies for Large Amounts (SocAcTs (1987)); and these reports are not sufficient detailed to enable one convincingly to model in detail this aspect of mortality.

As Klugman (1981) has pointed out, however, it is clear from these reports that mortality is generally lighter for policies with higher sums insured. This is consistent with the PA (90) and PL (90) mortality tables, both calculated from the same life insurance data for UK pensioners relating to 1967-1970. The acronyms stand for “pensioners' amounts” and “pensioners’ lives” respectively; the weights used are annual annuity payments on the one hand, and unity on the other. The PA (90) rates are substantially lower than the PL (90) rates, implying that wealthier pensioners experience better mortality than their less well-off neighbours, and the former table is used in preference to the latter simply for reasons of financial conservatism (CMIR3 (1978, p. 20)).

Mathematical methods cannot be expected to forecast mortality changes of the magnitude experienced in the UK in 1956, but one can at least expect them to track underwriting results which are changing less dramatically over time. In this paper we suggest the use of the ratio $R$ of the weighted mortality rate to the unweighted rate for the purpose of alerting insurance companies to general alterations in portfolio characteristics, and in particular to potential deterioration in underwriting results arising from changes in the interaction between the sums at risk and the underlying mortality rates. The statistic $R$ is very simple to estimate, and can be calculated every year for each age, or group of ages. Moreover, it is readily expressed in terms of the correlation between the weights and the mortality rates, as is seen in Section 6, and its use accords well with one’s intuition.
After preliminary definitions are given in Section 2, in Section 3 we find the asymptotic distribution of $R$ as the sample size becomes infinite, so that its estimated value can be inserted in an approximate confidence interval. A by-product of the calculations is that we also obtain the asymptotic distributions of the weighted and unweighted mortality rates, for both central and initial rates.

Simple simulations are carried out in Section 4 to obtain a rough idea of how large a sample is required before the asymptotic distributions of the ratio $R$ and the mortality rates become acceptable approximations; a further purpose is to illustrate the use of weighted mortality rates in tracking the effect of mortality changes in a portfolio over time. Some problems of using weighted mortality rates in this manner are then addressed in Section 5.

Finally a plea is entered in Section 6 to make a start to the collection of data to keep track of the problem of the inter-relationship between the weights and the underlying mortality, for the principal annuity and assurance classes. The data is no doubt investigated more thoroughly within individual insurance companies; but as stressed above, it seems that little attention has been devoted to this matter on a sectoral level.

2. PRELIMINARY DEFINITIONS

The (sample estimate of the) central weighted mortality rate is defined as

\[
\overline{M^A} = \frac{\sum_{j=1}^{N} S_j \theta_j}{\sum_{j=1}^{N} S_j (1 - \theta_j u_j)}
\]

where $N$ is the sample size; $S_j$ is the weight, labelled as the sum at risk, assigned to the $j$th individual in the sample; the superscript $A$ denotes amount, reflecting the fact that the weights will frequently be monetary amounts; and $\theta_j$ represents the fate of the $j$th individual, assuming the value 1 if the person dies during the year of observation, and 0 otherwise ($\theta$, for \textit{theta}, is a conventional symbol for death in actuarial work). Individuals who survive are given a full year’s exposure in the denominator, while deaths contribute to the exposure only while they live: the variable $u_j$ represents the shortfall in exposure by those dying, so labelled because in practice it would frequently be taken to have a uniform distribution. All those in the sample are assumed to come under observation at the same time: there is no attempt to model entrance into the sample at varying ages or times, as is done in ROBERTS (1992b).

The initial weighted mortality rate is similar, save that all individuals are given a full year’s exposure to risk regardless of whether they live or die; i.e.,
we set the variable \( u = 0 \) in (1):

\[
\overline{Q}^A = \frac{\sum_{j=1}^{N} S_j \theta_j}{\sum_{j=1}^{N} S_j} \tag{2}
\]

The unweighted rates are labelled as life rates, and are obtained by setting \( S = 1 \) in (1) and (2):

\[
\overline{M}^L = \frac{\sum_{j=1}^{N} \theta_j}{\sum_{j=1}^{N} (1 - \theta_j u_j)} ; \quad \overline{Q}^L = \frac{\sum_{j=1}^{N} \theta_j}{N}
\]

The population parameters corresponding to these latter sample estimates are customarily labelled \( m \) and \( q \), whence the notation used above.

For the ratios of weighted to unweighted rates we may use either the central or initial rates:

\[
\overline{R}^M = \frac{\overline{M}^A}{\overline{M}^L} ; \quad \overline{R}^Q = \frac{\overline{Q}^A}{\overline{Q}^L}
\]

3. ASYMPTOTIC DISTRIBUTIONS

All individuals are assumed to be mutually independent: i.e., \((S_j, \theta_j, u_j)\) are mutually independent random vectors, for \( j = 1, \ldots, N \); and we further assume that \( u_j \) is independent of both \( S_j \) and \( \theta_j \), for each \( j \). The variates \( S_j \) and \( \theta_j \) are not assumed independent: their possible dependence is the problem addressed in this paper.

We shall obtain the asymptotic distribution of \( \overline{R}^M \) as the sample size \( N \to \infty \), from which the asymptotic distributions of \( \overline{M}^A, \overline{M}^L, \overline{R}^Q, \overline{Q}^A \) and \( \overline{Q}^L \) may all be immediately inferred. Nevertheless, the general derivation is not so easy to follow, and we illustrate the procedure briefly by first finding the asymptotic distribution of \( \overline{M}^A \).

3.1. Asymptotic distribution of \( \overline{M}^A \)

Dividing the numerator and denominator of \( \overline{M}^A \) by the (deterministic) sample size \( N \), and letting \( N \to \infty \), both quantities tend to probability limits, say \( \xi_1 \) and \( \xi_2 \) respectively:

\[
T_1 = \sum_{j=1}^{N} S_j \theta_j / N \to \xi_1 = E(S\theta), \quad \text{and}
\]

\[
T_2 = \sum_{j=1}^{N} S_j (1 - \theta_j u_j) / N \to \xi_2 = E(S) - E(u) E(S\theta);
\]
whence \( \overline{M}^A = \frac{T_1}{T_2} \) in probability, as \( N \to \infty \).

Each of the numerator and denominator, suitably centred and normalised, possesses a limiting normal distribution by the Central Limit Theorem; and so will the mortality rate \( \overline{M}^A \), an intuitive argument for which is as follows.

\[
\overline{M}^A = \frac{\xi_1 (1 + \varepsilon_1)}{\xi_2 (1 + \varepsilon_2)} = \frac{\xi_1}{\xi_2} [1 + \varepsilon_1 - \varepsilon_2 + O(\varepsilon_1^2 + \varepsilon_2^2)],
\]

where \( \varepsilon_j \to 0 \) in probability, as \( N \to \infty \), for \( j = 1, 2 \); and the expansion is valid as long as \( |\varepsilon_2| < 1 \). Neglecting the remainder term, and denoting the covariance matrix of \((\xi_1, \varepsilon_1, \xi_2, \varepsilon_2)' \) by \( \Sigma_0 / N \), the asymptotic distribution of \( \sqrt{N} (\overline{M}^A - M^A) \) is \( \mathcal{N}(0, \sigma_0^2) \), with

\[
\frac{\sigma_0^2}{(M^A)^2} = \left( \frac{1}{\xi_1} - \frac{1}{\xi_2} \right) \Sigma_0 \left( \begin{array}{c} 1/\xi_1 \\ -1/\xi_2 \end{array} \right) = \frac{\sigma_{11}}{\xi_1} - \frac{2 \sigma_{12}}{\xi_1 \xi_2} + \frac{\sigma_{22}}{\xi_2}.
\]

where \( \sigma_{11} = E(S^2 \theta) - \xi_1^2; \sigma_{12} = E(S^2 \theta) [1 - E(u)] - \xi_1 \xi_2; \) and

\[
\sigma_{22} = E(S^2) - E(S^2 \theta) [2 E(u) - E(u^2)] - \xi_2^2.
\]

\( \sigma_{11} / N \) is the variance of \( T_1 \), \( \sigma_{22} / N \) the variance of \( T_2 \), and \( \sigma_{12} / N \) their covariance. The formal justification for our conclusion follows from a standard functional central limit theorem (RAO (1973), p. 387).

