Volume 27, No. 2 November 1997

ASTIN BULLETIN

A Journal of the International Actuarial Association

CONTENTS

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.

THE CONCEPTION(S) AND THE BIRTH OF ASTIN

A personal view by Gunnar Benktander, presented on the last day of the XXVIIth ASTIN Colloquium in Copenhagen 5 September 1996

(reprinted from ABaCusen No. 3/96, with kind permission of the Editor)

Thank you Mr. Chairman.

Dear friends, this has been a very fine Colloquium. For me the greatest thing has however been to meet an old friend for fifty years. Dr. Paul Johansen, now 86 years old. Last time we met was in Zurich in 1980.

In 1946 there was a Nordic Mathematical Congress in Copenhagen. Paul was there, we met Niels Bohr and his brother Harald whom I knew from the wartime when he was a refugee in Stockholm. Today I will deal with events of which only Paul, Teivo Pentikäiinen and myself have some direct experience.

Let us think back on Europe in the years after World War 2, a time when the optimists studied Russian and the pessimists Chinese and where an Anglo-American orientation was natural. In my country—which is Sweden—and several other countries there was a great admiration of the British, who had been able to resist on their island in the critical year 1940.

Furthermore London was the insurance centre of the world and it was natural for us, who were interested in a future ASTIN, to go there. A disappointment was that insurance managers were against using actuarial talent in the general lines and life actuaries did not like the idea of an ASTIN. One prominent exception was Bobbie Beard, who worked for The Pearl Assurance Company, and later became its General Manager.

An early pilgrim to London was Dr. Paul Johnsen, who met Mr. Beard in 1946. I had that pleasure in 1949.

Things certainly also happened on the continent of Europe: France, Italy, Belgium, Switzerland, Holland and up in the Nordic countries.

As the life actuaries were sceptical of the ASTIN idea, the Permanent Committee of what later became The International Actuarial Association wanted some supervision to be exercised. This task was given to the distinguished Government Actuary in England, Sir George Maddex, the Chris Daykin of those days.

At the actuarial congress in 1954 in Madrid, we—the dissidents—were allowed to meet unofficially for half a day for an exchange of ideas. Research data and statistical results were presented. In Madrid some papers of an ASTIN nature were published.

This was even more the case three years later in New York. 16 ASTIN papers from 8 countries were published. From each of the countries Belgium, Britain, France, Holland and Switzerland one paper, from each of Denmark and Italy two, and from Sweden no fewer than seven papers.

Why so many papers written by Swedish actuaries? Let me recall that at the beginning of this century Dr. Filip Lundberg presented his Collective Theory of Risk. In his work a central idea was Ruin Theory, the heart of Risk Theory.

Professor Harald Cramer of Stochholm interpreted the works of Dr. Lundberg and further developed the theory. During the thirties he gathered around himself a handful of talented and very brilliant researchers. Some of them became Chief Actuaries.

During the forties a further group of actuaries won respect by analyzing statistics from the general lines, mainly in order to find a realistic basis for the rating in motor, fire, etc.

Let me remind you that in the USA the Casualty Actuarial Society (CAS) existed since 1914. On the European scene the Permanent had agreed on the formation of a provisional ASTIN Committee with the following personalities, namely Mr. Beard, Prof. Franckx, Dr. Johansen, Mr. Monic, their activities to be supervised—as I said before—by Sir George Maddex.

Such was the situation when we prepared to go to New York in the autumn of 1957. A matter of decisive importance was the election of a chairman for ASTIN. Actually there were two prominent candidates and now the drama begins!

The Swedes—based on their above mentioned strength and with the support of the other Nordic countries—believed themselves to be able to determine the outcome of the election of the chairman. Impossible today, certainly.

Paul decided to come to Stockholm. I met him at the Bromma Airport and in my home four of us put relevant questions to him. The other three have passed away. One was Carl Philipson, known for writing papers very difficult to understand. Another was Ingvar Sternberg with his orientation towards statistical facts and his reservation regarding the so called "integral boys". Both served later on the ASTIN Committee. The third one was Carl-Otto Segerdahl who joined the publishing board of the ASTIN Bulletin.

The next step was that Sternberg and myself on behalf of the Swedish actuaries should meet and interview the other candidate in New York. So there were two prominent and distinguished candidates. We ourselves were clearly in favour of Paul and we took care to make our preference known.

And now to the final act of the drama.

Those with an ASTIN interest met on the 16th and 17th of October 1957. I think there were some thirty of us. Unexpectedly the French launched a third candidate — a prominent "President et Directeur General"! This was in a way natural as statistical research in the general lines was highly developed in France. Remember that Buonaparte—his original spelling—had started L'Ecole Polytechnique! Several leading managers in France were Polytechniciens. So it was by no way Napoleonic that a third candidate appeared. The voting took place. Sir George counted the votes, put them in his pocket and stated, "Dr. Paul Johansen has been elected with a clear majority"! My interpretation of this was that Paul had received more than 50% of the votes, in other words, he did not profit from the francophonic split.

Paul became an excellent chairman, even better than we had expected. He could rely upon his great linguistic abilities, his Danish charm and his general insurance knowledge.

Paul gave 23 years of devoted service to ASTIN before retiring from the Committee in 1981. It has been a great pleasure to me personally—and I know that feeling has been shared by the other actuaries here who have fond memories of Paul from earlier days—to see him back among us here in Copenhagen.

I shall end with a quotation from Paul. In a short sentence he distinguished between life and non-life mathematics: "You only die once — and totally".

THE ACTUARY: THE ROLE AND LIMITATIONS OF THE PROFESSION SINCE THE MID-19th CENTURY

HANS BÜHLMANN*

ETH Zürich

The history of any profession relates to the history of the division of labour in society. This is generally true and can also be attributed to expertise in the financial world — a world which generates with increasing complexity growing numbers of specialists offering know-how and expertise. The competition amongst these different specialists as to whose services are the more useful, valuable or promising is a phenomenon which is more than simply personal. Individuals with similar training and a common base of knowledge form groups which promote *in corpore* their competence in certain specialised services. Such a group, which also assumes the responsibility of guaranteeing the qualifications of its members, is usually called a "profession".

The profession of the actuary is one of the oldest in the financial world. There are also accountants, underwriters, statisticians, demographers, operations researchers and financial engineers to name but a few. Allow me now to attempt to place the actuarial profession in this area of professional conflict. There can be no doubt that we actuaries have constantly increased our area of competence since our professional beginnings in the last century. But to maintain that we are the only ones to have done so would be inappropriate.

Phase one: *Consolidation of life insurance*

To understand what actuaries are is not possible without knowing where their roots lie. Actuarial organisations owe their existence to a unique service to society which originated in the second half of the last century. I would like to point out that I consider this to be a worldwide achievement and not one limited to a British context, although history books emphasise this aspect. The latter was indeed chronologically first: in Victorian England—the world's leading nation—developments occurred somewhat earlier than in other countries. But the same developments took place in Germany, France, Scandinavia, USA and even in Japan and Australia, for example, a few decades later. What is this co-called unique service? Actuaries of the second half of the last century managed to gain recognition for their mathematical doctrine from the entire life insurance industry worldwide. The essential elements of this doctrine had already been around for almost a century and in fact James Dodson formed the Equitable Life Assurance Society on its principles in 1762.

* Translation of a talk held at a meeting of the "Deutsche Aktuarvereinigung" in Würzburg, April 30, 1997.

ASTIN BULLETIN, Vol. 27, No. 2, 1997, pp. 165-171

For such an astounding achievement, conditions at the time had to be favourable. The first condition being that in the first half of the last century life insurance was to a large extent speculative business. An insurance chronicler wrote: "Insurance companies writing life business were breeding like flies in the summer sky, and disappearing just as fast". The reason for this disappearance was the non-application of a proven base for premium calculations on the one hand, and above all the lack of method in the excessive distribution of profits to the entrepreneurs on the other. To give a concrete example: between 1844 and 1853, in Great Britain 335 new insurance companies were planned, 149 were actually formed and a total of 59 survived this period. This called for legislative measures to protect the insured. Secondly, it is just as important to note that Equitable, founded in 1762 as I mentioned earlier, managed to weather all storms in good shape and flourished because of the scientific methods it employed. Other companies thus took over the insurance technique developed mainly by its actuary William Morgan to calculate reserves, create a technical balance sheet and develop mortality tables using statistical data.

This professional know-how had already begun to spread by word of mouth beyond national boundaries. Legislative bodies in most industrial nations were thus able to adopt a tried and tested scientific and mathematical technique. What had once been only a competitive edge for some life insurance companies was suddenly the norm, and the pioneers of this technique became members of the leading industrial profession.

It was this high status of actuaries which led to the formation of the first national actuarial bodies, the Institute of Actuaries in London in 1848, the Faculty of Actuaries in Edinburgh in 1856, and the Institut des Actuaires Français in 1889. This was also the year that the first professional actuarial body was formed in America which was the forerunner to today's American Society of Actuaries. It is also interesting to note early developments in Germany. As early as 1860, a group of mathematicians met regularly to discuss problems related to the insurance industry. This group published its ideas mainly in the Masius Rundschau der Versicherungen, with extremely innovative contributions by Zillmer and Hattendorf for instance. My efforts to find the exact date on which the Deutsche Gesellschaft für Versicherungsmathematik was formed were unfortunately unsuccessful, though it is known that the Institut für Versicherungswissenschaften with an actuarial department at the University of Göttingen was founded in 1895. Looking to the Far East, I would like to add 1897 and 1899 as the dates when the Australian Institute of Actuaries and the Institute of Actuaries of Japan were formed respectively. As you all know from the recent anniversary celebrations, the first International Congress of Actuaries took place in Brussels in 1895, which additionally led to the formation of the Belgian ARAB.

The circumstances which I have just described provided the basis for the golden age of the actuary spanning the first half of the 20th century. This period was characterised by the proud awareness that actuaries controlled the know-how and expertise in the area of life insurance, which formed the solid foundation for an entire industry. Managers often came from this profession and used their managerial position to maintain the high status of actuaries within the company.

Phase 2: ***Lack of challenge***

This portrayal of the golden age of the actuary is a little one-sided. I first came into contact with the profession in the 1950s as a young actuary and it would perhaps be interesting to hear my personal impressions on joining the world of life insurance. It was probably the same for most of my young colleagues of that time:

I was additionally fortunate in that I had a job with a company which was characterised by outstanding personalities, offered me excellent opportunities for development and, last but not least, paid very well. I soon realised that my mathematical knowledge gained at university was regarded as a welcome status symbol, but that my command of languages and other general knowledge gained at high school were far more important than mathematics. My dear teacher and supervisor at work, Professor Jecklin, summarised this so: actuarial science is a "ready-made" theory; all that remains to be done is the tuning of fine technical details. He could also have said that actuarial science has devoted itself to statics after it had mastered the dynamic conditions of the last century. For a young actuary this meant that a career in the insurance industry may well have offered an attractive entrepreneurial challenge, but the creative side of mathematics would be nothing more than a hobby. Despite the undertone in these statements, I hope that you will not find my impressions too negative; I later came to believe that *viewed from the inside*, i.e. from the point of view of someone who joined the insurance industry, working in this branch definitely had its attractions; admittedly these were highly dependent on individual situations.

But, as you know all too well, this looked different *viewed from the outside.* It was difficult to convince students or anyone from academic institutions in general that a profession which uses mathematics statically, but produces no dynamic scientific development of its own, can be attractive. In the first half of the century, the recruiting of young academics interested in actuarial science began to suffer because of this. The situation was made more desperate by the fact that the universities asked themselves—from an academic angle quite justifiably—whether there was still a need to continue with actuarial science chairs; the declining interest shown by students was accompanied by the academic questioning of the lack of innovative ideas in actuarial science. This led to the bizarre situation, for example in Germany, where, on the one hand, the insurance industry wanted to promote actuarial chairs—even offering to support them financially—but on the other, the universities accepted these offers rather reluctantly, if at all. This situation could not only be observed in Germany: it was mirrored the world over. In the Anglo-Saxon countries it merely took another form: universities assumed the role of preparatory schools for examinations without having a say in their content. It is well known that the examination content in the US and Great Britain for example is set by the professional bodies, which meant that until only 20 years ago actuarial exams contained little on mathematics from the 20th century.

In broader terms, not too long ago it could be ascertained that, professionally, the actuary was flourishing while, academically, his reputation was low.

Phase 3: ***New impulses***

In the 1960s and 1970s, the actuaries began to break out of the domain of life insurance. The scenario for the acceptance of actuarial methods in non-life insurance was interestingly similar to the one of a hundred years ago in the life branch.

Motor liability in particular was responsible for the enormous losses suffered by insurers in many countries. Even if companies took the trouble to collect statistical data thoroughly they had no clear concept of how premium discounts should be rated. It needed a sound, theoretical approach to bring order to the proliferation of the system.

As in life insurance, the theoretical approaches were once again already in place. In 1903 the Swede Filip Lundberg published the first paper on what was later to become known as collective risk theory. In this publication, he made the pioneering step towards modelling insurance events as a mechanism of chance developing over time. Without knowing, he had used the key concept of modern probability theory—the stochastic process—for the first time in an insurance context. His problem was that in 1903 there was no exact theory of stochastic processes in the strict mathematical sense. The relevant rigorous mathematical foundations were laid in the 1930s and 1940s, mainly by Russian mathematicians. Although Lundberg's work caught the attention of actuaries, it was understood by few, until Professor Harald Cramér's excellent didactic explanations made the relationship between collective risk theory and the theory of stochastic process apparent to a wide readership.

And yet Lundberg's contribution was the most important one. He dragged actuarial science form the intellectual "constraints" of the past into the 20th century. Through his efforts a milestone was reached: modelling the processes of insurance events on the basis of modern probability theory.

The mathematically precise structure of the bonus-malus system in motor liability insurance, based on the number of losses of individual insureds, is a prime example of such a model. This model became established worldwide during the 1960s. It was also the driving force which offered actuaries new opportunities for professional development outside the field of life insurance. This development soon spread to fire insurance, reinsurance and other branches.

It is interesting to consider whether the activities of actuaries outside the field of life insurance actually contributed to the widening of the limits of the existing profession or whether a second, new profession was thereby created. In the US, where this professional development took place quite earlier than in other countries, it led to a clear separation between Life Actuaries and Casualty Actuaries. In all other countries of the world, however, this widening of competencies has been kept under the same roof. We must nevertheless continually ask ourselves what actually keeps us under the same professional roof. Moreover, in comparison to his colleagues in life insurance, the boundaries of the non-life actuary's competence are a lot more blurred with other professions such as auditors or underwriters. We may regret this, but the pressure of always having to prove oneself in competition with other professions through particularly good performance also has its positive aspects.

Phase 4: ***The challenge from other professions***

In defining the limits of actuarial competence we have so far only considered the profession from the inside. We have seen how the actuary clearly marked his professional field as his own, foremost in life insurance, in order to then extend it into other insurance branches.

Another trend has also been evident over the last 40 years, namely the establishing of competencies in other professions whose areas of competence overlap with those of the actuary. Generally speaking, I believe that this development can be attributed to the fact that the concept of probability has, in modern times, become a universal way of thinking which is incorporated in all aspects of our life; in the last century, this concept was relevant to practically only two types of people: to professional gamblers — and actuaries. Another profession which has integrated the concept of probability increasingly since the middle of this century is that of accountants. They, too, often determine the present value of cash flows. Because they do this partly in accordance with principles which differ from actuarial ones, there is inevitably friction between the two professions. This also means competition in activities which were previously reserved exclusively for actuaries. From an economic point of view, there is nothing to be said against the existence of different professional doctrines to treat identical economic problems. On the contrary, the range of possible means available to business managers should be enhanced by this co-existence.

The example of actuary and accountant shows, however, that the overlapping of competencies from different professions can lead to disagreement on which one has an intellectual monopoly on the correct solution to a problem. I believe that it is precisely the role of actuaries to point out how restricting it is to suppose that for a given problem there is only one right answer. This is not even true in mathematics, let alone in economics (as an aside, the closeness of the actuary to mathematics may also be responsible for the fact that, mostly, the claims to a monopoly come from the other side).

The challenge actuaries face from another, newly created profession is however much greater than that which has come from the ranks of the accountants over the last few decades. This latest profession is so new that there is only an English name for it. "Financial Engineers", as they are known, have initially taken up activity in an area which was not previously occupied by any profession. We could therefore speak of unconquered territory or no-man's-land. The typical activities are: pricing of options, term structures for interest rates, optimisation of investment portfolios; in short, the problems of the modern financial world.

Looking back it is difficult to understand why the approaches and solutions developed for today's financial sector, which are clearly oriented towards mathematics, or to be more precise towards probability theory, did not originate from the breeding-ground of actuarial thinking. This is also true of the academic world; here, the innovation stemmed from the faculties of business administration, i.e. the business schools. The challenge from financial engineers is thus of a different type. It has shown us that a static profession can miss obvious chances of key importance to its professional development. The traditional thought patterns of financial mathematics kept us chained for too long to deterministic, non-variable, technical interest rates for the entire duration of an insurance contract. But the actuaries have woken up and are beginning to occupy the field of modern financial mathematics around the world.

The universities in particular have realised that finance and insurance mathematics should be presented to today's students as one discipline. I personally favour a fruitful co-existence of the paradigms which have grown from different breeding-grounds and which may lead to different practical solutions. In the resulting competition between the two professions, the actuaries have the disadvantage of having been somewhat late arrivals in the field. In many places, however, the training of actuaries concentrates more on mathematics than courses at business schools do. This long-term advantage should not be underestimated. And why can't the actuarial roof be big enough to cover financial engineers too?

The "common" roof for actuaries

What exactly is this "common" roof? How wide should it be and how diversified can the profession be which it covers? What does it mean: "we are all actuaries"? Is it not more helpful to say that you carry responsibility for the calculation of reserves in life insurance, for the rating of substandard risks in life and health, for loss statistics in fire insurance, for the development of derivatives at a reinsurance company or for demographical projections in social insurance? If we take our daily responsibility as a means of classification, we are—as are members of many other professions—an association of specialists. What does the common profession of actuary mean to us then? The first response to this question comes from the corporative world: the profession of actuaries defines the necessary code of conduct for its members; it ensures professionalism to others and controls conduct from within. Most national actuarial associations are now in the process of revising their rules of conduct. This will help clarify the professional standing of the actuary in today's business and insurance world and is thus a welcome move. Corporate understanding of the profession is, however, not sufficient to define "a common roof". Experience shows that the majority of guidelines developed by the profession is aimed at disjunct groups of specialists and thus contributes little to the unity of members of the profession. This unity requires a communicative understanding of what constitutes a profession; in addition to the know-how of these specialists, it requires in particular an integrated and mutual understanding of what these specialists do and on which basis their know-how is founded. Only in this way can we stop the "common" roof from turning into a tower of Babel. Those of us who work in universities can perhaps make the biggest contribution to our profession in this respect by teaching actuarial mathematics, financial mathematics, non-life and life insurance techniques, capitalisation and "pay-as-you-go" from a common base of understanding. I should perhaps add that the concept of a common basis for the understanding of the modern financial world (of which insurance is a part) is greeted with great enthusiasm by the students. The number of students who actively participate in our seminars on this subject has shot up. What is particularly pleasing is that it is often the best students who want to work in the area of integrated finance and insurance mathematics. The future role of the actuarial profession will depend essentially on how successful it is at supporting and encouraging these promising young people after they have graduated. If we can succeed in integrating them actively and professionally in the financial sector, this important sector of our economy will be able to look to the future with confidence.

HANS BÜHLMANN
Department of Mathematics
ETHZ
CH-8092 Zürich
Switzerland

A FREQUENCY DISTRIBUTION METHOD FOR VALUING AVERAGE OPTIONS

By Edwin H. Neave

School of Business,
Queen's University,
Kingston, Ontario

Abstract

This paper finds payoff frequency distributions for valuing European and American fixed strike average options on a discrete time, recombining multiplicative binomial asset price process. In comparison to other discrete valuation methods the distributions, obtained analytically from a generating function, greatly reduce the computational requirements needed for accurate valuation. Less data are needed to value geometric than arithmetic averages, but the magnitude of calculations is similar for both instruments. Calculations of order T^3 are needed to value European instruments, of order T^4 to value their American counterparts. A frequency distribution of a quantity called *pathsums* is used to value geometric average options, and a joint distribution of path sums and realized prices is used to value arithmetic average options. The frequency distributions give an exact value for geometric average, an approximate value for arithmetic average instruments. The method obtains additional information from the generating function to estimate approximation errors relative to the exact binomial solution. If the errors are significant they can be reduced using still further detail from the generating function. Error reduction can be performed selectively to minimize additional calculation.

Keywords

Binomial models; options pricing; average options; Gaussian binomial coefficients; numerical methods.

1. Introduction

Average options are instruments whose payoff depends on the average price of the underlying asset, determined over a prespecified period. The averages may be either arithmetic or geometric, and there are both fixed strike and floating strike average options, with the former being by far the most common. Jarrow and Turnbull (1996: 651, hereafter JT) state that average options are used in foreign exchange and commodity trading as well as in interest rate contracts. Commodity based options are written on such assets as oil or aluminium. JT note that the use of an average price reduces an option's sensitivity to price changes in the underlying commodity, especially to price changes occurring at or near contract maturity. Reduced sensitivity to prices can prove especially important in the case of illiquid commodities.

Because of their path dependence, average options are generally regarded as difficult to value, despite the following considerable progress. European fixed strike geometric average options have known analytic solutions for both continuously and discretely determined averages. Valuing European arithmetic average options on a continuous time process is more difficult, mainly because the conventional choice of process is a geometric diffusion for which the distribution of prices' arithmetic averages is not lognormal. Nevertheless, analytic solutions for European arithmetic average options have been found for continuously determined averages by Yor (1992), and Geman and Yor (1993); Geman and Eydeland (1995) report computational experience with these methods.

There are no analytic solutions for continuous time models with discrete averaging, although Turnbull and Wakeman (1991), Levy (1991), and Curran (1992) offer approximate solutions. Neither have discrete time models been studied analytically. Hull and White (1993) approximately value arithmetic average instruments on a binomial process. Neave (1993) uses a binomial model to calculate values for European and American arithmetic average options. Ho (1992) and Tilley (1993) propose simulation with bundling techniques for reducing calculations, and Tilley uses his approach to value both European and American average options. Ritchken, Sankarasubramanian and Vijh (1993) approximately value European arithmetic average options with up to 64 reset points, American options for up to 16 reset points, and compare their approximations to values obtained by simulation.

Methods such as Turnbull and Wakeman's are sufficiently accurate for processes with an annual volatility of 0.40 or less, but some price processes (e.g., those for aluminum and crude oil) exhibit higher volatilities. Moreover, for a fixed number of time periods T convergence of approximate to exact values becomes slower as volatility increases.This paper reduces the computational tasks in valuation for any volatility. It both offers new approximation methods with greater accuracy than those in the literature, and shows how the approximations can be amended to find exact valuations. It achieves these goals by organizing the data along lines indicated by a generating function.

While less data are needed to value geometric than arithmetic average instruments, in both cases the calculations are of order T^3 for European instruments, of order T^4 for American instruments, where T is the number of time periods. (Both n and T are used to denote time in the literature; T is used in at least two recent texts.) The calculations employ sets of paths called *bundles*, where a bundle is defined as the set of all possible paths of the same length and having a common end point. Each bundle can be broken into *sub-bundles*, where a sub-bundle consists of the paths in a bundle that have the same *path sum*, the latter being defined as the sum of path price indices. The number of paths in each sub-bundle is described by the so-called Gaussian binomial coefficients, for which analytic formulae are available. Distributions of path sums can be used to value European and American geometric average options exactly, the latter by recursive methods.

Another description of path characteristics, the joint frequency distribution of path sums and realized prices, is obtained from the same generating function in this paper. The joint distribution can be used to obtain good approximate values of both European and American arithmetic average options. In the European case, exact solutions can be found from the approximations with relatively little additional computing. Further experimentation is needed to determine the best way of refining the approximations to obtain exact values of the American options.

The methods can be applied to a variety of options, but for illustrative purposes the paper only values fixed strike average calls. The discussion is organized as follows. Section 2 specifies the asset price process and defines the options. Section 3 describes the problem structure, defines the generating function, and specifies the frequency distributions. Section 4 values a European and an American geometric average call. Section 5 values the corresponding arithmetic average calls; Section 6 concludes. Appendices detail some features of the methods.

2. The Price Process and the Options

This section defines the price process and formulates European call valuation problems. A recursive form of the European valuation problem is developed to show how bundling methods can be extended to value American as well as European instruments.

2.1. The Process and its Averages

Let $S_0 = 1$, and define $\{S_t\}$, the asset price process, by:

$$S_t = US_{t-1}; \tag{2.1}$$

where for $t \in \{1, 2, \ldots, T\}$,

$$U = \begin{cases} u; & p \\ u^{-1}; & q \end{cases}$$

with $u > 1$. The realized price cannot become negative, and remains finite for finite values of T and u. Cox and Rubinstein (1985) show that one continuous time limit of the binomial process is the lognormal; Feller (1957) provides parameter values for which the limiting distribution is the Poisson.

It is helpful to rewrite (2.1) as

$$S_t = u^{J_t}; \; J_t \equiv \sum_{s=1}^{t} X_s; \; t = 1, 2, \ldots, T \tag{2.2}$$

where the X_s, $s = 1, 2, \ldots, T$ are independent, identically distributed random variables:

$$X_s \equiv \begin{cases} 1; & p \\ -1; & q. \end{cases}$$

The values J_t, $t = 0, 2, \ldots, T$ are called *node* values. Since $S_0 \equiv 1$, $J_0 \equiv 0$. The cumulative sums of node values

$$V_t \equiv \sum_{s=0}^{t} J_s = \sum_{s=0}^{t} (t - s)\, X_{s+1} \tag{2.3}$$

are called *path sums*. Define the process averages, geometric and arithmetic respectively, by

$$\begin{aligned} G_t &\equiv \left[\prod_{s=0}^{t} S_s\right]^{1/(t+1)} \\ &= \left[\prod_{s=0}^{t} u^{J_s}\right]^{1/(t+1)} \\ &= u^{V_t/(t+1)} \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} H_t &\equiv \left[\sum_{s=0}^{t} S_s\right]/(t+1) \\ &= \left[\sum_{s=0}^{t} u^{J_s}\right]/(t+1) \end{aligned} \tag{2.5}$$

Given u, (2.4) shows that the V_t are needed to determine geometric averages, while (2.5) shows that the J_s; $s = 0, \ldots, t$ are needed to determine arithmetic averages.

2.2. Standard Indices

It is convenient to represent the possible outcomes of (2.1) as in Figure 1. For $T = 4$, the Figure arrays successive periods' outcomes along the main diagonals, starting with $t = 0$ in the lower left hand corner. Price increases are represented by upward moves, decreases by horizontal moves to the right.

$$\begin{array}{lllll}
u^4 & & & & \\
u^3 & u^2 & & & \\
u^2 & u^1 & u^0 & & \\
u^1 & u^0 & u^{-1} & u^{-2} & \\
u^0 & u^{-1} & u^{-2} & u^{-3} & u^{-4}
\end{array}$$

FIGURE 1: Original Indexing.

Price paths can be described either by a sequence of realized prices or, as Figure 1 and the subsequent discussion suggest, by the timing and signs of their first differences. For example, the path with price indices

$$0 - 1 - 2 - 3 - 4$$

is described by the vector of first differences

$$(X_1\ X_2\ X_3\ X_4)' = (-1\ -1\ -1\ -1)';$$

while the path with indices

$$0\ 1\ 0 - 1 - 2$$

is described by

$$(X_1\ X_2\ X_3\ X_4)' = (+1\ -1\ -1\ -1)'.$$

The respective path sums –10 and –2 can be calculated either from the node values or, using (2.3), from the X_t. *Using* (2.3), the path sum for the second path above is

$$4 - 3 - 2 - 1 = -2.$$

Information regarding paths and path sums can be determined systematically from a generating function that recognizes the sign and timing of first differences. For example, when $t = 4$ a suitable generating function is

$$f_4^0(y, w) = \prod_{j=1}^{4} \left(y^{-1}w^{-j} + yw^{j}\right) =$$

$$y^{-4}w^{-10} + y^{-2}\left(w^{-8} + w^{-6} + w^{-4} + w^{-2}\right) \tag{2.6}$$

$$+y^0\left(w^{-4} + w^{-2} + 2w^0 + w^2 + w^4\right)$$

$$+y^2\left(w^2 + w^4 + w^6 + w^8\right) + y^4w^{10}.$$

The terms y^{-1} in the first line of (2.6) record the effect on the time 4 price of any differences such $X_t = -1$, while the terms y^1 record the effect on the time 4 price of any differences such that $X_t = 1$, $t \in \{1, 2, 3, 4\}$. The terms w^{-5} record the effect on the path sum if $X_{4+1-s} = -1$, while the terms w^5 record the effect on the path sum if $X_{4+1-s} = 1$; $s \in \{1, 2, 3, 4\}$.

Lines 2 through 4 of (2.6) suggest grouping paths according to powers of y. Let a *bundle* $B(t, j)$ be the set of all paths ending at (t, j). For any given bundle defined by (2.6), the associated polynomial in w defines the distribution of path sums: powers of w indicate the values of the sums, coefficients of individual terms indicate the frequencies with which the sums are attained. For later use, let a *sub-bundle* $B(t, j, V)$ be the set of all paths in $B(t, j)$whose indices sum to V. The number of paths in each sub-bundle is given by the coefficients of the appropriate polynomial in w.

Function (2.6) and Figure 1 help both to structure valuation problems and to simplify path descriptions. With regard to structure, Figure 1 indicates that the attainable set of realized indices for paths in $B(4,0)$ is defined by the rectangle with lower left-hand corner at $(0,0)$ upper righthand corner at $(4,0)$. The Figure can be used to verify that $B(4,0)$ consists of $4!/2!2! = 6$ paths, all with the same probability of occurrence, and that the maximal and minimal path sums in $B(4,0)$ are 4 and –4 respectively. Accordingly, the set of possible values for path sums in $B(4,0)$ is

$$-4 \ -2 \ 0 \ 2 \ 4,$$

and (2.6) shows these values respectively occur with the frequencies

$$1 \ 1 \ 2 \ 1 \ 1.$$

With regard to simplifying path descriptions, the paths in a given bundle are distinguished by different orderings of price increases and decreases, but the timing of the increases implies the timing of the decreases. For example, since all paths in $B(4, 0)$ have two increases and two decreases, the path for which $X_1 = X_2 = 1$ must also have $X_3 = X_4 = -1$, from which it follows that the path's node values are

$$0 \ 1 \ 2 \ 1 \ 0.$$

Generalizing the example, the paths in any bundle can be described fully just by specifying the values of s, $s \in \{1, 2, ..., t\}$, for which $X_s = 1$. More formally, path characteristics can be inferred from a standardized process which replaces S_t in (2. 1) with S_t^*, where

$$S_t^* \equiv J_t^*;\ J_t^* \equiv \sum_{s=1}^{t} X_s^*; \tag{2.7}$$

$$X_s^* \equiv \begin{cases} 1; & X_s = 1 \\ 0; & X_s = -1. \end{cases}$$

A standardized process for $t = 4$ is displayed in Figure 2.

$$
\begin{array}{ccccc}
U^4 & & & & \\
U^3 & U^3 & & & \\
U^2 & U^2 & U^2 & & \\
U^1 & U^1 & U^1 & U^1 & \\
U^0 & U^0 & U^0 & U^0 & U^0
\end{array}
$$

FIGURE 2: Standard Indexing.

The generating function for the standardized process is obtained by rewriting (2.6) as:

$$
\begin{aligned}
f_4^0(y, w) &= \prod_{j=1}^{4} \left(y^{-1}w^{-1} + yw^j\right) = \\
&y^{-4}w^{-10} \prod_{j=1}^{4} \left(1 + xv^j\right),
\end{aligned}
\tag{2.8}
$$

where $x \equiv y^2$ and $v \equiv w^2$. When using the standardized process the multiplicative constant $x^{-4}y^{-10}$ is ignored and the generating function written in the simpler form

$$
\begin{aligned}
f_4(x, v) &= \prod_{j=1}^{4} \left(1 + xv^j\right) = \\
&1 + x\left(v^1 + v^2 + v^3 + v^4\right) \\
&+x^2\left(v^3 + v^4 + 2v^5 + v^6 + v^7\right) \\
&+x^3\left(v^6 + v^7 + v^8 + v^9\right) + x^4v^{10} = \\
&1 + xv\left[\left(v^4 - 1\right)/(v - 1)\right] + x^2v^3\left[\left(v^4 - 1\right)\left(v^3 - 1\right)/\left(v^2 - 1\right)(v - 1)\right] \\
&+x^3v^6\left[\left(v^4 - 1\right)/(v - 1)\right] + x^4v^{10} = \\
&\equiv g_{4,0}(v) + xvg_{4,1}(v) + x^2v^3g_{4,2}(v) + \\
&x^3v^6g_{4,3}(v) + x^4g^{10}g_{4,4}(v).
\end{aligned}
\tag{2.9}
$$

More generally, the generating function for the standardized process is:

$$f_t(x, v) = \prod_{j*=1}^{t} \left(1 + xv^{j*}\right) =$$

$$\sum_{j*=0}^{t} g_{t,j*}(v)\, v^{j*(j*+1)/2} x^{j*}. \tag{2.10}$$

The functions $g_{t,j*}(v)$ in (2. 10), known as *Gaussianpolynomials*, take the form

$$g_{t,j*}(v) = \prod_{k=1}^{j*} \frac{v^{t+1-k} - 1}{v^k - 1};\ 1 \leq j^* \leq t, \tag{2.11}$$

and $g_{t,0^*}(v) \equiv 1$; cf. Berman and Fryer (1972). The coefficients of the $g_{j^*}(v)$, which can be written as polynomials, define the so-called *Gaussianbinomialcoefficients*. In the present setting the Gaussian binomial coefficients $g_{t,j^*}(v)$ describe the frequencies of V^*, conditional on $J^* = j^*$. It is clear from comparing (2.6), (2.7) and (2.9) that

$$\begin{aligned} J_t &= 2 \cdot J_t^* - t; \\ V_t &= 2 \cdot V_t^* - t(t+1)/2. \end{aligned} \tag{2.12}$$

The possible values J_t^* are the integers from *0* to t, those of V_t^* the integers from zero to $t(t+1)/2$.

The sub-bundles defined by (2.10) can be used directly to value European average options, but recursion relations between the sub-bundles are needed to value American options. As Section 3 will show in greater detail, the necessary relations can be determined from (2.10)

$$f_t(x, v) = f_{t-1}(x, v)(1 + xv^t),$$

in conjunction with:

$$g_{t,0} = g_{t-1,0};$$

$$g_{t,j} = g_{t-1,j} + g_{t-1,j-1} v^{t-j},\ j = 1, 2, \ldots, t-1;$$

$$g_{t,t} = g_{t-1,t-1}$$

2.3. European Fixed Strike Average Calls

The payoff to a European fixed strike average call with exercise date T is

$$C_T \equiv (A_T - K)^+ \tag{2.13a}$$

where A_T is a random variable representing either a geometric or an arithmetic average and X_+ means $max\{X, 0\}$. The geometric average call uses $A_T \equiv G_T$, where G_T is defined by (2.4), and the arithmetic average uses $A_T \equiv H_T$, where H_T is defined by (2.5). Given a probability measure p, the time zero values of the European options are

$$C_0 \equiv R^{-T}E(A_T - K)^+ \tag{2.13b}$$

where E denotes expectation under p and $R^t \equiv (1+r)^t$ indicates the t-period accumulation of \$1 at the single-period risk free interest rate r. Recursive approaches can be used with either a martingale or with objective probability measures; cf. Dixit and Pindyck (1994). Schwartz (1994) discusses the theoretical correctness of using the different measures. In consistency with option pricing theory, we assume no arbitrage opportunities, market completeness, and that transactions costs are zero. Then a unique martingale measure $p^* = \left(R - u^{-1}\right)/\left(u - u^{-1}\right)$ can be obtained from the normalized process $S_t^* = S_t/R^t$. The paper uses p rather than p^*, reserving the asterisks to denote standard indexing.

To value the American analogues to (2.13a) and (2.13b), it is convenient first to formulate (2.13b) with the states defined as individual price paths.The methods will then be adapted to find recursions between states defined as path sums. We first number paths according to

$$Z \equiv \sum_{s=1}^{T} X_s^* \cdot 2^{T-s}; \tag{2.14}$$

and note the state variable Z can assume the realized values $z \in \{0,\ 1,\ \ldots,\ 2^T - 1\}$. Identifying the paths using values of z (2.13b) can be written:

$$C_0(z) \equiv R^{-T}E\{(A(z) - K)^+\}, \tag{2.15}$$

where $A(z)$ indicates an average over path z. There are 2^T possible realized values of Z, making computation infeasible when T is large. The states are later redefined so that for computational purposes it is only necessary to recognize distinct values of order T^3. Since $H(z) \geq G(z)$ for all z, (2.15) immediately confirms the result, first pointed out by Kemna and Vorst (1990), that the value of a European arithmetic average call is never less than the value of the corresponding geometric average call.

A recursive formulation is not needed to solve problem (2.15), but will help relate our methods for valuing European options to those for their American counterparts. Suppose henceforth that the z are arranged in increasing order at time T i.e.,

$$0,\ 1,\ \ldots,\ 2^T - 1.$$

Examination of Z shows the path numbers are lexicographically ordered by the signs of path first differences. For example, the pair of paths $2^T - 1$ and $2^T - 2$ differ in the sign of the first difference taken between times $T-1$ and T. The same is true for the pair 2^T-3 and 2^T-4, and for all remaining pairs of adjacent paths. After the expected value at time $T-1$ is taken over pairs of paths that are adjacent in terms of z, the states then requiring to be distinguished are indicated by

$$z \in \left\{0,\ 2,\ \dots,\ 2\left(2^{T-1} - 1\right)\right\}.$$

Again adjacent pairs of the remaining paths differ in the sign of what is now the first difference between times $T-1$ and $T-2$. That is, the remaining states are

$$z \in \left\{0,\ 4,\ \dots,\ 4\left(2^{T-2} - 1\right)\right\}.$$

The process continues until time 0, when the single state denoted by $z = 0$ is reached. The path numbering method is further illustrated in Table 3 below.

Using the relations between values of Z, (2.15) can be written recursively as:

$$C_T(z) \equiv (A_T(z) - K)^+;$$

$$z \in \left\{0,\ 1,\ \dots,\ 2^T - 1\right\} \equiv Z_T;$$

$$C_{T-1}(z) \equiv R^{-1}\{pC_T(z+1) + qC_T(z)\};$$

$$z \in \left\{j \cdot 2 : j = 0,\ 2,\ \dots,\ 2^{T-1} - 1\right\} \equiv Z_{T-1}. \tag{2.16a}$$

In (2.16a) $C_T(z)$ is the value of the European call at time T if the price path from time 0 time time T is described by z. In general,

$$C_{T-t}(z) \equiv R^{-1}\left\{pC_{T-t+1}\left(z + 2^{t-1}\right) + qC_{T-t+1}(z)\right\}; \tag{2.16b}$$

$$z \in \left\{j \cdot 2^t : j = 0,\ 1,\ \dots\ 2^{T-t} - 1\right\} \equiv Z_{T-t}.$$

When $t = T$, (2.16b) defines the time zero call value.

2.4. American Fixed Strike Average Calls

To write the recursion for the American call, (2.16) is amended to recognize the effect of early exercise. Let $D_t(z)$ be the time t value of the call if the price process has followed path z from time 0 to time t:

$$D_T(z) \equiv C_T(z);$$

$$z \in Z_T;$$

$$D_{T-1}(z) \equiv max\left\{(A_{T-1}(z) - K)^+,\ R^{-1}[pD_T(z+1) + qD_T(z)]\right\};$$

$$z \in Z_{T-1};$$

and (2.17)

$$D_{T-t}(z) \equiv max\left\{(A_{T-t}(z) - K)^+,\ R^{-1}\left[pD_{T-t+1}\left(z + 2^{t-1}\right) + qD_{T-t+1}(z)\right]\right\};$$

$$z \in Z_{T-t}.$$

In conformity with the standard result that the value of an American call is never less than that of its European counterpart, equations (2.17) show immediately that $D_t(z) \geq C_t(z)$ for all feasible values of z and t.

Since they recognize 2^T distinct paths, computations based on (2.16) and (2.17) increase exponentially in T. To reduce computation, the rest of this paper defines state variable values as the values defining path sub-bundles. In the American case the paper further determines how sub-bundles at a given time point are related to sub-bundles for the immediately preceding time. This approach reduces the number of calculations to cubic or fourth degree polynomials in T, according to whether European or American options are being valued. (The higher degree of polynomial for American options results from having to repeat the calculations at each of the T stages in the problem.) The approach gives exact values for geometric average options, approximate values for arithmetic average options. In the latter case, approximation error can be estimated and eliminated using relatively little additional calculation.

3. Problem Data and Valuation Methods

This section states process parameters, then discusses how paths can be bundled for valuation purposes. The methods use properties of (2.10) to adapt (2.16) and (2.17).

3.1. Process Parameters; Option Specifications

To enhance comparisons among different types of instruments, the same process parameters are used to value examples of four options – European and American geometric and arithmetic fixed strike average calls. As specified in Table 2 below, the examples value instruments on (2.1) with $T = 6$ quarterly time intervals, an annual volatility $\sigma = 0.40$, and a risk free rate $r = 0.10$ per annum. The initial asset price is $S_0 = 1.00$. Let k be defined as the solution to $u^k = K$ and take $K = 1.00$, so that $k = 0$. All options are assumed to expire at time T. If an option is exercised at time t, its path averages are defined over times $0, \ldots, t$. For European options $t = T$, for American options the early exercise feature means $t \leq T$.

Section 6.1, reporting computational experience, values European arithmetic average calls for $t \in \{6, 12, ..., 48\}$, $\sigma \in \{0.40, 0.60, \text{and } 0.80\}$, and the remaining parameters as indicated in Table 2.

TABLE 2

DATA FOR VALUE COMPARISONS

$\Delta t = 0.250000$		
$T = 6$	$\sigma = 0.400000$ per annum	
$r = 0.100000$ per annum		
$R = 1.100000^{.25}$	$= 1.024114$ per quater	
$u = 1.221403$	$= \exp(0.400000/(0.250000^{.5})$	
$p = 0.510051$	$= (R - u^{-1})/(u - u^{-1})$	
$q = 0.489949$	$= 1 - p$	$K = u^0 = 1.000000$

3.2. Ordering and Bundling Paths

Valuing a fixed strike average call involves finding a probability distribution of paths whose averages exceed K. These calculations' efficiency can be enhanced by organizing the data as indicated in Figure 3. Figure 3 shows the relations between a bundle and its sub-bundles, as well as the behaviour of the bundle's arithmetic averages when the paths are organized as shown. Each cell in Figure 3 represents a path in the bundle B(8,0), and the cell height indicates a path arithmetic average, in this case when $\sigma = 0.80$. Each (horizontal) bar of cells represents one of the sub-bundles of B(8, 0), with the length of the bar indicating the number of paths in the sub-bundle. The different heights within a bar indicate distinct sub-bundle arithmetic averages, the number of which is generally very much less than the sub-bundle' s number of paths. All the information conveyed by the graph can be obtained analyticaly, and all features of the graph except the cell heights are invariant with respect to volatility. Grouping paths into sub-bundles as indicated by the graph orders both the sub-bundle geometric means and the minima and maxima of sub-bundle arithmetic averages, properties used to advantage in the subsequent valuations.

Using the approach suggested by Figure 3, Table 4 organizes the data needed to obtain sub-bundle means of arithmetic averages in$B(6, 0)$. Each line of the Table 4 records, in the first thirteen columns, data needed to obtain such a sub-bundle mean. (Table 4 is shown with more columns than would normally be used in practice.) Column g indicates the numbers of paths in each sub-bundle, column V the path sum defining each sub-bundle, and column V^* the standardized path sums. Column M/g calculated from the index frequency section of Table 4, defines the sub-bundle mean of path arithmetic averages. For example, in the row for$B(6, 0, -5)$ $M/g = 6.126512$, indicating the mean of the arithmetic averages of the two paths in the subbundle is *6.126512/7*. When the sub-bundles are ordered by V, the values in column M/g increase monotonically, as illustrated by the example.

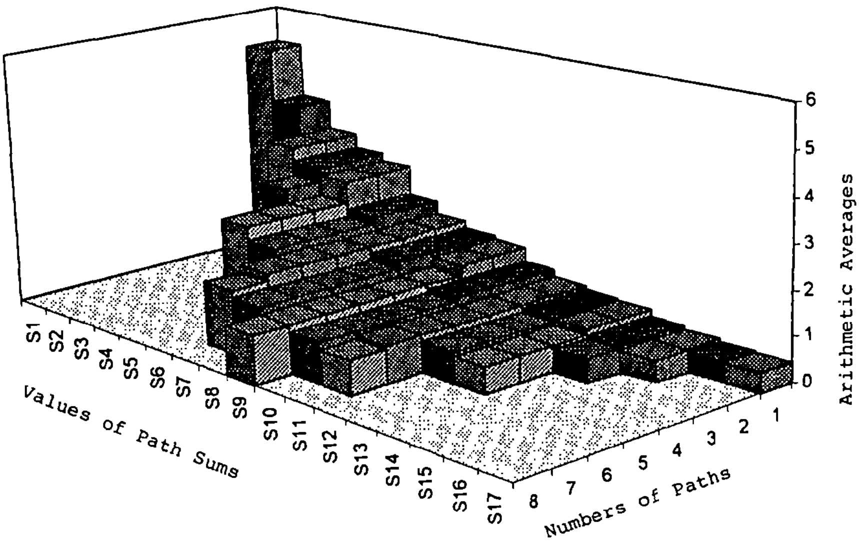


FIGURE 3. Arithmetic Averages in B(8,0); $\sigma = 0.80$.

TABLE 4

FREQUENCY DISTRIBUTIONS FOR B(6, 0)

Indices																
−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	g	V	V^*	M/g
0	0	0	1	2	2	2	0	0	0	0	0	0	1	−9	6	5.526912
0	0	0	0	2	3	2	0	0	0	0	0	0	1	−7	7	5.796831
0	0	0	0	2	6	6	0	0	0	0	0	0	2	−5	8	6.126512
0	0	0	0	2	7	10	2	0	0	0	0	0	3	−3	9	6.504853
0	0	0	0	0	6	12	3	0	0	0	0	0	3	−1	10	6.858864
0	0	0	0	0	3	12	6	0	0	0	0	0	3	1	11	7.261537
0	0	0	0	0	2	10	7	2	0	0	0	0	3	3	12	7.723644
0	0	0	0	0	0	6	6	2	0	0	0	0	2	5	13	8.156035
0	0	0	0	0	0	2	3	2	0	0	0	0	1	7	14	8.647860
0	0	0	0	0	0	2	2	2	1	0	0	0	1	9	15	9.248576

3.3. Frequency Distributions

Let the vector g_{t,j^*} represent the coefficients $g_{t,j^*}(v)$. It follows from (2.11) that g_{t,j^*} describes the frequency distributions of path sums for both $B(t, j^*, V^*)$, and $B(t, j, V)$. Columns g and V of Table 4 can thus be written directly from the $g_{t,j*}(v)$. The index frequency data in the first thirteen columns of Table 4 can be obtained using two-fold convolutions of (2.11). Consider each in turn.