Instead of further developing this formula, we shall obtain the analogous result for the statistic \( \overline{R^M} \), from which the asymptotic distributions of the other statistics, including \( \overline{M}^A \) itself, are easily derived.

### 3.2. Asymptotic distribution of \( \overline{R^M} \)

As above, we define the following random variables and probability limits.

\[
T_3 = \sum_{1}^{N} \theta_j / N \to \xi_3 = E(\theta), \quad T_4 = \sum_{1}^{N} (1 - \theta_j u_j) / N \to \xi_4 = 1 - E(u) E(\theta),
\]

so that \( \overline{R^M} = \frac{T_1 / T_2}{T_3 / T_4} \to R^M = \frac{\xi_1 \xi_4}{\xi_2 \xi_3} \) in probability, as \( N \to \infty \).

In analogy with the previous case, the limiting distribution of \( \sqrt{N} (\overline{R^M} - R^M) \) is \( \mathcal{N}(0, (R^M)^2 \Sigma \xi' \xi) \), where \( \xi' = (\xi_1^{-1}, -\xi_2^{-1}, -\xi_3^{-1}, \xi_4^{-1}) \), and \( \Sigma \) contains
elements $\sigma_{ij} = N \times \text{cov}(T_i, T_j)$. The leading $2 \times 2$ submatrix of $\Sigma$ is the covariance matrix $\Sigma_0$ previously employed in (3).

Defining

$$\mu = \frac{E(S\theta)}{E(S) E(\theta)}; \quad \nu = \frac{E(S^2 \theta)}{E(S) E(S\theta)};$$

$$\alpha = E(u) E(\theta); \quad \text{and} \quad \beta = E(u^2) - 2 E(u),$$

and making the innocuous assumption that the mortality rate $q_j$ for the $j$th individual is bounded away from unity, so that $q_j < q < 1$ for all $j$ and some $q$, the asymptotic mean and variance of $R^M$ are given by

$$\text{asmean}(R^M) = R^M = \mu \left[ 1 + \frac{\alpha (\mu - 1)}{1 - \alpha \mu} \right] = \mu + O(q)$$

$$N \times \text{asvar}(R^M) = \left( \frac{R^M}{E(S^2)} \right)^2 = \frac{1}{E(\theta)} \left[ \frac{\nu}{\mu} - 1 \right] - 2 \frac{1 - E(u)}{1 - \alpha \mu} (\nu - \mu)$$

$$+ \frac{E(S^2) + \beta E(S^2 \theta)}{[E(S)]^2 [1 - \alpha \mu]^2} + \frac{1 + \beta E(\theta)}{[1 - \alpha]^2} - 2 \frac{1 + \mu \beta E(\theta)}{[1 - \alpha \mu] [1 - \alpha]}$$

$$= \frac{1}{E(\theta)} \left[ \frac{\nu}{\mu} - 1 \right] + O(1) = \frac{E(S^2 \theta)}{[E(S\theta)]^2} - \frac{1}{E(\theta)} + O(1)$$

$$= \frac{1}{E(\theta)} \left[ \frac{\nu}{\mu} - 1 \right] - 2 [1 - E(u)] (\nu - \mu) + \frac{V(S)}{[E(S)]^2} + O(q)$$

in an obvious, if rather cumbersome, notation. Details are given in ROBERTS (1992a).

That $\nu/\mu \geq 1$ is easily shown from Schwarz' inequality; the inequality is strict unless $S$ reduces to a constant (disallowing the pathological case in which $E(\theta) = 1$). The factor $(\nu - \mu)$ appearing in the expression for the variance is therefore positive unless all the sums at risk are equal.

### 3.3. Asymptotic distributions for the remaining statistics

From the asymptotic distribution for $R^M$, the limiting distribution of $M^A$ is found by setting $T_3 = T_4 = 1$, causing the corresponding terms in the covariance matrix to vanish, while that of $M^L$ is found by setting $T_1 = T_2 = 1$. The limiting distributions of $R^Q_0$, $Q^A$ and $Q^L$ are then deduced by setting the variable $u = 0$. 
For completeness we gather the remaining results together below.

\[ \frac{\mu E(\theta)}{1-\alpha \mu} \]

\[ \frac{E(S^2 \theta)}{[E(S\theta)]^2} - 2 \frac{1-E(u)}{1-\alpha \mu} + \frac{E(S^2) + \beta E(S^2 \theta)}{E(S)^2 [1-\alpha \mu]^2} \]

\[ \frac{E(S^2 \theta)}{[E(S\theta)]^2} + O(1) \]

\[ \frac{E(\theta)}{1-\alpha} \]

\[ \frac{1}{E(\theta)} - \frac{2}{1-\alpha} \frac{1-E(u)}{1-\alpha} + \frac{1+\beta E(\theta)}{1-\alpha} = \frac{1}{E(\theta)} + O(1) \]

\[ \mu \]

\[ (v-\mu) \left[ \frac{E(S)}{E(S\theta)} - 2 \right] + \frac{V(S)}{[E(S)]^2} \]

\[ \frac{E(S^2 \theta)}{[E(S\theta)]^2} - \frac{1}{E(\theta)} + O(1) \]

\[ \frac{E(S^2 \theta)}{[E(S\theta)]^2} - 2 \frac{E(S^2)}{[E(S)]^2} = \frac{E(S^2 \theta)}{[E(S\theta)]^2} + O(1) \]

\[ \frac{1}{E(\theta)} - 1 \]

When deaths are uniformly distributed over the time period considered, i.e. when \( u \) has the uniform distribution between 0 and 1, the expressions given involving the central rates simplify substantially. In that case, \( \alpha = E(\theta)/2 \), \( \beta = -2/3 \) and \( E(u) = .5 \).

Regardless of the sums at risk, when all individuals are subject to the same survival curve, say with central rate \( m_0 \) and initial rate \( q_0 \), then \( E(M^L) = M^L = m_0 \), \( E(Q^L) = Q^L = q_0 \). The relationship between the two rates is given by \( m_0 = q_0/[1-E(u)q_0] \); see, \textit{i.a.}, \textsc{Elandt-Johnson and Johnson} (1980).

The asymptotic bias arising from the use of the weighted rates to estimate the unweighted rates, or vice versa, is seen to vanish when \( \mu = 1 \); i.e., when \( S \) and \( \theta \) are uncorrelated. In this case, it is to be expected that the variance is larger for the weighted rate, and this is verified for the initial rates below when \( S \) and \( \theta \) are independent.

We note that the expressions given for the means and variances of the central and initial rates agree as far as the highest order terms, as they also do for \( R^M \) and \( R^Q \).

The asymptotic distribution of \( Q^A \) is found in \textsc{Klugman} (1981), the variance agreeing with his formula for \( \text{Var}(q_c) \). The numerical dominance of the first
term \( E(S^2 \theta) [E(S\theta)]^{-2} \) in the variance is evident from the succeeding calculations in Klugman's paper. The ratio of the variances of \( \bar{Q}^A \) and \( \bar{Q}^L \) is 
\[
E(S^2 \theta) E(\theta) [E(S\theta)]^{-2} = v/\mu, \text{ neglecting lower order terms, which reduces to } E(S^2) [E(S)]^{-2} \text{ when } S \text{ and } \theta \text{ are independent, in agreement with formulae in CODY (1941, p. 71) and KLUGMAN (1981, Section II).}
\]
The asymptotic variance of the weighted rate estimate exceeds that of the unweighted estimate unless all sums at risk are equal, because \( v > \mu \), as noted above.

The variance of \( \bar{Q}^L \) can be written as the "binomial" variance \( Q(1 - Q)/N \), with \( Q = E(\theta) \), conforming with POLLARD (1970, eqn. (8)) (in the notation of that paper, \( n_i \) and \( q_i \) are both deterministic). Pollard actually deals with the number of deaths, but this is precisely a multiple of the unweighted initial mortality rate.

Finally suppose that the \( j \)th group has sum at risk \( S_j \); that the individuals in that group have a common mortality rate of \( q_j \); and that the probability of an individual selected at random from the portfolio belonging to the \( j \)th group is \( \pi_j \), where \( \sum \pi_j = 1 \). Then
\[
\bar{Q}^A = \frac{E(S\theta)}{E(S)} = \frac{\sum S_j \pi_j q_j}{\sum S_j \pi_j}
\]

Thus \( \bar{Q}^A \) is the weighted mean of the group mortality rates, with weights the sum at risk times the probability of the policyholder belonging to that category. There are comparable results for the central rates. Armed with the further definitions 
\[
d_j = [1 - E(u)q_j]/[1 - E(u)q_0],
\]
where the 0th group has been chosen arbitrarily as a reference group, we have that
\[
\bar{M}^A = \frac{E(S\theta)}{E(S) - E(u) E(S\theta)} = \frac{\sum S_j \pi_j d_j m_j}{\sum S_j \pi_j d_j}
\]
The asymptotic mean is again a weighted sum of the central rates for each group; the weights, while slightly more complicated, differ in value but little from the weights used for the initial rates, the parameters \( d_j \) being close to unity.

4. SIMULATION RESULTS

The calculation of mortality rates was simulated in order firstly to obtain a rough idea of how large a sample needs to be before the asymptotically valid normal distribution becomes a reasonable approximation to the finite sample distribution. A second purpose was to investigate how effective the calculation of the ratio of amounts' and lives' mortality rates is at tracking the behaviour of mortality within the portfolio over time.

The sample investigated was considered to consist of two groups, with equal numbers of individuals in each group. These numbers were assumed stationary through time, so that deaths are replaced by new lives in the investigation. Mortality rates were initially identical for the two groups, as were the sums at
risk; over the 10 year sampling period, however, while the characteristics of the first group remained unchanged, the sum at risk of the second group increased by 5% p.a.; and the initial unweighted rates for the second group were successively assumed to increase, then to decrease, by 5% p.a. For each scenario, 500 simulations were carried out.

Situations sampled ranged over mortality rates of .005, .01 and .05 at the beginning of the simulated period, and for sample sizes between 100 and 100,000 life years per annum.

For each run, the intervals bounded by the 2.5 and the 97.5 percentiles (95% highest density regions, or HDRs) were found for both the empirical sample distribution and the approximating normal distribution; when these intervals were close to coincident, the asymptotic distribution was taken to be a good approximation to the finite sample distribution. Using this criterion, we conclude that the limiting normal distribution is an adequate approximation for all of the mortality rates as soon as the expected numbers of deaths reaches about 5; for the ratio $R$, whether using initial or central rates, this stage is not reached until expected deaths reaches about 10. The larger sample size required for $R$ is probably due to the fact that a default value must be assigned to $R$ when there are no deaths (for which purpose unity was chosen here): there is little possibility of approaching the limiting distribution unless there is a negligible chance of no deaths arising.

Once the expected number of deaths reaches about 20, the limiting normal distribution would seem to be a good approximation for all statistics, in the sense that the two HDRs almost overlapped. Some typical results are shown below, for an initial mortality rate of 5%, and with mortality increasing for half the population at 5% p.a. We note that exact moments of the mortality rates for small sample sizes can be calculated from formulae given in Roberts (1992a); but this is impractical for other than very small samples.

As regards the second purpose of the simulation, viz. that related to keeping track of changes in the portfolio over time, one wants the ratio $R$ to differ significantly from unity for moderate sample sizes, when the weighted and unweighted rates are behaving differently from one another. The criterion chosen for deciding when $R$ differs from unity is that the asymptotic mean of $R$ should differ from unity by more than 4 (asymptotic) standard deviations. When this criterion is satisfied, there is at least a 97.5% probability that the 95% confidence interval surrounding the point estimate of $R$ will not contain the value unity; thus the conventional two sided statistical test of whether $R = 1$ will be rejected at the 5% level, with probability at least .975.

To see this, recall that $E(R^0) = \mu$; let $\text{Var}(R^0) = \sigma^2$, and suppose that $\mu - 1 > 4\sigma$. Using the normal approximation, $\text{Prob}(R^0 > \mu - 2\sigma) \approx .975$. On the assumption that the standard error estimated from the sample is close to the true standard deviation, the width of the 95% confidence interval encasing $\overline{R}^0$ will be close to $4\sigma$, and $\text{Prob}(\overline{R}^0 - 2\sigma > 1) > .975$.

The length of time, and/or the sample size, required to ascertain that $R$ is diverging from unity is surprisingly large. The heavy unbroken lines in Figure 1 map the asymptotic 95% HDR, viz. $\mu \pm 2\sigma$, over time. For a sample size of
10,000 life years p.a., 7 years is needed for the mean to exceed unity by four standard deviations; for a sample size of 1,000 p.a., the criterion is nowhere near being satisfied even after the 10 years shown, and in fact 18 years are required. Even for a sample size of 100,000 p.a., the criterion is not met until 3 years have elapsed.

For just two groups of identical, independent individuals, \( N \) times the variance of \( \hat{R} \) reduces to

\[
\frac{\mu^2}{E(\theta)} \left[ \frac{v}{\mu} - 1 \right] = \frac{bc(a-1)^2}{\pi_0^3 q_0 (1+ac)^2 (1+bc)^3},
\]

where \( a = S_1/S_0 \); \( b = q_1/q_0 \); \( c = \pi_1/\pi_0 \); and only the highest order terms are retained (so that it is immaterial whether we use \( R^M \) or \( \bar{R}^2 \)). The criterion which we are using for assessing when \( R \) departs from 1 becomes

\[
|\mu - 1| = \frac{1+abc}{\pi_0(1+ac)(1+bc)} - 1 > \frac{4}{\sqrt{N q_0} \pi_0^{3/2}(1+ac)(1+bc)^{3/2}} |a-1| \sqrt{bc}
\]

Now set \( c = 1 \), \( \pi_0 = 1/2 \). The criterion reduces to

\[
\left| \frac{(1-a)(1-b)}{(1+a)(1+b)} \right| > \frac{4 \times 2^{3/2}}{\sqrt{N q_0}} \frac{|a-1| \sqrt{b}}{(1+a)(1+b)^{3/2}}; \quad \text{or}
\]

\[
N q_0 > \frac{128 b}{(1+b)(1-b)^2}
\]
Note that when $b = 1$, the requisite sample size is infinite: we cannot distinguish between two groups with equal mortality rates. When $b = 0$, the sample size required becomes zero because $R^0$ reduces to a constant, distinct from unity unless $a = 1$.