The function (2.10) generates the data in columns g and V directly. (Subscripts are omitted when the context permits.) Consider*B(6, 0)*; i.e. *B(6, 3*)* in standardized notation. Using (2.10) and (2.11), the range of values for V^* is from *6* to *15*, and their frequencies are obtained from

$$\begin{aligned} g_{6,3^*}(v) &= \left(v^6 - 1\right)\left(v^5 - 1\right)\left(v^4 - 1\right)/(v - 1)\left(v^2 - 1\right)\left(v^3 - 1\right) \\ &= \left(v^4 + v^3 + v^2 + v + 1\right)\left(v^3 + 1\right)\left(v^2 + 1\right). \end{aligned}$$

Expanding the last line, it follows immediately that

$$g_{6,0} = g_{6,3^*} = (1\ 1\ 2\ 3\ 3\ 3\ 3\ 2\ 1\ 1)',$$

the values reported in column g of Table 4.

To derive the indices columns in Table 4, consider any price attained by one or more paths in *B(6, 0)*, and any one of the times at which that price can be attained. Then, consider the twofold convolution describing how those time-index combinations are related to the path sums at time T. A term from this convolution gives a frequency distribution of path sums for paths attaining the given time-price combination. Finally, since a given price can be attained at more than one point in time, the frequency distributions are summed across time to find the frequency distribution of path sums associated with the price index. Calculating these distributions for all attainable prices gives the joint frequency distribution for the bundle. The frequency data are generated column by column, as shown in Appendix II. Effectively, this method circumvents the analytical difficulty that the sum of lognormal variables is not lognormal.

In practice the data of Table 4 are computed using a forward recursion.The manner of constructing the data means they remain the same for all options of the type discussed here, so the valuation problem involves a setup cost that only needs to be incurred once.

4. Valuing Geometric Average Calls

This section values the European and then the American geometric average call.

4.1. Valuing the European Geometric Average Call

European geometric average options can be valued from just columns g and V of arrays like Table 4. Table 5, organized in a fashion similar to Table 4, shows all the data needed to value the European geometric average call. That is, Table 5 displays the frequency distributions for all sub-bundles

$$B(6,\ j,\ V); j \in \{-6,\ -4,\ \ldots,\ 6\};\ V \in \{-21,\ -19,\ \ldots,\ 21\}.$$

As in Table 4, blanks indicate unattainable combinations.

TABLE 5
NUMBERS OF PATHS BY SUB-BUNDLE

V/J	-6	-4	-2	0	2	4	6
-21	1						
-19		1					
-17		1					
-15		1	1				
-13		1	1				
-11		1	2				
-9		1	2	1			
-7			3	1			
-5			2	2			
-3			2	3			
-1			1	3	1		
1			1	3	1		
3				3	2		
5				2	2		
7				1	3		
9				1	2	1	
11					2	1	
13					1	1	
15					1	1	
17						1	
19						1	
21							1
Totals	1	6	15	20	15	6	1

The Table 5 data and the parameters of Table 2 are used to calculate the call payoffs shown in Table 6. For example, the contribution to call value of $B(6,0,5)$ is:

$$2 \cdot \left(1.221403^{5/7} - 1,\ 0\right)^{+} = 0.307130.$$

The 2 is the number of paths in $B(6,0,5)$, 1.221403 is the value of u, 5/7 is the index of the geometric average over the periods 0 through 6, and 1 is the exercise price.

The entries in each column of Table 6 are summed and multiplied by the appropriate probabilities as shown in the Table's last three lines. For example, column 2 adds to 3.489205 and the probability for each of its paths is $p^4q^2 = 0.016246$ when $p = .510051$ and $q = 1 - p$. The third line, the product of sums and probabilities, is summed over all columns and multiplied by R^{-6} to obtain the time 0 discounted call value of 0.121869.

TABLE 6
EVALUATING THE EUROPEAN GEOMETRIC AVERAGE CALL

V\J	−6	−4	−2	0	2	4	6
−21	0.000000						
−19		0.000000					
−17		0.000000					
−15		0.000000	0.000000				
−13		0.000000	0.000000				
−11		0.000000	0.000000				
−9		0.000000	0.000000	0.000000			
−7			0.000000	0.000000			
−5			0.000000	0.000000			
−3			0.000000	0.000000			
−1			0.000000	0.000000	0.000000		
1			0.028984	0.086951	0.028984		
3				0.268485	0.178990		
5				0.307130	0.307130		
7				0.221403	0.664209		
9				0.293230	0.586460	0.293230	
11					0.738563	0.369281	
13					0.449805	0.449805	
15					0.535064	0.535064	
17						0.625336	
19						0.720918	
21							0.822120
Column Sums							
	0.000000	0.000000	0.028984	1.177199	3.489205	2.993634	0.822120
Probabilities							
	0.013833	0.014400	0.014991	0.015606	0.016246	0.016913	0.017607
Expected Values at Time 6							
	0.000000	0.000000	0.000343	0.018371	0.056687	0.050631	0.014475

European Geometric Average Call	0.121869

4.2. Valuing the American Geometric Average Call: Recursions

The American geometric average call is valued using a specialized version of (2.16) that defines recursions between sub-bundles:

$$D_{T-t}(j,\ z) \equiv g_{T-t,j,z} \max\{(G_{T-t}(j,\ z) - K)^+,$$

$$R^{-1}[pD_{T-t+1}(j+1,\ z+j+1)/g_{T-t+1,j+1,z+j+1}+$$

$$+qD_{T-t+1}(j-1,\ z+j-1)/g_{T-t+1,j-1,z+j-1}]\}; \tag{4.1}$$

$$j \in \{-(T-t), -(T-t)+2,\ \ldots,\ T-t\};$$

$$z \in \{V_{T-t,j}\};\ t \in \{0,\ \ldots,\ T\},\ D_{T+1}(\cdot) \equiv 0,$$

where $g_{T-t,j,z}$ is the number of paths in $B(T\text{-}t, j, z)$ and $\{V_{T-t,j}\}$ is the set of values defined by the coefficients of $g_{T-t,j*}$. The recursion relations between sub-bundles can be derived from (2.10). For example, time 6 and time 5 frequency distributions are related by:

$$f_6(x, v) =$$

$$\left(1 + xv^6\right) \sum_{j^*=0}^{5} g_{5,j^*}(v) v^{j^*(j^*+1)/2} x^{j^*}. \tag{4.2}$$

A bundle defined at time 6 combines paths from adjacent end points at time 5. In terms of standard notation $B(6, j^*)$, has a distribution of path sums $g_{6,j*}$ determined by summing the generating function terms $v^6 g_{5,(j-1)*}$ and $g_{5,j*}$. A backward recursion to a bundle at time 5 must employ the relevant path sums and their frequencies taken from adjacent end points at time 6; again cf. (4.2). To perform the backward induction calculations at time 5 for an American option, the time 5 payoffs (with frequency distribution $g_{5,j*}$) are compared with the expected value of the time 6 payoffs (with frequency distributions determined by $v^6 g_{5,j*}$and $g_{5,j*}$ respectively).

To illustrate the recursions using the original indices, consider *B(5, –5, –15)*. This subbundle's single path extends to the single path in *B(6,–6,–21)* if the price decreases between times 5 and 6, to a path in *B(6,–4,–19)* if the price increases. (Remaining paths in *B(6,–4,–19)* are reached from *B(5,–3)*, and form a part of the calculation of expected payoffs for *B(5,–3).)* For *B(5,–5,–15)*, the payoff to holding the option is the expected value of the payoff from proceeding either to $B>$(6, –6, –21) or to *B(6,–4,–19)*. The payoff to immediate exercise for *B(5,–5, –15)* is zero, determined by comparing the geometric average $u^{-15/6}$ to the exercise price of u^0. In this case no further calculation is necessary: the expected value of continuing from *B(5, –5, –15)* cannot be less than the value of immediate exercise, and therefore it is only necessary to record the expected value of continuing. Table 7 shows in greater detail how the frequencies at time 6 are generated from the relevant frequencies at time 5, and thus also shows how time 6 frequencies can be divided to carry out the backward inductions just described.

Sub-bundles can contain many paths, but examining (4. 1) for *T, T-1, ..., 0* shows that each sub-bundle is defined to contain only paths whose payoffs are the same (for geometric average instruments) regardless of time point or nature of optimal policy. (The result is not true for arithmetic average instruments; see Section 5.) Thus bundling methods can be used for valuing both European and American geometric average options. In the latter case, for each of the two time 6 parts of Table 7, a payoff table similar to Table 6 is constructed. The payoff tables for time 6 are then used to construct a table of expected discounted payoffs at time 5, and these are compared to the payoffs for immediate exercise at time 5.

For example, there is one path ending at (6, 4) with a path sum of 19 and one path ending at (6, 6) with a path sum of 21. Both these paths emanated from a single path at (5, 5) with a path sum of 15. Since the payoffs at time 6 are *0.720918* and *0.822120* respectively, the expected discounted payoff at time 5 is

$$.754346 = [(.489949)(.720918) + (.510051)(.822120)]/1.024114$$

TABLE 7

RELATIONS BETWEEN PATH SUMS, TIMES 5 AND 6

Frequencies at time 5

V\J	−5	−3	−1	1	3	5
−15	1					
−13		1				
−11		1				
−9		1	1			
−7		1	1			
−5		1	2			
−3			2	1		
−1			2	1		
1			1	2		
3			1	2		
5				2	1	
7				1	1	
9				1	1	
11					1	
13					1	
15						1

Frequencies at time 6

V\J	−6	−4	−4	−2	−2	0	0	2	2	4	4	6
−21	1											
−19		1										
−17			1									
−15			1	1								
−13			1	1								
−11			1	1	1							
−9			1	1	1	1						
−7				1	2	1						
−5					2	2						
−3					2	2	1					
−1					1	2	1	1				
1					1	1	2	1				
3						1	2	2				
5							2	2				
7							1	2	1			
9							1	1	1	1		
11								1	1	1		
13									1	1		
15									1	1		
17										1		
19											1	
21												1

To allow for early exercise, these expected values are compared to the values of immediate exercise at time 5, and for each comparison the maximum is recorded. In the present example, the payoff to *B(5, 5, 15), 0.648722*, is calculated just as in Table 6. Since *0.648722* < 0.754346, the optimal policy for this sub-bundle is not to exercise, and the value of *0.754346* is recorded in the payoffs to the optimal policy at time 5.

The complete set of time 5 optimal decisions is given in Table 8, where C means it is optimal not to exercise, X means it is optimal to exercise immediately, and 0 means the payoff is zero whether the option is exercised or not. (The zero payoffs are recorded to display the form of the time 5 optimal policy for all time 5 sub-bundles.) Note that while the paths in $B(5, 1, -1)$ have an immediate payoff of zero – their time 5 geometric average is less than the strike price – there is still a positive payoff to continuing, as shown by the C in the position $(-1, 1)$, referring to the payoffs to $B(5, 1, -1)$.

TABLE 8

OPTIMAL DECISIONS AT $t = 5$

V\J	−5	−3	−1	1	3	5
−15	0					
−13		0				
−11		0				
−9		0	0			
−7		0	0			
−5		0	0			
−3			0	0		
−1			0	C		
1			X	C		
3			X	C		
5				C	C	
7				X	C	
9				X	C	
11					C	
13					C	
15						C

To continue with the backward induction, a time 5 frequency distribution, organized as in the second part of Table 7, is used to divide the optimal payoffs at the prices −3, −1, 1, and 3 into payoffs for upward and downward moves. (As before extreme prices are reached in only one way; price -5 by a downward move, price 5 by an upward move.) The backward induction then proceeds from time 5 to time 4, now comparing the discounted expected value of the optimal payoffs at time 5 with the payoffs to immediate exercise at time 4. Continuing the backward induction procedure until time zero is reached, choosing an optimal exercise policy at each time, gives a value for the American call of .126932.

European Geometric Average Call	0.121869
American Geometric Average Call	0.126932

5. Valuing Arithmetic Average Calls

Both European and American arithmetic average calls can be valued approximately using the joint frequency distribution of Table 4. The approximation error can then be estimated, and if it is small enough no further calculation will be needed. If greater accuracy is desired some parts of the joint distribution must be elaborated. Obtaining further detail requires the procedures described in Appendix I, but can be done selectively and typically does not require extensive additional calculations.

5.1. Initial Approximate Solution for the European Arithmetic Average Call

Approximate values of arithmetic average options can be obtained by using the kinds of data reported in the body of Table 4. Each line of Table 4 is be used to find the mean of the arithmetic averages for all paths in a given sub-bundle. The approximation is based on assuming that the arithmetic average for each path in a given sub-bundle is exactly equal to the sub-bundle mean. With this approximation, both European and American arithmetic average instruments can be valued in a manner analogous to that used for geometric average instruments in Section 4. Of course, the assumption introduces approximation error, but the error can be estimated and reduced with relatively few additional calculations as discussed in the next section.

To obtain the approximate value of a European arithmetic average option, the methods of section 4 are adapted as illustrated in Table 10. The only difference between Table 6 and Table 10 is that the latter now contains payoffs determined from the means of sub-bundle arithmetic averages. The analogous payoffs in Table 6 were determined from geometric averages, known to be equal for all paths in any given sub-bundle.

Table 10 shows a positive value for B(6, 2, −1), whereas the corresponding value in Table 6 was zero. The difference reflects the fact that arithmetic averages exceed geometric averages. The path in question is 0-1-2-1012, and has an arithmetic average of 1.003001 for $u = 1.221403$.

European Geometric Average Call	0.121869
American Geometric Average Call	0.126932
European Arithmetic Average Call	0.136520[1]
[1] Approximate Value	

TABLE 10

APPROXIMATE VALUE. EUROPEAN AVERAGE CELL

	-6	-4	-2	0	2	4	6
-21	0.000000						
-19		0.000000					
-17		0.000000					
-15		0.000000	0.000000				
-13		0.000000	0.000000				
-11		0.000000	0.000000				
-9		0.000000	0.000000	0.000000			
-7			0.000000	0.000000			
-5			0.000000	0.000000			
-3			0.000000	0.000000			
-1			0.000000	0.000000	0.003001		
1			0.060526	0.112087	0.050098		
3				0.310133	0.215246		
5				0.330296	0.343032		
7				0.235409	0.734518		
9				0.321225	0.626894	0.368517	
11					0.782972	0.426042	
13					0.477303	0.496303	
15						0.582119	
17						0.686936	
19						0.814960	
21							0.971328
Column Sums							
	0.000000	0.000000	0.060526	1.30915	3.815184	3.374877	0.971328
Probabilities							
	0.013833	0.014400	0.014991	0.015606	0.016246	0.016913	0.017607
Time 6 Expected Values							
	0.000000	0.000000	0.000907	0.020431	0.061983	0.057079	0.017102
Time 0 Approx Value of European Arithmetic Average Call							0.136520

5.2. Assessing and Reducing Approximation Error

Approximation errors can be introduced by the methods of 5.1, because the arithmetic averages of paths in a sub-bundle do not generally equal their mean. The present method could be expected to give good approximations even without additional refinements. First, it can only introduce error in a limited way, as the next section shows in greater detail. Second, the approximation itself should be at least as accurate as that of Curran (1992). The present approximation is actually based on both geometric averages and path end points, whereas Curran's is only based on the former. Some computational experience supporting the claim is given in section 6.1.

More importantly, the approximation can only introduce error for a limited number of sub-bundles. Whenver the maximal path average in a sub-bundle is less than the strike price, the subbundle contributes nothmg to the value of a European call and can be ignored. Whenever the minimal path average in a sub-bundle exceeds the strike price, every path in the sub-bundle contributes to the value of the European call, and using the mean of sub-bundle arithmetic averages introduces no error. Since all paths in a sub-bundle have equal probability, the sum of the individual path averages is n times their mean, where n is the number of paths in the subbundle. Thus the individual paths' contributions to option value are n times the contribution calculated using the sub-bundle mean.

The only sub-bundles for which error can be introduced in a European option are those for which the maximal path average exceeds the strike price and the minimum falls strictly below it. Such sub-bundles (which must have more than a single path) are said to be *cut* by the strike price. The number of sub-bundles which can be cut by the strike price is relatively small, and the subbundles in question can readily be identified; see Neave and Stein (1997) for a method. To eliminate all approximation error, it is necessary to examine the sub-bundles which are actually cut by the strike price, and to correct the approximation calculations for those cases.

To illustrate error estimation and reduction, consider $B(6, 0, -1)$. The aggregate data reported in Table 4 are:

				g	V	M/g
Indices	−1	0	1			
Frequencies	6	12	3	3	−1	6.858864

The value of M in the above extract from Table 4 is found using

$$M = 60u^{-1} + 12u^{0} + 3u^{1} = 20.576593$$

when $u = 1.221403$. Since all three paths in $B(6, 0, -1)$ have the same probability, the mean of the sub-bundle arithmetic averages is $M/3g = 0.979838$. Approximation error could arise if one or more of the paths in $B(6, 0, -1)$ had an arithmetic average in excess of 1, the strike price.

To eliminate approximation error, it is necessary to determine the frequency distribution of distinct arithmetic averages in any sub-bundle which can be cut by the strike price. Using the methods of Appendix I, it can be shown that the maximal path average in $B(6, 0, -1)$ is less than the strike price, which eliminates any need to examine it further. Nevertheless, to illustrate the issues more fully, it is useful to write out the individual paths according to methods outlined in Appendix I:

Indices	−1	0	1	z	V	N
Frequencies	2	4	1	37	−1	6.858864
	2	4	1	25	−1	6.858864
	2	4	1	21	−1	6.858864

In this case the sub-bundle has only one distinct arithmetic average. However, if the sub-bundle had more than one distinct average, and if it were cut by the strike price, only the averages above the price would contribute to call value and the original approximation would have to be corrected. In the present example, checking the remaining sub-bundles shows that no other subbundle can actually be cut by the strike price, and the exact value of the European arithmetic average equals the approximate value determined earlier.

European Geometric Average Call	0.121869
American Geometric Average Call	0.126932
European Arithmetic Average Call	0.136520 [1,2]
[1] Approximate Value [2] Exact Value	

5.3. Initial Approximate Solution for the American Arithmetic Average Call

The approximate value of the American arithmetic average call is obtained by using the methods of 5.2 recursively. A recursion relation identical to (4.1), except in its use of arithmetic averages, is used:

$$D_{T-t}(j,\ z) \equiv g_{T-t,j,z} max\{H_{T-t}(j,\ z) - K)^{+},$$

$$R^{-1}\left[pD_{T-t+1}(j+1,\ z+j+1)/g_{T-t+1,\,j+1,\,z+j+1}+\right.$$

$$\left.+qD_{T-t+1}(j-1,\ z+j-1)/g_{T-t+1,j-1,z-j-1};\right]\} \qquad (5.1)$$

$$j \in \{-(T-t),\ -(T-t)+2,\ \ldots,\ T-t\};\ z \in \{V_{tj}\};\ t \in \{0,\ \ldots,\ T\},$$

where $g_{T-t,j,z}$ is the number of paths in the sub-bundle defined by j and z. Equations (5.1) give an approximate value because they assume that arithmetic averages are equal for all paths in each sub-bundle. Recursive relations between joint frequency distributions are determined using exactly the same methods as in Table 8.

Using the mean value of payoffs for each sub-bundle, backward induction calculations can be performed just as in 3.4. The calculations give an approximate value of *0.14109.3* for the American call.

European Geometric Average Call	0.121869
American Geometric Average Call	0.126932
European Arithmetic Average Call	0.136520[1,2]
American Arithmetic Average Call	0.141093[1]
[1] Approximate Value [2] Exact Value	

5.4. Reducing approximation error

The section 5.3 assumption that all paths in a sub-bundle have equal arithmetic averages can lead to calculating a sub-optimal option value just as with the European option. However, it is possible both to assess the approximation error and to reduce it in much the same way as before.

In the backward induction calculations, the assumption of equal averages is used to divide payoffs according to the number of paths in each sub-bundle. To reduce approximation errors, it is necessary to evaluate which recursive calculations are affected by this approximation. The simplest way to eliminate all approximation error is to divide sub-bundles further on the basis of individual arithmetic averages, and then proceed exactly as in valuing the geometric average options. Unless computing resources are severely limited, this is probably the simplest way to eliminate all approximation error, since experiments indicate the number of divided sub-bundles is roughly described by a fourth-degree polynomial in T.

If the procedure of the foregoing paragraph is not followed, sub-bundles can contain differing arithmetic averages, and care needs to be taken in assessing and reducing the resulting approximation error. (The tradeoff between the two approaches is best assessed in the context of a given valuation problem.) A good rule of thumb is to begin by examining payoffs near the exercise boundary at some time period near $2T/3$, and continue backwards to earlier times if significant errors are detected. Section 5.2 demonstrated the importance of examining payoffs near the exercise boundary; the reason for choosing a time period around $2T/3$ is that typically more exercise decisions are made as option expiry nears.

In the present example, suppose it is desired to find the details of the two paths in $B(5,1,3)$, to check whether the assumption of equal arithmetic averages, which implies dividing payoffs in a 1:1 ratio, gives a nearly optimal value. Using Appendix I, the two time 5 paths are:

$$\begin{array}{cccccc} 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 2 & 1. \end{array}$$

The two paths' arithmetic averages are 1.110702 and 1.125660 respectively, and thus their contribution to value under a policy of immediate exercise is .110702 and .125660 respectively. However, if the two paths are extended to time 6, their expected values are

$$(.165418p + .094887q)/R = .127646$$

for the first path, and

$$(.177884p + .107623q)/R = .140082$$

for the second. Clearly, the optimal policy for $B(5, 1, 3)$ is not to exercise at time 5. Even so, the optimal sub-bundle payoff of

$$.267728 = .127646 + .140082$$

should be divided on the basis of expected values rather than by numbers of paths as in 5.3. In the present example, this is the only refinement to the approximation needed to determine an optimum; all other divisions based on numbers of path are already optimal. Thus, a single modification suffices to obtain the exact American arithmetic average option value of 0.141269. The difference between exact and approximate valuations indicates that before relying on approximations in practice, model-based evaluations of their accuracy should be established.

European Geometric Average Call	0.121869
American Geometric Average Call	0.126932
European Arithmetic Average Call	0.136520[1,2]
American Arithmetic Average Call	0.141093[1] 0.141269[2]
[1] Approximate Value [2] Exact Value	

The literature does not stress the importance of assessing approximations in relation to a model determined optimum. However as evidenced by the approximate and exact values for the American arithmetic average call, the present example indicates that even plausible approximations can create significant valuation errors. More computational experience of the sort described in 6.1 is is needed to determine the likely incidence of errors for the American option. In practice it may prove useful to find an exact solution for a set of typical parameter values and use that value to estimate approximation errors for American instruments when they are valued accordmg to the quick methods of Section 5.3.

6. Extensions and Conclusions

This section sketches computational experience to date and also remarks on how the paper's methods can be extended to valuing other path dependent instruments.

6.1. Computational Experience

While computations using the method are still in the early stages, experience to date is encouraging. The data in Table 11, taken from Neave and Stein (1997), report our results for European arithmetic average calls with relatively large volatilities.

TABLE 11

APPROXIMATE VALUES OF ARITHMETIC AVERAGE OPTION

T	$\sigma = 0.40$	$\sigma = 0.60$	$\sigma = 0.80$	CPU Secs
6	.136520	.184712	.231945	0
12	.137026	.185367	.232823	0
18	.137214	.185685	.233290	2
24	.137322	.185862	.233547	7
30	.137392	.185972	.233710	22
36	.137441	.186046	.233822	52
42	.137476	.186100	.233901	107
48	.137502	.186142	.233963	204

All unstated parameters are the same as in Table 2. Approximation errors are discussed below.

Table 11 suggests that, in the context of discrete models, the present method both increases accuracy and reduces calculation time. With respect to accuracy, Ritchken, Sankarasubramanian and Vijh (1993) use an Edgeworth approximation to value European arithmetic average options, benchmarking their results usmg simulated values. For volatilities of 0.20 and 0.30 respectively, the standard errors in simulations for 16 to 64 periods are on the order of 0.004 to 0.005. For 25 reset points, the relative approximation errors of this paper' s method are 0.0002 and 0.0009 for volatilities of 0.40 and 0.80 respectively (Stein, 1996), and exact valuations can be found with modest amounts of additional calculation.

An examination of Figure 3 suggests the present approximation is also likely to give greater accuracy than that of Hull and White (1993). Our approach approximates arithmetic averages using sub-bundle means, while Hull and White use nonlinear interpolation between arithmetic averages determined by the maximum and minimum path averages in a bundle. Our approach only introduces error in sub-bundles cut by the strike price, whereas nonlinear interpolation can introduce error at a greater number of sub-bundles. Finally, we can estimate and reduce the error created by a sub-bundle's being cut, while Hull and White offer no way of either estimating or reducing the error of their method.

With respect to computation time, Table 11 reports the number of CPU seconds needed to set up and obtain the valuations. In addition to the data reported in Table 11, we have been able to find exact values for European geometric average calls, and approximate values for European arithmetic averages calls, for values of T up to 100. The computation times for these experiments have been about one hour on a SunSparc workstation. Computation times are comparable to recent unpublished work using the Hull and White approximation, but as already mentioned the present method gives greater accuracy. Finally, computation time is independent of the process volatility.

With respect to memory requirements, experiments with the European arithmetic average call, conducted on a SunSparc work station, mdicate the procedure uses 0.7MB (megabytes) of RAM when T = 30, 9.3 MB when T = 60, and 68.8MB when $T = 100$. As rough comparisons, MicroSoft Word '97 uses 2.6MB, Netscape Navigator 3.0 uses 4.5MB. Personal computers with 64MB of RAM are now standard, and some work stations offer up to 110MB.

Additional experiments are needed to assess the approach's accuracy and memory requirements in valuing the American arithmetic average option. Nevertheless, the framework organizes and reduces the numbers of computations in new ways, and also permits comparing approximations with exact optima for the same problem.

6.2. Time Weighted Averages

The methods developed above can readily be modified to value instruments whose averages are computed on a subset of the time points. For example, if arithmetic averages are computed on a subset of time points, the joint frequency distributions used in this paper need only be modified to record the frequencies with which indices are realized at chosen reset points. They can also be modified relatively easily to value instruments with time weighted averages. The approach can be extended to average strike options by determining joint distributions of the averages and path ends, readily available from the information developed in this paper.

6.3. Path Sums and Time Dependent Probabilities

Since the present model uses a constant value of u, valuation under a martingale with time varying interest rates requires using time dependent probabilities. Given time dependent probabilities, exact values can be found recursively, but the calculations are exponential in T. The task can be simplified by using the generating function to define a joint frequency distribution of path sums and time dependent probabilities, using a procedure much like that of Appendix II. Then depending on the relations between probabilities at each point in time, the difference between maximal and minimal probabilities for the paths in a sub-bundle can be assessed. If the difference is unimportant for the problem at hand, an average path probability can be used; otherwise individual probabilities need to be enumerated using methods similar to those outlined above.

6.4. Conclusions

This paper valued European and American fixed strike average calls on a discrete time, recombining multiplicative binomial asset price process. Using generating functions to find frequency distributions of option payoffs, the paper showed how to eliminate much of the calculation previously thought to be involved in valuing path dependent options. The procedures value European geometric average options analytically, and use relatively few computations to value European arithmetic average options. Both types' American counterparts are valued using recursive relations between frequency distributions.

ACKNOWLEDGEMENTS

Versions of this paper have been presented at the International Meeting on Risk Theory, Mathematisches Forschungsinstitut, Oberwolfach, Germany, 1994; the 5th AFIR International Colloquium, Brussels, 1995; and at the Northern Finance Association Meetings, London, Ont., 1995. I am grateful to Phelim Boyle, Freddy Delbaen, Daniel Dufresne, Helyette Geman, Anders Martin-Lof, Jim Tilley and Stuart Turnbull for constructive comments. David Arthurs showed me how to create Figure 3 using Excel.

REFERENCES

BERMAN, GERALD, and K. D. FRYER, Introduction to Combinatorics, New York: Addison-Wesley, 1972.

CARVERHILL, A., and L. CLEWLOW, 1990, "Flexible Convolution," Risk 3, Apr. 25-29.

CHEN, CHUANG-CHONG and KOH KHEE-MENG, Principles and Techniques in Combinatorics, Singapore: World Scientific Publishing, 1992.

CURRAN, M., "Beyond Average Intelligence. " Risk, November 1992, p. 60.

DIXIT, AVINASH K. and ROBERT S. PINDYCK, Investment Decisions Under Uncertainty, Princeton: Princeton University Press, 1994.

FELLER, WILLIAM, An Introduction to Probability Theory and its Applications, 2nd Edition, New York: Wiley, 1957.

GEMAN, HELYETTE, and A. EYDELAND, "Domino Effect," Risk, 1995, 65-67.

GEMAN, HELYETTE, and MARC YOR, "Bessel Processes, Asian Options, and Perpetuities," Mathematical Finance 3, 1993, 349 - 375.

GOLDMAN, M. Barry, HOWARD B. SOSM and M. GATTO, "Path Dependent Options: Buy at the Low, Sell at the High," Journal of Finance 34, 1979, 1111-1127.

HOO, Thomas S. Y., "Managing Illiquid Bonds and the Linear Path Space," Journal of Fixed Income, June 1992, 80-94.

HUANG, CHI-FU, and ROBERT H. LITZENBERGER, Foundations for Financial Economics, New York: North-Holland, 1988.

HULL, JOHN, and ALAN WHITE, "Efficient Procedures for Valuing European and American Path-Dependent Options," The Journal of Derivatives, Fall 1993, 21-31.

JARROW, ROBERT, and STUART M. TURNBULL, Derivative Securities, Cincinnati: South-Western Publishing Company, 1996.

KEMNA, A. and VORST, TON "A Pricing Method for Options Based on Average Asset Values," Journal of Banking and Finance, 1990.

LEVY, EDMOND, "A Note on Pricing European Average Options," Nomura Bank International Working Paper, 1991.

LEVY, EDMOND, and STUART M. TURNBULL, "Average Intelli.gence," Risk Magazine, 1992.

NEAVE, EDWM H., "Valuing European Average Options on a Recombining Binomial Random Walk," Actuarial Approach For Financial Risks, vol. I, pp. 153-180, Third AFIR International Colloquium, Rome, 1993.

NEAVE, EDWIN H., and CHRISTOPHER L. STEIN "Approximate and Exact Valuation of European Arithmetic Average Spot Price Calls," Queen's University Working Paper, 1997.

RITCHKEN, P., L. SANKARASUBRAMANIAN, and A. VIJH, "The Valuation of Path Dependent Contracts on the Average," Management Science 39 (10), 1993, 1202-1213.

ROGERS, L. C. G. and SHI, Z., "The Value of an Asian Option," Journal of Applied Probability, 1995.

SCHWARTZ, EDUARDO S., Review of Dixit and Pindyck, Investment Under Uncertainty, Journal of Finance 49, 1994, 1924-1928.

STEIN, CHRISTOPHER L., "Computing the Exact Value of European Arithmetic Average Spot Price Calls," Queen's University Working Paper, 1996.

TILLEY, JAMES A., "Valuing American Options in a Path Simulation Model," Actuarial Approach For Financial Risks, vol. 1, pp. 249-266, Third AFIR International Colloquium, Rome, 1993.

TURNBULL, STUART M. and LEE MACDONALD WAKEMAN, "A Quick Algorithm for Pricing European Average Options," Journal of Financial and Quantitative Analysis 26, 1991, 377-389.

YOR, M., "On Some Exponential Functions of Brownian Motion," Advances in Applied Probability 24, 1992, 509-531.

ZHANG, P., "Flexible Arithmetic Asian Options," The Journal of Derivatives, Spring 1995, 53-63.

Appendix I. Relating Z and V*

The generating function can be used to study both the mapping from Z to V^* and the inverse mapping from V^* to Z. Consider the term of $f_4(x, 7, v)$ equal to $x^4z^{43}v^{13}$, which describes the path defined by $z=43$, ending price j^*, and path sum $v^* = 13$.

Suppose it is known that $z = 43$. The value of j^* can be determined by expressing 43 as a sum of powers of 2, and counting the number of terms. That is,

$$43 = 2^5 + 2^3 + 2^1 + 2^0$$

so that $j^*=4$. Since Z is defined as a sum of terms 2^{j-1} while v^* is determined by a sum of indices j,

$$v^* = 6 + 4 + 2 + 1 = 13$$

To study the inverse mapping from V^* to Z, let the values of j^* and v^* be given. To continue the previous example, if $j^*=4$ and $v^*=13$, then from (2.3) v^* must be the sum of four integers chosen from $1, \ldots, 6$. There are only two such combinations; either the foregoing or

$$v^* = 5 + 4 + 3 + 1 = 13$$

for which $Z=29$. The two paths in the sub-bundle $B(6,2,5)$ are thus

0101012 and *0-101212*

The maximal arithmetic average in any sub-bundle is defined by one of the extremal values of Z associated with the sub-bundle. Moreover, the maximal arithmetic average increases as the term V defining the sub-bundles increases. Finally, a minimal path average can be characterized in terms of Z. However the geometric average is also a lower bound on the arithmetic averages, and is in any case recorded as a part of the valuation method.

Appendix II. Finding the Joint Frequency Distribution: Example

This Appendix develops an analytic method of finding joint frequency distributions of indices and path sums. While in practice it is usually convenient to calculate the joint frequency distributions recursively, the analytic approach of this Appendix makes it possible to organize the computations efficiently. The joint frequency distribution is obtained analytically using two-fold convolutions of (2.17) taking the form $f^3(x, v) * f^{T-5}(x, v)$. A term of the convolved functions can be interpreted as follows. Consider any feasible index in $B(T, j)$, say (s, k). The number of paths through (s, k) is readily shown to be $b(s, k) \cdot b(T - s, j - k)$, where $b(T, j) \equiv T!/(T - j)!j!$. From (2.16) the distribution of path sums at (s, k) is $g_{s,k}(v) v^{k(k+1)/2} x^k$. Since any path in in $B(T, j)$ arriving at index (s, k) must still take T-s steps, the distribution of path sums at (t, j) is

$$v^{k(T-s)} \cdot g_{s,k}(v) v^{k(k+1)/2} x^k \cdot g_{T-s,j-k}(v) v^{(j-k)(j-k+1)/2} x^{j-k} = \tag{A.1}$$

$$v^{k(T-s)+[k(k+1)+(j_k)(j-k+1)]/2} x^j g_{s,k}(v) \cdot g_{T-s,j-k}(v)$$

The term $v^{k(T-s)}$ compensates for the fact that the remaining T-s steps begin at (s, k), while $g_{T-s,j-k}(v)$ begins its counting from $(0, 0)$.

To illustrate the calculations, Table A.I repeats the joint frequency distribution of path sums and indices realized reported in Table 4 for $B(6, 3^*)$. Blanks indicate combinations which cannot be realized by paths in $B(6, 3^*)$.

TABLE A.I

Joint Frequency Distribution for B(6, 3*)

V\j	−3	−2	−1	0	1	2	3	Row Totals
−9	1	2	2	2				7
−7		2	3	2				7
−5		2	6	6				14
−3		2	7	10	2			21
−1			6	12	3			21
1			3	12	6			21
3			2	10	7	2		21
5				6	6	2		14
7				2	3	2		7
9				2	2	2	1	7
	1	8	29	64	29	8	1	

Note that the row totals equal the product of the 7 indices in each path and the path frequencies. The second section of Table A. I supplements the column totals at the bottom of the first section by showing the frequencies with which individual indices are realized at different times. As before, blanks represent unattainable combinations.

Terms of the two-fold convolutions are used to calculate the joint frequency distributions are employed for each time-index combination, as shown in the detailed calculations of Table A.II. Each column of Table A.II represents a time-index combination; for example, the index -2 can be realized at time 2 or at time 4. These two columns then indicate the frequency distributions of path sums at time 6 for paths attaining the index -2 at either time 2 or time 4. Similarly, the index -1 can be realized at times 1, 3, and 5. The convolution describes only a single frequency distribution at times 1 and 5, but three at time 3. This is because paths arriving at index -1, time 3 can have three values at that point, and each path from that point to the end can also take on any one of three incremental values. The positioning of the frequencies within the columns is determined by the range of path sums, as described both in Section 3.2 and at the beginning of this Appendix.

TABLE A.II

Obtaining the Joint Frequency Distribution

Indices	−3	−2	−2	−1	−1	−1	−1	−1	0	0	0	0	0	0
Times	3	2	4	1	3	3	3	5	0	2	2	4	4	6
V														
−9	1	1	1	1				1	1					1
−7		1	1	1	1			1	1					1
−5		1	1	2	1	1		2	2	1		1		2
−3		1	1	2	1	1	1	2	3	1	1	1	1	3
−1				2		1	1	2	3	2	1	2	1	3
1				1			1	1	3	1	2	1	2	3
3				1				1	3	1	1	1	1	3
5									2		1		1	2
7									1					1
9									1					1
	1	4	4	10	3	3	3	10	20	6	6	6	6	20

Appendix III: Finding distinct arithmetic averages in a sub-bundle

To see how the distinct arithmetic averages in a sub-bundle can be found, consider the subbundles (8, 3, 15), (8, 3, 14), (8, 3, 13), and (8, 3, 12). Each sub-bundle contains six paths, as shown by the path numbers m the following rows:

8 6 1	8 5 2	8 4 3	7 6 2	7 5 3	6 5 4
8 5 1	8 4 2	7 6 1	7 5 2	7 4 3	6 5 3
8 4 1	8 3 2	7 5 1	7 4 2	6 5 2	6 4 3
8 3 1	7 4 1	7 3 2	6 5 1	6 4 2	5 4 3

The arithmetic averages for the foregoing paths are shown next for $\sigma = 0.40$.

0.9056653	0.8909149	0.8909149	0.8909149	0.8909149	0.8828047
0.8609239	0.8609239	0.8542838	0.8461735	0.8461735	0.8461735
0.8242928	0.8242928	0.8161825	0.8095424	0.8095424	0.8095424
0.7943017	0.7795514	0.7795514	0.7795514	0.7795514	0.7729113

The example shows the need, when exact valuation is desired, for carefully investigating any particular sub-bundles cut by the strike price. In the present each sub-bundle has exactly three distinct averages, but the frequency distributions of the three distinct averages vary. Thus, if one or more of these sub-bundles were cut by the strike price, the valuation effect would depend on the particular sub-bundle or sub-bundles affected. So far, it seems necessary to determine the frequency distribution of the distinct averages in any such sub-bundle.

Frequency distributions of distinct averages can be found either by enumerating the subbundle's path numbers or by using a dynamic programming search to find the distinct path averages, then determining the frequency of each distinct average using linear programming. In large subbundles, the second method is more efficient than complete enumeration, because the number of paths can be large while the number of distinct averages is very much less than the number of paths.

RELATIVE REINSURANCE RETENTION LEVELS

DAVID C.M. DICKSON
University of Melbourne

HOWARD R. WATERS
Heriot-Watt University
Edinburgh

ABSTRACT

The problem of determining optimal retention levels for a non-life portfolio consisting of a number of independent sub-portfolios was first discussed by de Finetti (1940). He considered retention levels as optimal if they minimised the variance of the insurer's profit from the portfolio subject to the constraint of a fixed level of expected profit. In this paper we consider a similar problem, changing the criterion for optimality to minimising the probability of ruin, either in discrete or continuous time. We investigate this problem through a series of case studies based on a real portfolio.

KEYWORDS

Reinsurance; optimal retention levels; finite time ruin; translated gamma process.

1. INTRODUCTION

This paper is a risk-theoretic discussion of the problem of determining the relative reinsurance retention levels for a non-life portfolio consisting of a number of independent sub-portfolios. We consider only simple forms of proportional and excess loss reinsurance. Our discussion will be based largely on numerical results derived from a "pseudo-real" portfolio. The characteristics and construction of this portfolio are described in detail in Section 2 below.

The classical results in this area are due to de Finetti (1940) (see also Bühlmann (1970, *section* 5.2)). De Finetti derived relative retention levels which have simple forms by considering the insurer's net (of reinsurance) profit from the portfolio at the end of a given time period. He then minimised the variance of this profit subject to its expected value being fixed. A summary of de Finetti's results is given in Section 3 below.

In Section 4 we discuss some alternative criteria for determining relative retention levels. These alternatives are to minimise the insurer's probability of ruin over a finite time horizon, either in continuous or in discrete time. Questions of interest to us are:

(1) Do some or all of our probability of ruin criteria produce relative retention levels close to those given by de Finetti's approach?
(2) Are the relative retention levels produced by a probability of ruin in continuous time criterion close to those produced by a discrete time criterion?
(3) How do the relative retention levels produced by our probability of ruin criteria depend on:
   (i) the insurer's expected net profit?
   (ii) the time horizon for ruin?
   (iii) the insurer's initial surplus?

These questions are investigated in Section 5 (proportional reinsurance) and Section 6 (excess loss reinsurance). Our conclusions are set out in Section 7.

2. THE PORTFOLIO

In order to investigate the problems outlined in the previous section, we have constructed a non-life insurance portfolio based on a study by Ramlau-Hansen of data supplied by a Danish insurance company. Ramlau-Hansen's work is detailed in a series of working papers (1986a, 1986b, 1986c and 1986d) and a conference paper (1983) and summarised in two papers (1988a and 1988b).

Ramlau-Hansen analysed data from the Nye Danske Lloyd insurance company covering the period 1977 to 1981. The data related to policies on:
- single-family houses, and,
- dwellings (mainly apartment buildings, but also some office buildings).

These policies covered the buildings, but not their contents, against:
- glass damage, i.e. damage to windows and sanitary fittings,
- fire damage, and,
- windstorm damage.

Claims from these three sources will have very different characteristics:
- Glass claims: these will be relatively numerous but for rather small amounts.
- Fire claims: these will be far less frequent than glass claims but will be for far greater amounts.
- Windstorm claims: the number of windstorms will be very small but each windstorm will produce a large number of individual claims.

In terms of claims experience, we would expect glass claims to be relatively stable, fire claims to be less stable and windstorm claims to be even less stable over time.

Our portfolio is based on Ramlau-Hansen's "Standard Portfolio" (1986d, section 4.3). It consists of three sub-portfolios covering glass, fire and windstorm claims, each of which can be reinsured separately. However, within each subportfolio, single-family houses and dwellings cannot be reinsured separately. The total annual expected claim amount, before reinsurance, is 500×10^6 of which 25% (125×10^6) is expected to come from glass claims, 70% (350×10^6) from fire claims and the remaining 5% (25×10^6) from windstorm claims. (Ramlau-Hansen's monetary unit was Danish Kroner at 1981 values. For our purposes only relative monetary values are important, not absolute values.)

Ramlau-Hansen modelled in some detail the annual claim numbers and amounts distributions for each sub-portfolio. We have adopted Ramlau-Hansen's models for our portfolio with some very minor simplifications. Our models are as follows:

Glass claims: Since glass claim amounts are almost always relatively small, we have assumed that this sub-portfolio would not be reinsured under an excess loss treaty, but would be reinsured under a proportional reinsurance treaty. (This agrees with Ramlau-Hansen's study (1988b, section 3.2).) For this reason we need to specify a model for the aggregate annual glass claims but not for claim numbers and claim amounts separately. We have assumed that the aggregate annual glass claims have a normal distribution. This is a slight simplification of Ramlau-Hansen's model but his analysis (1986a, Table 12) does show that the skewness of aggregate annual glass claims is very small. The expected aggregate annual glass claims are 125×10^6, as explained above, and we have taken the standard deviation to be 4.3×10^6. The standard deviation has been inferred from the information given by Ramlau-Hansen (1986a, Table 14).

Fire claims: The annual fire claim rate for dwellings is about 0.0885. (See Ramlau-Hansen (1986b, Tables 1 and 2).) The annual fire claim rate for single-family houses is 0.0127. (See Ramlau-Hansen (1983, Tables 1 and 7).) In 1981, the numbers of dwellings and single-family houses in Ramlau-Hansen's data were 12,318 and 83,699, respectively. These figures indicate that the expected number of claims each year is approximately the same for dwellings and single-family houses. Ramlau-Hansen (1988a, section 2.1) assumes claim numbers have a Poisson distribution. We have assumed the Poisson parameter for dwellings and for single-family houses is 7,893.9. (This value, when combined with the claim amount distributions specified below, gives a mean aggregate annual fire claim amount of 350×10^6, as required.)

We use different claim amount distributions for dwellings and for single-family houses. In each case, the distribution is loggamma, truncated at an expected maximum loss (EML), with a density function of the form:

$$f(x;\alpha,\gamma) = \frac{\alpha^\gamma}{\Gamma(\gamma)}\frac{1}{x_0}(\log(x/x_0))^{\gamma-1}(x/x_0)^{-(\alpha+1)} \quad \text{for } x_0 < x < EML$$

where in each case the lower limit x_0 is 100. The other parameters and the resulting moments are:

	Dwellings	Single-family houses
EML	35×10^6	402,500
α	1.4177	1.1220
γ	5.1003	3.2477
Mean	33,611	10,727
St. Dev.	490,721	42,560
Skewness	51.64	7.338

Ramlau-Hansen (1988a, section 2.2 and 1983, section 3) uses parameter values which depend on the floor area of the dwelling or house. We have selected a "typical" distribution for each type of property. Let $F(x;\ \alpha,\ \gamma)$ denote the distribution function corresponding to the density function $f(x;\ \alpha,\ \gamma)$. Then the aggregate annual fire claims have a compound Poisson distribution with Poisson parameter 15,787.8 and individual claim amount distribution $F(x)$, where:

$F(x) = 0 \quad \text{for } x < 100$

$F(x) = (F(x;\ 1.4177, 5.1003) + F(x;\ 1.1220, 3.2477))/2 \quad \text{for } 100 \le x < 402{,}500$

$F(x) = (1 + F(x;\ 1.4177, 5.1003))/2 \quad \text{for } 402{,}500 \le x < 35 \times 10^6$

$F(x) = 1 \quad \text{for } x \ge 35 \times 10^6$

For our model, aggregate annual fire claims have the following moments:

Mean	350×10^6
St. Dev.	43.875×10^6
Skewness	0.571

Windstorm claims: Ramlau-Hansen (1988a) developed a complicated model for windstorms. He modelled the number of storms per annum, the number of claims from each storm and the amount of the individual claims. For the purposes of proportional reinsurance we need model only the aggregate annual windstorm claims. When we consider excess loss reinsurance, we shall assume the insurer protects the windstorm (sub-)portfolio with a catastrophe excess loss treaty whereby the reinsurer reimburses the insurer for the amount by which the total claim amount caused by a storm exceeds a given retention. See Ramlau-Hansen (1986c, p. 42). This means that we need model only the annual number of windstorms and the total claim amount from each windstorm.