5. PRACTICAL CONSIDERATIONS

Despite the simplicity of the results in this paper, several problems may arise in their application to tracking the inter-relationship between mortality and weights over time.

In view of our simulation results, one could first admit that for smaller companies, or for the less popular classes of insurance, the expected number of deaths in one portfolio over a year may be well short of the 10 or 20 that seem to be needed for the limiting normal distribution to be valid. The second point arising concerns the large sample size necessary effectively to track the behaviour of mortality over time, even for only broad subgroupings of the exposure. On both counts, amalgamation of experience, either over time or over companies, may be necessary.

Even assuming that companies are willing to undertake the calculation of weighted rates over time, however, and that there is sufficient data for this to be worthwhile, there remain technical problems to be overcome. Of primary importance is the bias introduced by selective withdrawals in the measurement of mortality experience. This source of possible bias assumes some importance given the very high level of early withdrawals experienced by some classes of life insurance: in Australia, for example, 15% of ordinary life insurance policies lapse within the first year, and some 50% have surrendered within about 6 years (ISC (1992, p. 96)). The use of central mortality rates will decrease the potential bias, but the formulae derived above are far simpler for initial rates. Even if such problems are satisfactorily resolved within the individual company, moreover, there is still the problem at the sectoral level of obtaining data from the various companies on a comparable basis.

It is theoretically straightforward to derive the asymptotic distribution of rates similar to those used above, allowing for other decrements besides death (such rates are set out in ROBERTS (1992b), although the asymptotic distributions are not given there); but it is to be expected that there will be some correlation between sums at risk and lapse rates, and the accurate estimation of lapse and mortality rates for many different bands of sums at risk may present difficulties.

There is also the question of what weight to choose as the “sum at risk”. For the purposes of the insurance sector as a whole, this may best be chosen as the sum assured plus accrued bonus, or annual annuity payment in force. Within an individual company, however, the difference between the sum assured and the reserve may be a more sensible choice for insurance policies for sufficiently large portfolios. BATTEN (1978, ch. 7) considers several possibilities for the weighting variable.
Finally, as Klugman (1981) points out, the reason that weights used in practice are generally sums insured or numbers of policies is that of practicality: the information is easily available. The possibility arises then of using such a ratio as $R$, with both the numerator and denominator containing weighted mortality rates. The weights used in the numerator could remain sums insured, or other appropriate monetary weights; those in the denominator could be numbers of policies. The analysis in this paper would still be valid: the variable $\theta$ would now refer to the death of a policy, and the various terms like $E(S\theta)$ would need to be interpreted as expected values over a population of policies.

6. CONCLUSION

Referring to Section 3.3, recall that the moments for the $M$ and $Q$ rates and their ratios agree as far as terms of the highest order, and the following remarks will hold true whether one works with central or initial rates. The weights are assumed to be monetary weights, although comparable statements could be made for general weights.

Consider the ratio $R^Q$, the asymptotic mean of which is:

$$R^Q = \mu = \frac{E(S\theta)}{E(S)E(\theta)} = 1 + \frac{\text{cov}(S, \theta)}{E(S)E(\theta)} = 1 + \text{constant} \times \text{corr}(S, \theta).$$

We first note that the quantity $\mu$ is dimensionless, so that inflation affecting all members of the portfolio equally should not alter its value, as long as the underlying mortality rates in the portfolio remain unaltered (or the changes in mortality act uniformly on the whole portfolio). The second point to note is the ease with which $\mu$ is estimated, or rather $R^Q$ calculated: e.g., $E(S\theta) = \sum S_j/N$, where the summation is taken over deaths.

A third aspect is the appeal to our intuition engendered by the quantity $\mu$, which is simply related to the correlation between $S$ and $\theta$ as shown: the constant is the product of the coefficients of variation (standard deviation divided by the mean) of the two variates. A complete lack of correlation, i.e. no differential mortality for wealthier individuals in the portfolio, means that $\mu$ will remain as unity regardless of what happens separately to sums at risk and mortality.

The insurance company will wish to estimate both the ratio $\mu$ and its components $Q^Q = E(S\theta)/E(S)$ and $Q^L = E(\theta)$, for various classes of business and for different age groups, and to track these quantities over time. The expected sum to be paid out is the exposure to risk times the former quantity, while the second factor follows changes in the underlying mortality over time. Estimates should be encased in a confidence interval, standard errors being estimated from the above expressions for the asymptotic variance. Insurance companies should also commence to calculate these quantities on a sectoral level, to oversee changes in the insurance market as a whole.
ACKNOWLEDGEMENT

The author is grateful to the referees and the editor for constructive criticism, resulting in considerable improvement to the paper.

REFERENCES


Leigh Roberts
Institute of Statistics and Operations Research,
Victoria University of Wellington,
P.O. Box 600, Wellington, New Zealand.
THE INCIDENCE OF RISK UNDER CREDIT INSURANCE

BY GREG TAYLOR

December 1991
Revised May 1993

A Coopers & Lybrand research report

ABSTRACT

The incidence of risk under a credit insurance policy depends on the original term of the policy and the policy duration at which the incidence of risk is considered. Section 3 of the paper describes the procedure used to fit a bivariate function to this incidence. Section 4 gives the numerical detail of this model. Section 5 makes a comparison of the model with the data from which it was developed. Section 6 adds some general comments.

KEYWORDS

Credit insurance; incidence of risk; claim frequency.

1. INTRODUCTION

Credit insurance provides coverage of insureds who are subject to obligations to repay credit advances by means of periodic instalments. Coverage may be provided against the events of sickness/accident and/or unemployment.

Often, the credit to be insured will take the form of a fixed term loan, such as under a hire purchase contract. In this event the credit insurance policy would usually have a fixed term matching that of the credit contract. Typically, the fixed terms vary between 6 and 84 months, most commonly 24 or 36 months. Terms outside the range 6 to 84 months are occasionally encountered.

Typically, a single premium is paid at inception of the policy. It is then necessary for the insurer to spread this premium over the term of the policy for premium earning purposes. As the policies are of rather longer term than the conventional one-year coverages provided in the non-life market, it is by no means evident that premiums should be earned at a uniform rate over the whole term of a policy. Indeed, experience with this type of insurance invariably indicates a concentration of the risk of claim under the policy in the early part of the term.

The construction of a premium earning formula therefore involves the modelling of the incidence of risk over the term of such a policy. Although this is not particularly difficult in relation to any particular policy term, each separate policy term requires its own model, as will be seen in Section 4. There
is, therefore, the potential for one to finish with an inelegant plethora of premium earning models.

Section 4 illustrates how this situation may be averted. The final model there contains only a handful of parameters and deals with policies of any term.

2. DATA AND NOTATION

Data were provided in respect of the fixed term policies of an Australian credit insurer.

The insurer provided an analysis of its database giving, separately for each policy term, a tabulation of:

(a) number of claims, subdivided simultaneously according to:
   (i) month of attachment of the policy generating the claim;
   (ii) the development month (i.e. the number of months after month of attachment) of occurrence of the claim;

and

(b) number of premium transactions, subdivided simultaneously according to:
   (i) month of attachment of the policy generating the premium transaction;
   (ii) the development month of the premium transaction.