The number of windstorms per annum (in Denmark) in Ramlau-Hansen's model has a Poisson distribution with mean 4.36 and the expected cost of a single windstorm is 9.3×10^6. Since we require the expected aggregate annual cost of windstorms to be 25×10^6, we need to scale down either the expected number of windstorms or the expected cost of a single windstorm. We decided to do the latter, which is equivalent to an insurer (in Denmark) having fewer windstorm policies than in Ramlau-Hansen's portfolio.

Our model for windstorm claims is as follows:

The number of storms per annum has a Poisson distribution with mean 4.36.

The total claim amount from a single windstorm has the following moments:

Mean	5.734×10^6
St. Dev.	13.14×10^6
Skewness	2.649

We have assumed that the total claim amount from a single windstorm has a translated gamma distribution with the above moments, i.e. has the distribution of $\kappa + Y$, where Y has a $\Gamma(\alpha,\beta)$ distribution. The parameters of this distribution are:

$$\alpha = 0.5700$$
$$\beta = 5.746 \times 10^{-8}$$
$$\kappa = -4.187 \times 10^6$$

This model gives the following moments for the aggregate annual claims from windstorms:

Mean	25×10^6
St. Dev.	29.936×10^6
Skewness	1.49

Following Ramlau-Hansen, we assume that all random variables in our model are independent unless specified otherwise, so that, for example, aggregate claims from the three sub-portfolios are independent and aggregate claims in separate years are independent. In addition, we assume that the distributions do not change from year to year. It would not be difficult to relax this assumption, for example by incorporating inflation and business growth, but this would complicate the presentation without adding significantly to the study.

For the remainder of the paper we will work in units of one million, so that the expected aggregate annual claim amount from the portfolio is 500.

3. A REVIEW OF DE FINETTI'S RESULTS

This section contains a brief summary of the essential points of de Finetti's results. More details, and proofs, can be found in de Finetti (1940) (see also Bühlmann (1970)). The basic idea underlying these results is as follows. An insurer has a portfolio on n independent risks and wishes to effect the same type of reinsurance for each risk. The insurer's profit level from these risks clearly depends on the level of reinsurance. The insurer fixes a level for its expected profit from the portfolio over a given time period, say one year, and chooses retention levels to minimise the variance of the profit from the portfolio over this period. De Finetti's results state how retention levels for proportional and excess loss reinsurance should be calculated under this criterion, which we shall refer to as the minimum variance criterion.

Consider first proportional reinsurance. For a portfolio of n independent risks, let S_i denote aggregate claims from the ith risk in a fixed time period for $i = 1, 2, \ldots, n$, and let P_i denote the premium received by the insurer to cover this risk. The insurer effects proportional reinsurance for each risk with proportion a_i retained for the ith risk, paying a reinsurance premium of $(1 + \theta_i)(1 - a_i)E(S_i)$ for this reinsurance cover. Thus, the reinsurance premium is calculated by the expected value principle with a loading θ_i for the ith risk. The insurer's profit over the period is

$$Z(\underline{a}) = \sum_{i=1}^{n} (P_i - (1+\theta_i)(1-a_i)E(S_i) - a_i S_i)$$

Subject to the constraint $E[Z(\underline{a})] = k$, where k is a constant, $V[Z(\underline{a})]$ is minimised by

$$a_i = \frac{c\theta_i E(S_i)}{V(S_i)} \quad \text{for } i = 1,2,\ldots,n$$

where c is a constant which is determined by the condition $E[Z(\underline{a})] = k$. If this procedure produces a value of $a_i > 1$, the solution is to set that value of a_i equal to 1, with the remaining retentions being of the above form.

In the case of excess loss reinsurance, let S_i and P_i have the same meaning as above. We assume that each S_i has a compound Poisson distribution. The insurer effects excess loss reinsurance with retention level M_i for the ith risk and pays a reinsurance premium of $(1 + \theta_i)\,E(S_i - S_i^I)$ where S_i^I denotes the insurer's aggregate retained claim amount from the ith risk. The insurer's profit over the period is

$$Z(\underline{M}) = \sum_{i=1}^{n} (P_i - (1+\theta_i)E(S_i - S_i^I) - S_i^I)$$

Subject to the constraint $E[Z(\underline{M})] = k$, where k is a constant, $V[Z(\underline{M})]$ is minimised by

$$M_i = c\theta_i \quad \text{for } i = 1, 2, \ldots, n$$

where c is a constant which is determined by the condition $E[Z(\underline{M})] = k$.

Tables 1 and 2 show optimal retention levels for the portfolio described in Section 2 for proportional and excess loss reinsurance respectively. In the case of proportional reinsurance, the loadings in the reinsurance premiums are 10% (glass), 40% (fire) and 80% (windstorm), while for excess loss reinsurance they are 40% (fire) and 80% (windstorm). The tables also show the mean and variance of the insurer's retained aggregate claims. We can see in each case that these quantities increase as the expected net profit increases. We note that for each level of expected net profit, the values of mean retained aggregate claim amounts under each type of reinsurance are similar. However, for a given level of expected net profit, the variance of the retained aggregate claim amount is considerably smaller under excess loss reinsurance. For example, when the expected net profit is 90, a reduction of just 10 from its maximum value, the variance of the insurer's retained aggregate claim amount can be reduced by 44% using excess loss reinsurance, compared to a reduction of only 24% using proportional reinsurance.

TABLE 1

OPTIMAL RETENTIONS – PROPORTIONAL REINSURANCE

Expected Net Profit	*Glass Retention*	*Fire Retention*	*Windstorm Retention*	*Mean*	*Variance*
50	1	0.753	0.231	394	1,157
60	1	0.821	0.252	419	1,373
70	1	0.890	0.273	443	1,609
80	1	0.958	0.294	468	1,863
90	1	1	0.5	488	2,168
100	1	1	1	500	2,840

Note that in the case of proportional reinsurance, there is in fact no reinsurance for the glass sub-portfolio, nor for the fire sub-portfolio as the expected net profit increases. In all other cases in Table 1, the retentions for the fire and windstorm portfolios are in the same proportion. In Table 2, the retention levels for windstorm claims are twice those for fire claims since the reinsurance premium loading factors are in the ratio 2:1.

TABLE 2

OPTIMAL RETENTIONS – EXCESS LOSS REINSURANCE

Expected Net Profit	*Fire Retention*	*Windstorm Retention*	*Mean*	*Variance*
50	2.08	4.15	397	213
60	3.55	7.09	418	351
70	5.86	11.72	438	582
80	9.66	19.32	458	961
90	16.88	33.77	478	1,602
100	∞	∞	500	2,840

Thus, de Finetti's results provide simple formulae from which optimal retention levels can be calculated. In the case of proportional reinsurance, the optimal retention levels depend on the first two moments of aggregate claims from each sub-portfolio. This is perhaps not surprising since the problem is specified in terms of the first two moments of profit from the n sub-portfolios considered together. In the case of excess loss reinsurance, the optimal retention level for each sub-portfolio depends only on the reinsurer's loading for that sub-portfolio. An interesting feature of this result is that the distribution of individual claims for a sub-portfolio has no bearing whatsoever on the retention level.

The results are independent of the insurer's premium income (before reinsurance) and of the amount of the insurer's surplus. Intuitively we would expect these factors to play a part. We also note that these results hold for a single period analysis. If we assume that claims in successive time periods are independent, then a change in the time period considered does not alter the optimal retention levels.

Finally, we note that if the optimality criterion is altered from minimising $V[Z(\underline{b})]$ subject to the constraint $E[Z(\underline{b})] = k$ (where $\underline{b}$ denotes the vector of retention levels) to minimising $V[Z(\underline{b})]$ subject to the constraint $E[Z(\underline{b})] \geq k$ then it is not difficult to prove that the solution to the problem is unchanged. In our case studies in Sections 5 and 6, where we apply different criteria for optimality, we will see that a change in the constraint from $E[Z(\underline{b})] = k$ to $E[Z(\underline{b})] \geq k$ can make a considerable difference.

4. AN ALTERNATIVE CRITERION FOR OPTIMALITY

In this section we consider an alternative criterion for optimality. We will consider a vector of retention levels to be optimal if those retentions minimise the insurer's probability of ruin (net of reinsurance) subject to the constraint that the insurer's expected profit per unit time is greater than or equal to some constant. Thus we have not only changed the objective function from variance of profit to probability of ruin, but we have also altered the constraint. It will be clear in the examples in the next sections why it is sensible to do this. In our examples we will consider finite time ruin, both in discrete and in continuous time.

Since the probability of ruin depends on all the characteristics of the surplus process, we might expect this new criterion to produce different optimal retention levels to those produced by the minimum variance criterion. However, the following examples suggest that this new criterion may not produce very different results.

Example 1: It is well-known that if the adjustment coefficient, denoted R, for a risk exists, it can be approximated as

$$R \approx \frac{2 \times \text{Expected Profit}}{\text{Variance of Profit}}$$

Let us treat profit in this approximation as being the net of reinsurance profit from a portfolio of risks over a fixed time period. A natural (and approximate) way of obtaining retention levels to minimise the insurer's probability of ultimate ruin would be to find retention levels that maximise this approximation to R. When we apply the constraint that the expected profit is constant, maximising R is equivalent to minimising the variance of profit, i.e. minimising the variance of net retained claims.

Example 2: Suppose that an insurer has a portfolio of n risks and receives a total premium of P per annum to cover these risks. Suppose further that the insurer effects some form of reinsurance for each of these risks, defined by a vector $\underline{b}$ of retention levels. Let $\Pi(\underline{b})$ denote the total premium paid by the insurer for this reinsurance, and let $S_n(\underline{b})$ denote the aggregate claims, net of reinsurance, paid by the insurer up to time n. Finally, let U denote the insurer's initial surplus.

We assume that the insurer's expected net profit per unit time, $P - \Pi(\underline{b}) - [S_1(\underline{b})]$, is positive. Assuming that $S_n(\underline{b})$ has a normal distribution, and that aggregate claims are independent and identically distributed from year to year, the insurer's probability of ruin at the end of n years is

$$1 - \Phi\left(\frac{nP - n\Pi(\underline{b}) - nE(S_1(\underline{b})) + U}{[nV(S_1(\underline{b}))]^{1/2}}\right)$$

where Φ denotes the standard normal distribution function. Minimising this probability of ruin (as a function of $\underline{b}$) subject to the insurer's expected net profit per unit time being fixed is equivalent to minimising the variance of the insurer's net profit per unit time subject to the same constraint.

Example 3: Now let us extend the previous example by assuming in addition that the insurer's aggregate gain process $\{G_t(\underline{b})\}_{t\geq 0}$ is a Brownian motion with (positive) drift. Let $\Psi(U,T\,|\,\underline{b})$ denote the probability of ruin in continuous time before time T, which may be finite or infinite. Let $\underline{b}_1$ and $\underline{b}_2$ be two reinsurance retention vectors which result in the same expected net profit for the insurer, say μ per unit time, but different variances. Then using a coupling argument, i.e. regarding $G_t(\underline{b}_1)$ as equivalent to

$$\mu t + (G_t(\underline{b}_2) - \mu t)(V[G_t(\underline{b}_1)] / V[G_t(\underline{b}2)])^{1/2}$$

it is easy to see that $\Psi(U,T\,|\,\underline{b}_1) > \Psi(U,T\,|\,\underline{b}_2)$ is equivalent to $V[G_t(\underline{b}_1)] > V[G_t(\underline{b}_2)]$. Hence, minimising the probability of ruin in continuous and finite or infinite time subject to the insurer's expected net profit per unit time being fixed is equivalent to minimising the variance of the insurer's net profit subject to the same constraint.

Each of these last two examples relies on being prepared to approximate the insurer's net surplus process by a process determined by just its mean and variance (see, for example, Grandell (1977)). They also apply the constraint that the expected net profit equals some constant, rather than is greater than or equal to that constant. Nevertheless, they suggest that a change in the optimality criterion from minimising variance to minimising a ruin probability may not result in very different retention levels. We shall see in Sections 5 and 6 that this can be the case, although we shall also see that the change in optimality criterion can lead to very different results.

Since our new optimality criterion is to minimise a probability of ruin, we need to be able to calculate ruin probabilities. Our approach to this problem will not be to attempt to calculate exact ruin probabilities. Rather, we will use an approximation. We will approximate the retained aggregate claims process by a translated gamma process. There are two reasons for using this approximation. First, formulae exist from which ruin probabilities can be calculated. Second, recent evidence shows that this approach provides very good approximations to ruin probabilities, particularly in problems involving reinsurance. See Dickson and Waters (1993 and 1996).

We conclude this section by describing how we calculated ruin probabilities. Consider first the discrete time ruin problem. We require probabilities of the form

$$\Psi_1(u,t) = \Pr(u + Pn - X_n < 0 \quad \text{for some } n,\ n = 1,2,\ldots,t)$$

where P represents the insurer's premium income, net of reinsurance, per unit time, and X_n denotes aggregate claims up to time n, again net of reinsurance. We approximated X_n by $Y_n + kn$ where Y_n has a gamma distribution with parameters $n\alpha$ and β and calculated probabilities from

$$\Psi_1^*(u,t) = \Pr(u + P^*n - Y_n < 0 \quad \text{for some } n,\ n = 1,2,\ldots,t)$$

where $P^* = P - k$. The parameters α, β and k are found by matching the first three moments of X_n and $Y_n + kn$. Let $G(x)$ and $g(x)$ respectively denote the distribution function and density function of a gamma distribution with parameters α and β, so that the mean of the distribution is α/β. Then

$$\Psi_1^*(u,1) = 1 - G(u + P^*)$$

and for $t = 1, 2, 3, \ldots$

$$\Psi_1^*(u,t+1) = \Psi_1^*(u,t) + \int_0^{u+P^*} \Psi_1^*(x,t)g(u + P^* - x)dx$$

Values of $\Psi_1^*(u, 1)$ were calculated directly from computer routines which compute the gamma distribution function. Values of $\Psi_1^*(u, t)$ for $t > 1$ were calculated by numerical integration. For each value of u required we performed numerical integration on the interval $(0, [u + P^*])$, where $[u + P^*]$ denotes the greatest integer less than or equal to $u + P^*$, by applying the repeated trapezoidal rule on unit steps. The integral over the range $([u + P^*], u + P^*)$ was calculated by the trapezoidal rule. Thus, except for the integral over the final part of the range, $\Psi_1^*(x, t)$ values were required only for integer values of x. For the integral over $([u + P^*], u + P^*)$ values of $\Psi_1^*(x, t)$ were required for non-integer x. These were obtained by linear interpolation. For our numerical examples, a unit step size was deemed to be sufficiently large in view of the parameter values in our examples. In particular, the value of P^* was typically between 300 and 500.

In the case of continuous time ruin probabilities, we require probabilities of the form

$$\Psi(u,t) = \Pr(u + P\tau - S(\tau) < 0 \text{ for some } \tau, 0 < \tau \le t)$$

where P is as above and $\{S(t)\}_{t\ge 0}$ denotes the aggregate claims process, net of reinsurance. We approximate the process $\{S(t)\}_{t\ge 0}$ by the translated gamma process $\{S_G(t) + kt\}_{t\ge 0}$ where $\{S_G(t)\}_{t\ge 0}$ is a gamma process with parameters α and β. The parameters α, β and k are found by matching the first three moments of the two processes. Ruin probabilities for the translated gamma process were calculated by the method described by Dickson and Waters (1993).

5. Proportional Reinsurance

In this section we consider the problem of choosing proportional reinsurance retention levels for each of the three sub-portfolios, glass, fire and windstorm, of the portfolio described in Section 2. We will discuss two case studies which reveal rather different features.

Case Study 1: We have set the insurer's premium income (before reinsurance) to be 600 per unit time, i.e. 120% of the expected aggregate claims. The insurer's initial surplus has been set at 20. The initial surplus was chosen so that the one-year discrete time ruin probability is about 1% when the vector of retentions $\underline{a}$ is given by the solution under the minimum variance criterion with an expected net profit of 50. The reinsurer's premium loading factors are $\underline{\theta} = (0.044, 0.1605, 1.533)$. These loading factors are in proportion to the standard deviation of aggregate claims per unit time for the three sub-portfolios and are such that, if the insurer reinsured the whole of each sub-portfolio, the reinsurance premium would be 600.

Table 3A shows for the time horizons t = 1, 2, 5, 10 and 20, the probability of ruin in continuous time and in discrete time assuming the insurer does not effect any reinsurance. In this case the insurer's expected net profit per unit time is 100, as shown in the final column of Table 3A.

TABLE 3A

CASE STUDY 1 – NO REINSURANCE

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 1, 1)	0.2413	0.0237	100
2	(1, 1, 1)	0.2484	0.0262	100
5	(1, 1, 1)	0.2494	0.0267	100
10	(1, 1, 1)	0.2495	0.0267	100
20	(1, 1, 1)	0.2495	0.0267	100

The proportional reinsurance retention levels which minimise the variance of the insurer's net (of reinsurance) aggregate claims subject to the constraint that the insurer's expected net profit per unit time should be 50 are $\underline{a} = (1, 0.396, 0.581)$. Table 3B shows the insurer's probabilities of ruin with these retention levels.

TABLE 3B

CASE STUDY 1 – MINIMUM VARIANCE

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 0.396, 0.581)	0.0898	0.0103	50
2	(1, 0.396, 0.581)	0.0948	0.0115	50
5	(1, 0.396, 0.581)	0.0955	0.0117	50
10	(1, 0.396, 0.581)	0.0955	0.0117	50
20	(1, 0.396, 0.581)	0.0955	0.0117	50

Table 3C shows for each time horizon, the retention levels which minimise the insurer's probability of ruin in continuous time subject to the insurer's expected net profit being at least 50, the corresponding minimum probability of ruin, the probability of ruin in discrete time for these retention levels and finally the insurer's expected net profit. In this case, the optimal retention levels are such that the insurer's expected net profit is equal to 50 for each of the five time horizons.

TABLE 3C

CASE STUDY 1 – MINIMUM PROBABILITY OF RUIN IN CONTINUOUS TIME

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 0.438, 0.519)	0.0882	0.0095	50
2	(1, 0.438, 0.519)	0.0929	0.0106	50
5	(1, 0.439, 0.518)	0.0935	0.0108	50
10	(1, 0.439, 0.518)	0.0935	0.0108	50
20	(1, 0.439, 0.518)	0.0935	0.0108	50

Table 3D is similar to Table 3C except that for each time horizon, the retention levels are those which minimise the insurer's probability of ruin in discrete time subject to the insurer's expected net profit being at least 50.

TABLE 3D

CASE STUDY 1 – MINIMUM PROBABILITY OF RUIN IN DISCRETE TIME

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 0.456, 0.493)	0.0885	0.0094	50
2	(1, 0.456, 0.493)	0.0933	0.0105	50
5	(1, 0.456, 0.493)	0.0939	0.0107	50
10	(1, 0.456, 0.493)	0.0939	0.0107	50
20	(1, 0.456, 0.493)	0.0939	0.0107	50

Case Study 2: We have again set the insurer's premium income to be 600 but have increased the initial surplus to 35. This initial surplus gives a one-year discrete time probability of ruin of about 1% when there is no reinsurance. We have set the reinsurance premium loading factors as $\underline{\theta}$ = (0.1, 0.4, 0.8). These are somewhat arbitrary choices but are designed to reflect the relative risk for the three sub-portfolios. With these loadings, the premium for reinsuring the whole portfolio is greater than 600. Adopting the same constraints as for Case Study 1, the retention levels which minimise the variance of the insurer's net claims are (1, 0.753, 0.231).

Tables 4A, 4B, 4C and 4D give the information relating to Case Study 2 which corresponds to the information relating to Case Study 1 in Tables 3A, 3B, 3C and 3D.

TABLE 4A

CASE STUDY 2 – NO REINSURANCE

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 1, 1)	0.1282	0.0146	100
2	(1, 1, 1)	0.1347	0.0164	100
5	(1, 1, 1)	0.1357	0.0167	100
10	(1, 1, 1)	0.1357	0.0167	100
20	(1, 1, 1)	0.1357	0.0167	100

TABLE 4B

CASE STUDY 2 – MINIMUM VARIANCE

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 0.753, 0.231)	0.0746	0.0147	50
2	(1, 0.753, 0.231)	0.0861	0.0185	50
5	(1, 0.753, 0.231)	0.0894	0.0199	50
10	(1, 0.753, 0.231)	0.0895	0.0199	50
20	(1, 0.753, 0.231)	0.0895	0.0199	50

TABLE 4C

CASE STUDY 2 – MINIMUM PROBABILITY OF RUIN IN CONTINUOUS TIME

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 0.749, 0.257)	0.0745	0.0147	50
2	(1, 0.749, 0.257)	0.0860	0.0184	50
5	(1, 0.749, 0.257)	0.0893	0.0198	50
10	(1, 0.749, 0.257)	0.0894	0.0199	50
20	(1, 0.749, 0.257)	0.0894	0.0199	50

TABLE 4D

CASE STUDY 2 – MINIMUM PROBABILITY OF RUIN IN DISCRETE TIME

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(1, 1, 0.42)	0.0957	0.0103	88.4
2	(1, 1, 0.43)	0.1011	0.0115	88.6
5	(1, 1, 0.43)	0.1018	0.0118	88.6
10	(1, 1, 0.43)	0.1018	0.0118	88.6
20	(1, 1, 0.43)	0.1018	0.0118	88.6

Comparison of Tables 3A-D and 4A-D:

(a) Comparing the ruin probabilities in Table 3A (no reinsurance) with those in Tables 3B-D, and also those in Table 4A with those in Tables 4B-D, it is apparent that proportional reinsurance can reduce the probability of ruin considerably, although in many cases 50% of the maximum expected profit has been sacrificed to achieve this reduction.

(b) A feature of Tables 3C-D and Tables 4C-D is that the optimal reinsurance retentions are not very sensitive to changes in the time horizon for ruin. This suggests that if we wish to choose proportional reinsurance retentions which minimise the insurer's probability of ruin in either continuous or discrete time, subject to a minimum level for the insurer's expected net profit, it may be sufficient to calculate the optimal retentions for a short time horizon.

(c) A feature of Case Study 1 is that the optimal retentions in Tables 3C (1, 0.438/9, 0.519/8), and 3D, (1, 0.456, 0.493), are close to each other and not too far from those in Table 3B, (1, 0.396, 0.581). Also, the corresponding probabilities of ruin in Tables 3B-D are all very close to each other. This suggests that, in this example, if we wish to choose retention levels which minimise a probability of ruin, in either continuous or discrete time, an approximation can be obtained by calculating retention levels using the minimum variance criterion. This could be a significant point since the computational effort required for the latter is considerably less than that required for the former.

(d) The comments in (c) above, all of which related to Case Study 1, do not apply to Case Study 2. For Case Study 2, the optimal retentions, and ruin probabilities, calculated using a minimum variance criterion, Table 4B, and a continuous time ruin criterion, Table 4C, are very close to each other. Also, the optimal retentions in Table 4C give an expected net profit for the insurer of exactly 50. However, the optimal retentions and ruin probabilities calculated using the discrete time ruin criterion, Table 4D, are very different from those in Tables 4B and 4C. A noticeable feature of Table 4D is that the optimal retentions give expected net profits, 88.4/6, well in excess of the constrained minimum value of 50.

(e) A common feature of Tables 3A-D and 4A-D is that, for a given set of retentions and a given time horizon, the probaility of ruin in continuous time is a factor of almost 10 times greater than the probability of ruin in discrete time. To see why this is the case, consider Table 3B. The insurer's initial surplus is 20 and the expected surplus at the end of the first year is 70. This indicates that if ruin occurs in continuous time, it is likely to occur soon after time 0, so that there will be a large part of the year remaining in which the surplus can recover to a positive value. In fact, the probability of ruin in continuous time within the first half year is 0.0758 so that the probability of ruin in the following half year, having not been ruined in the first half year, is 0.0140. In general we would expect the probabilities of ruin within a given time period (continuous) and at the end of the time period (discrete) to be much closer if either the insurer's initial surplus were larger and/or the expected net profit in the time period were smaller. Referring again to the example in Table 3B, the probability of ruin at the end of 0.1 years is 0.0166. The important feature in this case is that the insurer's expected net profit in the time period is only 5.

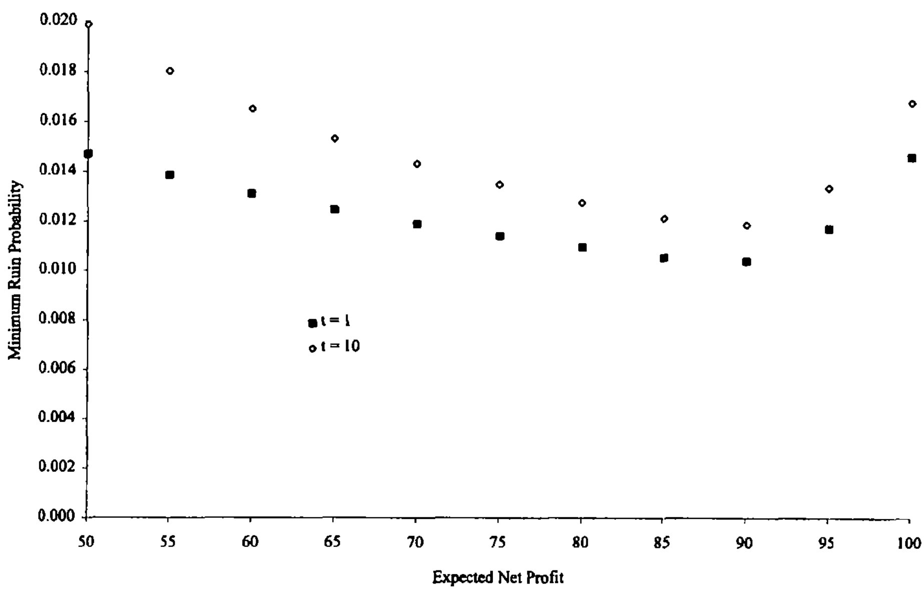


FIGURE 1: Proportional reinsurance, discrete time ruin, U = 35, loadings are 10% , 40% and 80%.

Further discussion of Case Study 2:

Figure 1 shows minimum discrete time ruin probabilities as a function of the insurer's expected net profit for $t = 1$ and $t = 10$. This figure shows the advantage to the insurer of constraining the expected net profit to be at least 50. In particular, when $t = 10$ we see that any expected net profit greater than 50 results in a lower probability of ruin than when the expected net profit equals 50. Results showing the effect of different values for the initial surplus are shown in Tables 5A, 5B, 6A and 6B, in all cases the reinsurance premium loadings are as in Case Study 2. Tables 5A and 6A show figures for an initial surplus of 20 and Tables 5B and 6B show figures for an initial surplus of 50. Tables 5A and 5B show for each of the five time horizons the optimal retention levels calculated using a continuous time ruin criterion, together with the resulting expected net profit for the insurer and the minimum value of the ruin probability. These values should be compared with those in Table 4C. Tables 6A and 6B show the optimal retention levels calculated using a discrete time ruin criterion. These values should be compared with those in Table 4D.

The optimal retentions in Table 5B are very close to those in Table 4C, indicating that increasing the insurer's initial surplus from 35 to 50 has had little effect in terms of optimal retention levels and the insurer's expected net profit. However, Table 5A displays different features. The optimal retention levels change as the time horizon increases, appearing to converge to (1, 0.827, 0.256), and the insurer's expected net profit moves away from the constrained minimum value. Table 5A indicates that the optimal retentions under a continuous time ruin criterion may depend on the time horizon and, by comparison with Tables 4C and 5B, on the insurer's initial surplus. Turning to Tables 6A and 6B, we see that a change in initial surplus has only a small impact on optimal retention levels and the insurer's expected net profit.

TABLE 5A

MINIMUM PROBABILITY OF RUIN IN CONTINUOUS TIME, $U = 20$

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Expected net profit*
1	(1, 0.753, 0.231)	0.1883	50.0
2	(1, 0.799, 0.247)	0.2025	56.8
5	(1, 0.827, 0.256)	0.2050	60.9
10	(1, 0.827, 0.256)	0.2050	60.9
20	(1, 0.827, 0.256)	0.2050	60.9

TABLE 5B

MINIMUM PROBABILITY OF RUIN IN CONTINUOUS TIME, $U = 50$

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Expected net profit*
1	(1, 0.747, 0.271)	0.0288	50
2	(1, 0.747, 0.271)	0.0362	50
5	(1, 0.748, 0.264)	0.0387	50
10	(1, 0.748, 0.264)	0.0387	50
20	(1, 0.748, 0.264)	0.0387	50

TABLE 6A

MINIMUM PROBABILITY OF RUIN IN DISCRETE TIME, $U = 20$

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Expected net profit*
1	(1, 1, 0.460)	0.0186	89.2
2	(1, 1, 0.470)	0.0205	89.4
5	(1, 1, 0.470)	0.0208	89.4
10	(1, 1, 0.470)	0.0208	89.4
20	(1, 1, 0.470)	0.0208	89.4

TABLE 6B

MINIMUM PROBABILITY OF RUIN IN DISCRETE TIME, $U = 50$

t	$\underline{a}$	*Prob'y of ruin (continuous)*	*Expected net profit*
1	(1, 1, 0.390)	0.0055	87.8
2	(1, 1, 0.400)	0.0063	88.0
5	(1, 1, 0.405)	0.0065	88.1
10	(1, 1, 0.405)	0.0065	88.1
20	(1, 1, 0.405)	0.0065	88.1

6. Excess loss reinsurance

Case Study 3: In this Case Study we investigate different optimal retention levels for excess loss reinsurance of the fire and windstorm sub-portfolios. For the reasons given in Section 2, we assume that the glass sub-portfolio is not reinsured under an excess loss treaty. The insurer's premium income is 600, as in the previous two Case Studies, and the initial surplus is 35. The reinsurance premium loading factors are 100% (fire) and 200% (windstorm). These factors are higher than those in the previous two Case Studies, a consequence of the fact that excess loss, by its very nature, should be more expensive than proportional reinsurance.

The probabilities of ruin, for continuous and discrete time, and for different time horizons, when there is no reinsurance are as in Table 4A. We will assume that the insurer wishes to find the optimal excess loss retentions subject to the constraint that the expected net profit is at least 50. The minimum variance solution to this problem is $\underline{M} = (\infty, 9.66, 19.32)$. The ruin probabilities with this set of retention levels are shown in Table 7B. Table 7C shows the optimal continuous time retentions and ruin probabilities for different time horizons, together with the discrete time ruin probabilities for these retentions and the insurer's expected net profit, which in every case is 50. Table 7D shows the optimal discrete time retentions and ruin probabilities for different time horizons, together with the continuous time ruin probabilities for these retentions and the insurer's expected net profit.

TABLE 7B

CASE STUDY 3 – MINIMUM VARIANCE

t	$\underline{M}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(∞, 9.66, 19.32)	0.0420	0.0068	50
2	(∞, 9.66, 19.32)	0.0485	0.0083	50
5	(∞, 9.66, 19.32)	0.0499	0.0087	50
10	(∞, 9.66, 19.32)	0.0499	0.0087	50
20	(∞, 9.66, 19.32)	0.0499	0.0087	50

TABLE 7C

CASE STUDY 3 – MINIMUM PROBABILITY OF RUIN IN CONTINUOUS TIME

t	$\underline{M}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(∞, 10.43, 17.39)	0.0414	0.0066	50
2	(∞, 10.39, 17.48)	0.0479	0.0081	50
5	(∞, 10.38, 17.50)	0.0492	0.0085	50
10	(∞, 10.38, 17.50)	0.0493	0.0085	50
20	(∞, 10.38, 17.50)	0.0493	0.0085	50

TABLE 7D

CASE STUDY 3 – MINIMUM PROBABILITY OF RUIN IN DISCRETE TIME

t	$\underline{M}$	*Prob'y of ruin (continuous)*	*Prob'y of ruin (discrete)*	*Expected net profit*
1	(∞, 11.52, 19.09)	0.0451	0.0066	54.7
2	(∞, 12.56, 20.78)	0.0543	0.0078	58.8
5	(∞, 12.91, 21.37)	0.0564	0.0081	60.1
10	(∞, 12.91, 21.37)	0.0564	0.0081	60.1
20	(∞, 12.91, 21.37)	0.0564	0.0081	60.1

A comparison of Tables 7B-D shows that the ruin probabilities in these tables, either continuous or discrete time, do not change significantly from one table to the next. This indicates that for many practical purposes the probability of ruin, in either discrete or continuous time, can be assumed to attain its minimum value at the solution to the minimum variance problem. However, the extra computational effort required to compute the optimal retentions for discrete time ruin in Table 7D may be considered worthwhile since they result in an expected net profit for the insurer in excess of 60, for $t \geq 5$, rather than 50 for the minimum variance optimal retentions.

Other features of Tables 7B-D are:

(a) the different time horizons in Tables 7C and 7D have little effect on the values of the optimal retention levels, and no effect for $t \geq 5$, and,

(b) optimal retentions for continuous time ruin, Table 7C, are closer to the minimum variance solution than are the optimal retentions for discrete time ruin, Table 7D. In particular, the former give an expected net profit for the insurer of 50, i.e. on the boundary of the constraint, as for the minimum variance solution, whereas the latter give an expected net profit away from the boundary.

Figure 2 shows the minimum discrete time ruin probabilities as a function of the insurer's expected net profit for t = 1 and t = 10. As in Figure 1, we can again see the advantage of constraining the expected net profit to be at least 50 rather than exactly 50.

The effect of altering the insurer's initial surplus is shown in Table 8. This table shows for $U = 20$ and $U = 50$ the optimal retentions for both the continuous time and the discrete time ruin criteria, together with the minimum value for the probability of ruin and the resulting expected net profit for the insurer. In all cases the time horizon for ruin is 20 years.

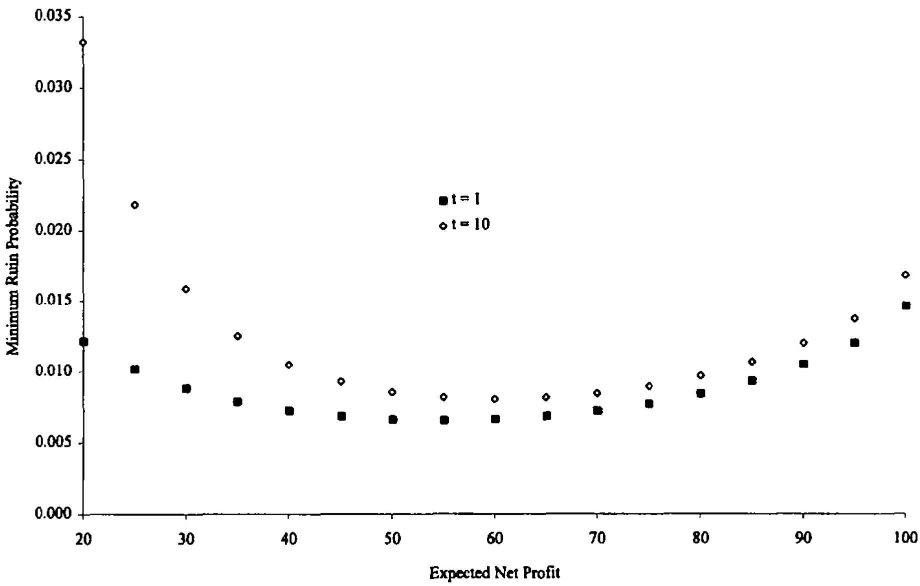


FIGURE 2: Excess loss reinsurance, discrete time ruin, U = 35, loadings are 100% and 200%.

TABLE 8

CASE STUDY 3 – DIFFERENT VALUES FOR THE INITIAL SURPLUS: t = 20

t	*Continuous / discrete*	$\underline{M}$	*Prob'y of ruin*	*Expected net profit*
20	Continuous	(∞, 10.08, 18.22)	0.1569	50
50	Continuous	(∞, 10.49, 17.25)	0.0155	50
20	Discrete	(∞, 16.18, 27.00)	0.0182	70.5
50	Discrete	(∞, 10.89, 17.89)	0.0031	51.8

The important point revealed by Table 8 is that changing the insurer's initial surplus has little effect, in terms of the optimal retentions or the insurer's expected net profit, in the case of continuous time ruin but makes a considerable difference in the case of discrete time ruin.

7. Conclusions

Our purpose in this paper has been to investigate different criteria for determining the optimal relative retention limits for a non-life portfolio consisting of a number of independent sub-portfolios. For the reasons discussed in Examples 1, 2 and 3 in Section 4, the minimum variance criterion could be regarded as a proxy for a probability of ruin criterion. The advantages of the minimum variance criterion are:

(a) it is possible to express the retention levels in closed form,

(b) the optimal retention levels depend only on the reinsurance premium loadings and, in the case of proportional reinsurance, on the first two moments of aggregate claims for the sub-portfolios, and,

(c) the optimal retention levels can be calculated very easily. In contrast, the optimal retention levels using a ruin probability criterion cannot be expressed in closed form and can be time consuming to compute, particularly for the longer time horizons.

Our method of investigation has been to carry out several "case studies" for a single portfolio. Using this method it can be difficult to draw any conclusions. Nevertheless, we consider that the numerical results in Sections 5 and 6, and the other examples we have investigated in the course of this study, enable us to reach the following tentative answers, for both proportional and for excess loss reinsurance, to the questions posed in Section 1:

(1) The minimum variance criterion produces optimal relative retention levels close to those produced by the continuous time ruin criterion (see Tables 3B and 3C, Tables 4B, 4C, 5A and 5B and Tables 7B, 7C and 8 (Continuous)) but not necessarily similar to those produced by the discrete time ruin criterion (see Tables 4B, 4D, 6A and 6B and Tables 7B, 7D and 8 (Discrete)). The three examples in Section 4 all indicated that optimality with respect to the minimum variance criterion might be approximately the same as optimality with respect to the probability of ruin in continuous time (Examples 1 and 3) and the probability of ruin in discrete time (Example 2). Specifically, we assumed in Examples 2 and 3 that the (retained) aggregate claim amount distribution could be reasonably approximated by a normal distribution, and hence is symmetric. However, with an expected net profit of at least 50 the coefficient of skewness of the retained aggregate claim amount distribution in Case Studies 1 and 2 turns out to be above 0.5 for all combinations of retention levels, and hence the distribution is not symmetric. For this reason it should not be surprising that optimality with respect to the minimum variance criterion can produce different results to optimality with respect to the probability of ruin in discrete time (Case Study 2). What may be considered surprising is the clo-

seness of the results in all three case studies under the minimum variance criterion and the continuous time ruin criterion.

(2) As indicated in (1) above, the discrete time ruin criterion can produce very different optimal retentions from those produced by the continuous time ruin criterion. This should not be too surprising since these two probabilities are rather different both in nature and, in our examples, numerically. See comment (e) in Section 5. That these two probabilities behave differently has already been observed in a somewhat different setting. See Dickson and Waters (1996, Section 8 and 9).

(3) (i) In most cases we investigated, the optimal retention levels for continuous time ruin give an expected net profit for the insurer on the boundary of its constrained values (see Tables 3C, 4C, 5B, 7C and 8 (Continuous)). In one example this was not the case (see Table 5A). The exact reverse is true for the optimal retentions for discrete time ruin (see Table 3D for the former case and Tables 4D, 6A, 6B, 7D and 8 (Discrete) for the latter case).
(ii) A marked feature of all our calculations is that the time horizon for ruin, for one year and longer, has very little effect on the optimal retention levels in either continuous time or discrete time. In all cases the optimal retention levels are unchanged to three significant figures as the time horizon increases from five years to twenty years.
(iii)The insurer's initial surplus, which is not considered by the minimum variance criterion, can have a considerable effect on the optimal retention levels using a probability of ruin criterion (see Tables 7D and 8 (Discrete)).

REFERENCES

BUHLMANN, H. (1970) *Mathematical methods in risk theory.* Spinger Verlag, New York.

DE FINETTI, B. (1940) Il problema dei pieni. *Giorn. Ist. Ital. Attuari* 11, 1-88.

DICKSON, D.C.M. and WATERS, H.R. (1993) Gamma processes and finite time survival probabilities. *ASTIN Bulletin* 23, 259-272.

DICKSON, D.C.M. and WATERS, H.R. (1996) Reinsurance and ruin. *Insurance: Mathematics & Economics* 19, 61-80.

GRANDELL, J. (1977) A class of approximations of ruin probabilities. *Scandinavian Actuarial Journal Supplement*, 37-52.

RAMLAU-HANSEN, H. (1983) Fire claims for single-family houses. *Presented at the XVII ASTIN Colloquium, Lindau.*

RAMLAU-HANSEN, H. (1986a) Statistical analysis of policy and claims data in non-life insurance: A solvency study. Part 1. Introduction to the study and analysis of glass claims. *Working Paper No. 60, Laboratory of Actuarial Mathematics, University of Copenhagen.*

RAMLAU-HANSEN, H. (1986b) Statistical analysis of policy and claims data in non-life insurance: A solvency study. Part 2. Analysis of fire claims. *Working Paper No. 61, Laboratory of Actuarial Mathematics, University of Copenhagen.*

RAMLAU-HANSEN, H. (1986c) Statistical analysis of policy and claims data in non-life insurance: A solvency study. Part 3. Analysis of windstorm claims. *Working Paper No. 62, Laboratory of Actuarial Mathematics, University of Copenhagen.*

RAMLAU-HANSEN, H. (1986d) Statistical analysis of policy and claims data in non-life insurance: A solvency study. Part 4. Analysis of windstorm claims. *Working Paper No. 63, Laboratory of Actuarial Mathematics, University of Copenhagen.*

RAMLAU-HANSEN, H. (1988a) A solvency study in non-life insurance. Part 1. Analyses of fire, windstorm and glass claims. *Scandinavian Actuarial Journal, 3-34.*

RAMLAU-HANSEN, H. (1988b) A solvency study in non-life insurance. Part 2. Solvency margin requirements. *Scandinavian Actuarial Journal*, 35-59.

DAVID C.M. DICKSON
Centre for Actuarial Studies
Faculty of Economics and Commerce
University of Melbourne
Victoria 3052
Australia
e-mail: ddickson@cupid.ecom.unimelb.edu.au

HOWARD R. WATERS
Department of Actuarial Mathematics and Statistics
Heriot-Watt University
Midlothian EH14 4AS
Great Britain
e-mail: H.R.Waters@ma.hw.ac.uk

A NEW DISTRIBUTION OF POISSON-TYPE FOR THE NUMBER OF CLAIMS

By MICHEL DENUIT

Université Libre de Bruxelles, Bruxelles, Belgium

ABSTRACT

This paper is concerned with two methods to estimate the parameters of the Poisson-Goncharov distribution introduced recently by Lefèvre and Picard (1996). These methods are applied to fit, *inter alia*, the six observed claims distributions, from automobile insurance third party liability portfolios, studied by Gossiaux and Lemaire (1981) and analysed afterwards by several authors.

KEYWORDS

Abel-Goncharov polynomials; Poisson-Goncharov distribution; Generalized Poisson law; fitting; claims distribution.

1. INTRODUCTION

The importance of the third party liability automobile insurance has not to be demonstrated. In most of the industrialized countries, this branch represents a considerable share of the yearly premium collection (for instance, in Belgium, 29.5% during the year 1993). In view of this, many attempts have been made in the actuarial literature to find a probabilistic model for the distribution of the number of auto-accidents (see the review contained in Section 3). Our purpose here is to show that the Poisson-Goncharov distribution introduced recently by Lefèvre and Picard (1996) provides an appropriate probability model to describe the annual number of claims incurred by an insured motorist. We will then propose two methods to estimate the parameters and we will apply them to fit the six observed claims distributions in Gossiaux and Lemaire (1981), as well as recent data sets from Belgium.

Let us briefly survey the paper. In Section 2, we will shortly present the Poisson-Goncharov distribution, establishing two new properties of it. In Section 3, we will first introduce the problem under study and then bring up the Poisson-Goncharov model for the number of claims in automobile insurance. Section 4 will be devoted to the Maximum Likelihood method to estimate the parameters of the Poisson-Goncharov distribution. We will see that this method usually yields implicit Likelihood equations which have to be solved numerically. The starting values of the parameters will be obtained using the so-called "Ad Hoc" method developed in Panjer and Willmot (1992). In Section 5, we will propose a specific

ASTIN BULLETIN, Vol. 27, No. 2, 1997, pp. 229-242

Least Squares type estimation method, specially built for the Poisson-Goncharov distribution. We will see that this method provides explicit expressions for the estimators and, in most of the cases, accurate fits. Section 6 will be concerned with concluding remarks. Finally, Appendix will take all the numerical illustrations up.

2. THE POISSON-GONCHAROV DISTRIBUTION

Recently, Lefèvre and Picard (1996) introduced a new discrete probability distribution on the set of the non-negative integers (subsequently denoted by $\mathbb{N}$), called the Poisson-Goncharov law, which is constructed in terms of Abel-Goncharov polynomials and which extends the classical Poisson law as well as the Generalized Poisson distribution proposed by Consul and Jain (1973).

Very briefly, le $U = \{u_i,\ i \in \mathbb{N}\}$ be any given family of real numbers. To U is attached a unique family of Abel-Goncharov polynomials, $\{G_n(x|U),\ n \in \mathbb{N}\}$, of degree n in x, defined recursively, starting from $G_0(x|U) \equiv 1$, by

$$G_n(x|U) = \frac{x^n}{n!} - \sum_{i=0}^{n-1} \frac{u_i^{n-i}}{(n-i)!} G_i(x|U),\ n \geq 1. \tag{2.1}$$

The reader is referred to Oskolkov (1988) (and the references therein) for a presentation of these polynomials (AG polynomials, in short). Note that the only particular case in which an explicit expression is known for the $G_n(x|U)'s$ is the Abel one. Specifically, if $u_i = a + bi,\ i \in \mathbb{N}$, a and b being real constants, then

$$G_n(x|U) = (x-a)\frac{(x-a-nb)^{n-1}}{n!},\ n \in \mathbb{N}. \tag{2.2}$$

When $u_i = a,\ i \in \mathbb{N}$, (2.2) reduces to $G_n(x|U) = (x-a)^n/n!,\ n \geq 0$. In order to have all the $G_n(x|U)'s$ positive for $x \geq 0$, it suffices for U to be negative and non-increasing, i.e. $0 > u_0 \geq u_1 \geq \ldots \geq u_i \geq u_{i+1} \geq \ldots$ (this condition will be retained subsequently). Now, the Poisson-Goncharov distribution associated with U, negative and non-increasing, is the family $\{\mathcal{PG}_n(U),\ n \in \mathbb{N}\}$ defined by

$$\mathcal{PG}_n(U) = G_n(0|U)e^{u_n},\ n \in \mathbb{N}. \tag{2.3}$$

It is denoted by $\mathcal{PG}(U)$.

As announced earlier, the $\mathcal{PG}(U)$ law can be viewed as a distribution of Poisson-type. If the u_i's are linear in i, $u_i = -\theta - i\lambda$ say, with $\theta \in \mathbb{R}_0^+$ and $\lambda \in \mathbb{R}^+$, then, using (2.2),

$$\mathcal{PG}_n(U) = \frac{\theta(\theta + n\lambda)^{n-1}}{n!} e^{-\theta - n\lambda},\ n \in \mathbb{N}. \tag{2.4}$$

The distribution (2.4) is nothing else than the Lagrangian or Generalized Poisson law introduced by Consul and Jain (1973) (see the book by Consul (1989)). It is non-defective if and only if $\lambda \in [0,\ 1]$. In particular, if all the u_i's are equal to $-\theta$, say, with $\theta \in \mathbb{R}_0^+$, then $\mathcal{PG}(U)$ becomes the usual Poisson distribution with parameter θ.

It is worthwhile recalling that the Generalized Poisson law belongs to the wide class of discrete Lagrangian probability distributions, defined by Consul and Shenton (1972), by means of the Lagrange expansion formula. Moreover, it is also part of the Abel series distributions family introduced by Charalambides (1990). It has various fields of applications, in particular biostatistics (see, e.g., Janardan *et al.* (1979)) as well as actuarial sciences where it has been proposed initially by Consul (1990), and then by Ter Berg (1996), to model the annual number of accidents incurred by a motorist. See also Gerber (1990) for an application linking to the ruin model. We mention that recursive algorithms to evaluate compound Generalized Poisson probabilities have recently been developed, e.g. by Goovaerts and Kaas (1991) and Sharif and Panjer (1995).

Coming back to the $\mathcal{PG}(U)$ law, this corresponds typically to the distribution of the first crossing level L of a Poisson process $\mathcal{N} = \{N(t), t \in I\!R^+\}$ (with parameter 1, say) in a lower non-decreasing boundary $\mathcal{B}_U$ (such as represented in Figure 2.1). More precisely, we first observe that $\mathcal{B}_U$ may be reduced to the set of points that are eligible as levels of first-crossing, i.e. points with integer ordinate. Denoting this set of points by $\{(-u_i, i), i \in I\!N\}$, where $U = \{u_i, i \in I\!N\}$ is negative and non-increasing, it can then be proved that the law of L is provided by (2.3) (see Lefèvre and Picard (1996)). We notice that the Poisson law for L is obtained when $\mathcal{B}_U$ is vertical, and the Generalized Poisson law when $\mathcal{B}_U$ is linear.

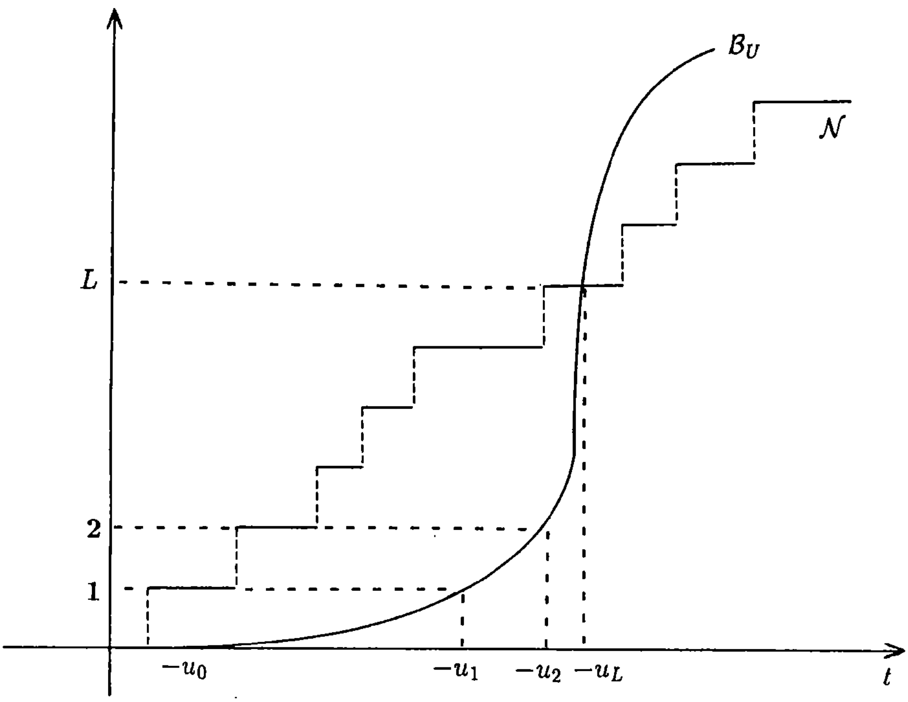


FIGURE 2.1. First-crossing level L of the Poisson process $\mathcal{N}$ in the lower boundary $\mathcal{B}_U$

Lefèvre and Picard (1996) have also pointed out the relationship between the $\mathcal{PG}(U)$ law and the number of customers served in the first busy period of a *M/D/1* queueing system in which customers arrive according to a Poisson process with rate 1, a unique customer is present initially in the queue and the service times are deterministic but differ from a customer to another (let $-u_0$ be the service time of the initial customer and, for $i \geq 1$, let $u_{i-1} - u_i$ denote the service time of the i-th customer). In such a system, the number of new customers served during the first busy period is of law $\mathcal{PG}(U)$.

Now, the family U may represent any negative and non-increasing sequence of real numbers, so that the $\mathcal{PG}(U)$ law may depend on a large, even infinite, number of parameters. For statistical estimation, however, it is necessary to specify for the u_i's a simple analytical expression, depending on one, two or three parameters for example (like $u_i = \theta_1 + \theta_2 i + \theta_3 i^2,\ i \in \mathbb{N}, \theta_1 < 0, \theta_2 \leq 0, \theta_3 \leq 0$).

It is possible to show that the $\mathcal{PG}(U)$ law belongs to the Sundt's family, i.e. satisfies, for some fixed values of k and ω,

$$\mathcal{PG}_n(U) = \sum_{i=1}^{k}\left(\alpha_i + \frac{\beta_i}{n}\right)\mathcal{PG}_{n-i}(U),\ n = \omega+1,\ \omega+2,\ \ldots \tag{2.5}$$

Indeed, starting from the following identity,

$$\mathcal{PG}_n(U) = e^{-u_0}\sum_{i=1}^{n}\mathcal{PG}_i(U)\left(\frac{2i}{n} - 1\right)\mathcal{PG}_{n-i}(U),\ n \geq 1, \tag{2.6}$$

we obtain (2.5) with $k = n$, $\omega = 0$ and, for $i \geq 1$, $\alpha_i = -e^{u_i - u_0}G_i(0|U)$, $\beta_i = 2ie^{u_i - u_0}G_i(0|U)$. These α_i's and β_i's are in fact those suggested in Theorem 3 of Sundt (1992). Nevertheless, if we desire to obtain the distribution of the compound $\mathcal{PG}(U)$ sum, as pointed out by Panjer and Wang (1995), when $k=n$ in (2.5), the computing effort using Sundt's recursive formula is of the same order as that needed by a direct convolution approach (Sundt's recursion is interesting only when the claim frequency distribution satisfies (2.5) with small values for k and ω).

Let us recall that, given two random variables X and Y valued in $\mathbb{N}$, Y is said to stochastically dominate X, denoted by $X \preceq_{st} Y$, when $P[X \leq n] \geq P[Y \leq n]\ \forall\, n \in \mathbb{N}$ (see, e.g., the recent books by Shaked and Shanthikumar (1994) and by Kaas *et al.* (1994)). It is well-known that $X \preceq_{st} Y$ if and only if there exist two random variables $\tilde{X}$ and $\tilde{Y}$, defined on the same probability space, such that X and $\tilde{X}$, as well as Y and $\tilde{Y}$, are identically distributed, and $P[\tilde{X} \leq \tilde{Y}] = 1$.

It is easy to prove that, if X (resp. Y) is distributed according to the $\mathcal{PG}(U)$ (resp. $\mathcal{PG}(V)$) law, with $u_i > v_i\ \forall\, i \in \mathbb{N}$, then $X \preceq_{st} Y$. This follows immediately from a decomposition formula of the Raikov type, obtained by Lefèvre and Picard (1996), which states that it is always possible to decompose Y into the sum $Z_1 + Z_2$ such that Z_1 follows the $\mathcal{PG}(U)$ law and the law of Z_2, given $Z_1 = j$, $j \in \mathbb{N}$, is $\mathcal{PG}(V(j))$, where $V(j) = \{v_i(j), i \in \mathbb{N}\}$ with $v_i(j) = v_{i+j} - u_j$. Thus, with X and Y described above and using $\tilde{X} = Z_1$ and $\tilde{Y} = Z_1 + Z_2$, we have $X \preceq_{st} Y$.

If we come back to the representation of the $\mathcal{PG}(U)$ as the law of the first-crossing level L, the above result becomes obvious. Indeed, if X (resp. Y) is the first-crossing level of the Poisson process $\mathcal{N}$ in the lower boundary $\mathcal{B}_U$ (resp. $\mathcal{B}_V$) described by U (resp. by V), and if $v_i < u_i \forall i \in I\!N$ (that is $\mathcal{B}_V$ lies on the right of $\mathcal{B}_U$), it is clear that $X \preceq_{st} Y$ (since $X \leq Y$ almost surely). Choosing families like $\{-\theta,\ i \geq 0\}$ or $\{-\theta - i\lambda,\ i \geq 0\}$, with $\theta > 0$ and $\lambda \in [0, 1]$, yields straight corollaries for the Poisson or the Generalized Poisson laws.

3. THE POISSON-GONCHAROV MODEL FOR THE NUMBER OF CLAIMS IN AUTOMOBILE INSURANCE

Let us first introduce the problem under investigation in the present paper. In order to see if there exist some probability law applicable to claims distributions in automobile insurance third party liability portfolio, Gossiaux and Lemaire (1981) examined six observed claims distributions. Those came from five countries and were studied before by other researchers. Gossiaux and Lemaire (1981) fitted the Poisson distribution, the Generalized Geometric distribution, the Negative Binomial distribution and the mixed Poisson distribution to each of them by the Maximum Likelihood method and the method of Moments. They concluded that no single probability law seems to emerge as providing a good fit to all of them. Moreover, there was at least one example where each model gets rejected by a chi-square test (at the level 10%). Seal (1982) supplemented the paper by Gossiaux and Lemaire (1981) with an analysis of some automobile accidents data from California. He concluded that it supports the mixed Poisson hypothesis for the distribution of the number of claims. Kestemont and Paris (1985), using mixtures of Poisson processes, defined a large class of probability distributions and developed an efficient method of estimating its parameters. For the six data sets in Gossiaux and Lemaire (1981), they proposed a law depending on three parameters and they always obtained extremely good fits. Willmot (1987) showed that the Poisson-Inverse Gaussian law deserves consideration as a model for the claims distribution due to its good fit to the data. Furthermore, this law enjoys abundance of convenient mathematical properties. Willmot (1987) compared the Poisson-Inverse Gaussian distribution to the Negative Binomial one and concluded that the fits are superior with the Poisson-Inverse Gaussian in all the six cases studied by Gossiaux and Lemaire (1981). See also the note by Lemaire (1991) about the confrontation between Negative Binomial and Poisson-Inverse Gaussian on the basis of six data sets not related to insurance. Ruohonen (1987) considered a model for the claim number process. This model is a weighted Poisson process with a three-parameters Gamma distribution as the structure function and is compared with the two-parameters Gamma model giving the Negative Binomial distribution. He fitted his model to some data that can be found in the actuarial literature and the results were satisfying. Panjer (1987) proposed the Generalized Poisson-Pascal distribution, which includes three parameters, for the modelling of the number of automobile claims. The fits obtained were satisfactory, too. Note that the Pólya-Aeppli, the Poisson-Inverse

Gaussian and the Negative Binomial are special cases of this distribution. Willmot (1988) enumerated completely the class of claim frequency distributions discussed by Sundt and Jewell (1981). He demonstrated the good fit to automobile claim frequency data of one member (in fact, the Modified Extended Truncated Negative Binomial distribution), using the six data sets analyzed by Gossiaux and Lemaire (1981). Consul (1990) tried to fit the same six data sets by the Generalized Poisson distribution. Although the Generalized Poisson law is not rejected by a chi-square test, the fits obtained by Kestemont and Paris (1985), for instance, are always much better. Furthermore, Elvers (1991) reported that the Generalized Poisson distribution did not fit very well the data observed in an automobile third party liability insurance portfolio (the distribution hypothesis was, according to his note, in almost every case rejected by a chi-square test). More recently, Ter Berg (1996) considered a slightly different model, involving the Generalized Poisson, too. Moreover, he introduced a loglinear model, which is able to incorporate explanatory variables. The fits were found satisfactory. Islam and Consul (1992) suggested the Consul distribution as a probabilistic model for the distribution of the number of claims in automobile insurance. These authors approximated the chance mechanism which produces vehicle accidents by a branching process. They fit the model to the data sets used by Panjer (1987) and by Gossiaux and Lemaire (1981). Note that this model deals only with autos in accident. Consequently, the zero-class has to be excluded. The fitted values seem good. However, this has to be considered cautiously, due to the comments by Sharif and Panjer (1993). Indeed, these found serious flaws embedded in the fitting of the Consul model; in particular, the very restricted parameter space and some theoretical problems in the derivation of the Maximum Likelihood estimators. They refer to other simple probability models, as the Generalized Poisson-Pascal or the Poisson-Inverse Gaussian, whose fit were found quite satisfying. We end this brief review with two books. The first one is due to Panjer and Willmot (1992) in which Chapter 9 is devoted to the fitting risk model problem. In the second one, by Lemaire (1995), Chapter 3 focus on models for the claims number distribution. These authors give a remarkable account to the problem under investigation.

The probabilistic model for the number of claims incurred by a motorist introduced here extends both the classical Poisson and the Generalized Poisson models. We will use extensively the decomposition formula of the Raikov-type for the $\mathcal{PG}(U)$ recalled above. We split the total number of claims N_{tot} caused by an individual during a fixed period of time (say one year), which is distributed according to the $\mathcal{PG}(U)$ law, where $u_i = \theta_1 + \theta_2 i + \theta_3 i^2, i \in I\!N, \theta_1 < 0,\ \theta_2,\ \theta_3 \leq 0$ into $N_{Poisson}$ and N_{extra}, that is

$$N_{tot} = N_{Poisson} + N_{extra}, \tag{3.1}$$

where $N_{Poisson}$ follows a Poisson law with parameter $-\theta_1$ and N_{extra}, given $N_{Poisson} = j_1$, is distributed according to a $\mathcal{PG}$ $(V(j_1))$ law, with $v_i(j_1) = \theta_2(i+j_1) + \theta_3(i+j_1)^2, i \in I\!N$. Note that $[N_{extra}|N_{Poisson} = j_1]$ increases in the stochastic dominance with j_1. By splitting up the extra claims, it is easily seen that (3.1) consists in fact in breaking up N_{tot} into

$$N_{tot} = N_{Poisson} + N^{(1)}_{extra} + N^{(2)}_{extra} + N^{(3)}_{extra} + \dots \tag{3.2}$$

where $N^{(1)}_{extra}$, given $N_{Poisson} = j_1$, follows a Poisson law with parameter $-j_1\theta_2 - j_1^2\theta_3$, $N^{(2)}_{extra}$, given $N_{Poisson} = j_1$ and $N^{(1)}_{extra} = j_2$, follows a Poisson law with parameter $-j_2\theta_2 - j_2^2\theta_3 - 2j_1j_2\theta_3$, and so on. Considering N_{tot} as distributed according to the $\mathcal{PG}$ (U) law comes thus down to distinguish among the claims whether they are produced by one or another source, each source adduceing a number of accidents conditionnally distributed as a Poisson law, so that the model (3.2) seems intuitively acceptable.

4 MAXIMUM LIKELIHOOD ESTIMATORS

4.1 A general approach

Let us suppose that the u_i's, $i \in I\!N$, depend on m parameters, i.e. $u_i = u_i(\theta_1, \dots, \theta_m)$, where $(\theta_1, \dots, \theta_m) \in \Theta_1 \times \dots \times \Theta_m \subseteq I\!R^m$. We want to find the Maximum Likelihood estimators (MLE, in short) of the parameters $\theta_1, \dots, \theta_m$.

Let a random sample of size n, $(X_1, \dots, X_n)$, be taken from a population with the $\mathcal{PG}$ (U) law. The corresponding observations are $(x_1, \dots, x_n)$. Let $kmax$ be the largest observation; n_k, $0 \le k \le kmax$, the number of occurences for k; and $f_k = n_k/n$, $0 \le k \le kmax$, the observed frequencies for the different classes. The Likelihood function is

$$L_{\theta_1, \dots \theta_m}(n_0, \dots, n_{k\,max}) = \prod_{k=0}^{kmax} (e^{u_k} G_k(0|U))^{n_k}. \tag{4.1}$$

The MLE $\hat{\theta}_1, \dots, \hat{\theta}_m$ of the parameters $\theta_1, \dots, \theta_m$ are such that they maximize the Log-Likelihood function. This leads to the Likelihood equations

$$0 = \sum_{k=0}^{kmax} n_k \left[\frac{\partial u_k}{\partial \theta_j}\right]_{(\theta_1, \dots, \theta_m) = (\hat{\theta}_1, \dots, \hat{\theta}_m)} +$$

$$\sum_{k=0}^{kmax} n_k \left[\frac{\Gamma_{j,k}(0|U)}{G_k(0|U)}\right]_{(\theta_1, \dots, \theta_m) = (\hat{\theta}_1, \dots, \hat{\theta}_m)}, \; 1 \le j \le m, \tag{4.2}$$

where $\Gamma_{j,k}(x|U), 0 \le k \le kmax$, $1 \le j \le m$, denotes the first partial derivative of $G_k(x|U)$ with respect to the parameter θ_j. From (2.1), we see that the $\Gamma_{j,k}(0|U)'s$ satisfy the following recurrence relations: starting from $\Gamma_{j,0}(0|U) = 0$, $1 \le j \le m$, we have that

$$\Gamma_{j,k}(0|U) = -\sum_{i=0}^{k-1} \frac{u_i^{k-i-1}}{(k-i-1)!}\left(\frac{\partial u_i}{\partial \theta_j} G_i(0|U) + \frac{u_i}{k-i}\Gamma_{j,i}(0|U)\right), \tag{4.3}$$

which allow us to compute them recursively.

Unfortunately, the *MLE* $\hat{\theta}_1, \ldots, \hat{\theta}_m$, which are solutions of (4.2) cannot be obtained in a closed form (except in the Poisson case). Hence, they are computed *via* numerical maximization of the Log-Likelihood function. Let us quote in the next subsection some particular cases of special interest.

4.2 Particular cases

4.2.1. If $U = \{-\theta_1, i \in \mathbb{N}\}, \theta_1 > 0$, then $G_k(0|U) = \theta_1^k/k!, \Gamma_{1,k}(0|U) = \theta_1^{k-1}/(k-1)!$, and (4.2) gives $\hat{\theta}_1 = \bar{x}$, which is the classical result for the Poisson law.

4.2.2. If $U = \{\theta_1 + \theta_2 h_1(i), i \in \mathbb{N}\}, \theta_1, \theta_2$, and $h_1(.)$ such that U is negative and non-increasing, let us establish the two following results, which will give us expression for $\Gamma_{1,k}(x|U)$ and $\Gamma_{2,k}(x|U), k \in \mathbb{N}$. First of all, we recall two interesting operational properties of the AG polynomials. For any integer n,

$$\frac{d^k}{dx^k} G_n(x|U) = \begin{cases} G_{n-k}(x|E^k U), & \text{if n} \geq \text{k}, \\ 0, & \text{otherwise}, \end{cases} \tag{4.4}$$

where $E^k U = \{u_{k+i}, i \in \mathbb{N}\}$ denotes the family U without its first k elements. We also have that, for $a, b \in \mathbb{R}$, and for $n \in \mathbb{N}$,

$$G_n(ax + b|aU + b) = a^n G_n(x|U), \tag{4.5}$$

where $aU + b = \{au_i + b, i \in \mathbb{N}\}$.

Lemma 4.1 *For a, $b \in \mathbb{R}$ and $n \in \mathbb{N}$,*

$$\frac{\partial}{\partial a} G_n(x|\{a + bf(i), i \in \mathbb{N}\}) = \begin{cases} -G_{n-1}(x|\{a + bf(i+1), i \in \mathbb{N}\}), & \text{if } n \geq 1, \\ 0, & \text{if } n = 0. \end{cases}$$

Proof. The result is obvious for $n = 0$. For $n \geq 1$, using (4.4) and (4.5) yields

$$\begin{aligned} \frac{\partial}{\partial a} G_n(x|\{a + bf(i), i \in \mathbb{N}\}) &= \frac{\partial}{\partial a} G_n(x - a|\{bf(i), i \in \mathbb{N}\}) \\ &= -G_{n-1}(x - a|\{bf(i+1), i \in \mathbb{N}\}) \\ &= -G_{n-1}(x|\{a + bf(i+1), i \in \mathbb{N}\}), \end{aligned}$$

hence the announced result. □

Lemma 4.2 *For a, $b \in I\!R$ and $n \in I\!N$,*

$$\frac{\partial}{\partial b} G_n(x|\{a+bf(i), i \in I\!N\}) = \begin{cases} \frac{n}{b} G_n(x|\{a+bf(i), i \in I\!N\}) \\ + \frac{a-x}{b} G_{n-1}(x|\{a+bf(i+1), i \in I\!N\}), & \text{if } n \geq 1, \\ 0, & \text{if } n = 0. \end{cases}$$

Proof. The result is obvious when $n = 0$. For $n \geq 1$, using (4.4) and (4.5) yields

$$\frac{\partial}{\partial b} G_n(x|\{a+bf(i), i \in I\!N\}) = \frac{\partial}{\partial b}[b^n G_n((x-a)/b|\{f(i), i \in I\!N\})]$$

$$= nb^{n-1} G_n((x-a)/b|\{f(i), i \in I\!N\})$$

$$+ b^{n-2}(a-x) G_{n-1}((x-a)/b|\{f(i+1), i \in I\!N\}),$$

and this achieves the proof by (4.5). □

From (4.2) together with Lemma's 4.1 and 4.2, the *MLE* $\hat{\theta}_1$ and $\hat{\theta}_2$ of the parameters θ_1 and θ_2 are thus solutions of the following system:

$$\begin{cases} n = \sum_{k=1}^{kmax} n_k \frac{G_{k-1}\left(0\left|\left\{-\hat{\theta}_2\overline{H}_1 - \overline{x} + \hat{\theta}_2 h_1(i+1), i \in I\!N\right\}\right.\right)}{G_k\left(0\left|\left\{-\hat{\theta}_2\overline{H}_1 - \overline{x} + \hat{\theta}_2 h_1(i), i \in I\!N\right\}\right.\right)}, \\ \hat{\theta}_1 = -\hat{\theta}_2\overline{H}_1 - \overline{x}, \end{cases} \tag{4.6}$$

where $\overline{x} = \frac{1}{n}\sum_{k=1}^{kmax} n_k k$ and $\overline{H}_1 = \frac{1}{n}\sum_{k=0}^{kmax} n_k h_1(k)$. The second equation of (4.6), with $h_1(i) = i$, gives $\hat{\theta}_1 = -\overline{x}\left(\hat{\theta}_2 + 1\right)$, which is the one obtained by Consul and Shoukri (1984) for the Generalized Poisson distribution. On the other hand, it is possible to show that, when $h_1(i) = i$, $\theta_1 < 0$ and $\theta_2 \in [-1,\ 0]$, the system (4.6) is equivalent to the one derived by these authors.

The first equation of (4.6), which provides the *MLE* for the parameter θ_2, is implicit. So, we have to use numerical methods to obtain the solution. To get the initial approximation of $\hat{\theta}_2$, we refer to the method described in paragraph 4.2.3.

4.2.3. If $U = \{\theta_1 + \theta_2 h_1(i) + \theta_3 h_2(i),\ i \in I\!N\}$, θ_1, θ_2, θ_3, $h_1(.)$, $h_2(.)$ such that U is negative and non-increasing, it is possible to obtain numerically the *MLE* of θ_1, θ_2 and θ_3, for instance using the method of Scoring, that can be found, e.g., in Panjer and Willmot (1992), pp. 326-328. We will utilize the starting values obtained by the "Ad Hoc" method (*ibidem*, pp. 303-305). The idea of "Ad Hoc" estimation is to equate sample statistics where "most of probability" is to corresponding theoretical quantities. Since there is a high proportion of zeros in the data sets concerning automobile claims, we propose estimators based upon lower classes frequencies, that is we equate the observed lower classes frequencies with the corresponding probabilities. The "Ad hoc" estimators $\ddot{\theta}_1$, $\ddot{\theta}_2$ and $\ddot{\theta}_3$ of the parameters θ_1, θ_2 and θ_3 are given by

$$\left\{\begin{array}{l} \ddot{\theta}_1 = ln(f_0) - \ddot{\theta}_2 h_1(0) - \ddot{\theta}_3 h_2(0)\,;\ \ddot{\theta}_2 = \frac{\alpha - \ln(f_0) - \ddot{\theta}_3(h_2(1) - h_2(0))}{h_1(1) - h_1(0)}\,; \\ \ddot{\theta}_3 = \frac{(\beta - ln(f_0))(h_1(1) - h_1(0)) + (h_1(0) - h_1(2))(\alpha - ln(f_0))}{(h_2(1) - h_2(0))(h_1(0) - h_1(2)) + (h_2(2) - h_2(0))(h_1(1) - h_1(0))}, \end{array}\right. \qquad (4.7).$$

where $\alpha = ln(f_1) - ln(-ln(f_0))$ and $\beta = ln(f_2) - ln\left[-0.5(ln(f_0))^2 + \alpha\, ln(f_0)\right]$.

5. LEAST SQUARES TYPE ESTIMATORS

Let us first establish the following result. For any discrete observed distribution defined on a subset of $I\!N$, $\{f_k;\ 0 \le k \le kmax\}$, such that $f_k > 0$ for all k, there exists a unique family $\tilde{U}$ such that $f_k = e^{\tilde{u}_k} G_k(0|\tilde{U})$, for $k = 0, 1, \ldots, kmax$. The $\tilde{u}_i$'s are defined recursively as $\tilde{u}_i = ln(f_i) - ln(G_i(0|\tilde{U}))$ $i = 0, 1, \ldots, kmax$ (let us quote that, by definition, the AG polynomial $G_i(x|\tilde{U})$ depends only on $\tilde{u}_0, \tilde{u}_i, \ldots, \tilde{u}_{i-1}$). Note that $\tilde{U}$ built above is not always negative and non-increasing. In practice, it is often preferable to only consider those $\tilde{u}_i$'s which form a negative non-increasing family.

The idea is to minimize the function $\mathcal{S}(\theta_1, \ldots, \theta_m)$ defined by

$$\mathcal{S}(\theta_1, \ldots, \theta_m) = \frac{1}{n}\sum_{k=0}^{k\max} n_k(u_k(\theta_1, \ldots, \theta_m) - \tilde{u}_k)^2,$$

where $\tilde{U}$ is the optimal family constructed above. The Least Square type estimators ($LSTE$, in short) proposed here are thus those which minimize $\mathcal{S}(\theta_1, \ldots, \theta_m)$, the weighted sum of the squared differences between the $\tilde{u}_i$'s and the u_i's having a specified parametric form. Let us mention that if we want to fit an observed distribution $\{(k, n_k),\ 0 \le k \le kmax\}$, we must at first group the classes in order to have all the n_k's positive.

The main advantage of this method is that it often provides explicit expressions for the estimators, as well as accurate fits. We give below the estimators in the case $U = \{\theta_1 + \theta_2 h_1(i) + \theta_3 h_2(i),\ i \in I\!N\}$, parameters θ_1, θ_2 and θ_3, and functions $h_1(.)$ and $h_2(.)$ such that U is negative and non-increasing. The estimators $\breve{\theta}_1$, $\breve{\theta}_2$ and $\breve{\theta}_3$ of the parameters θ_1, θ_2 and θ_3 are those which minimize

$$\mathcal{S}(\theta_1, \theta_2, \theta_3) = \frac{1}{n}\sum_{k=0}^{kmax} n_k(\theta_1 + \theta_2 h_1(k) + \theta_3 h_2(k) - \tilde{u}_k)^2.$$

They are given by

$$\breve{\theta}_3 = \frac{H_{12}H_{1u} - H_{11}H_{2u}}{(H_{12})^2 - H_{11}H_{22}}\,;\ \breve{\theta}_2 = -\theta_3\frac{H_{12}}{H_{11}} + \frac{H_{1u}}{H_{11}}\,;\ \breve{\theta}_1 = \overline{U} - \breve{\theta}_2\overline{H}_1 - \breve{\theta}_3\overline{H}_2, \qquad (5.1)$$

where $\overline{U} = \frac{1}{n}\sum_{i=0}^{kmax} n_i\tilde{u}_i$; $\overline{H}_j = \frac{1}{n}\sum_{i=0}^{kmax} n_i h_j(i), j = 1, 2$; $H_{j\times k} = \frac{1}{n}\sum_{i=0}^{kmax} n_i h_j(i) h_k(i), j, k = 1, 2$; $H_{j\times u} = \frac{1}{n}\sum_{i=0}^{kmax} n_i h_j(i)\tilde{u}_i, j = 1, 2$; $H_{jk} = H_{j\times k} - \overline{H}_j\overline{H}_k$, $j, k = 1, 2$; $H_{ju} = H_{j\times u} - \overline{H}_j\overline{U}, j = 1, 2$.

6. CONCLUDING REMARKS

Looking at the numerical results presented in Appendix, we could say that the $\mathcal{PG}$ (U) law seems to be suitable to fit the discrete data sets met in automobile insurance. The fits are more accurate than most of the ones discussed before, and applied to recent data sets coming from Belgium, the methods proposed here provide good fits. Moreover, the underlying probabilistic model is intuitively acceptable. Nevertheless, other authors, like for instance Kestemont and Paris (1985), also provided accurate fits, but sometimes with more intricate models. On the other hand, the Least Squares type method is easy to understand and provides explicit expressions for the estimators of the parameters, while it yields satisfying results. We also mention that the simulation method proposed in Devroye (1992), which consists in the partial recreation of the queueing system described in Section 2, can easily be used to simulate the number of claims that affect some automobile insurance portfolio.

ACKNOWLEDGEMENTS

We greatefully thank Professor Claude Lefèvre for all the time he spent to guide us, and Professor Jean-Marie Reinhard for his precious advices. The author also acknowledges the insightful comments of the two anonymous referees. Their interesting suggestions really improved the paper. We also express our gratitude to the belgian insurance companies for providing us the two last data sets of Appendix.

APPENDIX: NUMERICAL RESULTS

The reader will find here the fits of the six data sets in Gossiaux and Lemaire (1981), as well as of two recent belgian data sets, obtained with the $\mathcal{PG}$ (U), $U = \{\theta_1 + \theta_2 i + \theta_3 i^2, \, i \in I\!N\}$. To measure the goodness-of-fit, standard χ^2-statistics is used, with the following grouping procedure: the outside classes are gathered together in order to get theoretical class sizes greater or equal to 5 (that is, Rule B in Lemaire (1995)).

Belgium 1975-76				Zaïre 1974			
MLE: $\hat{\theta}_1 = -0.0981$, $\hat{\theta}_2 = -0.0250$; $\hat{\theta}_3 = -0.0037$ *LSTE*: $\check{\theta}_1 = -0.0981$; $\check{\theta}_2 = -0.0212$. $\check{\theta}_3 = -0.0069$				*MLE*. $\hat{\theta}_1 = -0.0728$; $\hat{\theta}_2 = -0.1546$; $\hat{\theta}_3 = -0.0005$ *LSTE*: $\check{\theta}_1 = -0.0728$; $\check{\theta}_2 = -0.1429$; $\check{\theta}_3 = -0.0110$			
k	n_k	ML	LST	k	n_k	ML	LST
0	96 978	96 978.16	96 975.53	0	3 719	3 719.00	3 719.06
1	9 240	9 244.4	9 252.45	1	232	231.98	232.22
2	704	693.27	684.44	2	38	38.00	37.06
3	43	53.19	55.48	3	7	8.22	8.27
4	9	4.50	5.37	4	3	2.03	2.26
≥ 5	0	0.49	0.73	5	1	0.55	0.71
				≥ 6	0	0.22	0.43
χ^2_{obs}		0.82	4.76	χ^2_{obs}		0.00	0.06

Belgium 1958			
MLE: $\hat{\theta}_1 = -0.1879$; $\hat{\theta}_2 = -0.1045$; $\hat{\theta}_3 = -0.0078$			
LSTE: $\tilde{\theta}_1 = -0.1883$; $\tilde{\theta}_2 = -0.0699$; $\tilde{\theta}_3 = -0.0337$			
k	n_k	ML	LST
0	7 840	7 840.00	7 836.81
1	1 317	1 316.97	1 330.77
2	239	239.00	221.81
3	42	49.31	48.47
4	14	11.48	13.77
5	4	2.98	4.84
6	4	0.85	2.01
7	1	0.26	0.95
≥ 8	0	0.14	1.57
χ^2_{obs}		3.38	2.36

Great-Britain 1968			
MLE: $\hat{\theta}_1 = -0.1285$; $\hat{\theta}_2 = -0.0182$; $\hat{\theta}_3 = -0.0048$			
LSTE: $\tilde{\theta}_1 = -0.1289$; $\tilde{\theta}_2 = -0.0183$; $\tilde{\theta}_3 = -0.0048$			
k	n_k	ML	LST
0	370 412	370 444.17	370 412.38
1	46 545	46 519.72	46 544.74
2	3 935	3 928.66	3 934.60
3	317	316.79	317.53
4	28	27.60	27.68
5	3	2.71	2.72
≥ 6	0	0.35	0.35
χ^2_{obs}		0.03	0.00

Switzerland 1961			
MLE: $\hat{\theta}_1 = -0.1447$; $\hat{\theta}_2 = -0.0555$; $\hat{\theta}_3 = -0.0099$			
LSTE: $\tilde{\theta}_1 = -0.1447$; $\tilde{\theta}_2 = -0.0571$; $\tilde{\theta}_3 = -0.0078$			
k	n_k	ML	LST
0	103 704	103 706.62	103 708.19
1	14 075	14 056.34	14 061.18
2	1 766	1 778.12	1 781.10
3	255	256.80	251.79
4	45	43.68	40.95
5	6	8.69	7.64
6	2	1.99	1.62
≥ 7	0	0.72	0.53
χ^2_{obs}		1.17	0.91

Germany 1960			
MLE: $\hat{\theta}_1 = -0.1359$; $\hat{\theta}_2 = -0.0387$; $\hat{\theta}_3 = -0.0154$			
LSTE: $\tilde{\theta}_1 = -0.1358$; $\tilde{\theta}_2 = -0.0414$; $\tilde{\theta}_3 = -0.0130$			
k	n_k	ML	LST
0	20 592	20 592.00	20 593.44
1	2 651	2 650.60	2 648.58
2	297	297.19	298.96
3	41	40.30	39.76
4	7	6.90	6.52
5	0	1.46	1.30
6	1	0.37	0.31
≥ 7	0	0.18	0.13
χ^2_{obs}		0.11	0.06

Belgium 1993			
MLE: $\hat{\theta}_1 = -0.1017$; $\hat{\theta}_2 = -0.0165$; $\hat{\theta}_3 = -0.0185$			
LSTE: $\tilde{\theta}_1 = -0.1017$; $\tilde{\theta}_2 = -0.0183$; $\tilde{\theta}_3 = -0.0168$			
k	n_k	ML	LST
0	57 178	57 178.02	57 179.55
1	5 617	5 615.00	5 613.30
2	446	448.56	450.23
3	50	48.25	47.39
4	8	7.22	6.82
≥ 5	0	1.95	1.72
χ^2_{obs}		0.23	0.22

Belgium 1994			
MLE: $\hat{\theta}_1 = -0.1000$; $\hat{\theta}_2 = -0.0253$; $\hat{\theta}_3 = -0.0095$			
LSTE: $\tilde{\theta}_1 = -0.1000$; $\tilde{\theta}_2 = -0.0243$; $\tilde{\theta}_3 = -0.0091$			
k	n_k	ML	LST
0	118 700	118 698.38	118 697.78
1	11 468	11 463.87	11 481.01
2	930	921.22	908.95
3	70	86.61	83.24
4	14	10.13	9.43
≥ 5	0	1.77	1.59
χ^2_{obs}		3.64	3.41

REFERENCES

CHARALAMBIDES, Ch.A. (1990). Abel series distributions with applications to fluctuations of sample functions of stochastic processes. *Commun. Statist. A, Theory Methods*, **19**, 255-263.

CONSUL, P.C. (1989). *Generalized Poisson Distributions, Properties and Applications*. Marcel Dekker. New York.

CONSUL, P.C. (1990). A model for distributions of injuries in auto-accidents. *MVSV*, 161-168.

CONSUL, P.C. and JAIN, G.C. (1973). A generalization of the Poisson distribution. *Technometrics*, **15**, 791-799.

CONSUL, P.C. and SHENTON, L.R. (1972). Use of Lagrange expansion for generating discrete generalized probability distributions. *SIAM J. Appl. Math.*, **23**, 239-248.

CONSUL, P.C. and SHOUKRI, M.M. (1984). Maximum Likelihood estimation for the Generalized Poisson distribution. *Comm. Statist. A, Theory Methods*, **13**, 1533-1547.

DEVROYE, L. (1992). The branching process method in Lagrange random variate generation. *Commun. Statist., Simula.*, **21**, 1-14.

ELVERS, E. (1991). A note on the Generalized Poisson distribution. *ASTIN Bulletin*, **21**, 167.

GERBER, H.U. (1990). When does the surplus reach a given target? *Insurance: Mathematics and Economics*, **9**, 115-119.

GOOVAERTS, M.J. and KAAS, R. (1991). Evaluating compound Generalized Poisson distributions recursively. *ASTIN Bulletin*, **21**, 193-198.

GOSSIAUX, A. M. and LEMAIRE, J. (1981). Méthodes d'ajustement de distributions de sinistres. *MVSV*, 87-95.

ISLAM, M.N. and CONSUL, P.C. (1992). A probabilistic model for automobile claims. *MVSV*, 85-93.

JANARDAN, K.G., KERSTER, H.W. and SCHAEFFER, D.J. (1979). Biological applications of the Lagrangian Poisson distribution. *Bioscience*, **29**, 599-602.

KAAS, R., VAN HEERWAARDEN, A.E. and GOOVAERTS, M.J. (1994). *Ordering of Actuarial Risks*. CAIRE (Education Series). Brussels.

KESTEMONT, R.-M. and PARIS, J. (1985). Sur l'ajustement du nombre de sinistres. *MVSV*, 157-164.

LEFÈVRE, Cl. and PICARD, Ph. (1996). On the first crossing of a Poisson process in a lower boundary. In *Athens Conference on Applied Probability and Time Series, Vol. I, Applied Probability*. C.C. Heyde, Yu V. Prohorov, R. Pyke and S.T. Rachev Editors. Lecture Notes in Statistics, **114**, pp. 159-175. Springer. Berlin.

LEMAIRE, J. (1991). Negative Binomial or Poisson-Inverse Gaussian? *ASTIN Bulletin*, **21**, 167-168.

LEMAIRE, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Klumer Academic Publishers. Boston.

OSKOLKOV, V.A. (1988). Abel-Goncharov problem. In *Encyclopaedia of Mathematics, Vol. I*, pp. 4-5. Kluwer Academic Publishers. Dordrecht.

PANJER, H.H. (1987). Models of claim frequency. In *Advances in the Statistical Sciences, Vol. VI, Actuarial Sciences*. I.B. Mac Neill and G.J. Umphrey Editors. The University of Western Ontario Series in Philosophy of Science, **39**, pp. 115-122. Reidel. Dordrecht.

PANJER, H.H. and WILLMOT, G.E. (1992). *Insurance Risk Models*. Society of Actuaries. Schaunburg III.

PANJER, H.H. and WANG, S. (1995). Computational aspects of Sundt's generalized class. *ASTIN Bulletin*, **25**, 5-17.

RUOHONEN, M. (1987). On a model for the claim number process. *ASTIN Bulletin*, **18**, 57-68.

SEAL, H.L. (1982). Mixed Poisson – an ideal distribution of claim number? *MVSV*, 293-295.

SHAKED, M. and SHANTHIKUMAR, J.G. (1994). *Stochastic Orders and their Applications*. Academic Press. New York.

SHARIF, A.H. and PANJER, H.H. (1993). A probabilistic model for automobile claims: a comment on the article by M.N. Islam and P.C. Consul. *MVSV*, 279-282.

SHARIF, A.H. and PANJER, H.H. (1995). An improved recursion for the compound Generalized Poisson distribution. *MVSV*, 93-98.

SUNDT, B. and JEWELL, W.S. (1981). Further results on recursive evaluation of compound distributions. *ASTIN Bulletin*, **12**, 27-39.

SUNDT, B. (1992). On some extension of Panjer's class of counting distributions. *ASTIN Bulletin*, **22**, 61-80.

TER BERG, P. (1996). A loglinear Lagrangian Poisson model. *ASTIN Bulletin*, **26**, 123-129.
WILLMOT, G.E. (1987). The Poisson-Inverse Gaussian distribution as an alternative to the Negative Binomial. *Scandinavian Actuarial Journal*, 113-127.
WILLMOT, G.E. (1988). Sundt and Jewell's family of discrete distributions. *ASTIN Bulletin*, **18**, 17-29.

MICHEL DENUIT
Institut de Statistique et de Recherche Opérationnelle
Universite libre de Bruxelles, CP 210
Boulevard du Triomphe
B-1050 Bruxelles
Belgium
e-mail: mdenuit@ulb.ac.be

ON ERROR BOUNDS FOR APPROXIMATIONS TO AGGREGATE CLAIMS DISTRIBUTIONS

BY JAN DHAENE

Katholieke Universiteit Leuven, Belgium

BJØRN SUNDT

University of Bergen, Norway

ABSTRACT

In the present paper we discuss error bounds for approximations to aggregate claims distributions. We consider approximations to convolutions by approximating each of the distributions and taking the convolution of these approximations. For compound distributions we consider two classes of approximations. In the first class we approximate the counting distribution, but keep the severity distribution unchanged, whereas in the second class we approximate the severity distribution, but keep the counting distribution unchanged. We finally look at some examples.

1. INTRODUCTION

During the last two decades there has developed a large literature on approximations to aggregate claims distributions and related functions, in particular their stop loss transforms. In the present paper we give bounds for some measures of errors caused by such approximations. These measures can also be applied as measures for the distance between two distributions.

In Section 2 we introduce some notation and conventions, and in a short Section 3 we present some simple inequalities for error bounds.

Approximations to convolutions of distributions is the topic of Section 4. We approximate a convolution by approximating each of the distributions in the convolution and then taking the convolution of the approximations.

Approximations to compound distributions is the topic of Section 5. We consider two classes of approximations. In the first class we approximate the counting distribution, but keep the severity distribution unchanged, whereas in the second class we approximate the severity distribution, but keep the counting distribution unchanged. Error bounds for approximations where both the counting distribution and the severity distribution are approximated, can be found by application of triangle inequalities.

In Section 6 we finally consider some applications. Further applications of results from the present paper are given in Dhaene & Sundt (1996).

ASTIN BULLETIN, Vol 27, No. 2, 1997, pp 243-262

The main topic of the present paper is approximations to probability distributions. These approximations are not necessarily distributions themselves. Sometimes one would apply an approximation that could be naturally split into more than one step, e.g. approximating a compound distribution by first approximating its counting distribution and then its severity distribution. In this situation one could first give bounds for the approximation error of the approximation with correct severity distribution and approximated counting distribution, then for the final approximation considered as an approximation to this intermediary approximation, and finally use triangle inequalities to assess the approximation error of the aggregate approximation. In such a procedure, the intermediary approximation would not necessarily be a distribution, and thus in our frame-work it is also of interest to discuss approximations to functions. On this background we have sometimes in our results assumed that the quantity to be approximated is a more general function than a probability distribution. Such generalisations are also possible in some of the other results where we for simplicity have made more restrictive assumptions.

2. Notation and conventions

In the present paper we shall be concerned with probability distributions on the non-negative integers. We shall approximate such distributions by approximating their discrete densities. Thus we identify a distribution by its discrete density, and for convenience we shall usually mean its discrete density when we talk about a distribution.

Let $\mathcal{P}$ denote the class of (discrete densities of) probability distributions on the non-negative integers. When discussing approximations to compound distributions, we shall restrict the severity distribution to the positive integers, and we therefore also introduce $\mathcal{P}_+$ as the class of distributions on the positive integers. As we shall approximate distributions in $\mathcal{P}$ and $\mathcal{P}_+$ by functions which are not necessarily distributions themselves, we shall also need the classes $\mathcal{F}$ and $\mathcal{F}_+$, being respectively the class of functions on the non-negative integers and the class of functions on the positive integers. We see that $\mathcal{P}_+ \subset \mathcal{P} \subset \mathcal{F}$ and $\mathcal{P}_+ \subset \mathcal{F}_+ \subset \mathcal{F}$.

For a function $f \in \mathcal{F}$ we introduce

$$\mu_j(f) = \sum_{x=0}^{\infty} x^j f(x) \qquad (j = 0, 1)$$

$$\Gamma_f(x) = \sum_{y=0}^{x} f(y) \qquad \Pi_f(x) = \sum_{y=x+1}^{\infty} (y - x) f(y). \qquad (x = 0, 1, 2, \ldots)$$

When the quantities $\mu_0(f)$ and $\mu_1(f)$ appear, it will always be silently assumed that they exist and are finite. When $\Pi_f(x)$ appears, it is assumed that $\mu_0(f)$ and $\mu_1(f)$ converge so that $\Pi_f(x)$ is well defined and has a finite value.

If $f \in \mathcal{P}$, then Γ_f is the corresponding cumulative distribution, Π_f the stop loss transform, $\mu_1(f)$ the mean, and $\mu_0(f) = 1$.

As the main purpose of this paper is to study the approximation error for approximations to a distribution, we introduce the following measures for the distance between two functions $f, g \in \mathcal{F}$:

$$\varepsilon_j(f,\ g) = \sum_{x=0}^{\infty} x^j |f(x) - g(x)| \qquad (j = 0,\ 1)$$

$$\eta(f,\ g) = \sup_{x \geq 0} \left| \Pi_f(x) - \Pi_g(x) \right|.$$

For evaluating the quality of an approximation only considered as an approximation to the discrete density, $\varepsilon_0(f,\ g)$ is a natural measure for the approximation error. If we want to evaluate the corresponding approximation to the stop loss transform, then $\eta(f, g)$ is a natural measure. We see that $\varepsilon_0(f,\ g)$, $\varepsilon_1(f,\ g)$, and $\eta(f, g)$ are equal to zero if and only if $f = g$

By the notation x_+ we shall mean the maximum of x and zero.

We denote by I the indicator function defined by $I(A) = 1$ if the condition A is true and $I(A) = 0$ if it is false.

We shall interpret $\Sigma_{i=a}^{b} v_i = 0$ and $\Pi_{i=a}^{b} v_i = 1$ when $b < a$.

3. SOME USEFUL INEQUALITIES

The following lemma gives some useful inequalities that we shall need later.