A "premium transaction" is any payment of premium by the insured, refund of premium to the insured, or error correcting journal entry by the insurer. A dissection of recorded transactions according to these categories was not available. For the purpose of the present study, and subject to the qualifications below, the transactions of development month 0 (i.e. the month of attachment) were taken to count 1 for each new premium, whereas those of subsequent development months to count 1 for each premium refund in respect of policy cancellation. The implications of this interpretation for the analysis are pointed out in Section 3.1.

As a consequence of a transfer of the portfolio from one insurer to another, the database contained:

(a) no information in respect of claims which occurred more than two years prior to the transfer;
(b) incomplete information in respect of policies effected more than two years prior to the transfer.

As a result, there was incomplete experience at policy durations in excess of 39 months. As will be seen in the graphs of Section 5, however, this proved to be a relatively minor shortcoming of the data, since few claims occur at such high durations.

Let

\[ p'_{ij} = \text{number of premium transactions in development month } j \text{ of policies with month of attachment } i \text{ and policy term } t \text{ months}; \]
\( N^r_{ij} \) = number of claims notified to date in respect of policies with month of attachment \( i \) and policy term \( t \) months, and with date of occurrence within development month \( j \).

Note that \( N^r_{ij} \) includes only claims notified to date. It follows that more recent development months will involve a systematic under-statement of the number of claims incurred in those months. Collateral data indicate that the great majority of claims within this line of business are reported within a few months. Thus, provided that the most recent few months of experience are excluded from consideration, any under-estimation of claim frequencies should be relatively small. Detail of the manner in which this question was treated is given in Section 3.1.

3. ANALYSIS

3.1. Observed incidence of risk

The term incidence of risk has been used to this point without strict definition. It is now defined.

Let

\[
e^r_{ij} = p_{i0} - \sum_{k=1}^{j-1} p_{ik} - \frac{1}{2} p_{ij}.
\]

According to the interpretation of the recorded premium transactions \( p_{ij} \) set out in Section 2, \( e^r_{ij} \) gives the average number of policies, with month of attachment \( i \) and policy term \( t \), which remain in force during development month \( j \). Equation (3.1) involves an implicit assumption that cancellations occur uniformly over development months.

Since the interpretation of the \( p_{i0} \) effectively ignores cancellations in the month of attachment, any such cancellations will lead to overstatement of \( e^r_{ij} \) for all \( j \). This effect was thought to be small.

Correction journals in month of attachment would tend to overstate \( e^r_{ij} \) for all \( j \), while those in development month \( k (>0) \) would tend to understate \( e^r_{ij} \) for all \( j \geq k \). Such corrections are generally concentrated in the early development months. To the extent that over- and under-statements fail to balance, there will be a systematic over- or under-statement of \( e^r_{ij} \) for the higher values of \( j \) (for given \( i \) and \( t \)).

To the extent that policy cancellations are also biased towards the front end of policy term, implying approximately constant exposure at the higher development months, the net result of the errors in computation of policies in force will be a proportionately constant over- or under-statement of claim frequency \( \lambda^r_{ij} \), defined below, at the higher values of \( j \). This has little effect on the incidence of risk, as defined below.

To the extent that policy cancellations are more dispersed over development months than are correction journals, the approximately constant error arising from the latter at higher \( j \) will relate to an exposure which reduces with
increasing \( j \). The under- or over-statement in exposure will therefore grow if relative terms as \( j \) increases. Correspondingly, the error in incidence of risk, as defined below, will also grow in relative terms.

It was not possible to correct the data for such effects, but they were thought to be relatively small.

Let

\[ I_{ij} = \text{number of claims incurred (i.e. notified to date \((N'_{ij}) + \text{IBNR}\) in respect of policies with month of attachment \( i \) and policy term \( t \) months, and with date of occurrence within development month \( j \);} \]

and suppose that

\[ E[I_{ij}] = e'_{ij} \lambda'_j, \]

where \( \lambda'_j \) is evidently a claim frequency parameter applying to development month \( j \) within policy term \( t \). Thus claims will be expected to occur (by number) over the term of the policy in proportion with \( \lambda'_j \).

In the particular portfolio considered here, there was little variation in claim size as \( j \) varied. The majority of claims involved sickness, injury or unemployment of only a few months. Unemployment coverage was so limited by terms of the insurance. As a result, tabulations of average claim sizes by \( j \) (for specific \( i \) and \( t \)) yielded no discernible trends until \( j \) reached values close to \( t \) (about \( t-1 \) or \( t-2 \)), when average claim sizes were reduced. Claim frequencies are generally small at these values of \( j \), and so this variation of claim size has been ignored.

With no trend in claim size as \( j \) varies, amounts of claims will be incurred in proportion with \( \lambda'_j \). The distribution

\[ \lambda'_j \sum_{k=0}^{t} \lambda'_k \]

will be referred to as the **incidence of risk** over policy term \( t \). Premium should be spread over the term of the policy in these proportions for earning purposes.

Let \( \hat{\lambda}'_j \) be estimated by:

\[ \hat{\lambda}'_{ij} = N'_{ij}/e'_{ij}, \]

which provides a separate estimator of \( \lambda'_j \) for each month of attachment \( i \). It may be calculated that:

\[ E[\hat{\lambda}'_{ij}] = \lambda'_j (N'_{ij}/I'_{ij}), \]

the bracketed term on the right indicating the factor by which the estimator (3.3) is biased. As noted above, the bias tends to be small:

\[ N'_{ij}/I'_{ij} \sim 1, \]

provided that the development month \( j \) is not too close to the end of the period of investigation.
Accordingly, the estimator (3.3) should either be used only for those values of \( j \) sufficiently separated from the end of the investigation, or should be replaced by:

\[
\hat{\lambda}_j = \hat{\lambda}_j \hat{\theta}_j,
\]

where \( \hat{\theta}_j \) is an unbiased estimator of \( E[I_{ij}/N_{ij}] \). Combination of (3.4) and (3.6) then yields that

\[
\hat{\lambda}_j
\]

is an unbiased estimator of \( \lambda_j \) provided that the random variables \( N_{ij} \) and \( I_{ij}/N_{ij} \) can be assumed stochastically independent.

In order to avoid notational inconvenience, the following Sections will use \( \hat{\lambda}_j \) to denote either of the estimators of \( \lambda_j \) defined in (3.3) and (3.6).

Examination of values of \( \hat{\lambda}_j \) for \( j \) and \( t \) fixed but \( i \) varying revealed no trend in claim frequency with month of occurrence (actually quarter of occurrence was used in the investigation described here). Hence, these estimators for different values of \( i \) were collapsed into a single estimator independent of \( i \):

\[
\hat{\lambda}_j^* = \sum_i \hat{\lambda}_{ij} / m_j^*,
\]

where \( m_j^* \) is the number of periods of occurrence included in the averaging. It would be possible to replace (3.7) by some form of weighted average if this were seen as desirable.

### 3.2. Model of incidence of risk

The estimators (3.7) depend on both policy term \( t \) and development month \( j \). As the graphs of Section 5 demonstrate, the parameter \( \lambda_j \) is genuinely bivariate. That is, its shape as a function of \( j \) changes with \( t \); similarly its shape as a function of \( t \) changes with \( j \).

In order to avoid an over-abundance of parameter estimates, it is necessary to estimate \( \lambda_j \) simultaneously as a function of \( t \) and \( j \). Preliminary plots of \( \lambda_j^* \) against \( j \) for various fixed values of \( t \) indicated that \( \lambda_j^* \) broadly followed a gamma or Hoerl curve:

\[
\lambda_j^* = C(t) (j+1)^A(t) [B(t)]^{j+1},
\]

where \( A(t), B(t) \) and \( C(t) \) depend only on \( t \). These preliminary plots appear (in conjunction with plots of the fitted model) in Section 5. From (3.8),

\[
\log \lambda_j = \log C(t) + A(t) \log (j+1) + (j+1) \log B(t).
\]

It is evident from this last equation that the numerical parameters of the formula for incidence of risk may be estimated, for any particular \( t \), by linear regression of the logged observations on incidence of risk against the variables \( (j+1) \) and \( \log (j+1) \). This has been done for each of a number of policy terms, and the results appear in Section 4.
The estimation procedure described above produces estimates of the numerical parameters $A(t), B(t)$ and $C(t)$ separately for each policy term. Formulas have then been fitted to $A(t)$ and $B(t)$, treating these as functions of policy term $t$. Examples of the fitted Hoerl curves appear in the plots of Section 5.

Although regression has been used here to obtain estimates of the parameters $A(t), B(t)$ and $C(t)$, other forms of estimation are possible. It should be noted, however, in choosing an estimation criterion, that the empirical distributions of incidence of risk are incomplete for policies of terms exceeding 39 months, for the reason given in Section 2. Hence, any estimation method which requires the entire empirical distribution, such as method of moments, will not be feasible.

No modelling of $C(t)$ is necessary, since this parameter cancels in the quantity

$$\lambda_t = \sum_{k=0}^{t} \lambda_k$$

and so does not influence the earning of premium. Effectively, $C(t)$ is a parameter setting the level of claim frequency over the full term of the policy but not affecting its distribution over term.

Note that, if $j$ is regarded as continuous variable, then the gamma distributed variable, policy duration (months) at claim occurrence, will have a mean (see e.g. HOGG and CRAIG, 1970, Section 3.3),

$$\lambda_t = M(t) [\text{say}] = \frac{[A(t) + 1]/[-\log B(t)]]}{M(t)}.$$

The parameters $A(t)$ and $M(t)$ are equivalent to $A(t)$ and $B(t)$ in characterizing the distribution of incidence of risk over policy term. In fact, $M(t)$ is a more convenient estimand than $B(t)$ since the former varies quite close to linearly with $t$. This is illustrated graphically in Section 4.

Experimentation with the estimates of $A(t)$ as a function of $t$ suggests that it takes the following approximate form:

(3.11) $A(t) = k + \exp(\alpha + \beta t),$

where $k, \alpha$ and $\beta$ are numerical constants.

Equation (3.11) can be reduced to the linear form

(3.12) $\log [A(t) - k] = \alpha + \beta t,$

if the value of $k$ is known. In this case, the parameters $\alpha$ and $\beta$ can be estimated by linear regression. Simple simultaneous regression of $k, \alpha$ and $\beta$ is not possible however. Experimentation with trial values of $k$ indicated a value of $-0.3$ as suitable. It then follows from (3.10), (3.12) and the linearity of $M(t)$ that:

(3.13) $\log [A(t) + 0.3] = \alpha + \beta t,$

(3.14) $M(t) = \gamma t,$
The linear expression for $M(t)$ in (3.14) has been forced through the origin. On purely logical grounds, it is reasonable to assume that the average policy duration at claim occurrence must approach zero as the term of the policy approaches zero. As the relevant graph of Section 4 shows, this assumption is supported empirically. Estimates of the parameters $\alpha$, $\beta$ and $\gamma$ may now be found by regression of:

(a) the estimates of $\log [A(t)+0.3]$ on $t$; and
(b) the estimates of $M(t)$ on $t$, forced through the origin.

The numerical detail derived from performing the procedures described above is given in Section 4.

4. NUMERICAL DETAIL OF MODEL

The first set of regressions described in Section 3.2, fitting separate distributions of incidence of risk to various terms $t$ yielded the following estimates of parameters of the Hoerl curve (3.8).

<table>
<thead>
<tr>
<th>Policy term $t$</th>
<th>Parameter estimate</th>
<th>Fitted mean policy duration $M(t)$ at claim occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Months</td>
<td>$A(t)$</td>
<td>$B(t)$</td>
</tr>
<tr>
<td>12</td>
<td>1.127</td>
<td>0.700</td>
</tr>
<tr>
<td>18</td>
<td>1.171</td>
<td>0.772</td>
</tr>
<tr>
<td>24</td>
<td>0.501</td>
<td>0.857</td>
</tr>
<tr>
<td>36</td>
<td>0.224</td>
<td>0.917</td>
</tr>
<tr>
<td>48</td>
<td>-0.054</td>
<td>0.959</td>
</tr>
<tr>
<td>60</td>
<td>-0.165</td>
<td>0.977</td>
</tr>
<tr>
<td>72</td>
<td>-0.247</td>
<td>0.977</td>
</tr>
</tbody>
</table>

In this table, the values of $M(t)$ have been computed from $A(t)$ and $B(t)$ according to (3.10).

The following graphs exhibit the forms $A(t)$ and $M(t)$ as functions of $t$. They also display the curves fitted to these values by means of (3.13) and (3.14). The parameters $\alpha$, $\beta$ and $\gamma$ in those equations have been estimated as:

$$
\alpha = 1.21014,
\beta = -0.05531,
\gamma = 0.49063.
$$
Parameter $A(t)$

- Fitted to particular policy term
- Fitted to all policy terms

Parameter - mean policy duration at claim occurrence

- Fitted to particular policy term
- Fitted to all policy terms
Thus, by insertion of these numerical values in (3.8), (3.13) and (3.14), the final model of incidence of risk is:

\[
\lambda_j^* = C(t) \times ((j + 1)^{\exp[1.21014 - 0.05531t] - 0.3} \\
\times \exp \left[ \frac{-(j + 1)\{\exp[1.21014 - 0.05531t] + 0.7\}}{0.49063t} \right] .
\]

It may be noted here that (4.1) involves a small bias. This arises from the informal procedure of logging the data, fitting a linear model and then exponentiating the fit. The effect of this can be seen as follows.

Suppose \( X \) is a lognormal random variable:

\[
\log X \sim N(\mu, \sigma^2),
\]

whence

\[
E[X] = \exp (\mu + \frac{1}{2}\sigma^2).
\]

Let \( \hat{\mu} \) be an estimator of \( \mu \) such that

\[
\hat{\mu} \sim N(\mu, v^2),
\]

and estimate \( E[X] \) by \( \xi = \exp (\hat{\mu}) \). Then

\[
E[\xi] = \exp (\mu + \frac{1}{2}v^2) = E[X] \exp \frac{1}{2}(v^2 - \sigma^2).
\]

This means that, as an estimator of \( E[X] \), \( \xi \) is biased by a factor of \( \exp \frac{1}{2}(v^2 - \sigma^2) \). Such bias arises in the inversion of both (3.9) and (3.13).

In the former case, constant factor biases such as this cancel out of the incidence of risk as defined in Section 3.1. In the latter case, however, a positive or negative bias in \( A(t) \) will arise, with corresponding distortion of the resulting estimated incidence of risk. While correction of this distortion is possible by means of estimates of \( \sigma^2 \) and \( v^2 \), it is small in the example of this section and has been disregarded.

It is appropriate at this point to interpolate a couple of remarks in interpretation of the above model. The basic Hoerl curve (3.8) is characterized by a location parameter \( M(t) \) and a shape parameter \( A(t) \).

The empirically justified form (3.14) of \( M(t) \) means that, on average, a constant proportion of policy term has elapsed at claim occurrence, whatever the term of the policy. If \( A(t) \) did not vary with \( t \), the last observation would mean that all incidence of risk distributions would be the same except for a change of scale by reference to \( t \).