Lemma 3.1 *For $f,\ g,\ h \in \mathcal{F}$ and $j = 0, 1$, we have*

$$\varepsilon_j(f,\ g) \leq \varepsilon_j(f,\ h) + \varepsilon_j(h,\ g) \tag{3.1}$$

$$\eta(f,\ g) \leq \eta(f,\ h) + \eta(h,\ g) \tag{3.2}$$

$$|\mu_1(f) - \mu_1(g)| \leq \varepsilon_1(f,\ g); \tag{3.3}$$

and for $f,\ g \in \mathcal{P}$

$$|f(0) - g(0)| \leq \frac{1}{2}\varepsilon_0(f,\ g) \leq \varepsilon_1(f,\ g). \tag{3.4}$$

Proof. The inequalities (3.1)-(3.3) are obvious.

For (3.4) we have

$$\varepsilon_0(f,\ g) - 2|f(0) - g(0)| = \sum_{x=1}^{\infty} |f(x) - g(x)| - \left| \sum_{x=1}^{\infty} (f(x) - g(x)) \right| \geq 0,$$

which proves the first inequality. Furthermore,

$$\varepsilon_0(f,\ g) = \left|\sum_{x=1}^{\infty}(f(x) - g(x))\right| + \sum_{x=1}^{\infty}|f(x) - g(x)| \le$$

$$2\sum_{x=1}^{\infty} x|f(x) - g(x)| = 2\varepsilon_1(f,\ g),$$

which proves the second inequality.

This completes the proof of Lemma 3.1. Q.E.D.

4. CONVOLUTIONS

4A. When for $i = 1, ..., m$ approximating $f_i \in \mathcal{P}$ by $g_i \in \mathcal{F}$, which is not necessarily in $\mathcal{P}$ itself, it is also natural to approximate the convolution $*_{i=1}^{m} f_i$ by $*_{i=1}^{m} g_i$. The convolution $h_1 * h_2$ of two functions h_1 and h_2 on the non-negative integers is defined by

$$(h_1 * h_2)(x) = \sum_{y=0}^{x} h_1(y)h_2(x-y); \qquad (x = 0,\ 1,\ ...)$$

we also define $h^{0*}(x) = I(x = 0)$ for a function h on the non-negative integers.

The following well-known properties of convolutions of distributions in $\mathcal{P}$ also hold for convolutions of functions in $\mathcal{F}$:

$$h_1 * h_2 = h_2 * h_1$$

$$(h_1 * h_2) * h_3 = h_1 * (h_2 * h_3)$$

$$h_1 * h_3 + h_2 * h_3 = (h_1 + h_2) * h_3.$$

Furthermore, we easily see that

$$|h_1 * h_2| \le |h_1| * |h_2|$$

$$\mu_j(h_1) \le \mu_j(h_2). \qquad (h_1 \le h_2;\ j = 0,\ 1)$$

Lemma 4.1 *If $h_1,\ h_2 \in \mathcal{F}$ such that $\mu_0(|h_i|) < \infty$ for* i $= 1, 2$, *then*

$$\mu_0(h_1 * h_2) = \mu_0(h_1)\mu_0(h_2).$$

Proof. We have

$$\mu_0(h_1 * h_2) = \sum_{x=0}^{\infty} (h_1 * h_2)(x) = \sum_{x=0}^{\infty} \sum_{y=0}^{x} h_1(y) h_2(x-y) =$$

$$\sum_{y=0}^{\infty} h_1(y) \sum_{x=y}^{\infty} h_2(x-y) = \mu_0(h_1)\mu_0(h_2).$$

Q.E.D.

4B. We shall first consider bounds for $\varepsilon_0\left(*_{i=1}^{m} f_i, *_{i=1}^{m} g_i\right)$. For the proof of our main result we shall need the following lemma.

Lemma 4.2 *For $f,\ g,\ h \in \mathcal{F}$ we have*

$$\varepsilon_0(f * h,\ g * h) \leq \mu_0(|h|)\varepsilon_0(f,\ g).$$

Proof. We have

$$\varepsilon_0(f * h,\ g * h) = \sum_{x=0}^{\infty} |(f * h)(x) - (g * h)(x)| =$$

$$\sum_{x=0}^{\infty} \left| \sum_{y=0}^{x} h(y)(f(x-y) - g(x-y)) \right| \leq \sum_{x=0}^{\infty} \sum_{y=0}^{x} |h(y)||f(x-y) - g(x-y)| =$$

$$\sum_{y=0}^{\infty} |h(y)| \sum_{x=y}^{\infty} |f(x-y) - g(x-y)| = \mu_0(|h|)\varepsilon_0(f,\ g).$$

Q.E.D.

Theorem 4.1 *For $f_i,\ g_i \in \mathcal{F}$ $(i = 1, \ldots, m)$, we have*

$$\varepsilon_0\left(\mathop{*}_{i=1}^{m} f_i, \mathop{*}_{i=1}^{m} g_i\right) \leq \sum_{i=1}^{m} \varepsilon_0(f_i, g_i)\left(\prod_{j=1}^{i-1} \mu_0(|f_j|)\right)\left(\prod_{j=i+1}^{m} \mu_0(|g_j|)\right). \tag{4.1}$$

Proof. If $\mu_0(|f_i|) = \infty$ or $\mu_0(|g_i|) = \infty$ for some i, then the theorem obviously holds. Let us therefore assume that $\mu_0(f_i)$ and $\mu_0(g_i)$ are finite for all i. Under this assumption we shall prove (4.1) by induction on m. For $m = 1$ it trivially holds. We now assume that it holds for $m = 1, \ldots, n$. By using successively (3.1), Lemma 4.2, Lemma 4.1, and (4.1), we obtain

$$\varepsilon_0\left(\mathop{*}_{i=1}^{n+1} f_i, \mathop{*}_{i=1}^{n+1} g_i\right) \leq \varepsilon_0\left(\mathop{*}_{i=1}^{n+1} f_i, \left(\mathop{*}_{i=1}^{n} f_i\right) * g_{n+1}\right) + \varepsilon_0\left(\left(\mathop{*}_{i=1}^{n+1} f_i\right) * g_{n+1}, \mathop{*}_{i=1}^{n+1} g_i\right) \leq$$

$$\mu_0\left(\left|\mathop{*}_{i=1}^{n} f_i\right|\right)\varepsilon_0(f_{n+1}, g_{n+1}) + \mu_0(|g_{n+1}|)\varepsilon_0\left(\mathop{*}_{i=1}^{n} f_i, \mathop{*}_{i=1}^{n} g_i\right) \leq$$

$$\left(\prod_{i=1}^{n} \mu_0(|f_i|)\right)\varepsilon_0(f_{n+1}, g_{n+1}) +$$

$$\mu_0(|g_{n+1}|)\left\{\sum_{i=1}^{n} \varepsilon_0(f_i, g_i)\left(\prod_{j=1}^{i-1} \mu_0(|f_j|)\right)\left(\prod_{j=i+1}^{n} \mu_0(|g_j|)\right)\right\} =$$

$$\sum_{i=1}^{n+1} \varepsilon_0(f_i, g_i)\left(\prod_{j=1}^{i-1} \mu_0(|f_j|)\right)\left(\prod_{j=i+1}^{n} \mu_0(|g_j|)\right),$$

that is, (4.1) also holds for $m = n + 1$. By induction it holds for all m. Q.E.D.

One somewhat disappointing aspect of Theorem 4.1 is that the upper bound in (4.1) is not in general invariant against permutations of the pairs (f_i, g_i) $(i = 1, \ldots, m)$. However, in the special case when $f_i, g_i \in \mathcal{P}$, (4.1) reduces to

$$\varepsilon_0\left(\mathop{*}_{i=1}^{m} f_i, \mathop{*}_{i=1}^{m} g_i\right) \leq \sum_{i=1}^{m} \varepsilon_0(f_i, g_i),$$

which is invariant.

4C. For $\eta\left(*_{i=1}^{m} f_i, *_{i=1}^{m} g_i\right)$ we have the following result.

Theorem 4.2 *For $f_i, g_i \in \mathcal{P}$ $(i = 1, \ldots, m)$, we have*

$$\Pi_{*_{i=1}^{m} f_i}(x) - \Pi_{*_{i=1}^{m} g_i}(x) \leq \sum_{i=1}^{m} \sup_{y \geq 0}\left(\Pi_{f_i}(y) - \Pi_{g_i}(y)\right) \qquad (x = 0, 1, 2, \ldots) \tag{4.2}$$

$$\eta\left(\mathop{*}_{i=1}^{m} f_i, \mathop{*}_{i=1}^{m} g_i\right) \leq \sum_{i=1}^{m} \eta(f_i, g_i). \tag{4.3}$$

Proof. Formula (4.2) follows from Lemma 6 in De Pril & Dhaene (1992), and (4.3) follows immediately from (4.2). Q.E.D.

In (4.2) we gave an upper bound for the difference between the two stop loss transforms. By symmetry we can immediately obtain an analogous lower bound. Similarly, we shall also in the following often present our results only with upper bounds when the analogous lower bounds follow immediately by symmetry.

5. COMPOUND DISTRIBUTIONS

5A. In this section we shall discuss approximations to compound distributions. For simplicity we assume that the severity distribution is in $\mathcal{P}_+$.

We denote the compound distribution with counting distribution $p \in \mathcal{P}$ and severity distribution $h \in \mathcal{P}_+$ by $p \vee h$, that is,

$$(p \vee h)(x) = \sum_{n=0}^{x} p(n) h^{n*}(x), \qquad (x = 0,\ 1,\ 2,\ \ldots)$$

and we extend this definition of the function $p \vee h$ to the case when $p \in \mathcal{F}$ and $h \in \mathcal{F}_+$.

5B. We first consider the case when we approximate a compound distribution by approximating the counting distribution and keeping the severity distribution unchanged.

Theorem 5.1 *For $p,\ q \in \mathcal{F}$ and $h \in \mathcal{F}_+$ with $\mu_0(|h|) \leq 1$, we have*

$$\varepsilon_0(p \vee h,\ q \vee h) \leq \varepsilon_0(p,\ q). \tag{5.1}$$

Proof. We have

$$\varepsilon_0(p \vee h,\ q \vee h) = \sum_{x=0}^{\infty} |(p \vee h)(x) - (q \vee h)(x)| =$$

$$\sum_{x=0}^{\infty} \left| \sum_{n=0}^{\infty} (p(n) - q(n)) h^{n*}(x) \right| \leq \sum_{x=0}^{\infty} \sum_{n=0}^{\infty} |p(n) - q(n)| |h^{n*}|(x) =$$

$$\sum_{n=0}^{\infty} |p(n) - q(n)| \sum_{x=0}^{\infty} |h^{n*}|(x) = \sum_{n=0}^{\infty} |p(n) - q(n)| \mu_0(|h^{n*}|) \leq$$

$$\sum_{n=0}^{\infty} |p(n) - q(n)| \mu_0^n(|h|) \leq \sum_{n=0}^{\infty} |p(n) - q(n)| = \varepsilon_0(p,\ q).$$

Q.E.D.

To deduce bounds for the approximation error for approximations to stop loss premiums, we shall need the following lemma, which is proved as formula (38) in De Pril & Dhaene (1992).

Lemma 5.1 *For $f \in \mathcal{P}$ we have*

$$n\Pi_f(x) \leq \Pi_{f^{n*}}(x) \leq (n-1)\mu_1(f) + \Pi_f(x). \qquad (x = 0,\ 1,\ \ldots;\ n = 1,\ 2,\ \ldots)$$

Theorem 5.2 *For $h \in \mathcal{P}_+$, $p,\ q \in \mathcal{F}$ with $\mu_1(|p|) < \infty, \mu_1(|q|) < \infty$, and* $B(p,\ q) = \varepsilon_1(p,\ q) - \varepsilon_0(p,\ q) + 2(p(0) - q(0))_+ + \mu_1(p) - \mu_1(q) - \mu_0(p) + \mu_0(q)$, *we have*

$$\Pi_{p\vee h}(x) = \Pi_{q\vee h}(x) \leq \frac{1}{2}(\mu_1(h) - \Pi_h(x))B(p,\ q) + \Pi_h(x)(\mu_1(p) - \mu_1(q))$$

$$(x = 0,\ 1,\ 2,\ ...) \tag{5.2}$$

$$\eta(p \vee h,\ q \vee h) \leq \frac{1}{2}\mu_1(h)(\varepsilon_1(p,\ q) + |\mu_1(p) - \mu_1(q)|) \leq \mu_1(h)\varepsilon_1(p,\ q). \tag{5.3}$$

Proof. For $x = 0, 1, 2, \ldots$, we have

$$\Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) = \sum_{y=z+1}^{\infty} (y - x)((p \vee h)(y) - (q \vee h)(y)) =$$

$$\sum_{y=x+1}^{\infty} (y - x) \sum_{n=1}^{\infty} (p(n) - q(n))h^{n*}(n) = \sum_{n=1}^{\infty} (p(n) - q(n))\Pi_{h^{n*}}(x),$$

from which we obtain

$$\Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) = \sum_{n=1}^{\infty} (p(n) - q(n))(\Pi_{h^{n*}}(x) - n\Pi_h(x)) +$$

$$\Pi_h(x)(\mu_1(p) - \mu_1(q)). \tag{5.4}$$

Two applications of Lemma 5.1 give

$$\sum_{n=1}^{\infty} (p(n) - q(n))(\Pi_{h^{n*}}(x) - n\Pi_h(x)) \leq$$

$$\sum_{n=1}^{\infty} (p(n) - q(n))_+(\Pi_{h^{n*}}(x) - n\Pi_h(x)) \leq$$

$$\sum_{n=1}^{\infty} (p(n) - q(n))_+(n - 1)(\mu_1(h) - \Pi_h(x)) =$$

$$\tfrac{1}{2}(\mu_1(h) - \Pi_h(x)) \sum_{n=1}^{\infty} (|p(n) - q(n)| + p(n) - q(n))(n - 1) =$$

$$\tfrac{1}{2}(\mu_1(h) - \Pi_h(x))B(p,\ q),$$

which together with (5.4) proves (5.2).

As

$$B(p, q) = \varepsilon_1(p, q) + \mu_1(p) - \mu_1(q) - 2\sum_{n=1}^{\infty}(p(n) - q(n))_+ \le \varepsilon_1(p, q) + \mu_1(p) - \mu_1(q),$$

(5.2) gives

$$\Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \le \frac{1}{2}(\mu_1(h) - \Pi_h(x))(\varepsilon_1(p, q) + \mu_1(p) - \mu_1(q)) +$$

$$\Pi_h(x)(\mu_1(p) - \mu_1(q)) \le \frac{1}{2}\mu_1(h)(\varepsilon_1(p, q) + \mu_1(p) - \mu_1(q)).$$

Together with the analogous inequality with interchanging of p and q, this gives the first inequality in (5.3); the last inequality in (5.3) follows by (3.3).

This completes the proof of Theorem 5.2. Q.E.D.

The following theorem is a special case of Theorem 1 in Sundt & Dhaene (1996).

Theorem 5.3 *For $p, q \in \mathcal{P}$ and $h \in \mathcal{P}_+$, we have*

$$\Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \le \left(\mu_1(h) - \Pi_h(x)\right)\Pi_p(1) + \Pi_h(x)(\mu_1(p) - \mu_1(q)).$$

$$(x = 0, 1, 2, \ldots) \tag{5.5}$$

The bounds in (5.1), (5.2), and (5.3) become equal to zero when $p = q$. Unfortunately, this is not the case with the bound in (5.5) unless $\Pi_p(1) = 0$, that is, p is a Bernoulli distribution. On the other hand, we see that the bound in (5.5) is sharper than the bound in (5.2) when $\Pi_p(1) = 0$ and $p \neq q$. We shall discuss this case in more detail in subsection 6.2.

5C. Let us now consider the special case with $h \in \mathcal{P}_+$ and $p, q \in \mathcal{P}$ with $\mu_1(p) = \mu_1(q)$. In that case (5.2), (5.5), and (5.3) reduce to respectively

$$\Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \le$$

$$\frac{1}{2}(\mu_1(h) - \Pi_h(x))\left(\varepsilon_1(p, q) - \varepsilon_0(p, q) + 2(p(0) - q(0))_+\right) \qquad (x = 0, 1, 2, \ldots) \tag{5.6}$$

$$\Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \le \left(\mu_1(h) - \Pi_h(x)\right)\Pi_p(1) \qquad (x = 0, 1, 2, \ldots) \tag{5.7}$$

$$\eta(p\vee h, q\vee h) \le \frac{1}{2}\mu_1(h)\varepsilon_1(p, q). \tag{5.8}$$

From (5.6) we obtain

$$\eta(p \vee h,\ q \vee h) \leq \frac{1}{2}\mu_1(h)(\varepsilon_1(p,\ q) - \varepsilon_0(p,\ q) + 2|p(0) - q(0)|). \tag{5.9}$$

From (3.4) we see that this is sharper than or equal to the bound in (5.8).

We see that the bounds in (5.6) and (5.7) are non-decreasing in x. For $x = 0$ these bounds become equal to zero.

5D. In subsections 5B-C we discussed approximating a compound distribution by approximating the counting distribution and keeping the severity distribution unchanged. Let us now instead consider approximating the severity distribution and keeping the counting distribution unchanged. For such approximations we have the following theorem:

Theorem 5.4 *For $p \in \mathcal{F}$ and $h,\ k \in \mathcal{F}_+$ with $\mu_0(|h|) \leq 1$ and $\mu_0(|k|) \leq 1$, we have*

$$\varepsilon_0(p \vee h,\ p \vee k) \leq \mu_1(|p|)\varepsilon_0(h,\ k). \tag{5.10}$$

If in addition $h,\ k \in \mathcal{P}_+$, then

$$\eta(p \vee h,\ p \vee k) \leq \mu_1(|p|)\eta(h,\ k). \tag{5.11}$$

Proof. By application of Theorem 4.1 we obtain

$$\varepsilon_0(p \vee h,\ p \vee k) = \sum_{x=0}^{\infty} |(p \vee h)(x) - (p \vee k)(x)| =$$

$$\sum_{x=1}^{\infty}\left|\sum_{n=1}^{\infty} p(n)(h^{n*}(x) - k^{n*}(x))\right| \leq \sum_{x=1}^{\infty}\sum_{n=1}^{\infty} |p(n)||h^{n*}(x) - k^{n*}(x)| =$$

$$\sum_{n=1}^{\infty} |p(n)| \sum_{x=1}^{\infty} |h^{n*}(x) - k^{n*}(x)| = \sum_{n=1}^{\infty} |p(n)|\varepsilon_0(h^{n*}, k^{n*}) \leq$$

$$\sum_{n=1}^{\infty} |p(n)| n\varepsilon_0(h,\ k) = \mu_1(|p|)\varepsilon_0(h,\ k),$$

which proves (5.10).

We now assume that $h,\ k \in \mathcal{P}_+$. For $x = 0, 1, 2, \ldots$ we obtain

$$|\Pi_{p\vee h}(x) - \Pi_{p\vee k}(x)| = \left|\sum_{n=1}^{\infty} p(n)(\Pi_{h^{n*}}(x) - \Pi_{k^{n*}}(x))\right| \leq$$

$$\sum_{n=1}^{\infty} |p(n)||\Pi_{h^{n*}}(x) - \Pi_{k^{n*}}(x)| \leq \sum_{n=1}^{\infty} |p(n)|\eta(h^{n*}, k^{n*}).$$

Application of (4.3) gives

$$|\Pi_{p\vee h}(x) - \Pi_{p\vee k}(x)| \leq \sum_{n=1}^{\infty} |p(n)| n\eta(h, k) = \mu_1(|p|)\eta(h, k),$$

from which we obtain (5.11).

This completes the proof of Theorem 5.3. Q.E.D.

5E. We shall now discuss two classes of approximations that can be convenient both for the counting distribution and the severity distribution in a compound distribution.

For $f \in \mathcal{P}$ we define the approximation $f^{(r)}$ for a positive integer r by

$$f^{(r)}(x) = f(x) I(x \leq r). \qquad (x = 0, 1, 2, \ldots)$$

As

$$\mu_j\left(f^{(r)}\right) = \sum_{x=0}^{r} x^j f(x) \qquad \varepsilon_j\left(f, f^{(r)}\right) = \sum_{x=r+1}^{\infty} x^j f(x), \qquad (j = 0, 1)$$

we obtain

$$\varepsilon_j\left(f, f^{(r)}\right) + \mu_j\left(f^{(r)}\right) = \mu_j(f) \qquad (j = 0, 1) \tag{5.12}$$

$$\mu_0\left(f^{(r)}\right) = \Gamma_f(r) \tag{5.13}$$

$$\varepsilon_1\left(f, f^{(r)}\right) = \Pi_f(r) + r\left(1 - \Gamma_f(r)\right).$$

As $f(x) \geq f^{(r)}(x)$ for $x = 0, 1, 2, \ldots, \Pi_f(x) - \Pi_{f^{(r)}}(x)$ is non-negative and non-increasing in x, and we obtain

$$\eta\left(f, f^{(r)}\right) = \Pi_f(r) - \Pi_{f^{(r)}}(0) = \mu_1(f) - \mu_1\left(f^{(r)}\right) = \varepsilon_1\left(f, f^{(r)}\right).$$

We see that unless $\Gamma_f(r) = 1$, the approximation $f^{(r)}$ will not be a proper distribution as $\mu_0(f^{(r)}) < \mu_0(f) = 1$. To obtain a proper distribution, we can apply the modified approximation $\tilde{f}^{(r)}$ defined by

$$\tilde{f}^{(r)}(x) = \begin{cases} f(x) & (x = 0, 1, \ldots, r-1) \\ 1 - \Gamma_f(r-1) & (x = r) \\ 0. & (x = r+1, r+2, \ldots) \end{cases}$$

For $j = 0, 1$ we get

$$\mu_j\left(\tilde{f}^{(r)}\right) = \mu_j\left(f^{(r)}\right) + r^j\left(1 - \Gamma_f(r)\right)$$

$$\varepsilon_j\left(f, \tilde{f}^{(r)}\right) = \varepsilon_j\left(f, f^{(r)}\right) + r^j\left(1 - \Gamma_f(r)\right).$$

It is easily shown that

$$\Pi_f(x) - \Pi_{\tilde{f}^{(r)}}(x) = \Pi_f(\max(x, r)), \qquad (x = 0,\ 1,\ 2,\ \ldots)$$

from which we obtain

$$\eta\left(f, \tilde{f}^{(r)}\right) = \Pi_f(r). \tag{5.14}$$

If X is a random variable with distribution f, then $\tilde{f}^{(r)}$ is the distribution of $\tilde{X}^{(r)} = \min(X, r)$. As $\tilde{X}^{(r)} \leq X$, we immediately obtain inequalities like

$$\Gamma_{\tilde{f}^{(r)}}(x) \geq \Gamma_f(x) \qquad \Pi_{\tilde{f}^{(r)}}(x) \leq \Pi_f(x). \qquad (x = 0,\ 1,\ 2,\ \ldots)$$

Theorem 5.5 *If p, $h \in \mathcal{P}$ and r and x are positive integers, then*

$$0 \leq \Pi_{p \vee h}(x) - \Pi_{\tilde{p}^{(r)} \vee h}(x) \leq \mu_1(h)\Pi_p(r) \tag{5.15}$$

$$0 \leq \Pi_{p \vee h}(x) - \Pi_{p \vee \tilde{h}^{(r)}}(x) \leq \Pi_h(r)\mu_1(p). \tag{5.16}$$

Proof. Sundt (1991) proved (5.15). The last inequality in (5.16) follows from Theorem 5.4 and (5.14), and the first inequality is immediately seen by interpreting $\Pi_{p \vee h}(x) - \Pi_{p \vee \tilde{h}^{(r)}}(x)$ as the mean of a non-negative random variable.
This completes the proof of Theorem 5.5. Q.E.D.

We notice that

$$\varepsilon_j\left(f, \tilde{f}^{(r)}\right) \geq \varepsilon_j\left(f, f^{(r)}\right) \qquad (j = 0,\ 1)$$

$$\eta\left(f, \tilde{f}^{(r)}\right) \leq \eta\left(f, f^{(r)}\right).$$

5F. By combining the results from Section 5 with the results from Section 4, we can obtain error bounds for approximations to convolutions of compound distributions. For a simple illustration, let $p_i \in \mathcal{P}$ and $h_i \in \mathcal{P}_+$ ($i = 1, \ldots, m$). From Theorem 4.1, (5.1), (5.12), and (5.13), we obtain

$$\varepsilon_0\left(\mathop{*}_{i=1}^{m}(p_i \vee h_i), \mathop{*}_{i=1}^{m}\left(p_i^{(r)} \vee h_i\right)\right) \leq \sum_{i=1}^{m} \varepsilon_0\left(p_i \vee h_i, p_i^{(r)} \vee h_i\right) \leq$$

$$\sum_{i=1}^{m} \varepsilon_0\left(p_i, p_i^{(r)}\right) = \sum_{i=1}^{m} \left(1 - \Gamma_{p_i}(r)\right).$$

6. APPLICATIONS

6.1. Introduction

In this section we shall under various assumptions discuss approximations to compound distributions by approximating the counting distribution with another distribution with the same mean and keeping the severity distribution fixed, that is, we want to approximate $p \vee h$ with $q \vee h$ when $p, q \in \mathcal{P}$, $h \in \mathcal{P}_+$ and $\mu_1(q) = \mu_1(p)$.

6.2. Bernoulli distribution

Lemma 6.1 *If p is a Bernoulli distribution and $q \in \mathcal{P}$ with $\mu_1(q) = \mu_1(p)$, then*

$$q(0) \geq p(0) \qquad q(1) \leq p(1) \tag{6.1}$$

$$\varepsilon_0(p, q) = \varepsilon_1(p, q) = 2(p(1) - q(1)) \tag{6.2}$$

$$\Pi_p(1) = 0 \tag{6.3}$$

$$\Pi_q(1) = q(0) - p(0). \tag{6.4}$$

Proof. We have

$$1 - p(0) = p(1) = \mu_1(p) = \mu_1(q) = \sum_{n=1}^{\infty} nq(n) \geq \sum_{n=1}^{\infty} q(n) = 1 - q(0) \geq q(1),$$

which proves (6.1).

We have

$$\varepsilon_0(p, q) = \sum_{n=0}^{\infty} |p(n) - q(n)| = q(0) - p(0) + p(1) - q(1) + \sum_{n=2}^{\infty} q(n) =$$

$$2(p(1) - q(1))$$

$$\varepsilon_1(p, q) = \sum_{n=1}^{\infty} n|p(n) - q(n)| = p(1) - q(1) + \sum_{n=2}^{\infty} nq(n) =$$

$$p(1) - q(1) + \mu_1(q) - q(1) = 2(p(1) - q(1)),$$

which prove (6.2).

Formula (6.3) is obvious.
We have

$$\Pi_q(1) = \sum_{n=1}^{\infty}(n-1)q(n) = \mu_1(q) - (1-q(0)) = p(1) - 1 + q(0) = q(0) - p(0),$$

which proves (6.4).
This completes the proof of Lemma 6.1. Q.E.D.

By application of (6.2) to respectively (5.1) and (5.9), we obtain

$$\varepsilon_0(p \vee h, q \vee h) \leq 2(p(1) - q(1)) \tag{6.5}$$

$$\eta(p \vee h, q \vee h) \leq \mu_1(h)(q(0) - p(0)), \tag{6.6}$$

and insertion of (6.3) and (6.4) in (5.7) gives

$$-(\mu_1(h) - \Pi_h(x))(q(0) - p(0)) \leq \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \leq 0 \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.7}$$

the second inequality was proved by Bühlmann et al. (1977).

6.3. Binomial distribution

We now assume that

$$p(n) = \binom{t}{n}\pi^n(1-\pi)^{t-n}. \qquad (n = 0,\ 1,\ \ldots,\ t;\ t = 1,\ 2,\ \ldots;\ 0 < \pi < 1) \tag{6.8}$$

The Bernoulli distribution discussed in subsection 6.2 occurs as a special case with $t = 1$. However, unfortunately the situation becomes more complicated when $t > 1$.

In the general case we have

$$\mu_1(p) = t\pi$$

$$\Pi_p(1) = t\pi + (1-\pi)^t - 1 \qquad \Pi_q(1) = t\pi + q(0) - 1, \tag{6.9}$$

and insertion in (5.7) gives

$$-(\mu_1(h) - \Pi_h(x))(t\pi + q(0) - 1) \leq \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \leq$$

$$(\mu_1(h) - \Pi_h(x))\left(t\pi - (1-\pi)^t - 1\right). \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.10}$$

Unfortunately, when $t > 1$, the upper bound does not become equal to zero like in the case $t = 1$. However, as the present binomial distribution is the t-fold convolution of the Bernoulli distribution p_t given by

$$p_t(1) = 1 - p_t(0) = \pi,$$

it is tempting to apply the results of Theorems 4.1 and 4.2. To be able to do that, we have to assume that there exists a distribution $q_t \in \mathcal{P}$ such that $q = q_t^{t*}$. Under this assumption we have

$$p \vee h = (p_t \vee h)^{t*} \qquad q \vee h = (q_t \vee h)^{t*}.$$

From Theorem 4.1 and (6.5) we obtain

$$\varepsilon_0(p \vee h, q \vee h) \leq t\varepsilon_0(p_t \vee h, q_t \vee h) \leq 2t(\pi - q_t(1)). \tag{6.11}$$

We obviously have

$$q(0) = q_t(0)^t \tag{6.12}$$

$$q(1) = tq_t(0)^{t-1}q_t(1).$$

Thus

$$q_t(1) = \frac{1}{t}\frac{q(1)}{q(0)}q(0)^{\frac{1}{t}},$$

and insertion in (6.11) gives

$$\varepsilon_0(p \vee h, q \vee h) \leq 2\left(t\pi - \frac{q(1)}{q(0)}q(0)^{\frac{1}{t}}\right). \tag{6.13}$$

From Theorem 4.2, (6.7), and (6.10) we obtain

$$-(\mu_1(h) - \Pi_h(x))(t\pi + q(0) - 1) \leq \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \leq 0,$$

$$(x = 0,\ 1,\ 2,\ \ldots) \tag{6.14}$$

which implies

$$\eta(p \vee h, q \vee h) \leq \mu_1(h)(t\pi + q(0) - 1).$$

However, from Theorem 4.2, (6.6), and (6.12) we obtain

$$\eta(p \vee h, q \vee h) \leq t\mu_1(h)\left(\pi + q(0)^{\frac{1}{t}} - 1\right) \tag{6.15}$$

which gives a sharper bound when $t > 1$. This implies that the lower bound in (6.14) is sharper than the bound in (6.15) only for high values for $\Pi_h(x)$, that is, low values of x.

The distribution q is called *infinitely divisible* if there for each positive integer m exists a distribution q_m such that $q = q_m^{m*}$ (cf. e.g. Feller (1968)). In particular, this condition should hold for $m = t$, and thus (6.13)-(6.15) hold when q is infinitely divisible.

The condition that there has to exist a distribution q_t such that $q = q_t^{t*}$, may seem intuitively unnatural. However, the following example shows that the inequality $\Pi_{q\vee h} \leq \Pi_{p\vee h}$ does not necessarily hold when this condition is not fulfilled.

Example. Let $t = 2$, $\pi = \frac{1}{2}$, and

$$q(0) = q(2) = \frac{1}{8} \qquad q(1) = \frac{3}{4}.$$

Then $\mu_1(p) = \mu_1(q) = 1$, and application of (6.9) gives $\Pi_p(1) - \Pi_q(1) = \frac{1}{8} > 0$.

6.4. Two infinitely divisible distributions

We shall now assume that both p and q are infinitely divisible. From Theorem 4.2, (5.7), and (6.12) we obtain that for each positive integer m

$$-\mu_1(h)\left(\mu_1(p) + mq(0)^{\frac{1}{m}} - m\right) \leq \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \leq$$

$$\mu_1(h)\left(\mu_1(p) + mp(0)^{\frac{1}{m}} - m\right), \qquad (x = 0,\ 1,\ 2,\ \ldots)$$

and by letting m go to infinity we obtain

$$-\mu_1(h)(\mu_1(p) + \ln q(0)) \leq \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \leq$$

$$\mu_1(h)(\mu_1(p) + \ln p(0)). \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.16}$$

6.5. Poisson vs. infinitelv divisible distribution

We now assume that

$$p(n) = \frac{\lambda^n}{n} e^{-\lambda}, \qquad (n = 0,\ 1,\ 2,\ \ldots;\ \lambda > 0) \tag{6.17}$$

and that q is infinitely divisible. Then p is also infinitely divisible, and we have $\mu_1(p) = \lambda$.

Let

$$\bar{p}_t(n) = \binom{t}{n}\left(\frac{\lambda}{t}\right)^n\left(1-\frac{\lambda}{t}\right)^{t-n}. \qquad (n = 0,\ 1,\ \ldots,\ m;\ t = 1,\ 2,\ \ldots)$$

Then

$$p(n) = \lim_{t\uparrow\infty} \bar{p}_t(n). \qquad (n = 0,\ 1,\ 2,\ \ldots)$$

From (6.13) we obtain

$$\varepsilon_0\left(\bar{p}_t \vee h, q \vee h\right) \leq 2\left(\lambda - \frac{q(1)}{q(0)} q(0)^{\frac{1}{t}}\right),$$

and as this bound is decreasing in t, we obtain

$$\varepsilon_0(p \vee h, q \vee h) \leq 2\left(\lambda - \frac{q(1)}{q(0)}\right) \tag{6.18}$$

by letting t go to infinity. A similar limiting argument for (6.14) gives

$$-\left(\mu_1(h) - \Pi_h(x)\right)(\lambda + q(0) - 1) \leq \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \leq 0. \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.19}$$

From (6.16) and (6.19) we obtain

$$\eta(p \vee h, q \vee h) \leq \mu_1(h)(\lambda + \ln q(0)), \tag{6.20}$$

which could also have been found by a limiting argument in (6.15). As $\ln q(0) \leq q(0) - 1$, the lower bound in (6.19) in weaker than (6.20) for large values of x.

6.6. Binomial vs. negative binomial distribution

We now assume that p is the binomial distribution given by (6.8), and that q is given by

$$q(n) = \binom{\alpha + n - 1}{n}(1-\rho)^\alpha \rho^n. \qquad (n = 0,\ 1,\ \ldots;\ \alpha > 0; 0 < \rho < 1) \tag{6.21}$$

Then q is infinitely divisible with

$$\mu_1(q) = \alpha \frac{\rho}{1-\rho},$$

and from (6.13)-(6.15) we obtain

$$\varepsilon_0(p \vee h, q \vee h) \le 2t\pi\left(1 - (1-\rho)^{\frac{\alpha}{t}+1}\right) \tag{6.22}$$

$$-\left(\mu_1(h) - \Pi_h(x)\right)[t\pi + (1-\rho)^{\alpha} - 1] \le \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \le 0 \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.23}$$

$$\eta(p \vee h, q \vee h) \le t\mu_1(h)\left(\pi + (1-\rho)^{\frac{\alpha}{t}} - 1\right). \tag{6.24}$$

6.7. Binomial vs. Poisson distribution

We now assume that p is the binomial distribution given by (6.8) and q the Poisson distribution given by (6.17). Then (6.13)-(6.15) give

$$\varepsilon_0(p \vee h, q \vee h) \le 2t\pi(1 - e^{-\pi}) \tag{6.25}$$

$$-\left(\mu_1(h) - \Pi_h(x)\right)(t\pi + e^{-t\pi} - 1) \le \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \le 0 \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.26}$$

$$\eta(p \vee h, q \vee h) \le t\mu_1(h)(\pi + e^{-\pi} - 1), \tag{6.27}$$

which can also be deduced from (6.22)-(6.24) by a limiting argument.

6.8. Poisson vs. negative binomial distribution

We now assume that p is the Poisson distribution given by (6.17) and q the negative binomial distribution given by (6.21). Then (6.18)-(6.20) give

$$\varepsilon_0(p \vee h, q \vee h) \le 2\alpha\frac{\rho^2}{1-\rho} \tag{6.28}$$

$$-(\mu_1(h) - \Pi_h(x))\left(\alpha\frac{\rho}{1-\rho} + (1-\rho)^{\alpha} - 1\right) \le \Pi_{p\vee h}(x) - \Pi_{q\vee h}(x) \le 0 \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.29}$$

$$\eta(p \vee h, q \vee h) \le \alpha\mu_1(h)\left(\frac{\rho}{1-\rho} + \ln(1-\rho)\right), \tag{6.30}$$

which can also be deduced from (6.22)-(6.24) by a limiting argument.

The bound in (6.28) was deduced by Gerber (1984) and the bound in (6.30) by Dhaene (1991).

6.9. Collective approximation to individual model

For $i = 1, \ldots, m$, let $h_i \in \mathcal{P}_+$ and p_i be the Bernoulli distribution given by

$$p_i(1) = 1 - p_i(0) = \pi_i.$$

We approximate $p_i \vee h_i$ with the compound Poisson distribution $q_i \vee h_i$ with

$$q_i(n) = \frac{\pi_i^n}{n!} e^{-\pi_i}. \qquad (n = 0,\ 1,\ 2,\ \ldots)$$

It is well known that then $*_{i=1}^m (q_i \vee h_i) = q \vee h$ with

$$q(n) = \frac{\lambda^n}{n} e^{-\lambda} \qquad (n = 0,\ 1,\ 2,\ \ldots)$$

$$\lambda = \sum_{i=1}^m \pi_i \qquad h = \frac{1}{\lambda} \sum_{i=1}^m \pi_i h_i.$$

By a trivial generalisation of (6.25) and (6.27) we obtain

$$\varepsilon_0\left(\mathop{*}_{i=1}^m (p_i \vee h_i), q \vee h\right) \leq 2 \sum_{i=1}^m \pi_i (1 - e^{-\pi_i}) \tag{6.31}$$

$$\eta\left(\mathop{*}_{i=1}^m (p_i \vee h_i), q \vee h\right) \leq \sum_{i=1}^m \mu_1(h_i)(\pi_i - e^{-\pi_i} - 1). \tag{6.32}$$

Unfortunately we have not been able to generalise the first inequality in (6.26), but the second inequality is easily generalised to

$$\Pi_{*_{i=1}^m (p_i \vee h_i)}(x) \leq \Pi_{q \vee h}(x). \qquad (x = 0,\ 1,\ 2,\ \ldots) \tag{6.33}$$

The inequalities in (6.31)-(6.33) have been deduced by respectively Gerber (1984), De Pril & Dhaene (1992), and Bühlmann et al. (1977).

When π_i and h_i are the same for all i, we are back in the situation of subsection 6.7.

REFERENCES

BÜHLMANN, H., GAGLIARDI, B., GERBER, H.U., & STRAUB, E. (1977). Some inequalities for stop-loss premiums. *ASTIN Bulletin* **9**, 75-83.

DE PRIL, N. & DHAENE, J. (1992). Error bounds for compound Poisson approximations to the individual risk model. *ASTIN Bulletin* **22**, 135-148.

DHAENE, J. (1991). Approximating the compound negative binomial distribution by the compound Poisson distribution. *Bulletin of the Swiss Association of Actuaries*, 117-121.

DHAENE, J. & SUNDT, B. (1996). On approximating distributions by approximating their De Pril transforms. To be published in *Scandinavian Actuarial Journal*.

FELLER, W, (1968). *An introduction to probability theory and its applications*. (3. ed.). Wiley, New York.

GERBER, H.U. (1984). Error bounds for the compound Poisson approximation. *Insurance: Mathematics and Economics* **3**, 191-194.

SUNDT, B. (1991). On approximating aggregate claims distributions and stop-loss premiums by truncation. *Insurance: Mathematics and Economics* **10**, 133-136.

SUNDT, B. & DHAENE, J. (1996). On bounds for the difference between the stop loss transforms of two compound distributions. *ASTIN Bulletin* **26**, 225-231.

JAN DHAENE
Departement Toegepaste Economische Wetenschappen
Katholieke Universiteit Leuven
Huis Eighen Heerd
Minderbroederstraat 5
B-3000 Leuven
Belgium

BJØRN SUNDT
Department of Mathematics
University of Bergen
Johannes Brunsgate 12
N-5008 Bergen
Norway

CALCULATING RUIN PROBABILITIES VIA PRODUCT INTEGRATION

COLIN M. RAMSAY[1]
Actuarial Science Program
University of Nebraska-Lincoln
Lincoln NE 68588-0426, USA

MIGUEL A. USABEL[2]
Dto. Economia Financiera y Actuarial
UCM Campus de Somosaguas
Ftad. CC Economicas
Madrid, Spain

ABSTRACT[3]

When claims in the compound Poisson risk model are from a heavy-tailed distribution (such as the Pareto or the lognormal), traditional techniques used to compute the probability of ultimate ruin converge slowly to desired probabilities. Thus, faster and more accurate methods are needed. Product integration can be used in such situations to yield fast and accurate estimates of ruin probabilities because it uses quadrature weights that are suited to the underlying distribution. Tables of ruin probabilities for the Pareto and lognormal distributions are provided.

KEYWORDS

Integral equation, convergence, heavy-tailed distributions.

1. INTRODUCTION

Let us consider the classical compound Poisson risk model with nonnegative claims. Specifically, let u be the initial risk reserv, $F(.)$ be the cumulative distribution function of the nonnegative claim size random variable, p_1 be the expected claim size, $1 + \theta$ be the loading factor applied to the net premium rate, and $\psi(u)$ be the infinite time probability of ruin for an initial risk reserve of u.

Gerber (1979, p. 115, equation (3.7)) has shown that $\psi(u)$ satisfies the following Volterra integral equation of the second kind:

---

[1] Internet address: cramsay@unlinfo.unl.edu

[2] Internet address: econ109@sis.ucm.es

[3] This research was done at the University of Nebraska-Lincoln while Dr. Usabel was visiting his post-doctoral researcher. The authors gratefully acknowledge the financial support from Universidad Complutense de Madrid (Ayudas Postdoctorales en el Extranjero) and from the E.J. Faulkner Chair at the University of Nebraska-Lincoln.

$$\psi(u) = \frac{1}{1+\theta}\left[A(u) + \int_0^u K(u,t)\psi(t)dt\right], \quad u \geq 0 \tag{1}$$

where

$$a(u) = \int_u^\infty \frac{1-F(t)}{p_1}dt, \quad u \geq 0 \tag{2}$$

$$K(u,t) = \frac{1-F(u-t)}{P_1}, \quad 0 \leq t \leq u. \tag{3}$$

A classic problem (of interest mainly to academic actuaries) is the numerical evaluation $\psi(u)$. Numerous authors have studied this problem; see, for example, recent texts by Grandell (1991) and Panjer and Willmot (1992, Chapter 11) and references therein. In general, no explicit closed form solution to equation (1) exists except in the case where claims are mixtures of exponential distributions; see Bowers et al. (1986, Chapter 12.6).

There are, however, several broad approaches to the evaluation $\psi(u)$. The older approaches are ad hoc: focusing inverting the Laplace transform, or on matching the first few moments of the claim size distribution or on the Cramer-Lundberg approximation; see Ramsay (1992a) for a comparison of some of these methods.

Since the early 1980s, the shift has been to approaches based on discretizing some aspect of the risk process and deriving recursive expressions for $\psi(u)$; see, for example, Goovaerts and De Vylder (1984), Panjer (1986), Dickson (1989), Dickson and Waters (1991), Ramsay (1992b), and Dickson, Egidio dos Reis and Waters (1995). Panjer and Wang (1993) describe the conditions under which these recursions are stable.

Though these recursive approaches may be able to determine $\psi(u)$ to any desired degree of accuracy, they are not suitable for heavy-tailed distributions, such as the Pareto or lognormal distributions, for two main reasons:

1. To achieve a reasonable degree of accuracy, the interval of discretization must be at most one unit of the mean in lenght. If we standardize the unit of currency such that $p_1 = 1$, then to obtain $\psi(10)$ we must recursively estimate every intermediate unit point $\psi(u)$ for $k = 0,1,2,\ldots,9,10$. This may be acceptabel if we need only small values of u; however, for large values of u, say $u = 500$ units, this method can be slow. For the Pareto, $\psi(500)$ is not insignificant.
2. The quadrature rules inherent in the recursive schemes are usually of low order. This further reduces its accuracy and its rate of convergence. To improve accuracy, the intervals of discretization are made even smaller. This substantially increases the number of intermediate calculations required, making the process of finding $\psi(u)$ slower.

The objective of this paper is to present a method of evaluating $\psi(u)$ using so-called product integration. We show that this method can be fast and accurate when dealing with heavy-tailed distributions.

2. PRODUCT INTEGRATION

Consider the numerical solution of the Volterra integral equation

$$x(s) = y(s) + \int_a^s k(s,t)x(t)dt, \quad a \le s \le b \quad (4)$$

where $k(.,.)$ is the kernel (and is known) and $x(.)$ is the unknown function to be determined. Assume $k(.,.)$ or one of its low-order derivatives is badly behaved in one of its arguments. (For example, $k(.,.)$ may be singular or nearly singular). In such a situation, the Newton-Cotes integration (e.g., trapezoid rule, Simpson's rule, etc.) may produce inaccurate results or suffer a reduced rate of convergence.

Delves and Mohamed (1985) and Linze (1985) recommend the use of product integration[1] to take account of the fact that $k(.,.)$ may be badly behaved. Our development of the product integration quadrature rule follows the exposition and notation of Delves and Mohamed (Chapters 4.4 and 5.5). For a more detailed description of the product integration technique, see Linz (1985, Chapter 8).

First we factorize $k(s,t)$ as

$$k(s,t) = p(s,t)\bar{k}(s,t)$$

where $\bar{k}(.,.)$ is smooth and well-behaved and can be accurately approximated by a suitable Lagrangian interpolating polynomial, and $p(s,t)$ is badly behaved. Next we decompose the interval $[a,b]$ into n subintervals $\{h_i\}$ where

$$h_i = s_{i+1} - s_i, \quad i = 0,1,\ldots,n-1$$

and

$$a = s_0 < s_1 < \ldots < s_n = b.$$

Product integration proceeds by approximating the integral in equation (4) for $s = s_i$, $i = 1,2,\ldots,n$, using a quadrature rule of the form

$$\int_a^{s_i} p(s_i,t)\bar{k}(s_i,t)x(t)dt \approx \sum_{j=0}^{i} w_{ij}\bar{k}(s_i,t)x(t_j) \quad (6)$$

where $t_i = s_i$ for $i = 0,1,2,\ldots,n$. The weights are determined by insuring that the rule of equation (5) is exact when $\bar{k}(s,t)x(t)$ is a polynomial in t of degree $\le d$. Product integration is only applicable if the following $(d + 1)$ moments μ_{ij} exist and can be calculated for each i, where

$$\mu_{ij} = \int_a^{s_i} t^j p(s_i,t)dt, \quad j = 0,1,\ldots,d.$$

In this paper we assume $\bar{k}(s_i,t)x(t)$ is linear ($d = 1$) in t, i.e.,

$$\bar{k}(s_i,t)x(t) \approx \frac{(t_{j+1} - t)}{h_j}\bar{k}(s_i,t)x(t_j) + \frac{(t - t_j)}{h_j}\bar{k}(s_i,t_{j+1})x(t_{j+1}).$$

[1] Linz (1985, Chapter 8, p. 141) attributes the origin of the product integration technique to Young (1954).