However, \( A(t) \) does in fact vary with \( t \). Effectively, \( A(t) \) relates mode and mean since

\[
\text{mode} = \text{mean} \times [1 + 1/A(t)]^{-1}, \quad \text{if} \quad A(t) > 0, \\
= 0, \quad \text{otherwise}.
\]
Thus, the smaller is $A(t)$, the smaller is the mode as a percentage of the mean. The above graph of $A(t)$ indicates that the mode (relative to the mean) shifts left with increasing policy term, reaching zero (policy attachment) when the policy term reaches about 4 years.

The exponential form of $A(t)$ as a function of $t$ is quite empirical, though any other form would need to observe the fact that (3.8) requires $A(t)$ to be bounded below by $-1$.

5. COMPARISON OF MODEL WITH EXPERIENCE

The following three graphs provide illustrative comparisons of the experienced incidence of risk over policy term with that modelled. The three examples deal with a range of policy terms, from short (12 months) through the common medium term of 36 months to long (72 months).

Each graph, which corresponds to a particular $t$, displays three curves representing different functions over $j$. These functions are:

(a) the empirical incidence of risk defined by (3.7);
(b) the fitted incidence of risk (3.8), where fitting is by reference to just the empirical curve appearing in the graph, i.e. with fitting by reference to only the particular $t$ to which the graph applies (the parameters $A(t)$ and $B(t)$ are obtained from the table set out in Section 4);
(c) the fitted curve obtained from the final model (4.1), which applies to all values of $t$ and $j$.

The curves (a), (b) and (c) are designated in each graph “actual”, “fitted” and “model” respectively. Each curve has been normalized so that the sum of its masses is unity since, as remarked just before (3.3), this indicates the distribution of incidence of risk required for premium earning purposes.
The three graphs are representative of the efficiency of the model generally. It is seen to fit the data quite well, though perhaps with a slight tendency to understate earned premium, the feature being particularly noticeable at term 36 months.

This rather minor inefficiency needs to be assessed in the light of the fact that the many values of \( \lambda_j \) have been described with considerable parsimony in the final model (4.1), which depends on only four parameters.
6. EXTENSIONS AND QUALIFICATIONS

6.1. Estimation procedure

The estimation procedure described in Section 3 is to some extent *ad hoc*. It has the advantage of simplicity since each of the two stages of the procedure involves simple univariate regression. Because of this, the fitting may be carried out within a spreadsheet.

Although Section 5 suggests that it produces reasonable results, it might be preferable from a theoretical point of view to fit an expression of the form (4.1) to the data by some recognized optimization procedure such as maximum likelihood, instead of by the two-stage procedure used in Section 3.2.

Estimation by this means would of course be considerably more difficult. It is not possible to linearize (4.1) in all its parameters simultaneously and so formal methods of parameter estimation would need to involve non-linear computation. This has not been attempted.

The objective stated in Section 1 was effectively the fitting of an incidence of risk model appropriate to all policy terms and dependent on a small number of parameters. This is to be contrasted with a situation of searching for a model whose parameter set is reduced just to the point where predictive power of the model is optimized.

The latter case would be dealt with by optimization of some performance measure such as the Akaike information criterion. This statistic has not been computed here.

6.2. Economic conditions

The model (4.1) has been obtained from data summarizing past experience. Use of this model for premium earning purposes amounts to forecasting the future incidence of risk under policies now in force. This in turn involves an implicit assumption that the future incidence of risk will be a reproduction of the past.

Credit insurance is likely to be sensitive to economic conditions, particularly when it provides unemployment coverage. There is, therefore, a need for caution in applying the earning model (4.1). The reserve for unearned premium implied by this model should be supplemented by a margin for adverse experience if economic circumstances suggest this to be warranted.

6.3. Mortgage insurance

Mortgage insurance represents to some extent an extreme case of credit insurance, extreme in the sense that the sums insured involved are generally much larger and the policy terms much longer than is the case under conventional credit insurance.
The similarity between credit and mortgage insurance suggests that some of
the ideas and functional forms used in this paper may be transportable to the
latter line of business. As yet, no experimentation has been carried out in this
area.

REFERENCE

Company, New York.

GREG TAYLOR

IMPROVED ERROR BOUNDS FOR BERTRAM’S METHOD

BY BJØRN SUNDT

The Wyatt Company, Oslo

ABSTRACT

In an earlier note the present author deduced bounds for the approximation error of stop loss premiums when the aggregate claims distribution is calculated by a method introduced by Bertram. From the error bounds of the stop loss premiums we deduced bounds for the approximation error of the cumulative distribution and the discrete density of the aggregate claims. In the present note we shall improve the bounds for the cumulative distribution and the discrete density.

KEYWORDS

Aggregate claims distributions; approximations; error bounds.

Let $X$ be the aggregate claims occurred in an insurance portfolio within a given period and $G$ its cumulative distribution. We assume that $X$ is integer-valued and non-negative with finite mean. Let $g$ denote the discrete density and $G$ the stop loss transform of $G$, that is,

\begin{equation}
G(t) = E \max (X-t, 0) = EX - \sum_{x=0}^{t-1} (1-G(x)); 
\end{equation}

the latter quantity is the pure premium for an unlimited stop loss treaty with priority $t$.

For a positive integer $m$ we introduce

\[ X_m = X - mr_m(X) \]

with $r_m(x)$ denoting the largest integer less than or equal to $x/m$. Let $G_m$ be the cumulative distribution of $X_m$ and $g_m$ and $G_m$ respectively its discrete density and stop loss transform. We easily see that

\[ 0 \leq X_m \leq m-1 \]

\[ g_m(x) = \sum_{k=0}^{\infty} g(x + km). \quad (x = 0, 1, \ldots, m-1) \]

BERTRAM (1981) introduced a method for calculation of compound distributions, by which $g$ is approximated by $g_m$ on the range $\{0, 1, \ldots, m-1\}$. It is therefore of interest to study how well $g_m$, $G_m$, and $G_m$ approximate $g$, $G$, and $G$. 

ASTIN BULLETIN, Vol. 23, No. 2, 1993
Let
\[
S_m = E r_m(X) = \sum_{k=1}^{\infty} (1 - G(km - 1)) \\
D_m = E(X - X_m) = mS_m.
\]

In Sundt (1986) we showed the following inequalities:
\[
\begin{align*}
(3) & \quad \overline{G}_m(t) \leq \overline{G}(t) \leq \overline{G}_m(t) + D_m \quad (t = 0, 1, \ldots, m-1) \\
(4) & \quad G_m(x) - D_m \leq G(x) \leq G_m(x) \quad (x = 0, 1, \ldots, m-1) \\
(5) & \quad |g(x) - g_m(x)| \leq D_m. 
\end{align*}
\]

Formula (3) is a trivial consequence of Lemma 10.1 in Sundt (1991).

As \(EX < \infty\) and \(0 \leq mr_m(x) \leq x\) for all \(x\) and \(m\) with \(mr_m(x) = 0\) for \(m > x\), we see by bounded convergence that \(\lim_{m \to \infty} D_m = 0\). Thus \(\overline{G}_m\), \(\overline{G}_m\), and \(g_m\) converge uniformly towards respectively \(\overline{G}\), \(G\), and \(g\) when \(m\) goes to infinity. We see that if \(G(m-1) = 1\), then \(D_m = 0\). In that case \(g_m = g\).