It follows that

$$\int_a^{s_i} p(s_i,t)\bar{k}(s_i,t)x(t)dt \approx \sum_{j=0}^{i-1}\int_{t_j}^{t_{j+1}} p(s_i,t)\left[\frac{(t_{j+1}-t)}{h_j}\bar{k}(s_i,t_j)x(t_j) + \frac{(t-t_j)}{h_j}\bar{k}(s_i,t_{j+1})x(t_{j+1})\right]$$

$$= \sum_{j=0}^{i} w_{ij}\bar{k}(s_i,t_j)x(t_j)$$

where

$$w_{i0} = \int_{t_0}^{t_1} p(s_i,t)\frac{(t_1-t)}{h_0}dt \quad \text{for } j=0$$

$$w_{ij} = \int_{t_j}^{t_{j+1}} p(s_i,t)\frac{(t_{j+1}-t)}{h_j}dt + \int_{t_{j-1}}^{t_j} p(s_i,t)\frac{(t-t_{j-1})}{h_{j-1}}dt \quad \text{for } j=1,2,\ldots,i-1$$

$$w_{ij} = \int_{t_{i-1}}^{t_i} p(s_i,t)\frac{(t-t_{i-1})}{h_{i-1}}dt \quad \text{for } j=i$$

To facilitate easy computation of the weights, we introduce two new variables:

$$v_{ij} = \int_{t_j}^{t_{j+1}} (t_{j+1}-t)p(s_i,t)dt \tag{6}$$

$$c_{ij} = \int_{t_j}^{t_{j+1}} p(s_i,t)dt. \tag{7}$$

As $t-t_j = (t_{j+1}-t_j)-(t_{j+1}-t)$, then

$$w_{i0} = \frac{v_{i0}}{h_0} \tag{8}$$

$$w_{ij} = \frac{v_{ij}}{h_j} + c_{ij} - \frac{v_{i,j-1}}{h_{j-1}} \quad \text{for } j=1,2,\ldots,i-1 \tag{9}$$

$$w_{ii} = c_{i,i-1} - \frac{v_{i,i-1}}{h_{i-1}}. \tag{10}$$

Thus, the approximate solution to equation (4) is determined recursively using

$$\hat{x}_n(s_i) = y(s_i) + \sum_{j=0}^{i} w_{ij}\bar{k}(s_i,t_j)\hat{x}_n(t_j) \tag{11}$$

for $i = 1,2,\ldots,n$, with

$$\hat{x}_n(s_0) = y(a). \tag{12}$$

The resulting estimate of $x(s)$ is $\hat{x}_n(s_n)$.

3. ACCELERATING THE CONVERGENCE

We can improve the accuracy of our estimate $\hat{x}_n(s)$ by dividing the interval $[a, s]$ into smaller subintervals. Following the arguments of Ramsay (1992), Richardson's extrapolation technique can be used to accelerate the convergence of $\hat{x}_n(s)$ to $x(s)$ as $n \to \infty$. To this end, let us divide the interval $[a, s]$ into n_j intervals of equal length, where

$$n_j = \gamma \times 2^j \quad j = 0,1,2,\ldots \tag{13}$$

and γ is a positive integer. For given j and $[a, s]$, we have

$$\begin{aligned} s_{n_j} &= s \\ h &= (s-a)/n_j \quad \text{for } i = 0,1,2,\ldots,n_j - 1 \\ s_i = t_i &= a + ih \quad \text{for } i = 0,1,2,\ldots,n_j - 1 \end{aligned}$$

The Richardson extrapolation technique generates a lower diagonal matrix of approximations:

$$T_r^j = T_{r-1}^j + \frac{T_{r-1}^j - T_{r-1}^{j-1}}{2^r - 1} \tag{14}$$

for $r = 1, 2, \ldots, j$ and $j = 1, 2, \ldots$ with $T_0^j = \hat{x}_{n_j}(s)$. The final estimate of $x(s)$ is:

$$\hat{x}(s) = T_j^j. \tag{15}$$

4. THE MAIN RESULTS

Product integration is used to compute ruin probabilities for the Pareto and lognormal distributions. Without loss of generality, set $p_1 = 1$ for each distribution. Tables 1 and 2 show the final estimated values of the ruin probabilities after the Richardson extrapolation technique has been applied.

4.1 The Pareto Distribution

Consider the Pareto distribution defined on $(0, \infty)$ with unit mean, i.e.,

$$F(t) = 1 - \left(\frac{\alpha}{\alpha + t}\right)^{\alpha+1} \quad \alpha > 0 \text{ and } t > 0.$$

Equations (2) and (3) imply

$$A(u) = \left(\frac{\alpha}{\alpha + u}\right)^{\alpha}$$

$$K(u,t) = \left(\frac{\alpha}{\alpha + u - t}\right)^{\alpha+1}$$

Even though $K(u, t)$ and all of its derivatives are smooth and wellbehaved, they converge slowly as $u \to \infty$. As all of the moments μ_{ij} exist for any finite s, product integration can be used.

Next set

$$p(s,t) = K(s,t)$$

$$\bar{k}(s,t) = \begin{cases} 1 & \text{if } 0 \le t \le s; \\ 0 & \text{otherwise.} \end{cases}$$

To determine the product integration weights, we need v_{ij} and c_{ij} from equations (6) and (7).

$$v_{ij} = d_{ij} + (\alpha + s_i - t_{j+1})c_{ij}$$

where

$$d_{ij} = \begin{cases} \ln(1 + s_i - t_j) - \ln(1 + s_i - t_{j+1}) & \text{if } \alpha \neq 1; \\ \dfrac{\alpha^2}{\alpha - 1}\left[\left(\dfrac{\alpha}{\alpha + s_i - t_j}\right)^{\alpha-1} - \left(\dfrac{\alpha}{\alpha + s_i - t_{j+1}}\right)^{\alpha-1}\right] & \text{if } \alpha \neq 1. \end{cases}$$

$$c_{ij} = \left(\frac{\alpha}{\alpha + s_i - t_j}\right)^{\alpha} - \left(\frac{\alpha}{\alpha + s_i - t_{j+1}}\right)^{\alpha}.$$

Tabel 1 shows the ruin probabilities for the Pareto distribution with $\alpha = 1$ and several values of θ. From equation (13), we use $\gamma = 20$ and $j = 0, 1, 2, 3$ and 4. (Thus, $n_4 = 320$.)

TABLE 1

RUIN PROBABILITIES: PARETO DISTRIBUTION ($\alpha = 1$)

Ψ(u) for Various Values of θ ***u***	***$\theta = 0.10$***	***$\theta = 0.25$***	***$\theta = 0.50$***	***$\theta = 0.75$***	***$\theta = 1.00$***
10	0.627128	0.372677	0.206646	0.138242	0.102523
20	0.498142	0.245260	0.119274	0.075908	0.055049
30	0.411437	0.178338	0.081426	0.051056	0.036887
40	0.347893	0.137559	0.060856	0.038038	0.027509
50	0.299155	0.110519	0.048164	0.030142	0.021847
60	0.260646	0.091524	0.039650	0.024884	0.018080
70	0.229551	0.077594	0.033588	0.021150	0.015402
80	0204018	0.067029	0.029075	0.018369	0.013404
90	0.182761	0.058794	0.025596	0.016222	0.011859
100	0.164860	0.052227	0.022839	0.014517	0.010630
200	0.076323	0.023800	0.010860	0.007028	0.005194
300	0.046612	0.015154	0.007083	0.004621	0.003429
400	0.032827	0.011071	0.005247	0.003438	0.002557
500	0.025123	0.008708	0.004165	0.002737	0.002038
600	0.020273	0.007170	0.003451	0.002273	0.001694
700	0.016962	0.006092	0.002946	0.001943	0.001449
800	0.014566	0.005294	0.002569	0.001696	0.001266
900	0.012756	0.004681	0.002278	0.001505	0.001124
1000	0.011341	0.004194	0.002046	0.001353	0.001011

4.2 Lognormal Distibution

In this case things will be more complicated because of the presence of the normal cumulative distribution function. Again we assume that
$p_1 = 1$. This implies

$$A(u) = \int_u^\infty 1 - F(t)dt, \quad u \geq 0$$
$$K(u,t) = 1 - \Phi\left(\frac{\ln(u-t) - \mu}{\sigma}\right), \quad 0 \leq t \leq u$$
$$\mu = e^{-\sigma^2/2} \quad (\text{as } p_1 = 1)$$

where μ and σ are the parameters of the lognormal and

$$\Phi(u) = \int_{-\infty}^{u} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt.$$

A source of difficulty is in the computation of v_{ij} adn c_{ij}, i.e.,

$$v_{ij} = \int_{t_j}^{t_{j+1}} (t_{j+1} - t)(1 - \Phi(\frac{\ln(s_i - t) - \mu}{\sigma}))dt$$
$$c_{ij} = \int_{t_j}^{t_{j+1}} (1 - \Phi(\frac{\ln(s_i - t) - \mu}{\sigma}))dt.$$

As the function $\Phi(.)$ is known only approximately, these integrals must be computed numerically; see for example Abramowitz and Stegun (1964, Chapter 26) for several approximations. The approximation used in this paper is:

$$\Phi(u) = 1 - \frac{e^{-u^2/2}}{\sqrt{2\pi}}\left(\sum_{k=1}^{5} b_k t^k\right) + \varepsilon(u)$$

where $|\varepsilon(u)| < 7.5 \times 10^{-8}$, and

$t = 1/(1 + pu)$	$p = 0.2316419$
$b_1 = 0.319381530$	$b_4 = -1.821255978$
$b_2 = -0.356563782$	$b_5 = 1.330274429$
$b_3 = 1.781477937$	

Gaussian integration rules many be used to evalutate the integrals.

Table 2 shows the ruin probabilities for the lognormal distribution with $\sigma = 1.80$ and several values of θ. From equation (13), we use $\gamma = 10$ and $j = 0, 1, 2, 3$ and 4. (Thus, $n_4 = 160$. These values are very close to those of Thorin and Wikstad (1977), where appropriate.

TABLE 2

RUIN PROBABILITIES: LOGNORMAL DISTRIBUTION (σ = 1.80)

Ψ(u) for Various Values of u and θ					
u	$\theta = 0.10$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 1.00$
10	0,739768	0,518832	0,336874	0,245749	0,192154
20	0,656692	0,410781	0,240187	0,165669	0,125229
30	0,593553	0,339538	0,184539	0,122940	0,091161
40	0,541731	0,287396	0,147713	0,096077	0,070371
50	0,497634	0,247190	0,121512	0,077676	0,056424
60	0,459303	0,215164	0,101989	0,064361	0,046484
70	0,425505	0,189068	0,086956	0,054343	0,039091
80	0,395396	0,167437	0,075086	0,046580	0,033413
90	0,368362	0,149265	0,065528	0,040423	0,028940
100	0,343939	0,133830	0,057704	0,035446	0,025344
200	0,188093	0,055553	0,022128	0,013482	0,009651
300	0,113139	0,029147	0,011567	0,007112	0,005124
400	0,072445	0,017524	0,007067	0,004390	0,003180
500	0,048684	0,011534	0,004747	0,002974	0,002164
600	0,034048	0,008096	0,003397	0,002143	0,001565
700	0,024637	0,005960	0,002544	0,001614	0,001182
800	0,018360	0,004551	0,001971	0,001257	0,000922
900	0,014040	0,003577	0,001569	0,001004	0,000738
1000	0,010981	0,002878	0,001276	0,000819	0,000603

5. CONCLUDING COMMENTS

The important strength of the product integration technique in solving equation (1) is that it converges significantly faster and is more accurate than the Goovaerts and de Vylder (1984) technique, or the improved version proposed by Ramsay (1992b). This is acheived by using a quadrature rule that exploits some of the features of the kernel, thus requiring a reduced amount of recursions. Even though the weights w_{ij} (and hence c_{ij} and v_{ij}) have to be computed directly from the kernel, these extra computations are fast and easy to perform.

Because product integration converges relatively rapidly, it does not require the use of small intervals, thus reducing the possibilitiy of subtracting nearly equal numbers (and hence rounding errors). In addition, it requires a small fraction of the computations required by the Goovaerts-De Vylder-Ramsay approach to obtain the same degree of accuracy. This should not be surprising because product integration uses much more information from the integrand than do the common Newton-Cotes quadrature formulae.

A further area of research is the determination of the error bounds of the solutions generated via the product integration technique. Linz (1985, Chapter 8, p. 131) shows that the error bounds and orders of convergence for product integration follow the standard results of approximation theory. Thus, product integration based on the trapezoidal rule is of order $O(h^2)$.

Additionally, one may be able to use the Goovaerts-De Vylder-Ramsay approach and combine it with product integration to produce a faster scheme with explicit error bounds.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I.A. (1964). *Handbook of Mathematical Functions*. New York, N.Y.: Dover Publications.

BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A. and NESBITT, C.J. (1986). *Actuarial Mathematics*. Ithasca, Ill.: Society of Actuaries.

DELVES, L.M. and MOHAMED, J.L. (1989) *Computational Methods for Integral Equations*. Cambridge, England: Cambridge University Press.

DICKSON, D.C.M. (1989). "Recursive Calculation of the Probability and Severity of Ruin." *Insurance: Mathematics and Economics*, 8, pp. 145-148.

DICKSON, D.C.M., EGIDIO DOS REIS, A.D. and WATERS, H.R. (1995). "Some Stable Algorithms in Ruin Theory and Their Application." *Astin Bulletin* 25, pp. 153-175.

DICKSON, D.C.M. and WATERS, H.R. (1991) "Recursive Calculation of Survival Probabilities." *Astin Bulletin* 21, pp. 199-221.

GERBER, H.U. (1979) *An Introduction to Mathematical Risk Theory*. Huebner Foundation Monograph 8. Philadelphia, Pa.: University of Pennsylvania. (Distributed by Irwin, Homewood, IL.)

GOOVAERTS, M. and DE VYLDER, F. (1984) "A Stable Recursive Algorithm for Evaluation of Ultimate Ruin Probabilities." *Astin Bulletin*, 14, pp. 53-59.

GRANDELL, J. (1990). *Aspects of Risk Theory*. Springer Verlag, New York.

LINZ, P. (1985). *Analytical and Numerical Methods for Volterra Equations*. Philadelphia, Pa.: SIAM Studies in Applied Mathematics.

PANJER, H.H. (1986) "Direct Calculation of Ruin Probabilities." *Journal of Risk and Insurance*, 53, pp. 521-529.

PANJER, H.H. and WANG, S (1993) "On the Stability of Recursive Formulas." *ASTIN Bulletin* , 23, pp. 227-258.

RAMSAY, C.M. (1992a) "A Practical Algorithm for Approximating the Probability of Ruin." *Transactions of the Society of Actuaries*, XLIV, 443-459.

RAMSAY, C.M. (1992b) "Improving Goovaerts' and de Vylder's Stable Recursive Algorithm." *Astin Bulletin*, 22, pp. 51-59.

THORIN, O. and WIKSTAD, N. (1977) "Calculation of Ruin Probabilities When the Claim Distribution is Lognormal." *Astin Bulletin*, 9, pp. 231-246.

YOUNG, A. (1954) "Approximate Product Integration." *Proc. Royal Soc. London*, Ser. A, 224, pp. 561-573.

CREDIBILITY USING SEMIPARAMETRIC MODELS

BY VIRGINIA R. YOUNG

School of Business
University of Wisconsin-Madison

ABSTRACT

To use Bayesian analysis to model insurance losses, one usually chooses a parametric conditional loss distribution for each risk and a parametric prior distribution to describe how the conditional distributions vary across the risks. A criticism of this method is that the prior distribution can be difficult to choose and the resulting model may not represent the loss data very well. In this paper, we apply techniques from nonparametric density estimation to estimate the prior. We use the estimated model to calculate the predictive mean of future claims given past claims. We illustrate our method with simulated data from a mixture of a lognormal conditional over a lognormal prior and find that the estimated predictive mean is more accurate than the linear Bühlmann credibility estimator, even when we use a conditional that is not lognormal.

KEYWORDS

Kernel density estimation, claim estimation, Bayesian estimation.

1. INTRODUCTION

In a portfolio of insurance policyholders (also called risks), risks are heterogeneous; that is, the insurance losses of different risks follow different loss distributions. The premium an insurer charges a given risk depends on the information available concerning the loss distribution of that risk. If the insurer knew the exact loss distribution of a risk, then the appropriate net premium to charge would be the expectation of that loss distribution. On the other hand, if the insurer has no information about a specific policyholder, then the net premium is the expectation over the entire portfolio of policyholders. For the situation between these two extremes, suppose the insurer has prior claim data for the risk, then the net premium is the conditional expectation of future claims given the prior claims.

To use Bayesian analysis to model insurance losses, one usually chooses a parametric conditional loss distribution for each risk and a parametric prior distribution to describe how the conditional distributions vary across the risks. A criticism of this method is that the prior distribution can be difficult to choose and the resulting model may not represent the loss data very well. One method of circumventing this problem is to apply empirical Bayesian analysis in which one uses the data to estimate the parameters of the model (Klugman, 1992).

In this paper, we use a semiparametric mixture model to represent the insurance losses of a portfolio of risks: We choose a flexible parametric conditional loss distribution for each risk with unknown conditional mean that varies across the risks. This conditional distribution may depend on parameters other than the mean, and we use the data to estimate those parameters. Then, we apply techniques from nonparametric density estimation to estimate the distribution of the conditional means.

In Section 2, we describe a mixture model for insurance claims and estimate the prior density using kernel density estimation. In Section 3, we calculate the credibility estimator assuming squared-error loss and also give the projection of that estimator onto the space of linear functions. Finally, in Section 4, we apply our methodology to simulated data from a mixture of a lognormal conditional over a lognormal prior. We show that our method can lead to good credibility formulas, as measured by the mean squared error of the claim predictor, even when we use a gamma conditional instead of a lognormal conditional.

2. SEMIPARAMETRIC MIXTURE MODEL

2.1. Notation and Assumptions

Assume that the underlying claim of risk i per unit of exposure is a conditional random variable $Y|\theta_i$, $i = 1, 2, ..., r$, with probability density function $f(y|\theta_i)$. For each of the r risks, we observe the average claims per unit of exposur $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{in_i})$ with an associated exposure vector $\boldsymbol{w}_i = (w_{i1}, w_{i2}, ..., w_{in_i})$, $i = 1, 2, ..., r$. Thus, the observed average claim x_{ij} is the arithmetic average of w_{ij} claims, each of which is an independent realization of the conditional random variable $Y|\theta_i$. For example, if a risk is a group policyholder, then x_{ij} may be the average claim per insured member of the group in the j^{th} policy period and w_{ij} is the number of members in the group during the j^{th} policy period. For the data from Hachemeister (1975), a risk is the collection of insureds in a particular state covered by bodily injury automobile insurance, x_{ij} represents the average claim severity during period j, and w_{ij} is the corresponding number of claims.

Assume that the parameter θ is the conditional mean, $E[Y|\theta] = \theta$. There may be other parameters that characterize the conditional distribution, such as the shape parameter α for the gamma density. However, in this paper, we assume that parameters, other than the conditional mean, are fixed across the risks. The loss distribution of a given risk is, therefore, characterized by its conditional mean, although that mean is generally unknown. Denote the probability density function of θ by $\pi(\theta)$, also called the *structure function* (Bühlmann, 1970). The structure function characterizes how the conditional mean θ varies from risk to risk. We argue that assuming θ to be continuous is reasonable because in the

Bayesian paradigm, our uncertainty about θ for any particular risk would be represented by a continuous random variable. Also, if r is large, then the variable θ can be well approximated by a continuous random variable. Even if r is not large, the collection of r risks may be a sample from a larger population of risks whose distribution can be approximated by a continuous distribution. Assume that the experience of different risks is independent.

Note that our model is a special case of the one given by Bühlmann and Straub (1970). Because X_{ij} is the random variable of an average of w_{ij} iid claims $Y_1, Y_2, \ldots, Y_{w_{ij}}$, given θ_j, we have that $E[X_{ij}|\theta_i] = E[Y|\theta_i] = \theta_i$ is independent of the period j. It also follows that

$$Cov[X_{ij}, X_{ik}|\theta_i] = \begin{cases} \frac{Var[Y|\theta_i]}{w_{ij}} & if \;\; j = k, \\ 0, & if \;\; j \neq k, \end{cases}$$

as in the Bühlmann-Straub model. In the literature, $E[Y|\theta_i]$ is called the *hypothetical mean* and $Var[Y|\theta_i]$ the *process variance*. Note that we assume the observations for a risk arise as arithmetic averages of an underlying claim random variable $Y|\theta$, while Bühlmann and Straub (1970) do not assume this in their more general model.

The goal of credibility theory is to predict the future claim y (or an average of future claims) of a risk, given that the risk's claim experience is $\boldsymbol{x}$ and exposure $\boldsymbol{w}$. In this paper, we restrict our attention to credibility formulas that are functions of a single statistic because they are easier to estimate and to use. We choose the sample mean as our statistic, $\bar{x}_i = \frac{\sum_{j=1}^{n_i} w_{ij} x_{ij}}{\sum_{j=1}^{n_i} w_{ij}}$ because the claim experience $\boldsymbol{x}$ is a vector of averages. However, we do not restrict a claim estimator to be linear.

To pick a parametric conditional distribution for $Y|\theta$, we use the following criteria:

- $E[Y|\theta] = \theta$
- The sample mean is a sufficient statistic for θ.
- The functional form of $f(y|\theta)$ is closed under averaging. That is, if $\overline{X}$ is an average of w claims that follow the distribution given by $f(y|\theta)$, then the density of $\overline{X}$ has the same functional form as $f(y|\theta)$.

Three such families of densities are commonly used in actuarial science to model insurance losses—(1) the normal, with mean θ and fixed variance σ^2, (2) the gamma, with mean $\theta = \frac{\alpha}{\beta}$ and fixed shape parameter α, and (3) the inverse gaussian, with mean θ and fixed $\lambda = \frac{\theta^3}{Var[X|\theta]}$. Indeed, $Y|\theta \sim N(\theta, \sigma^2)$ implies that if $\overline{X}$ is an average of w iid claims $Y_1, Y_2, \ldots, Y_w$, given θ, then $\overline{X}|\theta \sim N\left(\theta, \sigma^2/_w\right)$. Similarly, if $Y|\theta \sim G(\theta, \alpha)$, then $\overline{X}|\theta \sim G(\theta, w\alpha)$, and the probability density function of $Y|\theta$ is

$$f(y|\theta) = \frac{\alpha^\alpha}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-\frac{\alpha y}{\theta}}, \quad y > 0.$$

Finally, if $Y|\theta \sim InvG(\theta, \lambda)$, then $\overline{X}|\theta \sim InvG(\theta, w\lambda)$ and the probability density function of $Y|\theta$ is

$$f(y|\theta) = \sqrt{\frac{\lambda}{2\pi y^3}}\exp\left[-\frac{\lambda}{2}\frac{(y-\theta)^2}{y\theta^2}\right], \quad y > 0.$$

We use the family of gamma conditional distributions in an example in Section 4. In practice, one might use the normal conditional if the conditional variance is assumed constant across the risks. One might use the gamma conditional if the conditional coefficient of variation is assumed constant across risks or the inverse gaussian conditional if one wanted to use a loss distribution with a long tail. Note that for these three families, the predictive mean is a function of the sample mean for any prior distribution π. See Young (1997) for examples of credibility estimators that are functions of a one-dimensional sufficient statistic, not necessarily the sample mean.

In the Bayesian spirit, for a given loss function $L = L(y, d(\overline{x}))$ of the future claim y and the claim predictor d, we propose that the credibility estimator d be the function that minimizes the expected loss

$$E[L(y, d(\overline{x}))],$$

in which we take the expectation with respect to the joint density of the sample mean and future claim. In our mixture model, this joint density is $\int f(y|\theta)\, f(\overline{x}|\theta)\, \pi(\theta)\, d\theta$ Therefore, we require an estimate of the density $\pi(\theta)$.

2.2. Kernel Density Estimation

We use kernel density estimation (Silverman, 1986) to estimate the probability density $\pi(\theta)$. A *kernel* K acts as a weight function and satisfies the condition

$$\int_{-\infty}^{\infty} K(t)dt = 1.$$

If we were to observe directly the conditional means $\theta_1, \theta_2, ..., \theta_r$, then the kernel density estimate of $\pi(\theta)$ with kernel K would be given by

$$\frac{1}{r}\sum_{i=1}^{r}\frac{1}{h_i}K\left(\frac{\theta-\theta_i}{h_i}\right), \tag{2.1}$$

in which h_i is a positive parameter called the *window width*, or *bandwidth*. Assume that the kernel is symmetric; therefore, the expectation of θ is the sample mean.

Because we observe only data x_i and w_i and not the true conditional means θ_i, we rely on the law of large numbers and use the sample mean $\overline{x}_i$ to estimate θ_i consistently, $i = 1, 2, ..., r$, (Serfling, 1980). In the expression in (2.1), one may wish to weight the terms in the sum according to the relative number of claims for the i^{th} risk so that the expectation of θ is the sample mean $\bar{\bar{x}} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} w_{ij}x_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{n_i} w_{ij}} = \frac{\sum_{i=1}^{r} w_i\bar{x}_i}{\sum_{i=1}^{r} w_i}$ in which $w_i = \sum_{j=1}^{n_i} w_{ij}$. We, therefore, propose the following kernel density estimator for $\pi(\theta)$

$$\hat{\pi}(\theta) = \sum_{i=1}^{r} \frac{w_i}{w_{tot}} \frac{1}{h_i} K\left(\frac{\theta - \bar{x}_i}{h_i}\right), \tag{2.2}$$

in which $w_{tot} = \sum_{i=1}^{r} w_i = \sum_{i=1}^{r}\sum_{j=1}^{n_i} w_{ij}$. See the Appendix for a discussion of the asymptotic mean square consistency of $\hat{\pi}(\theta)$.

Two commonly used symmetric kernels are (1) the Gaussian kernel, G,

$$G(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad -\infty < t < \infty,$$

and (2) the Epanechnikov kernel, Epa,

$$Epa(t) = \begin{cases} \frac{3}{4}\frac{\left(1 - \frac{t^2}{5}\right)}{\sqrt{5}}, & -\sqrt{5} < t < \sqrt{5}, \\ 0, & else. \end{cases} \tag{2.3}$$

In our example in Section 4, we use the Epanechnikov kernel because its domain is bounded, and we can, therefore, easily restrict the support of $\hat{\pi}(\theta)$ to lie in the positive real numbers.

Remark: The Epanechnikov is optimal with respect to mean integrated square error (Silverman, 1986). The efficiency of the Gaussian kernel with respect to the optimal Epanechnikov kernel is roughly 95% (Silverman, 1986), so one does not lose much efficiency by using the Gaussian kernel. Silverman, therefore, suggests that one choose the kernel according to auxiliary requirements, such as ease of computing. □

There are many techniques for choosing the window width h_i; see, for example, Silverman (1986, Section 3.4) and Jones, Marron, and Sheather (1996). In our example in Section 4, we use a (modified) fixed window width selected by reference to a standard distribution (Silverman, 1986, Section 3.4.2). The window width h that minimizes the mean integrated squared error is given by

$$h = \left\{\int t^2\, K(t)\, dt\right\}^{-2/5} \left\{\int K(t)^2\, dt\right\}^{1/5} \left\{\int \pi''(\theta)\, d\theta\right\}^{-1/5} r^{-1/5}. \tag{2.4}$$

To approximate this optimal window width h, ones assumes that $\pi(\theta)$ is say, normal, with mean 0 and standard deviation σ. In that case, the term $\int \pi''(\theta)\, d\theta$ equals $\frac{3}{8}\pi^{-1/2}\sigma^{-5}$. We modify the window width h at each point $\bar{x}_i$ to ensure that the density has support on the nonnegative real numbers. Specifically, we set h_i equal to h, if $h < \frac{\bar{x}_i}{\sqrt{5}}$ otherwise, we set h_i equal to $\frac{\bar{x}_i}{\sqrt{5}}$.

3. Credibility using Squared-Error Loss

In this section, we use squared-error loss to determine a credibility estimator, as is used in greatest accuracy credibility theory, (Willmot, 1994) or (Herzog, 1996). The squared-error loss function has the form

$$L(y,\ d(\bar{x})) = (y - d(\bar{x}))^2.$$

It is straightforward to show that the minimizer of the expected loss is the predictive mean (Bühlmann, 1967), which in this case is the posterior mean of θ given the sample mean $\bar{x}$ which we estimate by

$$\hat{\mu}(\bar{x}) = \int E[Y|\theta]\hat{\pi}(\theta|\bar{x})d\theta = \hat{E}[\theta|\bar{x}].$$

For a general kernel K and bandwidths h_i, this estimated posterior mean of θ can be written

$$\hat{E}[\theta|\bar{x}] = \frac{\int \theta\, f(\bar{x}|\theta)\hat{\pi}(\theta)d\theta}{\int f(\bar{x}|\theta)\hat{\pi}(\theta)d\theta}$$

$$= \frac{\sum_{i=1}^{r} \frac{w_i}{h_i} \int \theta\, f(\bar{x}|\theta) K\left(\frac{\theta - \bar{x}_i}{h_i}\right) d\theta}{\sum_{i=1}^{r} \frac{w_i}{h_i} \int f(\bar{x}|\theta) K\left(\frac{\theta - \bar{x}_i}{h_i}\right) d\theta} \tag{3.1}$$

Recall that $\bar{x}$ is an average of w iid claims, each of which follows the density $f(y|\theta)$, as in Section 2.1. If we constrain the estimator d to be linear, then it is well-known that the least-squares linear estimator of $E[Y|\bar{x}] = E[\theta|\bar{x}]$ is

$$d(\bar{x}) = (1 - Z)E[Y] + Z\bar{x}, \tag{3.2}$$

in which $Z = \frac{w}{w + k}$ with $k = \frac{\hat{E}Var[Y|\theta]}{V\hat{a}r[\theta]}$ (Bühlmann, 1967). Using our estimate for the prior density (2.2), we obtain $\hat{E}[Y] = \hat{E}[\theta] = \bar{\bar{x}}$, as noted in Section 2.2. In the case of the normal conditional; $k = \frac{\sigma^2}{\hat{E}[\theta^2] - \bar{\bar{x}}^2}$ in the case of the gamma conditional, $k = \frac{\hat{E}[\theta^2]}{\alpha\left(\hat{E}[\theta^2] - \bar{\bar{x}}^2\right)}$; and in the case of the inverse gaussian conditional,

$$k = \frac{\hat{E}[\theta^3]}{\lambda\left(\hat{E}[\theta^2] - \bar{\bar{x}}^2\right)}.$$

To end this section, we show that as w approaches ∞, $\hat{\mu}(\bar{x})$ approaches the true expected value θ_0, for the given risk. Because $\bar{X}|\theta$ has mean θ and variance $\frac{Var(Y|\theta)}{w}$ under certain regularity conditions, (DeGroot, 1970) and (Walker, 1969), the density $f(\bar{x}|\theta)$ approaches the delta function with its mass concentrated at the point $\bar{x} = \theta_0$ Then,

$$\lim_{w\to\infty} \hat{\mu}(\bar{x}) = \lim_{w\to\infty} \frac{\int \theta f(\bar{x}|\theta)\hat{\pi}(\theta)d\theta}{\int f(\bar{x}|\theta)\hat{\pi}(\theta)d\theta} = \frac{\theta_0\hat{\pi}(\theta_0)}{\hat{\pi}(\theta_0)} = \theta_0, \; w.p. 1.$$

Thus, as an actuary gets more claim information for a given policyholder (w gets large), the estimated expected claim approaches the true expected claim with probability 1.

4. SIMULATED DATA FROM A LOGNORMAL-LOGNORMAL MIXTURE

The lognormal distribution is used by actuaries to model the distribution of claim severity. It is also used to model the distribution of total claims in some lines of insurance, such as health insurance. In this section, we assume that we are given individual claim data; that is, $w_{ij} = 1$, for all risks i and policy periods j, and $X = Y$. We model the lognormal-lognormal mixture as follows:

$$f(x|\phi) = \frac{1}{\sigma x\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}\left[\ln\left(\frac{x}{\phi}\right)\right]^2\right\}, \quad x > 0,$$

in which $\sigma > 0$ is a known parameter, and

$$\pi(\phi) = \frac{1}{\tau\phi\sqrt{2\pi}}\exp\left\{-\frac{1}{2\tau^2}\left[\ln\left(\frac{x}{\phi}\right)\right]^2\right\}; \quad \phi > 0,$$

in which $\mu > 0$ and $\tau > 0$ are known parameters. That is, $(\ln X)|\phi \sim N(\ln\phi, \sigma^2)$, and $\ln\phi \sim N(\ln\mu, \tau^2)$. The marginal distribution of X is lognormal; $\ln X \sim N(\ln\mu, \sigma^2 + \tau^2)$.

Given claim data for a specific policyholder, $\boldsymbol{X} = \boldsymbol{x} = \langle x_1, x_2, \ldots, x_n \rangle \in [0, \infty]^n$, the posterior distribution of $\phi|\boldsymbol{x}$ is lognormal; $(\ln\phi)|\boldsymbol{x} \sim N\left(\ln\mu^*, \tau^{*^2}\right)$, in which

$$\mu^* = \exp\left(\frac{\sigma^2\ln\mu + \tau^2 t}{\sigma^2 + n\tau^2}\right),$$

$t = \sum_{i=1}^{n} \ln(x_i)$ and

$$\tau^{*^2} = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}.$$

Thus, the predictive distribution of $X_{n+1}|\boldsymbol{x}$ is lognormal; $(\ln X_{n+1})|\boldsymbol{x} \sim N\left(\ln \mu^*, \sigma^2 + \tau^{*^2}\right)$. It follows that the true predictive mean is a function of t

$$\mu(\boldsymbol{x}) = E(X_{n+1}|\boldsymbol{x}) = \exp\left(\frac{\sigma^2 \ln \mu + \tau^2 t}{\sigma^2 + n\tau^2} + \frac{\sigma^2(\sigma^2 + (n+1)\tau^2)}{2(\sigma^2 + n\tau^2)}\right). \tag{4.1}$$

We performed 200 simulations of a lognormal-lognormal mixture of claims. We let $\sigma^2 = 0.25$, $\tau^2 = 0.50$, and $\mu = 2000e^{-0.25}$. The marginal expectation of X is 2267, and the marginal standard deviation is 2395. For each simulation run, we simulated claim data from this lognormal-lognormal mixture for $r = 100$ risks (values of ϕ). For each of the 100 risks, we simulated $n_i = w_i = 5$ claims. To estimate the distribution of the conditional means, we used kernel density estimation with the Epanechnikov kernel, as given by (2.3). Also, we used a fixed window width h, chosen by reference to a normal distribution with mean 0 and standard deviation σ. We estimated the standard deviation by the interquartile range of the sample means, R, divided by 1.34 (Silverman, 1986, Section 3.4). The bandwidth h was calculated by $h = (1)^{-2/5}(0.268)^{1/5}(0.212)^{-1/5}\frac{R}{1.34}100^{-1/5}$ $\approx 0.312R$ as in (2.4). We truncated this bandwidth h for a given risk if, by otherwise using it, the prior density would have a negative support. Specifically, if $h > \frac{\overline{x}_i}{\sqrt{5}}$ then we set the bandwidth h_i equal to $\frac{\overline{x}_i}{\sqrt{5}}$ to guarantee that the support of the estimated density of θ be contained in the nonnegative real numbers, as described in Section 2.2.

Instead of assuming that the conditional is lognormal, we assumed that the coefficient of variation is constant from risk to risk and, therefore, fit a gamma conditional to each risk. In each simulation run, we estimated the parameter α by the median of the following sample statistic $\frac{\overline{x}_i^2}{\frac{1}{5-1}\sum_{j=1}^{5}\left(x_{ij} - \overline{x}_i\right)^2}$. We used the estimated prior density along with the gamma conditional to estimate the marginal density of X.

We used the estimated mixture model to estimate the predictive mean of X_{n+1} given claim data $\boldsymbol{x}$. We also computed the Bühlmann credibility estimator, $lin(\boldsymbol{x})$, for which we estimated the expected process variance by

$$E\hat{P}V = \frac{1}{100(5-1)}\sum_{i=1}^{100}\sum_{j=1}^{5}\left(x_{ij} - \overline{x}_i\right)^2$$

and the variance of the hypothetical means by

$$V\hat{H}M = \frac{1}{100-1}\sum_{i=1}^{100}\left(\overline{x}_i - \overline{\overline{x}}\right)^2 - \frac{E\hat{P}V}{5},$$

(Willmot, 1994, Section 5.1).

TABLE 4.1
DESCRIPTIVE STATISTICS OF h, MSE, MSEB, AND RATIO

Variable	Mean	Median	StDev	Q1	Q3
h	564.35	561.00	91.64	500.25	623.75
MSE	16,450	12,111	13,146	7,808	21,623
$MSEB$	74,559	69,595	37,539	44,466	94,878
$Ratio$	0.2984	0.1777	0.3239	0.0890	0.3819

For $n = w = 1$, we compared the estimated predictive mean, $\hat{\mu}(x)$ and the Bühlmann credibility estimator, $lin(x)$, with the true predictive mean, $\mu(x)$. To compare these credibility estimators numerically, for each of the 200 simulation runs, we calculated the mean squared errors up to the 95^{th} percentile of X, namely 6,500: $MSE = \int_0^{6500} (\hat{\mu}(x) - \mu(x))^2 f(x)dx$ and $MSEB = \int_0^{6500} (lin(x) - \mu(x))^2 f(x)dx$. See Table 4.1 for descriptive statistics of the bandwidth h; the mean squared errors, MSE and $MSEB$; and the ratio of MSE to $MSEB$, $Ratio$.

Thus, we see that up to the 95^{th} percentile, on average, our estimated predictive mean performs much better than the linear Bühlmann credibility estimator. See Figure 4.1 for a scatter plot of MSE versus h. Note the quadratic relationship between the two variables and that the minimum of MSE occurs near the average value of h, 564. We fit a quadratic to these observations by minimizing the sum of the absolute values of the errors and obtained the fitted model

$$\hat{MSE} = 196,603 - 691.36h + 0.6402h^2,$$

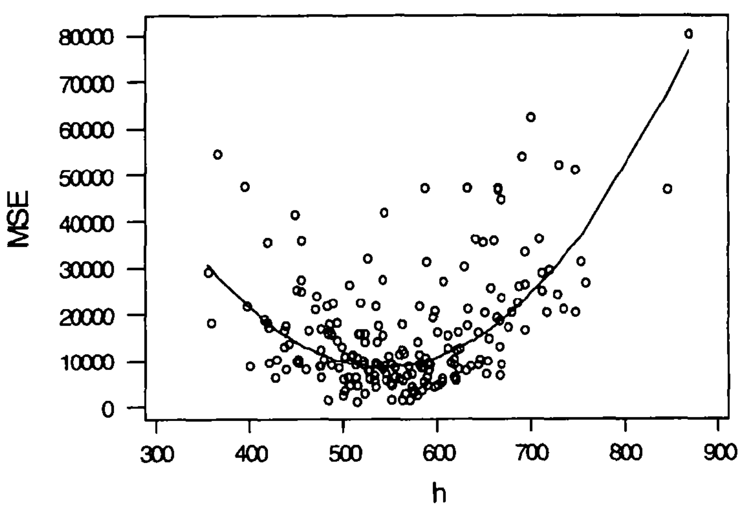


FIGURE 4.1: Scatter Plot of MSE versus h with Quadratic Superimposed.

with vertex at 542. See Figure 4.1 for a graph of this quadratic superimposed on a scatter plot of the observations.

We also computed some of the mean squared errors up to the 99^{th} percentile and found that the estimated predictive mean compared poorly relative to the Bühlmann credibility estimator. We conclude that our estimate of the prior density at larger conditional means may suffer. Silverman (1986) suggests a variable bandwidth approach for estimating densities with long tails which uses

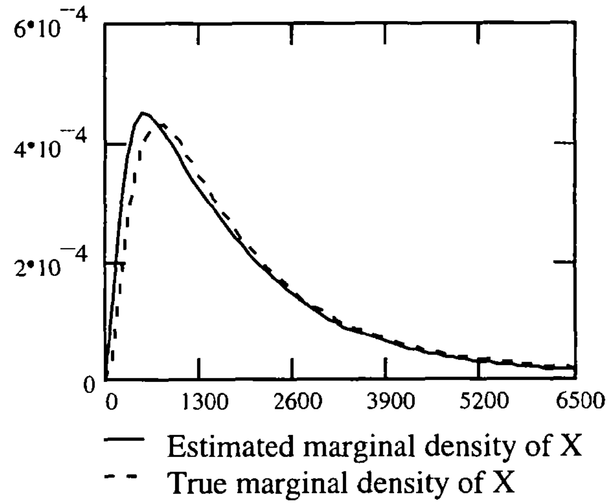


FIGURE 4.2: Estimated and True Marginal Densities of Claims.

larger bandwidths in the regions of lower density. We tried this method without increased accuracy in the upper percentiles of our claim estimator. We suspect that the poor fit at the higher percentiles may be due to our using a medium-tailed gamma conditional to model a heavy-tailed lognormal. We encourage the interested reader to investigate using an inverse gaussian instead of a gamma conditional to model the conditional claim distribution.

See Figure 4.2 for graphs of the estimated and true marginal densities of X for one of the simulations[1]. Of the graphs we plotted, Figure 4.2 is typical, in that the estimated marginal density of X is less skewed than the true density.

See Figure 4.3 for the corresponding graphs of the estimated and true predictive means. Notice how closely the estimated predictive mean follows the true predictive mean, compared with the linear Bühlmann estimator for claims less than 4000. Also note how the estimated predictive mean diverges upward for claims larger than 4000. This phenomenon occurred in all of the several graphs that we plotted and is due, we believe, to the fact that we used a gamma conditional to estimate a lognormal. It may also be due to computational errors

[1] In this run, h = 476, MSE = 12,076, and $MSEB$ = 84,571. Recall that $n = 1$ and that the claim amount 6,500 is the 95^{th} percentile of X.

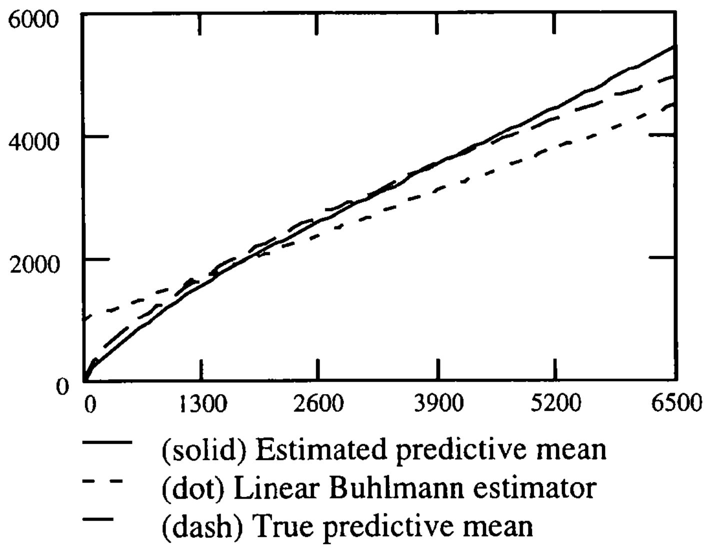


FIGURE 4.3: Credibility Estimators.

because there are only a few simulated claims in the right tail. One way to adjust the estimated predictive mean to eliminate this divergence is to extend it linearly beyond some large value of the sample mean. Another solution may be to use a conditional distribution with a longer tail, such as the inverse gaussian. Yet another solution may be to apply my method of blending the criteria of accuracy and linearity (Young, 1997).

5. SUMMARY AND CONCLUSIONS

The Bühlmann-Straub credibility method results in a linear estimator with a different slope (or credibility weight) for each risk. Therefore, to apply their method to a risk not used to construct the original model, one would be required to recalculate the model to obtain a linear estimator for the new risk. An advantage of our method is that it is applicable to risks outside the original data set, if one assumes that the average claims and corresponding exposures of the new risk come from the same parent (mixture) population as the data. Another advantage of our method is increased accuracy over a linear estimator, as demonstrated in the example in Section 4, even when we use an 'incorrect' conditional density.

One may wish to use the underlying mixture model and kernel density estimation in combination with other loss functions, such as a linear combination of a squared-error term and a second-derivative term to blend the goals of accuracy and linearity (Young, 1997). Also, it would be interesting if one were to extend the model to include a trend component, as in Hachemeister (1975), and apply kernel density estimation in the more general model.

ACKNOWLEDGMENTS

I thank the Committee for Knowledge Extension and Research of the Society of Actuaries (SOA) and the Actuarial and Education Research Fund for financial support. I especially thank my SOA Project Oversight Group (Hans Gerber and Gary Venter, led by Thomas Herzog and assisted by Warren Luckner of the SOA) for helpful guidance.

APPENDIX

ASYMPTOTIC MEAN SQUARE CONSISTENCY OF (2.2)

Let $\tilde{\pi}(\theta) = \sum_{i=1}^{r} \frac{w_i}{w_{tot}} \frac{1}{h_i} K\left(\frac{\theta-\theta_i}{h_i}\right)$ denote the kernel density estimator of π when we are given observations θ_i, $i = 1, 2, ..., r$. Consider the mean squared error of the density estimate $\hat{\pi}$ at a fixed value θ :

$$E\left[(\hat{\pi}(\theta) - \pi(\theta))^2\right] = E\left[(\hat{\pi}(\theta) - \tilde{\pi}(\theta))^2 + 2(\hat{\pi}(\theta) - \tilde{\pi}(\theta))(\tilde{\pi}(\theta) - \pi(\theta)) + (\tilde{\pi}(\theta) - \pi(\theta))^2\right]$$

$$= E\left[\sum_{i=1}^{r} \frac{w_i}{w_{tot}} \frac{1}{h_i}\left\{K\left(\frac{\theta - \bar{x}_i}{h_i}\right) - K\left(\frac{\theta - \theta_i}{h_i}\right)\right\}^2\right] + E\left[(\tilde{\pi}(\theta) - \pi(\theta))^2\right]$$

$$+2E\left[\sum_{i=1}^{r} \frac{w_i}{w_{tot}} \frac{1}{h_i}\left\{K\left(\frac{\theta - \bar{x}_i}{h_i}\right) - K\left(\frac{\theta - \theta_i}{h_i}\right)\right\} \cdot (\tilde{\pi}(\theta) - \pi(\theta))\right].$$

By the law of large numbers (Serfling, 1980), $\bar{x}_i$ approaches θ_i, with probability one, as w_i approaches infinity. Therefore, as w_i approaches infinity, the first term in the mean squared error goes to zero. By Silverman (1986) or Thompson and Tapia (1990), the second and third terms go to zero as r goes to infinity if $\lim_{r\to\infty} h_i = 0$ and $\lim_{r\to\infty} rh_i = \infty$.

REFERENCES

BÜHLMANN, H. (1967), Experience rating and credibility, *ASTIN Bulletin*, 4: 199-207.

BÜHLMANN, H. (1970), *Mathematical Models in Risk Theory*, Springer-Verlag, New York.

BÜHLMANN, H. and E. Straub (1970), Glaubwürdigkeit für Schadensätze, *Mitteilungen der Vereinigung Schweizerischer Versicherungs-Mathematiker*, 70: 111-133.

DE GROOT, M. H. (1970), *Optimal Statistical Decisions*, McGraw-Hill, New York.

HACHEMEISTER, C. A. (1975), Credibility for regression models with application to trend, 129-163, in *Credibility: Theory and Applications*, (ed. P. M. Kahn), 129-163, Academic Press, New York.

HERZOG, T. L. (1996), *Introduction to Credibility Theory*, second edition, ACTEX, Abington, Connecticut.

JONES, M. C., J. S. MARRON, and S. J. SHEATHER (1996), A brief survey of bandwidth selection for density estimation, *Journal of the American Statistical Association*, **91**: 401-407.

KLUGMAN, S. A. (1992), *Bayesian Statistics in Actuarial Science with Emphasis on Credibility*, Kluwer, Boston.

SERFLING, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

SILVERMAN, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

THOMPSON, J. R. and R. A. TAPIA (1990), *Nonparametric Function Estimation, Modeling, and Simulation*, Society for Industrial and Applied Mathematics, Philadelphia.

WALKER, A. M. (1969), On the asymptotic behaviour of posterior distributions, *Journal of the Royal Statistical Society*, Series B, **31**: 80-88.

WILLMOT, G. E. (1994), *Introductory Credibility Theory*, Institute of Insurance and Pension Research, University of Waterloo, Waterloo, Ontario.

YOUNG, V. R. (1997), Credibility using a loss function from spline theory: Parametric models with a one-dimensional sufficient statistic, to appear, *North American Actuarial Journal*.

VIRGINIA R. YOUNG
School of Business
975 University Avenue
University of Wisconsin-Madison
Madison, WI USA 53706

EXACT CREDIBILITY FOR WEIGHTED OBSERVATIONS

BY ROB KAAS, DENNIS DANNENBURG and MARC GOOVAERTS[1]

University of Amsterdam

ABSTRACT

This note generalizes Jewell's theorem on exact credibility from the classical Bühlmann model to the (weighted) Bühlmann-Straub model.

1. INTRODUCTION

A well-known theorem of Jewell (1974) states that exact credibility, which is the concurring of the Bayesian estimator (posterior mean) with the credibility estimator of a contract mean, is found for a class of examples which includes many common situations. In a nutshell, exact credibility obtains when the observations are drawn from distributions in the exponential family, with natural conjugate prior distributions for the risk parameter. Surprisingly, Jewell's theorem pertains only to the classical Bühlmann model, and does not hold in case different variances of the observations are allowed, as in the Bühlmann-Straub model. In this contribution we prove exact credibility to hold for the (weighted) Bühlmann-Straub model as well, thus allowing the observations to be averages of varying numbers of observations, also in case of Poisson and Binomial distributions. The parametrization used coincides with the one used in the theory of Generalized Linear Models. In the original form of Jewell's theorem, and in ours as well, rather cumbersome reparametrizations are required to prove that ordinary distributions like Poisson and Gamma are special cases of this theorem. This is remedied in Gerber (1995) by choosing a more convenient parametrization.

Our extension of Jewell's theorem still does not incorporate Jewell's hierarchical model. Exact credibility for this model, and also for even more complicated ones like Hachemeister's regression credibility model, can, however, easily be proven for the normal-normal model. This is because conditional expectations of multinormal random variables are linear in the conditions.

Consider a portfolio consisting of J contracts, for which we have data of the past claims. These observations are assumed to have been generated by a Bayesian chance mechanism: first a contract-specific risk parameter Θ_j, $j = 1, \ldots, J$, is drawn from a structure distribution with known parameters, called hyperparameters. In this contribution, we will concentrate on one particular contract j. Since the observations on other, independent, contracts do not appear in the estimators of $\mu(\Theta_j)$ used, we will simply write the observations as $X_1, \ldots, X_T$, and not incorporate the contract number j in our

[1] The authors acknowledge the contributions of Ria Kuzee.

notation. The unobservable random variable Θ represents the risk characteristics of the contract in question. These characteristics may be unobservable, or if they are, unusable for reasons such as social acceptability. The observations, conditionally given Θ, are independently drawn from some distribution of which the mean is a function of the parameter Θ. The quantity of interest is not Θ itself, but the risk premium for contract j, traditionally denoted by $\mu(\Theta) = E[X|\Theta]$. The risk variable Θ acts as a parameter of the distribution of the risks $X_1, X_2, \ldots, X_T$; conditionally on Θ, the risks will be independent with mean $\mu(\Theta)$. They are not necessarily identically distributed, since the conditional variance of X_t is taken inversely proportional to some known weight w_t, just as in the Bühlmann-Straub model.

The best estimator of $\mu(\Theta)$, in the least squares sense, in the class consisting of all random variables of type $g(X_1, X_2, \ldots, X_T)$ where $g(\cdot)$ is any function, is the one with minimal mean squared error $E[\{g(X_1, X_2, \ldots, X_T) - \mu(\Theta)\}^2]$. It is obtained by taking $g = g^*$ with

$$g^*(X_1, X_2, \ldots, X_T) = E[\mu(\Theta)|X_1, X_2, \ldots, X_T] \tag{1}$$

Thus, we see that the best predictor of $\mu(\Theta)$ is just the conditional mean of $\mu(\Theta)$, given the observations, or in the idiom of Bayesian estimation, the posterior mean. These posterior means may have a rather unpleasant form, which is the reason why in credibility theory the restriction to linear functions of the observations is imposed. The estimator thus obtained is not only the best approximation to $\mu(\Theta)$, but it is also closest to the posterior mean (1). It can be shown that if the simultaneous distribution of Θ and $X_1, \ldots, X_T$ is of particular type, $g^*(\cdot)$ happens to be a linear function of the data already, and thus is the credibility estimator. In the second section of this note, we investigate conditions for which this holds.

2. Exact credibility

When the optimal Bayes estimator (1) is linear, it is obviously equal to the credibility estimator, since they both minimize the mean squared error. In this case we say that the credibility estimator is exact Bayesian, or equivalently, that exact credibility holds. Jewell (1974, 1975) showed that exact credibility is found when the observations $X_1, \ldots, X_T$, given the value of the structure parameter Θ, are an iid random sample from the exponential family of distributions; moreover, the prior distribution of Θ must be the so-called natural conjugate prior, which ensures that the posterior distribution of Θ, given $X_1, \ldots, X_T$, is of the same type as the prior distribution. In Jewell's original theorem, the X_t are iid, given Θ, as is the case in the original Bühlmann model. To be able to apply the theorem to the more general Bühlmann-Straub model, we have to account for the observations having different variances (weights). For definitions and assumptions of these credibility models, consult e.g. Goovaerts et al. (1990).

The well-known exponential family of distributions contains many frequently used distributions. Prominent members are the Normal, Poisson, Binomial, Gamma and Inverse Gaussian distributions. The densities in it can be written as :

$$f_X(x;\theta) = \exp\left[\frac{x\theta - b(\theta)}{\phi/w} + c(x, \phi/w)\right], x \in A_w. \tag{2}$$

The parameter θ of the distribution of X will be regarded as a realization of a structure random variable Θ. The other parameter ϕ is a dispersion parameter, like σ^2 in normal distributions. It may be assumed known or unknown. For one-parameter distributions, for example the Poisson, ϕ is taken to be 1. The weight $w > 0$ is known. Since (2) involves only the ratio ϕ/w, we might also say that only the relative weights of the contracts are known. Just as in the Bühlmann-Straub model, the variance of X, given Θ, is proportional to $1/w$. This is the case when X is an average of w elementary claims (natural weight), as we will prove later on, but w is not necessarily an integer. The set A_w consists of possible values of the claims. If the elementary risks are for instance Poisson, then $A_w = \{0,\ 1/w,\ 2/w,\ ..\}$. In the sequel, we assume X to be continuous. In the discrete case, integrals over $x \in A_w$ below should be replaced by summations.

The above parametrization of the exponential family is sometimes called 'natural', in view of the fact that the part of it depending on both x and θ has the form $e^{x\theta}$. As we will see later on, it proves that the natural parametrization is not always the customary one, which is generally chosen because it is the most convenient. The one Gerber (1995) uses makes the reparametrizations much easier, but gives problems when incorporating weights. The parametrization we use closely resembles the one standard in the theory of Generalized Linear Models, see, e.g., McCullagh and Nelder (1989) or Nelder and Verrall (1995). Here ϕ is, without much gain of generality, replaced by $a(\phi)$.

We can evaluate the moment generating function with density (2) as follows:

$$\begin{aligned}
m_X(r) &= \int e^{rx} \exp\left[\frac{x\theta - b(\theta)}{\phi/w} + c(x,\ \phi/w)\right]\, dx \\
&= \int \exp\left[\frac{x\{\theta + r\phi/w\} - b(\theta + r\phi/w)}{\phi/w} + c(x,\ \phi/w)\right]\, dx\ \times \\
&\quad \times \exp\left[\frac{b(\theta + r\phi/w) - b(\theta)}{\phi/w}\right] \\
&= \exp\left[\frac{b(\theta + r\phi/w) - b(\theta)}{\phi/w}\right].
\end{aligned} \tag{3}$$

Note that the second integral in (3) equals one, because it is the integral over a density of type (2), with θ replaced by $\theta + r\phi/w$. Mean and variance of (2) follow easily from the cumulant generating function $\kappa_X(r) = \log m_X(r)$:

$$\begin{aligned}
\mu(\theta) &= E[X|\Theta = \theta] = \kappa_X'(0) = b'(\theta); \\
Var[X;\theta] &= \kappa_X''(0) = b''(\theta)\phi/w.
\end{aligned} \tag{4}$$

The function $b(\theta)$ is sometimes referred to as the cumulant function.

Having found the moment generating function, we can show that density (2) truly represents the density of an average $\overline{X} = \frac{1}{w}\sum_i X_i$ of w iid random variables $X_1, \ldots, X_w$, with the same density (2), but with weight 1. Indeed we have

$$m_{\overline{X}}(r) = m_{\sum X_i}(\frac{r}{w}) = \{m_{X_i}(\frac{r}{w})\}^w = \exp\left[\frac{b(\theta + r\phi/w) - b(\theta)}{\phi/w}\right]. \tag{5}$$

Assume that Θ has a prior density which is the so-called *natural conjugate prior*, i.e., of which the θ-dependent part is the same as in (2), and $x_0, \phi/w_0$ are parameters:

$$f_\Theta(\theta) = \exp\left[\frac{\theta x_0 - b(\theta)}{\phi/w_0} + d(x_0, \phi/w_0)\right]. \tag{6}$$

The normalizing function $d(x_0, \phi/w_0)$ is chosen in such a way that the density, which ranges over some θ-interval, integrates to one. Assume further that, conditionally given $\Theta = \theta$, the random variables X_t are independent drawings from density (2) with parameters θ, ϕ and weight w_t, $t = 1, \ldots, T$. Then the posterior density of Θ, given $X_1 = x_1, \ldots, X_T = x_T$, is found to be, apart from division by a normalizing constant equal to the integral over θ of the resulting expressions:

$$f_{\Theta|X_1,\ldots,X_T}(\theta|x_1, \ldots, x_T) \propto f_\Theta(\theta)\prod_{t=1}^{T}\exp\left[\frac{x_t\theta - b(\theta)}{\phi/w_t} + c(x_t, \phi/w_t)\right]$$

$$\propto \prod_{t=0}^{T}\exp\left[\frac{x_t\theta - b(\theta)}{\phi/w_t}\right] = \exp\sum_{t=0}^{T}\left[\frac{w_t x_t\theta - w_t b(\theta)}{\phi}\right] = \exp\left[\frac{\theta x_\bullet - b(\theta)}{\phi/w_\bullet}\right], \tag{7}$$

$$\text{where } w_\bullet = \sum_{t=0}^{T} w_t \text{ and } x_\bullet = \sum_{t=0}^{T}\frac{w_t}{w_\bullet}x_t. \tag{8}$$

Thus, posterior and prior distribution are of the same type, but with parameter x_0 replaced by $x_\bullet$ and w_0 by $w_\bullet$.

As a corollary to the above discussion we formulate the main theorem of exact credibility:

Theorem 2.1 (Posterior mean equals credibility estimator of exponential family with natural conjugate prior)
Suppose that, conditionally on $\Theta = \theta$, $X_1, \ldots, X_T$ are independent random variables with density (2) for fixed ϕ and weights $w_t, t = 1, \ldots, T$. Further, let Θ have a prior distribution (6) with parameters x_0 and w_0. Then the posterior mean $E[\mu(\Theta)|X_1, \ldots, X_T]$ is an inhomogeneous linear form in $X_1, \ldots, X_T$ provided the prior density (6) vanishes at the endpoints of the θ-interval.

Proof. We must prove that the following expression is linear in $x_1, \ldots, x_T$:

$$\int \mu(\theta) f_{\Theta|X_1,\ldots,X_T}(\theta|x_1, \ldots, x_T)d\theta. \tag{9}$$

Since $\mu(\theta) = b'(\theta)$ by (4) and the posterior density is proportional to (7), we must compute

$$\int \mu(\theta) f_\Theta(\theta; x_\bullet, w_\bullet)d\theta = \frac{\int b(\theta)\exp\{\frac{w_\bullet}{\phi}[\theta x_\bullet - b(\theta)]\}\,d\theta}{\int \exp\{\frac{w_\bullet}{\phi}[\theta x_\bullet - b(\theta)]\}\,d\theta}$$

$$= x_\bullet - \frac{\int \{x_\bullet - b'(\theta)\}\exp\{\frac{w_\bullet}{\phi}[\theta x_\bullet - b(\theta)]\}\,d\theta}{\int \exp\{\frac{w_\bullet}{\phi}[\theta x_\bullet - b(\theta)]\}\,d\theta} \tag{10}$$

$$= x_\bullet - \frac{\int \frac{\phi}{w_\bullet} d\,\exp\{\frac{w_\bullet}{\phi}[\theta x_\bullet - b(\theta)]\}}{\int \exp\{\frac{w_\bullet}{\phi}[\theta x_\bullet - b(\theta)]\}d\theta} = x_\bullet,$$

where the numerator vanishes because by assumption, $\theta x_\bullet - b(\theta) = -\infty$ at both endpoints of the integration interval. By (8), this expression is indeed inhomogeneous linear in $x_1, \ldots, x_T$. ∎

Remark 2.2 (Credibility factor and virtual experience)
By (8), we may write the estimator $X_\bullet$ resulting from (10) as follows:

$$X_\bullet = \frac{w_0 x_0 + \sum_{t=1}^T w_t X_t}{w_0 + \sum_{t=1}^T w_t} = zX_w + (1-z)m, \tag{11}$$

where $z = \dfrac{w_\Sigma}{w_0 + w_\Sigma}$ for $w_\Sigma = \sum_{t=1}^T w_t$ is the credibility factor,

$$X_w = \sum_{t=1}^T \frac{w_t}{w_\Sigma} X_t \tag{12}$$

and $m = E[\mu(\Theta)] = x_0$ (see (10)).

So the premium is the ratio of total claims and exposure, where a 'virtual experience' of m on average in w_0 exposure units is added to the actual experience of X_w on average, with a total weight (exposure) of w_Σ. ∎

Remark 2.3 (Credibility estimator equals posterior mode)
Under the same conditions of the previous theorem, the maximum of the posterior density is found when θ is such that $\mu(\theta) = x_\bullet$ as well, since the derivative of (7) is zero when $x_\bullet = b'(\theta)$, which equals $\mu(\theta)$ by (4). ∎

In the examples that follow, two special cases are given where the credibility estimators of the Bühlmann-Straub model are exact Bayesian. The third example shows that not all cases of exact credibility are covered by Theorem 2.1.

Example 2.4 (Poisson observations with Gamma prior)
Suppose that the risks X_t represent average numbers of claims in homogeneous cells with w_t policies in it, which, given $\Lambda = \lambda$, are Poisson(λ) distributed, for some positive structure random variable Λ. In automobile insurance, this risk parameter represents the 'accident-proneness' of the drivers in the cell considered. In general, the Gamma distribution proves to describe the spread of Λ rather well. The conditional density is

$$\begin{aligned} f_{X|\Lambda}(x|\lambda) &= \frac{e^{-\lambda w}(\lambda w)^{wx}}{(wx)!}, \; x \in A = \{0, \; 1/w, \; 2/w, \; ...\} \\ &= \exp[w\{x \log \lambda - \lambda\} + w\, x \; \log \; w - \log((wx)!)]. \end{aligned}$$

From the last expression we see that this density belongs to the exponential family (2), with

$$\theta = \log \lambda, \;\; \phi = 1, \;\; b(\theta) = e^{\theta}, \;\; c(x, \; \phi/w) = w\, x \; \log \; w - \log((wx)!). \tag{14}$$

By (6), the natural prior of $\Theta = \log \Lambda$ is, apart from the normalization constant $d(x_0\; 1/w_0)$:

$$f_{\Theta}(\theta) \propto \exp\left[\frac{\theta x_0 - e^{\theta}}{1/w_0}\right], -\infty < \theta < \infty, \tag{15}$$

for some parameters $x_0 > 0$ and $w_0 > 0$ The corresponding density for Λ is then

$$f_{\Lambda}(\lambda) = f_{\Theta}(\log(\lambda))\left|\frac{d\theta}{d\lambda}\right| \;\; \propto \;\; \lambda^{x_0 w_0 - 1} e^{-\lambda w_0}, \;\; \lambda > 0, \tag{16}$$

in which we immediately recognize the Gamma(α, β) distribution with $\alpha = x_0 w_0, \;\; \beta = w_0$.

It is easy to verify that the extra condition of Theorem 2.1 is met, since $\theta x_0 - b(\theta)$ tends to $-\infty$ both for $\theta \to -\infty$ and $\theta \to \infty$. Therefore we know that the original Bühlmann inhomogeneous credibility estimator of $\mu(\Theta) = E[X|\Theta](= \exp(\Theta))$ is exact Bayesian. As a consequence, the conditional mean of $\mu(\Theta)$, given $X_1, \; ..., \; X_T$, is linear in $X_1, \; ..., \; X_T$. Because Λ is a one-to-one function of Θ, we have also

$$\mu(\Theta) = E[X|\Theta] = E[X|\log(\Theta)] = E[X|\Lambda]. \tag{17}$$

So we conclude that the conditional expected value of $\mu(\Theta) = E[X|\Lambda]$, given $X_1, \; ..., \; X_T$, is linear as well. This means that if the claims are averages of Poisson distributions with the usual parametrization of the first expression in (13), and the prior distribution of Λ is Gamma, the inhomogeneous credibility estimator for the Bühlmann-Straub model is exact Bayesian. ∎

Example 2.5 (Normal distribution with Normal prior in the Bühlmann-Straub model)
Another example of exact credibility arises if the risks X_t are independent and $N(\theta, s^2/w_t)$ distributed, conditionally given $\Theta = \theta$, where Θ is an $N(m, a)$ distributed random variable. This model arises when $X_t = m + \Xi + \Xi_t$ for independent normal Ξ-components, with $\Xi \sim N(0, a)$ and $\Xi_t \sim N(0, s^2/w_t)$. Then $\Theta = m + \Xi$. To determine the credibility estimator, only the first and second order moments matter, and they are just those of the Bühlmann-Straub model. Recall that contracts of other cells appear neither in the posterior mean, nor in the inhomogeneous credibility estimator, by the independence between the cells.

The conditional density of the X_t can be written as

$$f_{X_t|\Theta}(x_t|\theta) = \frac{1}{\sqrt{2\pi s^2/w_t}} \exp \frac{-(x_t - \theta)^2}{2s^2/w_t}$$

$$= \exp\left[\frac{x_t\theta - {}^1\!/\!_2\theta^2}{s^2/w_t} - {}^1\!/\!_2\left\{\frac{x_t^2}{s^2/w_t} + \log(2\pi s^2/w_t)\right\}\right], \tag{18}$$

which is (2) when

$$b(\theta) = {}^1\!/\!_2\theta^2,\ \phi = s^2, c(x_t, s^2/w_t) = -{}^1\!/\!_2\{\frac{x_t^2}{s^2/w_t} + \log(2\pi s^2/w_t)\}. \tag{19}$$

The natural conjugate prior density is again normal, see (6) and (18), so

$$f_\Theta(\theta) = \frac{1}{\sqrt{2\pi a}} \exp \frac{-(\theta - m)^2}{2a} \propto \exp \frac{\theta m - {}^1\!/\!_2\theta^2}{s^2/(s^2/a)}, \tag{20}$$

which, apart from the normalization constant, equals expression (6) when

$$x_0 = m,\ \phi = s^2,\ w_0 = s^2/a. \tag{21}$$

Because $\theta x_w - {}^1\!/\!_2\theta^2$ again tends to $-\infty$ both for $\theta \to -\infty$ and $\theta \to \infty$, we find from Theorem 2.1 that the posterior mean equals the Bühlmann-Straub estimator

$$E[\Theta|X_1, \ldots, X_T] = \frac{w_0 x_0 + \sum_{t=1}^T w_t X_t}{w_0 + \sum_{t=1}^T w_t} = z\frac{\sum_{t=1}^T w_t X_t}{\sum_{t=1}^T w_t} + (1 - z)m,$$

with $z = \dfrac{a\sum_{t=1}^T w_t}{a\sum_{t=1}^T w_t + s^2}$. (22)

■

Other situations, apart from Examples 2.4 and 2.5, in which exact credibility holds are (Negative) Binomial data with Beta prior, and Inverse Gaussian data with the corresponding natural prior.

Example 2.6 (Exact credibility in normal-normal models)
Theorem 2.1, which extends Jewell's original theorem to the weighted case, cannot be applied to Jewell's hierarchical credibility model, see, e.g., Goovaerts et al. (1990). Written in the same additive components form of the previous example, the statistic for sector p, cell j, and time period t is

$$X_{pjt} = m + \Xi_p + \Xi_{pj} + \Xi_{pjt}, \tag{23}$$

where the Ξ-components of the risks are independent with mean zero and variances b, a and s^2/w_{pjt}. First we try to consider only one sector p. Then, as is required in Theorem 2.1, the observations of other sectors are independent of the ones considered. Conditionally on Θ, which in this case is Ξ_p, the observations in sector p have the same mean. They are, however, not independent, since the observations in cell j of this sector have a common risk component Ξ_{pj}. If on the other hand we only look at a specific cell j in some sector p, taking $\Theta = \Xi_p + \Xi_{pj}$ we do have that conditionally given Θ, the observations are independent and have equal mean. But in this case the other observations cannot be disregarded when estimating the risk premium of this cell, since observations in cell $i \neq j$ of sector p are dependent on those of cell j through the common component Ξ_p.

Still, when we assume in addition that the Ξ-components are normally distributed, the credibility estimators for the Jewell model can easily be shown to be exact Bayesian. This is because for each choice $M = m + \Xi_p$, $M = m + \Xi_p + \Xi_{pj}$ and $M = X_{pj,T+1}$, M has a multivariate normal joint distribution with the vector of observations $\vec{X}$. This, as is well-known and can be found in any statistics text of a reasonable level, implies that $E[M|\vec{X}]$is linear in $\vec{X}$.

Also under normality assumptions, the estimators in Hachemeister's regression credibility model can be shown to be exact Bayesian. ∎

Remark 2.7 (Variance components outlook on credibility theory)
In the authors' opinion, credibility is currently taught in an unnecessarily complicated way. For didactic reasons, models should not be formulated using a hard-to-explain risk variable Θ, a function $\mu(\Theta)$ of which is the variable of interest. Setting credibility in a Bayesian framework also isn't exactly helpful for the acceptance of credibility techniques by practitioners, especially in Europe. Since in most countries actuaries generally are not fully qualified mathematicians, formulating credibility estimation as a projection in a Hilbert space, however elegant mathematically, is also aiming too high. Rather, one should formulate the credibility models as additive independent variance components models such as (23). As argued in Dannenburg, Kaas and Goovaerts (1996), this presents no loss of generality, since only the first and second moments of the data and (weighted) averages thereof are needed for the calculation of credibility estimators. To calculate covariances and correlations is almost trivial in this framework, but a much more laborious process via the conditional expectations, given Θ, needed in the more usual model. ∎

REFERENCES

DANNENBURG, D.R., R. KAAS, M.J. GOOVAERTS (1996); Practical actuarial credibility models, *Institute of Actuarial Science and Econometrics, University of Amsterdam*, Amsterdam.

GERBER, H.U. (1995); A teacher's remark on exact credibility, *ASTIN Bulletin*, **25**, 189-192.

GOOVAERTS, M.J., R. KAAS, A.E. VAN HEERWAARDEN & T. BAUWELINCKX (1990); Effective Actuarial Methods; *North-Holland*, Amsterdam.

JEWELL, W.S. (1974); Credible means are exact Bayesian for exponential families; *ASTIN Bulletin*, **8**, 77-90.

JEWELL, W.S. (1975); Regularity conditions for exact credibility; *ASTIN Bulletin*, *8*, 336-341.

MCCULLAGH, P. and NELDER, J.A. (1989), Generalized Linear Models, 2nd edition. *Chapman & Hall*, London.

NELDER, J.A. and R.J. VERRALL (1997), Credibility Theory and Generalized Linear Models, *ASTIN Bulletin*, **27.1**, 71-82.

ROB KAAS, DENNIS DANNENBURG and MARC GOOVAERTS
Universiteit van Amsterdam
Instituut vor Actuariaat en Econometrie
Roetersstraat 11
1018 WB Amsterdam, The Netherlands

SIMULATION OF RUIN PROBABILITIES FOR SUBEXPONENTIAL CLAIMS

By S. ASMUSSEN and K. BINSWANGER

ABSTRACT

We consider the classical risk model with subexponential claim size distribution. Three methods are presented to simulate the probability of ultimate ruin and we investigate their asymptotic efficiency. One, based upon a conditional Monte Carlo idea involving the order statistics, is shown to be asymptotically efficient in a certain sense. We use the simulation methods to study the accuracy of the standard Embrechts-Veraverbeke [16] approximation for the ruin probability and also suggest a new one based upon ideas of Hogan [21].

KEYWORDS

Conditional Monte Carlo, corrected diffusion approximation, ladder heights, order statistics, Pollaczeck-Khinchine formula, probability of ultimate ruin, rare events, regular variation, subexponential distribution.

1. INTRODUCTION

This paper is concerned with the simulation of the probability $\psi(u)$ of ruin in a classical compound Poisson risk process $U(t)$ with initial (large) reserve $u = U(0)$ in the case where the claim size distribution B is heavy-tailed. Our main aim is to investigate ways to improve upon crude Monte Carlo simulation.

We assume that the claim arrival process $\{N(t),\ t \geq 0\}$ $(N(0)=0)$ is a homogeneous Poisson process with rate $\lambda > 0$. The claim sizes are assumed to independent and identically distributed non-negative random variables $\xi_i (i \in N)$ with cumulative distribution function $B(x)$ and finite mean μ_B, and independent of $\{N(t),\ t \geq 0\}$. The net premium is considered to be payable at a constant rate c over time, where

$$c = (1+\theta)\lambda\mu_B$$

and $\theta > 0$ is the relative security loading. The insurance surplus at time t is $U(t)$. The total claim process $R(t) = \sum_{i=1}^{N(t)} \xi_i$ is by the assumptions above a compound Poisson process and thus

$$U(t) = u + ct - R(t).$$

The probability of ruin is defined as

$$\psi(u) = P(\inf_{t \geq 0} U(t) < 0).$$

All simulation methods that we study are based upon representing the ruin probability as $\psi(u) = z = EZ$ for some r.v. Z that can be generated by simulation, simulate iid replicates $Z_1, \ldots, Z_n$ of Z, estimate $\psi(u)$ by $\hat{z} = (Z_1 + \ldots + Z_n)/n$ and use the empirical variance of the Z_i to produce confidence intervals. The performance measure of a particular simulation method is the relative error $\sigma_Z/\psi(u)$ where $\sigma_Z^2 = var(Z)$ (when comparing different simulation methods based upon $Z(1)$, $Z(2)$, say, this in only reasonable if the computer times needed to generate $Z(1)$, $Z(2)$ are roughly the same; we assume this to be the case without further discussion). We face two difficulties:

1) The ruin problem has infinite horizon so that it is not straightforward to find the desired representation $\psi(u) = z = \mathrm{E}[Z]$ for some simulatable Z.
2) Since u is large, the ruin probability $\psi(u)$ is small and hence we are in the framework of rare events simulation (see Heidelberger [20] or Asmussen & Rubinstein [7] for surveys). Neglecting problem 1) for a moment, assume that we can generate $Z = I(\tau(u) < \infty)$ where $I(\cdot)$ stands for the indicator function and $\tau(u)$ is the time of ruin with initial capital u. This procedure is known in the literature as the the crude Monte Carlo method and leads to a relative error

$$\frac{\sigma_Z}{\psi(u)} = \frac{\sqrt{\psi(u)(1-\psi(u))}}{\psi(u)} \approx \frac{1}{\sqrt{\psi(u)}} \to \infty, \ u \to \infty. \tag{1}$$

In the case where B is light-tailed, a solution to both problems was suggested by Siegmund [29] and Asmussen [4] who used importance sampling (Rubinstein [28] or Glynn & Iglehart [18]). One then performs a change of measure, replacing the given governing probability measure P by a different one $\tilde{P}$ satisfying $\tilde{P}(\tau(u) < \infty) = 1$ and takes $Z = dP/d\tilde{P}$ where the likelihood ratio (Radon-Nikodym derivative) is computed on $\mathcal{F}_{\tau(u)}$. More precisely, $\tilde{P}$ corresponds to an exponential change of measure involving the Lundberg exponent (adjustment coefficient) R, such that the Poisson intensity and the claim size distribution is changed in a certain way given by R. That problem 1) is solved follows from $\tilde{P}(\tau(u) < \infty) = 1$. Empirical evidence strongly suggests that also problem 2) is solved, and the theoretical verification of this has been the subject of much research. We follow here a standard current criterion (e.g. Heidelberger [20] or Asmussen [7]) for calling a rare events simulation estimator asymptotically (or logarithmically) efficient: one should have

$$\liminf_{u\to\infty} \frac{\log \sigma_Z}{\log \psi(u)} \geq 1. \tag{2}$$

In particular, it suffices that

$$\sigma_Z^2 \leq \psi(u)^2 p(|\log\ \psi(u)|) \tag{3}$$

for some polynomial p, and this is well-known to hold in the setting of Siegmund [29], Asmussen [4] with p constant. Note that the CMC method can never be efficient according to (2) because it always gives rise to the limit 1/2 rather than 1 there.

The present paper is concerned with the simulation of $\psi(u)$ in the case where B does not have exponential moments so that R does not exist and the method of Siegmund [29], Asmussen [4] is not applicable. Among such distributions we focus on the class of subexponential distributions $\mathcal{S}$. To be more precise:

Definition 1.1. *A non negative random variable X with distribution function F is called subexponential ($F \in \mathcal{S}$), if for all $n \geq 2$,*

$$\lim_{x\to\infty} \frac{P(X_1 + \dots + X_n > x)}{P(\max(X_1, \dots, X_n) > x)} = 1,$$

where $X_1, \dots, X_n$ are iid copies of X.

This class is quite broad and contains many of the common claim size distributions, i.e. longtailed distributions such as Pareto, Lognormal or Weibull with decreasing failure rate. Good summaries of the properties of this class are given in Embrechts and Veraverbeke [16] and Klüppelberg [23].

Our vehicle to deal with problem 1) in this setting is the Pollaczeck-Khinchine formula (see Asmussen [5])

$$\psi(u) = 1 - (1 - \rho) \sum_{n=0}^{\infty} \rho^n B_0^{*n}(u), \quad u > 0, \tag{4}$$

where $\rho = \frac{1}{1+\theta}$, $B_0(u) = \int_0^u b_0(s)ds$ and $b_0(s) = \frac{1}{\mu_B}\bar{B}(s)$ with $\bar{B}(s) = 1 - B(s)$; B_0^{*n} denotes the n-th convolution of B_0 with itself. Note that (4) means that $1 - \psi(u)$ is a compound geometric distribution function,

$$\psi(u) = P(S_K > u), \tag{5}$$

where $S_K = X_1 + \dots + \dots X_K$, K is geometric with parameter ρ, independent of the X_i's, and the $X_1, X_2, \dots$ are non-negative iid random variables with common density b_0. This means that the CMC method is applicable: $\psi(u) = z = E[Z]$ where $Z = I(S_K > u)$. The algorithm is as follows:

1. Generate $K_i \sim$ geometric (ρ) , i.e. $P(K_i = k) = (1 - \rho)\rho^k$ $(k = 0, 1, 2 \dots)$.
2. Generate $X_1^i, \dots, X_{K_i}^i$ from the density b_0 and let $S_{K_i} = X_1^i + \dots + X_{K_i}^i$.
3. If $S_{K_i} > u$ then $Z_i = 1$, otherwise $Z_i = 0$.
4. Repeat steps 1 to 3 n times.
5. Estimate E $[Z]$ by $\hat{z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

As a CMC algorithm, this procedure (referred to as Algorithm 1 in the following) cannot be efficient in the sense of (2). To deal with problem 2), we suggest (Section 2) two conditional Monte Carlo estimators. The idea is to replace the CMC estimator Z by $E(Z \mid \mathcal{G})$ for a suitable σ-field $\mathcal{G}$, which always leads to reduction in variance, cf. Rubinstein [28]. We show that one of the estimators is efficient in the sense of (2) in the particular case where the tail of B is regularly varying. This result is remarkable since, to our knowledge, it is the first example in the general area of rare events simulation of an asymptotically efficient solution to a problem involving heavy tails. It also has the unusual feature that the asymptotic efficient solution is not given in terms of importance sampling.

In addition to simulation methodology, we also discuss analytic approximations, of which the most standard ones are Panjer's recursion (cf. Section 4.1) and

$$\psi(u) \sim \frac{1}{\theta}\overline{B}_0(u), \quad u \to \infty \tag{6}$$

(Embrechts and Veraverbeke [16] and references therein) which will be refered to as $\psi_{EV}(u)$ in the sequel. The accuracy of (6) is for instance discussed in Abate, Choudhury and Whitt [1]. They computed exact values by transform inversion (for a summary of inversion methods and applicability of this approach see Abate and Whitt [2] and references therein). In the latter paper, a class PME (Pareto Mixtures of Exponentials, see further Section 4) with explicit Laplace transforms was constructed and numerical comparisons of exact values and (6) were given with rather negative results concerning the accuracy of (6). We present some further numerical results along the same lines, computing the exact values by simulation also for more general claim size distributions than the ones in PME. Motivated by these negative findings, we suggest an alternative approximation, essentially an adaptation of the correction due to Hogan [21] of the standard diffusion approximation

$$\psi(u) \approx \exp\left(-u\theta\frac{2\mu_B}{\mu_B^2+\sigma_B^2}\right), \tag{7}$$

where σ_B^2 denotes the variance of B. This approximation is introduced and discussed in more detail in Section 3.

2. Conditional Monte Carlo aralgorithms

In this section random variables are mostly denoted with capital letters (e.g. $Z, K, S_K, X_l, X_2, \ldots$), the realization of simulation i ($i = 1, \ldots, n$) with indexed capital letters (e.g. $Z_i, K_i, X_1^i, X_2^1, \ldots$).

Recall that we refer to the CMC method as Algorithm I and that a conditional Monte Carlo estimator always reduces variance.

The 95% asymptotic confidence intervals are given by:

$$\hat{\psi}(u) \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}},$$

where $\hat{\psi}(u)$ stands for the estimated ruin probability and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Z_i - \overline{Z}\right)^2$.

2.1. **Algorithm II.** Write

$$\psi(u) = P(X_1 + \dots + X_K > u)$$
$$= E[P(X_1 + \dots + X_K > u | X_1, \dots, X_{K-1})]$$
$$= E\left[\overline{B}_0(u - X_1 - \ \dots \ - X_{K-1})\right].$$

Thus we generate only $X_1, \dots, X_{K-1}$, compute $Y = u - X_1 - \dots - X_{k-1}$ and set $Z = \overline{B}_0(Y)$, the probability that the next claim causes ruin. More precisely:

1. Generate $K_i \sim$ geometric (ρ) i.e. $P(K_i = k) = (1 - \rho)\rho^k (k = 0, 1, 2, \dots)$.
2. Generate $X_1^i, \dots, X_{K_{i-1}}^i$ from the density b_o and let $Y_i = u - X_1^i - \dots - X_{K_i-1}^i$.
3. Let $Z_i = \overline{B}_0(Y_i)$ ($Z_i = 1$ if $Y_i < 0$)
4. Repeat steps 1 to 3 n times.
5. Estimate $E[Z]$ *by* $\hat{z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$

Again $\hat{z}$ is an unbiased estimator for $\psi(u)$. However, even if the variance must be smaller than for Algorithm I, the performance as measured by (2) is not asymptotically better:

Proposition 2.1. *Assume that $B \in \mathcal{S}$. Then for Algorithm II,*

$$\lim_{u \to \infty} \frac{\log \sigma_Z}{\log \psi(u)} = \frac{1}{2}$$

Proof.

$$E[Z^2] = E\left[\overline{B}_0^2(u - X_1 - \ \dots - X_{K-1})\right]$$
$$\geq E\left[\overline{B}_0^2(u); K \leq 1\right] + E\left[\overline{B}_0^2(u - X_1); K \geq 2\right]$$
$$= (1 - \rho^2)\overline{B}_0^2(u) + E\left[\overline{B}_0^2(u - X_1); K \geq 2\right]$$
$$\geq (1 - \rho^2)\overline{B}_0^2(u) + E\left[\overline{B}_0^2(u - X_1); X_1 > u, K \geq 2\right] = (1 - \rho^2)\overline{B}_0^2(u) + \rho^2 \overline{B}_0(u).$$

The last equality follows from the fact that the event $(X_I > u)$ occurs with probability $\overline{B}_0(u)$ and then $\overline{B}_0^2(u - X_1) = 1$. Since

$$E[Z]^2 = \psi(u)^2 \sim \frac{1}{\theta^2}\overline{B}_0^2(u),$$

it follows that σ_Z is of the order of magnitude at least $\overline{B}_0^{1/2} \sim (\theta\psi(u))^{1/2}$. Hence $\log \sigma_Z$ cannot go to $-\infty$ faster than $\log \psi(u)/2$ so that $1/2$ is an upper bound for lim inf in (2). That $1/2$ is also a lower bound for lim sup follows since the algorithm, based upon conditional Monte Carlo, is an improvement of the CMC algorithm

□

2.2. **Algorithm III.** The third algorithm is slightly more complicated. The main idea underlying this algorithm is that for subexponential distributions only the largest claim and not the sum of all claims causes ruin as stated in Definition 1.1. The following two lemmas will elaborate on this idea.

Lemma 2.1. *Let $X_1, X_2, ..., X_n \sim B_0$ be non negative iid random variables and denote by $X_{(1)} < X_{(2)} < ... < X_{(n)}$ the order statistic. Furthermore let $\mathcal{F}_{(n-1)} = \sigma(X_{(1)}, ..., X_{(n-1)})$.*
Then

$$P\left(X_{(n)} > x \middle| \mathcal{F}_{(n-1)}\right) = \frac{\overline{B}_0\left(X_{(n-1)} \vee x\right)}{\overline{B}_0\left(X_{(n-1)}\right)}$$

where $a \vee b$ stands for $\max(a, b)$.

Proof. Suppose $X_1, ..., X_n$ iid and X_i's are absolutely continuous, then the order statistics form a Markov chain.

$$P\left(X_{(n)} > x \middle| \mathcal{F}_{(n-1)}\right) = P\left(X_{(n)} > x \middle| X_{(n-1)}\right)$$

and

$$P\left(X_{(n)} > x \middle| X_{(n-1)} = y\right) = \begin{cases} 1, & x < y, \\ \int_x^\infty f_{X_{(n)}|X_{(n-1)}}(u|y)du, & x \geq y, \end{cases}$$

where

$$\int_x^\infty f_{X_{(n)}|X_{n-1}}(u|y)du = \frac{\overline{B}_0(x)}{\overline{B}_0(y)}$$

(see for instance Arnold, Balakrishnan and Nagaraja [3], p. 23). Hence

$$P\left(X_{(n)} > x \middle| \mathcal{F}_{(n-1)}\right) = \frac{\overline{B}_0\left(X_{(n-1)} \vee x\right)}{\overline{B}_0\left(X_{(n-1)}\right)}$$

□

Remark: If the X_i's are not absolutely continuous a different proof can be given using combinatorical arguments.

Lemma 2.2. *Let* $S_n = X_1 + \ldots + X_n$ and $S_{(k)} = X_{(1)} + \ldots + X_{(k)} (1 \leq k \leq n)$. *Then*

$$P(S_n > u) = E\left[\frac{\overline{B}_0\big((u - S_{(n-1)}) \vee X_{(n-1)}\big)}{\overline{B}_0\big(X_{(n-1)}\big)}\right]$$

Proof. By conditioning,

$$\begin{aligned} P(S_n > u) &= E\big[P\big(S_n > u \,|\, \mathcal{F}_{(n-1)}\big)\big] \\ &= E\big[P\big(X_{(n)} + S_{(n-1)} > u \,|\, \mathcal{F}_{(n-1)}\big)\big] \\ &= E\big[P\big(X_{(n)} > u - S_{(n-1)} \,|\, \mathcal{F}_{(n-1)}\big)\big] \end{aligned}$$

and applying Lemma 2.1 completes the proof. □

Algorithm III can then be written as:
1. Generate K_i as geometric (ρ), i.e. $P[K_i = k] = (1 - \rho)\rho^k (k = 0, 1, 2, \ldots)$.
2. Generate $X_1^i, \ldots, X_{K_i}^i$ from the density b_0 and set $Y^i = u - X_{(1)}^i - \ldots - X_{(K_i - 1)}^i$ and $m_i = X_{(K_i - 1)}^i$.
3. Set $Z_i = \dfrac{\overline{B}_0(Y_i \vee m_i)}{\overline{B}_0(m_i)}$.
4. Repeat steps 1 to 3 n times.
5. Estimate E [Z] by $\hat{z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

The main result of the paper is the following

Theorem 2.1. *Assume that* $\overline{B}_0(x) = L(x)/x^{\alpha} (\alpha > 1)$ *with L slowly varying (i.e.* $\lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1$ *for all* $\lambda > 0$*). Then Algorithm III satisfies*

$$\liminf_{u\to\infty} \frac{\log \sigma_Z}{\log \psi(u)} \geq 1.$$

In order to proof Theorem 2.1 we first give three Lemmas.

Lemma 2.3. *For Algorithm III we have*

$$\sigma_Z^2 \leq E\left[K^2\left(\frac{1}{2}\overline{B}_0^2\left(\frac{u}{2}\right) + \overline{B}_0^2\left(\frac{u}{K}\right)\left|\log \overline{B}_0\left(\frac{u}{2}\right)\right|\right)\right] \tag{8}$$

Proof. We first derive the conditional density $f_{X_{(K-1)}}(x)$ of the random variable $X_{(K-1)}$ given K:

$$\begin{aligned} P(X_{K-1} \leq x) &= P\big(X_{(1)} \leq x, \ldots, X_{(K-1)} \leq x\big) \\ &= \binom{K}{1} P(X_1 \leq x, \ldots, X_{K-1} \leq x, X_K > x) \end{aligned}$$

$$+ P(X_1 \le x, \ \ldots, \ X_{K-1} \le x, \ X_K \le x)$$

$$= K B_0^{K-1}(x)\overline{B}_0(x) + B_0^K(x).$$

Hence the density is:

$$f_{X_{(K-1)}}(x) = K(K-1)B_0^{K-2}(x)\overline{B}_0(x)b_0(x). \tag{9}$$

Next we calculate

$$E\left[Z^2|K\right] = E\left[\left(\frac{\overline{B}_0\big((u - S_{(K-1)}) \vee X_{(K-1)}\big)}{\overline{B}_0\big(X_{(K-1)}\big)}\right)^2 |K\right]$$

$$= E\left[\left(\frac{\overline{B}_0\big(u - S_{(K-1)}\big)}{\overline{B}_0\big(X_{(K-1)}\big)}\right)^2 ; X_{(K-1)} \le \frac{u}{K}|K\right] \tag{10}$$

$$+E\left[\left(\frac{\overline{B}_0\big((u - S_{(K-1)}) \vee X_{(K-1)}\big)}{\overline{B}_0\big(X_{(K-1)}\big)}\right)^2 ; \frac{u}{K} < X_{(K-1)} \le \frac{u}{2}|K\right] \tag{11}$$

$$+E\left[1; X_{(K-1)} > \frac{u}{2}|K\right]. \tag{12}$$

The first summand (10) can be bounded as follows. If $X_{(K-1)} \le \frac{u}{K}$ then $\overline{B}_0\big(u - S_{(K-1)}\big) \le \overline{B}_0\big(\frac{u}{K}\big)$, so that

$$E\left[\left(\frac{\overline{B}_0\big(u - S_{(K-1)}\big)}{\overline{B}_0\Big(X_{(K-1)}\Big)}\right)^2 ; X_{(K-1)} \le \frac{u}{K}|K\right]$$

$$\le \overline{B}_0^2\Big(\frac{u}{K}\Big) \int_0^{u/K} \frac{f_{X_{(K-1)}}(x)}{\overline{B}_0^2(x)} dx$$

$$\le K(K-1)\overline{B}_0^2\Big(\frac{u}{K}\Big) \int_0^{u/K} \frac{b_0(x)}{\overline{B}_0(x)} dx$$

$$= -K(K-1)\overline{B}_0^2\Big(\frac{u}{K}\Big) \log\Big(\overline{B}_0\Big(\frac{u}{K}\Big)\Big).$$

The second summand (11) can be bounded in the same way. For $\frac{u}{K} < X_{(K-1)} \le \frac{u}{2}$, $\overline{B}_0\big((u - S_{(K-1)}) \vee X_{(K-1)}\big) \le \overline{B}_0(\frac{u}{K})$, yielding

$$E\left[\left(\frac{\overline{B}_0\big((u - S_{(K-1)}) \vee X_{(K-1)}\big)}{\overline{B}_0\big(X_{(K-1)}\big)}\right)^2; \frac{u}{K} < X_{(K-1)} \le \frac{u}{2} | K\right]$$

$$\le \overline{B}_0^2\Big(\frac{u}{K}\Big) \int_{u/K}^{u/2} \frac{f_{X_{(K-1)}}(x)}{\overline{B}_0^2(x)} dx$$

$$\le K(K-1)\overline{B}_0^2\Big(\frac{u}{K}\Big) \int_{u/K}^{u/2} \frac{b_0(x)}{\overline{B}_0(x)} dx$$

$$= -K(K-1)\overline{B}_0^2\Big(\frac{u}{K}\Big)\Big(\log(\overline{B}_0(\frac{u}{2})) - \log(\overline{B}_0(\frac{u}{K}))\Big)$$

To find an upper bound for (12) we write

$$E\Big[1; X_{(K-1)} > \frac{u}{2} | K\Big] = \int_{u/2}^{\infty} f_{X_{(K-1)}}(x) dx$$

$$= K(K-1) \int_{u/2}^{\infty} B_0^{K-2}(x)\overline{B}_0(x) b_0(x) dx$$

$$\le K(K-1) \int_{u/2}^{\infty} \overline{B}_0(x) b_0(x) dx$$

$$= K(K-1)\frac{1}{2}\overline{B}_0^2\Big(\frac{u}{2}\Big).$$

Adding the above inequalities leads to

$$E[Z^2|K] \le K(K-1)\left(\frac{1}{2}\overline{B}_0^2\Big(\frac{u}{2}\Big) - \overline{B}_0^2\Big(\frac{u}{K}\Big)\log\Big(\overline{B}_0\Big(\frac{u}{2}\Big)\Big)\right)$$

$$\le K^2\left(\frac{1}{2}\overline{B}_0^2\Big(\frac{u}{2}\Big) + \overline{B}_0^2\Big(\frac{u}{K}\Big)|\log \overline{B}_0\Big(\frac{u}{2}\Big)|\right)$$

and hence

$$\sigma_Z^2 = E[E[Z^2|K]] - E[Z]^2$$

$$\le E\left[K^2\left(\frac{1}{2}\overline{B}_0^2\Big(\frac{u}{2}\Big) + \overline{B}_0^2\Big(\frac{u}{K}\Big)|\log \overline{B}_0\Big(\frac{u}{2}\Big)|\right)\right].$$

□

Lemma 2.4. *If* $\overline{B}_0(x) = \frac{L(x)}{x^\alpha}$, *L slowly varying then for any* $\varepsilon > 0$ *there exist constants* $C_-(\varepsilon)$ *and* $C_+(\varepsilon)$ *such that*

$$C_-(\varepsilon)d^{\alpha-\varepsilon}x^{-\alpha-\varepsilon} \le \overline{B}_0\left(\frac{x}{d}\right) \le C_+(\varepsilon)d^{\alpha-\varepsilon}x^{-\alpha-\varepsilon}, \ \forall x > 0 \ \forall\, d > 0.$$

Proof. From $\overline{B}_0(x)x^{\alpha-\varepsilon} = x^{-\varepsilon}L(x)$ it follows that $\lim_{x\to 0} x^{-\varepsilon}L(x) = 0$ and that L is a continuous function. Since L is slowly varying also $\lim_{x\to\infty} x^{-\varepsilon}L(x) = 0$. Hence there exists a constant $C_+(\varepsilon)$ such that $L(x) \le C_+(\varepsilon)x^\varepsilon$ for all x and hence

$$\overline{B}_0\left(\frac{x}{d}\right) = \frac{L(x/d)}{(x/d)^\alpha} \le \frac{C_+(\varepsilon)(x/d)^\varepsilon}{(x/d)^\alpha}$$

For the lower bound the proof is similar. Just note that if L is slowly varying then also 1/L is slowly varying.