From (1) we see that the stop loss transform \(\overline{G}\) satisfies the recursion
\[
\overline{G}(t) = \overline{G}(t-1) + G(t-1) - 1 \quad (t = 1, 2, \ldots)
\]
with initial value \(\overline{G}(0) = EX\). Analogously we have
\[
\overline{G}_m(t) = \overline{G}_m(t-1) + G_m(t-1) - 1 \quad (t = 1, 2, \ldots)
\]
with initial value \(\overline{G}_m(0) = EX_m\). It is interesting to note that by applying \(EX\) instead of \(EX_m\) as initial value, we obtain the upper bound in (3) instead of the lower bound. In particular we see that \(\overline{G}(t)\) is equal to the upper bound for \(t = 0\), and thus we believe that \(\overline{G}(t)\) is closer to the upper bound than to the lower bound for low values of \(t\).

We shall now show that we can replace \(D_m\) with \(S_m\) in (4) and (5). In practice \(m\) will be a relatively large number, and thus this replacement implies a considerable improvement of the inequalities.

For \(x = 0, 1, \ldots, m-1\) we have
\[
G_m(x) = G(x) + \sum_{k=1}^{\infty} (G(x+km) - G(km - 1)) \leq \frac{1}{1 - G(m - 1)}. \\
\]

On the other hand, we have \(G(x) \leq G_m(x)\), and thus
\[
G_m(x) - (1 - G(m - 1)) \leq G(x) \leq G_m(x). 
\]

Unfortunately \(G(m - 1)\) would normally be unknown, and thus we cannot immediately apply the lower bound in (6). However, as \(1 - G(m - 1) \leq S_m\), we can replace \(1 - G(m - 1)\) with \(S_m\) in the lower bound in (6), and thus we obtain
\[
G_m(x) - S_m \leq G(x) \leq G_m(x). 
\]

As
\[
G(x) - (G_m(x) - S_m) = \sum_{k=1}^{\infty} (1 - G(x + km)) 
\]
is decreasing in $x$, $G(x)$ is closest to the lower bound in (7) for high values of $x$. For low values of $x$, the lower bound might be less than zero, and then it will of course be of no practical interest.

We obviously have $g(x) \leq g_m(x)$. On the other hand, by (7)

$$g(x) = G(x) - G(x-1) \geq G_m(x) - S_m - G_m(x-1) = g_m(x) - S_m,$$

and we therefore obtain

$$g_m(x) - S_m \leq g(x) \leq g_m(x).$$

In practice $G$ is usually a compound distribution. In that case $EX$ can be calculated as the product of the mean of the counting distribution and the mean of the severity distribution. Unfortunately we will normally need the values of $G_m$ to calculate $EX_m$ (and thus $D_m$ and $S_m$), and therefore we cannot beforehand determine an $m$ that will give a desired accuracy. What we could do, is to first calculate a rough approximation or upper bound to $S_m$ to obtain an idea of how large we should choose $m$. When $g_m$ has been found, we calculate the correct value of $S_m$.

Let us look at the special case when the tail of $G$ is exponentially bounded, that is, there exist positive constants $C$ and $\kappa$ such that

$$1 - G(x) \leq Ce^{-\kappa x}$$

for all non-negative integers $x$. By applying this inequality to the sum in (2) we obtain

$$S_m \leq \frac{Ce^x}{e^{\kappa m} - 1}, \quad D_m \leq \frac{mCe^x}{e^{\kappa m} - 1}.$$

We see that these bounds approach zero when $m$ approaches infinity.

Willmot (1993) deduces an exponential bound for the tail of $G$ for the case when $G$ is a compound geometric distribution.

REFERENCES


Bjørn Sundt
The Wyatt Company A.S, P.O. Box 1508 Vika, N-0117 Oslo, Norway.
GUIDELINES TO AUTHORS

1. Papers for publication should be sent in quadruplicate to one of the Editors:
   Hans Bühlmann,
   Mathematik, ETH-Zentrum,
   CH-8092 Zürich, Switzerland.
   D. Harry Reid,
   Eagle Star Insurance Company Ltd,
   Eagle Star House, 9 Aldgate High Street,
   London EC3N 1LD, United Kingdom.

   or to one of the Co-Editors:
   Alois Gisler,
   "Winterthur" Swiss Insurance Company,
   P.O. Box 357, CH-8401 Winterthur, Switzerland.
   David Wilkie
   Messrs. R. Watson & Sons

   Submission of a paper is held to imply that it contains original unpublished work and is not
   being submitted for publication elsewhere.
   Receipt of the paper will be confirmed and followed by a refereeing process, which will take
   about three months.

2. Manuscripts should be typewritten on one side of the paper, double-spaced with wide
   margins. The basic elements of the journal's style have been agreed by the Editors and
   Publishers and should be clear from checking a recent issue of ASTIN BULLETIN. If variations
   are felt necessary they should be clearly indicated on the manuscript.

3. Papers should be written in English or in French. Authors intending to submit longer papers
   (e.g. exceeding 30 pages) are advised to consider splitting their contribution into two or more
   shorter contributions.

4. The first page of each paper should start with the title, the name(s) of the author(s), and an
   abstract of the paper as well as some major keywords. An institutional affiliation can be
   placed between the name(s) of the author(s) and the abstract.

5. Footnotes should be avoided as far as possible.

6. Upon acceptance of a paper, any figures should be drawn in black ink on white paper in a
   form suitable for photographic reproduction with lettering of uniform size and sufficiently
   large to be legible when reduced to the final size.

7. References should be arranged alphabetically, and for the same author chronologically. Use
   a, b, c, etc. to separate publications of the same author in the same year. For journal
   references give author(s), year, title, journal (in italics, cf. point 9), volume (in boldface, cf.
   point 9), and pages. For book references give author(s), year, title (in italics), publisher, and
   city.

   Examples:
   Barlow, R. E. and Proschan, F. (1975) Mathematical Theory of Reliability and Life
   Jewell, W. S. (1975a) Model variations in credibility theory. In Credibility: Theory and
   Jewell, W. S. (1975b) Regularity conditions for exact credibility. ASTIN Bulletin 8,
   336–341.

   References in the text are given by the author's name followed by the year of publication
   (and possibly a letter) in parentheses.

8. The address of at least one of the authors should be typed following the references.

   Continued overleaf.
COMMITTEE OF ASTIN

Björn ÅJNE Sweden Chairman
James N. STANARD USA Vice-Chairman
Bouke POSTHUMA Netherlands Secretary
Jean LEMAIRE Belgium/USA Treasurer
Hans BÜHLMANN Switzerland Editor/IAA-Delegate
D. Harry REID United Kingdom Editor
James MACGINNITIE USA Member/IAA-Delegate
Edward J. LEVAY Israel Member
Charles LEVI France Member
Thomas MACK Germany Member
Ermanno PITACCO Italy Member
Jukka RANTALA Finland Member
Gregory C. TAYLOR Australia Member
Alois GISLER Switzerland Co-Editor
David WILKIE United Kingdom Co-Editor

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

Guidelines to Authors continued from inside back cover.

9. Italics (boldface) should be indicated by single (wavy) underlining. Mathematical symbols will automatically be set in italics, and need not be underlined unless there is a possibility of misinterpretation.
   Information helping to avoid misinterpretation may be listed on a separate sheet entitled 'special instructions to the printer'. (Example of such an instruction: Greek letters are indicated with green and script letters with brown underlining, using double underlining for capitals and single underlining for lower case.)

10. Electronic Typesetting using Word Perfect 5.1 is available. Authors who wish to use this possibility should ask one of the editors for detailed instructions.

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.