□

Lemma 2.5. *If* $\overline{B}_0(x) = \frac{L(x)}{x^\alpha}$, *L slowly varying then for any* $\varepsilon > 0$ *there exist constants* $D_1(\varepsilon)$ *and* $D_2(\varepsilon)$ *such that*

$$E\left[Z^2\right] \le (D_1(\varepsilon) + D_2(\varepsilon)|\log u|)u^{2\varepsilon-2\alpha}.$$

Proof. From Lemma 2.3 we have

$$E\left[Z^2\right] \le E\left[K^2\left(\frac{1}{2}\overline{B}_0^2\left(\frac{u}{2}\right) + \overline{B}_0^2\left(\frac{u}{K}\right)|\log \overline{B}_0\left(\frac{u}{2}\right)|\right)\right]$$

Lemma 2.4 yields

$$\le E\left[K^2\right]\frac{1}{2}C_+^2(\varepsilon)2^{2\alpha-2\varepsilon}u^{-2\alpha+2\varepsilon}$$

$$+E\left[C_+^2(\varepsilon)K^{2\alpha-2\varepsilon+2}u^{-2\alpha+2\varepsilon}|\log(C_-(\varepsilon)2^{\alpha-\varepsilon}u^{-\alpha-\varepsilon})\ |\right]$$

$$\le (D_1(\varepsilon) + D_2(\varepsilon)|\log u|)u^{2\varepsilon-2\alpha}$$

where $D_1(\varepsilon) = E\left[K^2\right]\frac{1}{2}C_+^2(\varepsilon)2^{2\alpha-2\varepsilon} + E\left[K^{2\alpha-2\varepsilon+2}\right]C_+^2(\varepsilon)|\log(C_-(\varepsilon)2^{\alpha-\varepsilon})\ |$

and $D_2(\varepsilon) = E\left[K^{2\alpha-2\varepsilon+2}\right]C_+^2(\varepsilon)(\alpha+\varepsilon)$.

□

Now we have all the tools needed to prove Theorem 2.1.
Proof of Theorem 2.1. From Lemma 2.5 we get

$$\log\sigma_Z \le \log\sqrt{(D_1(\varepsilon) + D_2(\varepsilon)|\log u|)u^{2\varepsilon-2\alpha}}$$

$$= \tfrac{1}{2}\log(D_1(\varepsilon) + D_2(\varepsilon)|\log u|) + (\varepsilon-\alpha)\log u$$

and therefore

$$\lim_{u\to\infty}\frac{\log\sigma_Z}{\log\psi(u)}\geq\lim_{u\to\infty}\frac{\frac{1}{2}\log(D_1(\varepsilon)+D_2(\varepsilon)|\log u|)+(\varepsilon-\alpha)\log u}{\log\psi(u)}$$

using (6) yields

$$=\lim_{u\to\infty}\frac{\frac{1}{2}\log(D_1(\varepsilon)+D_2(\varepsilon)|\log u|)+(\varepsilon-\alpha)\log u}{\log\left(\overline{B}_0(u)/\theta\right)}$$

$$=\lim_{u\to\infty}\frac{\frac{1}{2}\log(D_1(\varepsilon)+D_2(\varepsilon)|\log u|)+(\varepsilon-\alpha)\log u}{-\log\theta+\log L(u)-\alpha\log u}$$

$$=\frac{\varepsilon-\alpha}{-\alpha}=1-\frac{\varepsilon}{\alpha}$$

Now let $\varepsilon\to 0$ which completes the proof.

□

Remark:

1. For lognormal claimsizes Algorithm III is also asymptotically efficient. The proof is given in Binswanger [8].

TABLE 1

SIMULATED RUIN PROBABILITIES AND THEIR PRECISION MEASURED BY (2) FOR PARETO DISTRIBUTED CLAIMS (ALL NUMBERS ARE ROUNDED TO THEIR LAST DIGIT)

$Pareto(1,2), \theta = 0.1,\ n = 1000$				
$\hat{\psi}(u)\pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}$ $\log(\hat{\sigma})/\log(\hat{\psi})$	Algorithm I	Algorithm II	Algorithm III	$\psi_P(u)$
u = 10	$(5.6 \pm 0.3)\cdot 10^{-1}$ 1.21	$(6.0 \pm 0.3)\cdot 10^{-1}$ 1.56	$(5.5 \pm 0.3)\cdot 10^{-1}$ 1.38	$5.5\cdot 10^{-1}$
u = 50	$(2.0 \pm 0.2)\cdot 10^{-1}$ 0.57	$(2.0 \pm 0.2)\cdot 10^{-1}$ 0.60	$(1.9 \pm 0.2)\cdot 10^{-1}$ 0.72	$1.9\cdot 10^{-1}$
u = 100	$(8.1 \pm 1.7)\cdot 10^{-2}$ 0.52	$(9.0 \pm 1.7)\cdot 10^{-2}$ 0.54	$(8.6 \pm 1.2)\cdot 10^{-2}$ 0.69	$8.5\cdot 10^{-2}$
u = 500	$(1.2 \pm 0.7)\cdot 10^{-2}$ 0.50	$(0.9 \pm 0.5)\cdot 10^{-2}$ 0.51	$(1.0 \pm 0.2)\cdot 10^{-2}$ 0.77	$1.2\cdot 10^{-2}$
u = 1000	$(6.0 \pm 4.8)\cdot 10^{-3}$ 0.50	$(9.5 \pm 5.9)\cdot 10^{-3}$ 0.51	$(5.3 \pm 0.6)\cdot 10^{-3}$ 0.88	$5.4\cdot 10^{-3}$

TABLE 2

SIMULATED RUIN PROBABILITIES AND THEIR PRECISION MEASURED BY (2) FOR PME DISTRIBUTED CLAIMS (ALL NUMBERS ARE ROUNDED TO THEIR LAST DIGIT)

PME(3), $\theta = 0.25$, $n = 1000$				
$\hat{\psi}(u) \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}$ $\log(\hat{\sigma})/\log\left(\hat{\psi}\right)$	Algorithm I	Algorithm II	Algorithm III	$\psi(u)$
u = 50	$(5.0 \pm 4.4) \cdot 10^{-3}$ 0.50	$(1.8 \pm 2.0) \cdot 10^{-3}$ 0.54	$(3.0 \pm 0.9) \cdot 10^{-3}$ 0.74	$3.1 \cdot 10^{-3}$
u = 60	$(3.0 \pm 3.4) \cdot 10^{-3}$ 0.50	$(4.3 \pm 3.9) \cdot 10^{-3}$ 0.51	$(2.4 \pm 2.0) \cdot 10^{-3}$ 0.57	$1.8 \cdot 10^{-3}$
u = 70	$(2.0 \pm 2.8) \cdot 10^{-3}$ 0.50	$(1.8 \pm 0.1) \cdot 10^{-3}$ 1.04	$(1.0 \pm 0.2) \cdot 10^{-3}$ 0.84	$1.2 \cdot 10^{-3}$
u = 80	0 –	$(1.4 \pm 0.1) \cdot 10^{-4}$ 1.05	$(8.8 \pm 2.0) \cdot 10^{-4}$ 0.82	$8.2 \cdot 10^{-4}$
u = 90	$(1.0 \pm 2.0) \cdot 10^{-3}$ 0.50	$(1.0 \pm 0.1) \cdot 10^{-4}$ 1.01	$(5.6 \pm 1.1) \cdot 10^{-4}$ 0.84	$6.1 \cdot 10^{-4}$
u = 100	$(1.0 \pm 2.0) \cdot 10^{-3}$ 0.50	$(8.2 \pm 0.3) \cdot 10^{-5}$ 1.06	$(4.1 \pm 0.7) \cdot 10^{-4}$ 0.87	$4.7 \cdot 10^{-4}$

TABLE 3

SIMULATED RUIN PROBABILITIES AND THEIR PRECISION MEASURED BY (2) FOR LOGNORMAL DISTRIBUTED CLAIMS (ALL NUMBERS ARE ROUNDED TO THEIR LAST DIGIT)

Lognormal (-1.62, 1.8), $\theta = 0.1$, $n = 1000$				
$\hat{\psi}(u) \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}$ $\log(\hat{\sigma})/\log\left(\hat{\psi}\right)$	Algorithm I	Algorithm II	Algorithm III	$\psi(u)$
u = 0	$(8.3 \pm 0.2) \cdot 10^{-1}$ 5.35	$(8.9 \pm 0.2) \cdot 10^{-1}$ 11.1	$(9.0 \pm 0.2) \cdot 10^{-1}$ 11.4	$9.1 \cdot 10^{-1}$
u = 100	$(3.5 \pm 0.3) \cdot 10^{-1}$ 0.70	$(3.9 \pm 0.3) \cdot 10^{-1}$ 0.82	$(3.4 \pm 0.3)\ 10^{-1}$ 0.84	$3.4 \cdot 10^{-1}$
u = 1000	$(1.2 \pm 0.7) \cdot 10^{-2}$ 0.50	$(7.4 \pm 4.8) \cdot 10^{-3}$ 0.52	$(8.0 \pm 2.2) \cdot 10^{-3}$ 0.69	$1.1 \cdot 10^{-2}$
u = 10 000	0 –	$(3.3 \pm 0.1) \cdot 10^{-6}$ 1.09	$(3.5 \pm 0.4) \cdot 10^{-5}$ 0.93	$4 \cdot 10^{-5}$

2. If B_0 (or B) is a Weibull distribution,

$$b_0(x) = \nu x^{\nu-1}e^{-x^\nu},\ \overline{B}_0(x) = e^{-x^\nu},$$

Algorithm III is not efficient in the sense of (2). Indeed, we get

$$\begin{aligned}
E[Z^2|K=2] &\geq \int_0^{u/2} \frac{\overline{B}_0^2(u-y)}{\overline{B}_0^2(y)} f_{X_{(1)}}(y)dy \\
&= 2\int_0^{u/2} \overline{B}_0^2(u-y)\nu y^{\nu-1}dy \\
&\geq 2\nu(u/2)^{\nu-1}\int_0^{u/2} \overline{B}_0^2(u-y)dy \\
&= 2\nu(u/2)^{\nu-1}\int_{u/2}^{u} \overline{B}_0^2(y)dy \\
&\geq \int_{u/2}^{u} 2\nu y^{\nu-1}\overline{B}_0^2(y)dy \\
&= e^{-2(u/2)^\nu} - e^{-2u^\nu} = e^{-2^{1-\nu}u^\nu}(1+o(1))
\end{aligned}$$

So we get

$$\begin{aligned}
\lim_{u\to\infty}\frac{\log\sigma_Z}{\log\psi(u)} &\leq \lim_{u\to\infty}\frac{\frac{1}{2}\log\left(E[Z^2|K=2]P(K=2) - E[Z^2]\right)}{\log\left(\overline{B}_0(u)/\theta\right)} \\
&\leq \lim_{u\to\infty}\frac{2^{1-\nu}u^\nu/2}{u^\nu} = \frac{1}{2^\nu} < 1
\end{aligned}$$

Of course we should mention that this does not imply that the algorithm does not work well in the Weibull setting; and indeed the numerical experience is convincing. It should be noted that, as a conditional MC algorithm, Algorithm III is always an improvement on the crude MC method, even in the light tailed case. (Though here we do not obtain any improvement of the asymptotic efficiency and the algorithms of Asmussen [4], Siegmund [29] are superior.)

3. THE CORRECTED DIFFUSION APPROXIMATION

The standard diffusion approximation (Iglehart [22] or Grandell [19]) is given by (7). For light-tailed random walk problems Siegmund [30] derived a correction which was adapted to ruin probabilities by Asmussen [6] and shown to be extremely accurate. An alternative covering also certain heavy-tailed cases was given in Theorem 2 of Hogan [21]. As in Asmussen [6], it requires some adaptation to ruin probabilities which we shall next present. The result will be an approximation of the type

$$\psi_H(u) = \exp(-c_1 u)(1 + c_2 u - c_3), \tag{13}$$

where

$$c_1 = \frac{2\theta m_1}{m_2},\ c_2 = \frac{4\theta^2 m_1^2 m_3}{3m_2^3},\ c_3 = \frac{2\theta m_1 m_3}{3m_2^2}$$

and m_i is the i-th moment of B. Note that formally the conditions of Theorem 2 in Hogan [21] lead to the requirement that $m_5 < \infty$ though our numerical experience indicates that this is not crucial.

To derive (13) from Hogan [21], substitute first $v = \xi/\vartheta$ to get

$$P_{-\vartheta}(\tau_v < \infty) \approx e^{-2\vartheta v}\left(1 + \frac{4\gamma\vartheta^2 v}{3} - \frac{\vartheta E_0 S_{\tau+}^2}{E_0 S_{\tau+}}\right). \tag{14}$$

Next we consider a RW with drift $-\mu$ and $\sigma^2 = E_0 X_1^2$ not necessarily $= 1$, and τ_u. The normalized RW S_n/σ has drift $-\vartheta = -\mu/\sigma$, $\gamma = E_0 X_1/\sigma^3, v = u/\sigma$. Similar substitutions for the ladder height moments yield

$$P_{-\mu}(\tau_u < \infty) \approx e^{-2\mu u/\sigma^2}\left(1 + \frac{4E_0 X_1^3/\sigma^3 \cdot \mu^2/\sigma^2 \cdot u/\sigma}{3} - \frac{\mu/\sigma \cdot E_0 S_{\tau+}^2/\sigma^2}{E_0 S_{\tau+}/\sigma}\right)$$

$$= e^{-(2\mu/\sigma^2)u}\left(1 + \frac{4E_0 X_1^3 \mu^2}{3\sigma^6} u - \frac{\mu E_0 S_{\tau+}^2}{\sigma^2 E_0 S_{\tau+}}\right). \tag{15}$$

In the next step, we take the RW as a discrete skeleton of the risk process. $S_n = R(nh) - cnh$. Then

$$\mu = h\theta\lambda m_1, \quad \sigma^2 = h\lambda m_2, \quad E_0 X_1^3 = h\lambda m_3.$$

Further the risk process corresponding to $\vartheta = 0$ has ladder height distribution B_0 so that

$$\frac{E_0 S_{\tau+}^2}{E_0 S_{\tau+}} \to \frac{\int_0^\infty x^2 B_0(dx)}{\int_0^\infty x\ B_0(dx)} = \frac{2m_3}{3m_2},\ \ h \downarrow 0.$$

Taking the limit $h \downarrow 0$ in (15) we thus get

$$\psi(u) \approx e^{-(2\theta\lambda m_1/\lambda m_2)u}\left(1 + \frac{4\lambda m_3(\theta\lambda m_1)^2}{3\lambda^3 m_2^3} u - \frac{2m_3\theta\lambda m_1}{3\lambda m_2^2}\right)$$

$$= e^{-c_1 u}(1 + c_2 u - c_3).$$

Various other approximations and bounds for $\psi(u)$ are known. For an overview see Embrechts and Klüppelberg [15], Feilmeier and Bertram [17], Panjer [26], Buchwald, Chevallier and Klüppelberg [9] and references therein.

4. NUMERICAL RESULTS

In this section we present the numerical evaluation of the algorithms for the PME, the Pareto and the Lognormal case. For the PME distributions Abate, Choudhury and Whitt [1] have calculated the exact values of the ruin probabilities. Therefore we choose the parameters in such a way that we can compare the simulation and the exact results. For the Pareto and the Lognormal case only few exact values are available. The Panjer approximation $\psi_P(u)$ (see below) is chosen as a benchmark.

The simulation has been done with MATLAB 4.2a. To construct B_0 distributed random variables we used the inversion method for the Pareto case and the inversion/rejection method by Newton-Raphson iteration for the other two. For more details see for instance Devroye [11].

4.1. **The Panjer recursion.** Panjer [27] suggested to use a recursion formula for calculating the probability of ultimate survival $\phi = 1 - \psi$. The recursion formula is based on a discretisation of the density ϕ' which we denote by ϕ^* leading to

$$\phi^*(u) = \frac{1}{1+\theta-g(0)} \sum_{y=1}^{u} g(y)\phi^*(u-y) \qquad u = 1,\ 2,\ \ldots$$

with

$$\phi^*(0) = \frac{\theta}{1+\theta-g(0)}$$

where g is a discretised version of the density b_0. Finally we get

$$\psi(u) \approx 1 - \sum_{y=0}^{u} \phi^*(y) \qquad u = 0,\ 1,\ 2,\ \ldots$$

The time to evaluate this procedure increases for large u since the recursion always has to start with $u = 0$. A great advantage of this method is that it leads to upper and lower bounds for $\psi(u)$ by choosing g in such a way that $g_l(x) \leq b_0(x)$ for the lower bound and $g_u(x) \geq b_0(x)$ for the upper bound. Since b_0 is a decreasing function we can set $g_l(x) = B_0([x]+1) - B_0([x])$ and $g_u(x) = B_0([x]) - B_0([x]-1)$ ($[x]$ stands for the integer part of x). For the approximation of $\psi(u)$ denoted by $\psi_P(u)$ we choose $g_a(x) = B_0([x]+1/2) - B_0([x]-1/2)$. Panjer's recursion method has meanwhile become the standard tool for actuaries; see for instance Dickson [13] for a comprehensive review.

4.2. **Pareto Distribution (PAR(a, b))**. The distribution function of the Pareto distribution is given by:

$$B(x) = \left(1 - \left(\frac{a}{x}\right)^b\right) I(x > a) \text{ where } a > 0,\ b > 1, \text{ and } x > 0.$$

The mean is $\mu_B = ab/(b-1)$, and the density b_0 and the cdf B_0 of the integrated tail distribution are respectively

$$b_0(x) = \frac{b-1}{a\,b}\left(I(x < a) + \left(\frac{a}{x}\right)^b I(x \geq a)\right),$$

$$B_0(x) = \frac{b-1}{a\,b} x I(x < a) + \left(1 - \frac{1}{b}\left(\frac{a}{x}\right)^{b-1}\right) I(x \geq a).$$

For the simulation with the inversion method we also need $B_0^{-1}(x)$ which is

$$B_0^{-1}(x) = \frac{a\,b}{b-1} x\, I\left(x < \frac{b-1}{b}\right) + \frac{a}{(b(1-x))^{\frac{1}{b-1}}} I\left(x \geq \frac{b-1}{b}\right).$$

4.3. **Pareto Mixture of Exponentials Distribution (PME(r))**. This class of distribution was defined in order to have subexponential distributions with an explicit Laplace transform. Starting from a Pareto distribution the PME is defined as follows.

Definition 4.1. *Let for* $r > 1$

$$f(x) = r\left(\frac{r-1}{r}\right)^r x^{-(r+1)}\ I(x > \frac{r-1}{r})$$

be the density function of a Pareto distribution with mean 1. Then the density of a Pareto Mixture of Exponentials is defined as

$$b(x) := \int_{\frac{r-1}{r}}^{\infty} f(y)\frac{1}{y}\exp(-\frac{x}{y})dy = \left(\frac{r-1}{r}\right)^r x^{-(r+1)}\gamma\left(r+1, \frac{r}{r-1}x\right),$$

where $\gamma(a, u) = \int_0^u t^{a-1}\exp(-t)\ dt$ *is the incomplete Gamma function.*

The tail behavior of the density of a PME distribution is the same as for the Pareto distribution, namely $\sim c_r x^{-(r+1)}$ (c_r a constant depending only on r). The distribution function $B_0(x)$ can be calculated explicitly for some values of r, for example for $r = 3$:

$$B_0(x) = 1 - \frac{1}{9x^2}\left(8 - (8 + 12x)\exp(-\frac{3x}{2})\right).$$

4.4. **Lognormal Distribution (LN(m, s))**. The density of a Lognormal distribution is given by

$$b(x) = \frac{1}{s\sqrt{2\pi}}\frac{1}{x}\exp\left(-\frac{1}{2}\left(\frac{\log(x) - m}{s}\right)^2\right) \quad I(x > 0)$$

and the k-th moment $\mu_B^{(k)} = \exp\left(km + \frac{1}{2}k^2s^2\right)$. $B(x) = \Phi(w(x))$ where $\Phi(\cdot)$ denotes the c.d.f. of a standard normal distribution and $w(x) = \frac{1}{s}(\log(x) - m)$. For efficient programming the following representation of $B_0(x)$ is useful:

$$\begin{aligned}
B_0(u) &= \frac{1}{\mu_B}\int_0^u (1 - B(x))\, dx \\
&= \frac{1}{\mu_B}\left(u - \frac{1}{\sqrt{2\pi}}\int_0^u \int_{-\infty}^{w(u)} \exp(-y^2/2)dy\, dx\right) \\
&= \frac{1}{\mu_B}\left(u - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w(u)} \int_{w^{-1}(y)}^{u} \exp(-y^2/2)dx\, dy\right) \\
&= \frac{1}{\mu_B}(u - u\Phi(w(u)) + \mu_B\Phi(w(u) - s)).
\end{aligned}$$

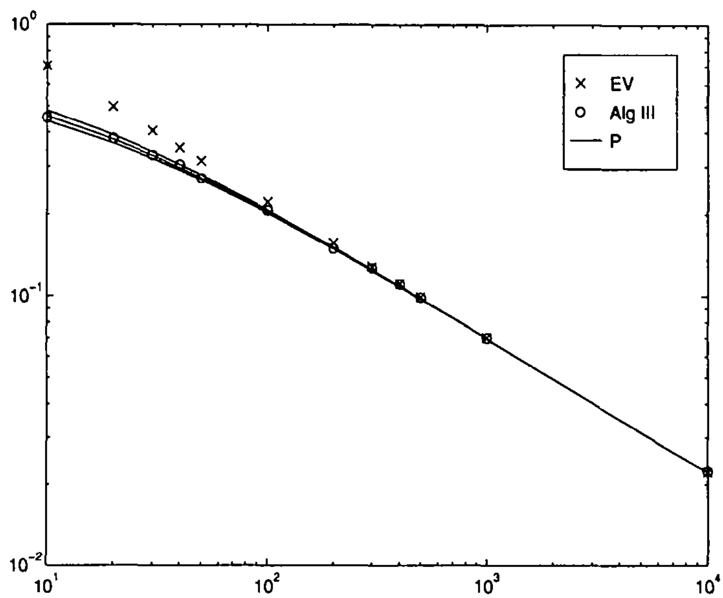


FIGURE 1: PAR(1,1.5), θ = 0.3.

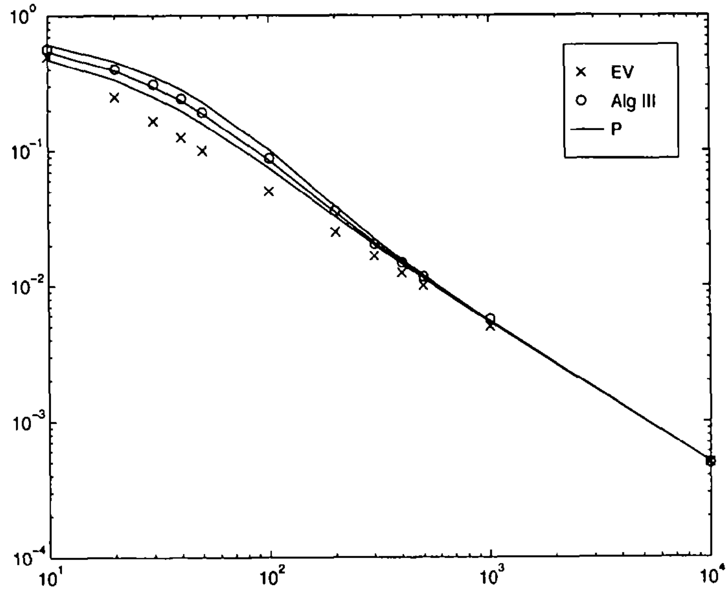


FIGURE 2: PAR(1,2), $\theta = 0.1$

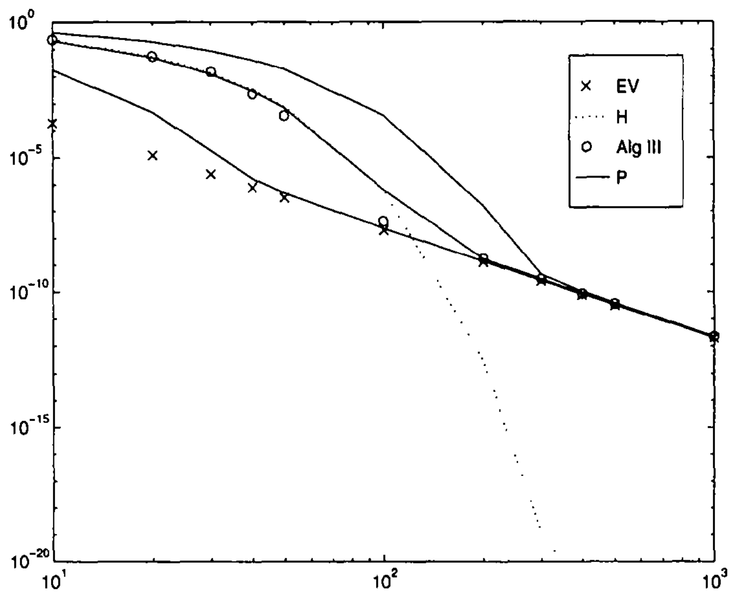


FIGURE 3: PAR(1,5), $\theta = 0.1$

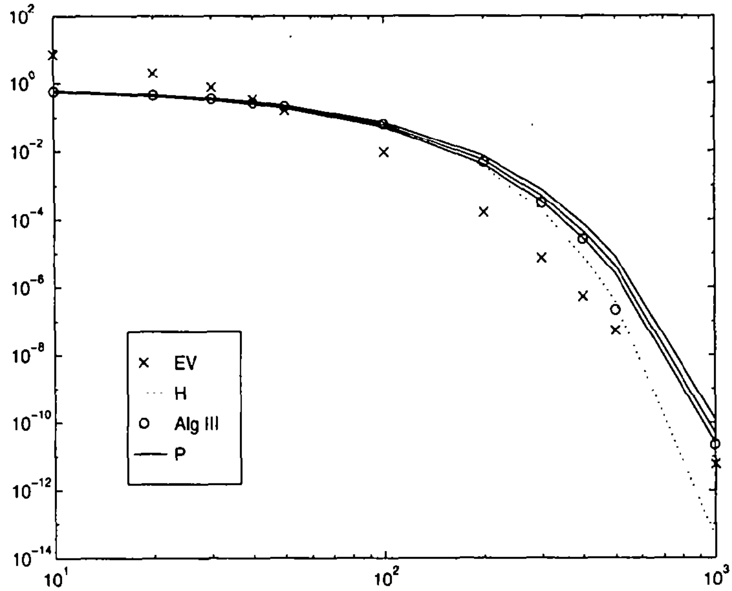


FIGURE 4: Weibull(1/2), $\theta = 0.2$

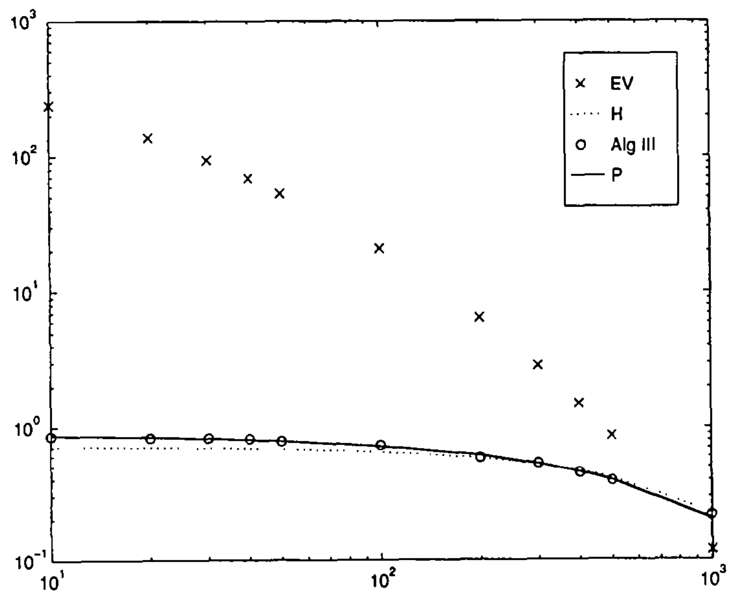


FIGURE 5: Weibull(1/3), $\theta = 0.1$

Thorin and Wikstad [31] have calculated the exact ruin probabilities for some values of u and θ. Therefore we compare our estimates with those values.

4.5. **Results.** The Tables 1-3 contain the estimates for different initial reserve u derived from the three algorithms together with their confidence intervals and the precision measured by (2). The estimates for PAR(1,2) distributed claims with security loading $\theta = 0.1$ are presented in Table 1. The results for PME(3) distributed claims with $\theta = 0.25$ are shown in Table 2 and for Lognormal (-1.62, 1.8) claims with $\theta = 0.1$ in Table 3.
In the Figures 1-5 we give the simulated values from Algorithm III based upon $n = 200$ replications, the approximation $\psi_{EV}(u)$ and $\psi_H(u)$ (if the third moment exists). These values are compared with the estimates, lower and upper bound derived from Panjer's approximation. Figure 1 shows the values for PAR(1,1.5), $\theta = 0.3$, Figure 2 for PAR(1,2), $\theta = 0.1$ and Figure 3 for PAR(1,5), $\theta = 0.1$. For the Weibull distribution we give the figure for $\nu = 1/2$, $\theta = 0.2$ and for $\nu = 1/3$, $\theta = 0.1$.

5. CONCLUSION

Below we give an overview of the most important properties of the algorithms and approximations we considered. The key observations from the above tables and figures as well as other examples, see Binswanger [8], are:

- O1 Algorithm I works fine for 'small' initial capital and underestimates $\psi(u)$ when u is 'large'.
- O2 Algorithm II usually overestimates $\psi(u)$ for 'small' u and underestimates for 'large' u.
- O3 Algorithm III is always of the right order of magnitude.
- O4 The precision measured by (2) is usually around $\frac{1}{2}$ for Algorithm I. For Algorithm II it is also around $\frac{1}{2}$ as long as the estimates are valid and around 1 when the estimates are wrong. The precision of the third algorithm is always around 1 even when the claim size distribution is Weibull.
- O5 The corrected diffusion approximation (13) gives very satisfactory results for 'small' initial capitals and is poor for 'large' initial reserves. The less heavy tailed the distribution of the claims is, the better the approximation is.
- O6 The asymptotic approximation (6) often requires u to be so large that the resulting ruin probability becomes extremely small, in fact much smaller than typical values of practical interest. The approximation turns out to be better the more heavy-tailed B is. In particular, it is much better for Pareto then for Weibull distributed claims.

Of course it would be nice to know what 'large' and 'small' initial capitals mean. The interpretation of 'large' or 'small' depends on the kind of distribution and on the choice of its parameter as well as on the security loading θ.

A comparative study of the accuracy of the various bounds and approximations in De Vylder and Goovaerts [10], Dickson [12], Omey and Willekens [24] and Omey and Willekens [25] is given by Binswanger [8]. In the latter, also alternative variance reduction techniques, like the use of regression-adjusted control variates, are to be found.

We point out also that Algorithm III applies to the total claims as well. That is, rather than the ruin probability, one wants to compute

$$G(x) = P\left(\sum_{i=1}^{M} \xi_i > x\right)$$

by simulation where M is the number of claims in a given period. The simplest case is where M is Poisson with parameter λ, say, and one can proceed just as for the ruin probability, generating M as Poisson rather than geometric. One again obtains the efficiency property (2). More generally, M could be allowed to have any distribution with finite second moment. For example, one could treat risk processes where the arrivals occur according to some Cox process in this way.

Besides Panjer's recursion also transform inversion via FFT offers an interesting estimation method. See for instance Embrechts, Grübel and Pitts [14] and Buchwald, Chevallier and Klüppelberg [9] for a discussion in the context of insurance.

For a broad overview of the application of numerical methods in risk theory, see Feilmeier and Bertram [17].

ACKNOWLEDGEMENT:

The authors would like to thank a referee for various remarks on earlier versions of the paper.

REFERENCES

[1] ABATE, J., CHOUDHURY, G. L., and WHITT, W. Waiting-time tail probabilities in queues with long-tailed service-time distributions. *Queueing Systems 16* (1994), 311-338.

[2] ABATE, J., and WHITT, W. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems 10* (1992), 5-88.

[3] ARNOLD, B. C., BALAKRISHNAN, N., and NAGARAJA, H. N. *A First Course in Order Statistics.* Wiley, New York, 1992.

[4] ASMUSSEN, S. Conjugate processes and the simulation of ruin problems. *Stochastic Processes and their Applications 20* (1985), 213-229.

[5] ASMUSSEN, S. *Applied Probability and Queues.* Wiley, Chichester, 1987.

[6] ASMUSSEN, S. Approximations for the probability of ruin within finite time. *Scandinavian Actuarial Journal p* 31-57, 1984, (p:64,1985).

[7] ASMUSSEN, S., and RUBINSTEIN, R. *Steady-state Rare Events Simulation in Queueing Models and its Complexity Properties, Advances in Queueing (J. Dhashalou, ed), 79-102.* CRC Press, Boca Raton, Florida, 1995.

[8] BINSWANGER, K. *Rare Events and Insurance.* ETH Zürich, PhD-thesis No. 12233.

[9] BUCHWALD, M., CHEVALLIER, E., and KLÜPPELBERG, C. Approximation methods for the total claimsize distribution – an algorithmic and graphical presentation. *Mitteilungen der Schweizerischen Vereinigung der Versicherungsmathematiker 2* (1993), 187-226.

[10] DE VYLDER, F., and GOOVAERTS, M. Bounds for classical ruin probabilities. *Insurance: Mathematics and Economics 3* (1984), 121 - 131.

[11] DEVROYE, L. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

[12] DICKSON, D. C. M. An upper bound for the probability of ultimate ruin. *Scandinavian Actuarial Journal 2* (1994), 131-138.

[13] DICKSON, D. C. M. A review of Panjer's recursion formula and its applications. *British Actuarial Journal 1* (1995), 107-124.

[14] EMBRECHTS, P., GRÜBEL, R., and PITTS, S. Some applications of the fast Fourier transform algorithm in insurance mathematics. *Statistica Neerlandica 47* (1993), 59-75.

[15] EMBRECHTS, P., and KLÜPPELBERG, C. Some aspects of insurance mathematics. *Theory of Probability and its Applications 39* (1994), 262-295.

[16] EMBRECHTS, P., and VERAVERBEKE, N. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics 1* (1982), 55-72.

[17] FEILMEIER, M., and BERTRAM, J. *Anwendungen numerischer Methoden in der Risikotheorie*. Verlag Versicherungswirtschaft E. V., Karlsruhe, 1987.

[18] GLYNN, P. W., and IGLEHART, D. L. Importance sampling for stochastic simulations. *Management Science 35* (1989), 1367-1392.

[19] GRANDELL, J. A class of approximations of ruin probabilities. *Scandinavian Actuarial Journal Suppl.* (1977; 1978: 77-78, 1978), 1977: 37-52.

[20] HEIDELBERGER, P. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Computer Modeling and Simulation 5* (1995), 43-85.

[21] HOGAN, M. L. Comment on corrected diffusion approximations in certain random walk problems. *Journal of Applied Probability 23* (1986), 89-96.

[22] IGLEHART, D. L. Diffusion approximations in collective risk theory. *Journal of Applied Probability 6* (1969), 285-292.

[23] KLÜPPELBERG, C. Estimation of ruin probabilities by means of hazard rates. *Insurance: Mathematics and Economics 8* (1989), 279-285.

[24] OMEY, E., and WILLEKENS, E. Second order behaviour of the tail of a subordinated probability distribution. *Stochastic Processes and their Applications 21* (1986), 339-353.

[25] OMEY, E., and WILLEKENS, E. Second order behaviour of distributions subordinate to a distribution with finite mean. *Communications in Statistics; Stochastic Models 3(3)* (1987), 311-342.

[26] PANJER, H., and DORAY, L. Further results on the probability of ruin with an absorbing upper barrier. *Transactions of the 23rd International Congress of Actuaries Helsinki* (1988), 159-177.

[27] PANJER, H. H., and WILLMOT, G. E. *Insurance Risk Models*. Society of Actuaries, 1992.

[28] RUBINSTEIN, R. Y. *Simulation and the Monte Carlo Method*. Wiley, New York, 1981.

[29] SIEGMUND, D. Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics 4* (1976), 673-684.

[30] SIEGMUND, D. Corrected diffusion approximations in certain random walk problems. *Advances in Applied Probability 11* (1979), 701-719.

[31] THORIN, O., and WIKSTAD, N. Calculation of the ruin probabilities when the claim distribution is lognormal. *ASTIN Bulletin 9* (1979), 231-246.

S. ASMUSSEN
Department of Mathematical Statistics
University of Lund
S-221 00 Lund
Sweden

K. BINSWANCER
Departement of Mathematics
ETH Zurich
CH-8092 Zurich
Switzerland

SETTING A BONUS-MALUS SCALE IN THE PRESENCE OF OTHER RATING FACTORS

GREG TAYLOR

Consultant, Tillinghast-Towers Perrin,
GPO Box 3279,
Sydney NSW 2001, AUSTRALIA
and
Professorial Associate, Centre for Actuarial Studies,
Faculty of Economics and Commerce,
University of Melbourne,
Parkville, Victoria 3052, AUSTRALIA

October, 1996
Revised April, 1997

SUMMARY

The operation of a bonus-malus system, superimposed on a premium system involving a number of other rating variables, is considered. To the extent that good risks are rewarded in their base premiums, through the other rating variables, the size of the bonus they require for equity is reduced. This issue is discussed quantitatively, and a numerical example given.

KEYWORDS

Bonus-malus, experience rating.

1. INTRODUCTION

A system of **bonus-malus** (BM) calculates the premium applicable to particular contract as a base premium, adjusted by a quantity (the bonus or malus) which depends on previous claims experience.

Consider a BM system in which the BM has J possible values, called the BM **levels**. These may be labelled 1, 2, ..., J, called the BM classes. The system is defined by the classes, levels, and the rules according to which claims experience is mapped to transitions between classes.

The collection of classes, together with their associated levels, will be referred to as the BM **scale**.

Over time, the portfolio will be distributed over the BM classes. In a typical BM system the distribution will ultimately stabilise. Because occupancy of each BM class is a function of claims experience, the individuals in the portfolio with low claim frequency parameters will tend to gravitate to the BM classes characterised by light claims experience. Conversely, for individuals with high claim frequencies.

The ultimate average claim frequency in each BM class defines the level to which that class is theoretically entitled. This, and related issues, have been dealt with many times in the literature. The two books of Lemaire (1985, 1995) provide a summary of a number of relevant matters.

It is common in such writings to assume that BM is the only means by which premiums are differentiated. In other words, all contracts are subject to the same base premium.

In practice, some portfolios, e.g. motor, are rated on a comparatively large (perhaps 10 or so) other variables. These will also differentiate individuals according to claim frequencies.

Consider the distribution of the portfolio over risk classes in the presence of these other rating variables. If they are used effectively by the premium system, then those BM classes with low average claim frequencies will tend to have low base premiums also.

In this event, the justifiable BM levels need to recognise the differentiation of underlying claim frequency by experience, but only to the extent that this differentiation is **not already recognised within base premiums**.

Subsequent sections of this paper examine the detail of this issue.

2. Notation

Let:

θ = vector of covariates (e.g. age, sex, etc.) with risk premium of an individual;

Λ = an individual's true underlying risk premium.

It is asumed that, for given θ, there is a distribution of values of Λ. Suppose that the pdf of Λ, conditioned on θ, takes the form:

$$f(\lambda \mid \theta) = g(\lambda \mid \mu(\theta)), \tag{2.1}$$

for some pdf g(.) and where

$$\mu(\theta) = E[\Lambda \mid \theta]. \tag{2.2}$$

The parameter θ will vary from one contract to antother, and hence so does $\mu(\theta)$. Let

$$h(\mu) = \text{pdf of } \mu \text{ over the whole portfolio.}$$

Now introduce a BM system with classes 1, 2, ..., J, and let

$\pi_j^{(t)}(\lambda)$ = probability that a policy owner with underlying risk premium λ occupies BM class j in the t-th period since commencement of the system.

Note that

$$\sum_{j=1}^{J} \pi_j^{(t)}(\lambda) = 1 \text{ for each } t, \lambda. \tag{2.3}$$

The system is initialised at $t = 1$. It is assumed **Markovian**.

For most realistic BM systems, the vector $[\pi_1^{(t)}(\lambda), \ldots, \pi_J^{(t)}(\lambda)]$, representing the distribution of BM levels in period t of risks characterised by λ, will approach a **steady state** with increasing t. It will be assumed here that such a steady state exists, and that convergence to it occurs over time. Let $\pi_j(\lambda)$ = the steady state value of $\pi_j^{(t)}(\lambda)$.

One can define the Bayesian posterior expectations:

$$\lambda_j^{(t)} = \mathrm{E}[\Lambda \mid \text{BM class in t-th period} = j], \tag{2.4}$$

$$\mu_j^{(t)} = \mathrm{E}[\mu \mid \text{BM class in t-th period} = j], \tag{2.5}$$

and let λ_j, μ_j be the steady state versions of $\lambda_j^{(t)}$, $\mu_j^{(t)}$.

One way of viewing these quantities is as follows. The portfolio consists of two levels of heterogeneity:

- different risk classes defined by different $\mu(\theta)$; and
- within these different risk classes, different individuals characterised by their personal values of λ.

The quantities $\mu_j^{(t)}$ indicate the extent to which the BM system differentiates the **risk classes** over time. The quantities $\lambda_j^{(t)}$ indicates the extent to which the BM system differentiates individuals over time.

3. SETTING THE BONUS-MALUS SCALE

The Bayesian expectation $\lambda_j^{(t)}$ can be represented as:

$$\lambda_j^{(t)} = \int \lambda \, p^{(t)}(\lambda, j) \, d\lambda \Big/ \int p^{(t)}(\lambda, j) \, d\lambda, \tag{3.1}$$

where $p(.)$ will be used generically to denote a pdf and in this case $p^{(t)}(.)$ is a pdf in the t-th interval.

Now the joint pdf in (3.1) can be expanded:

$$\begin{aligned} p^{(t)}(\lambda, j) &= \pi_j^{(t)}(\lambda) \, p^{(t)}(\lambda) \\ &= \pi_j^{(t)}(\lambda) \int g(\lambda \mid \mu) \, h(\mu) \, d\mu. \end{aligned} \tag{3.2}$$

By (3.1) and (3.2),

$$\lambda_j^{(t)} = \frac{\int d\mu \, d\lambda \, \lambda \, \pi_j^{(t)}(\lambda) \, g(\lambda \mid \mu) \, h(\mu)}{\int d\mu \, d\lambda \, \pi_j^{(t)}(\lambda) \, g(\lambda \mid \mu) \, h(\mu)}. \tag{3.3}$$

Similarly,

$$\mu_j^{(t)} = \frac{\int d\mu\, d\lambda\, \mu\, \pi_j^{(t)}(\lambda)\, g(\lambda \mid \mu)\, h(\mu)}{\int d\mu\, d\lambda\, \pi_j^{(t)}(\lambda)\, g(\lambda \mid \mu)\, h(\mu)}. \tag{3.4}$$

Define

$$r_j^{(t)} = \lambda_j^{(t)} / \mu_j^{(t)}. \tag{3.5}$$

As in Section 2, the absence of the time index indicates the steady state, i.e.

$$r_j = \lambda_j / \mu_j \tag{3.6}$$

To interpret $r_j^{(t)}$, first consider the degenerate case in which $h(.)$ concentrates all mass at a single value μ. That is, the portfolio contains only one value of θ; there is no variation of risk covariates, which in turn means that all policy owners are indistinguishable before the accumulation of claims experience. This is the case most commonly considered in the literature.
In this case (3.4) gives

$$u_j^{(t)} = \mu, \tag{3.7}$$

hence (3.5) becomes

$$r_j^{(t)} = \lambda_j^{(t)} / \mu. \tag{3.8}$$

The number $\lambda_j^{(t)}$, is effectively the Bayesian revision of μ taking into account the information that BM level is j in the t-th period. Thus $r_j^{(t)}$, is the factor by which the Bayesian revision adjusts the policy owners' prior expectation. Equivalently, $r_j^{(t)}$ is the factor by which t years of experience revises the prior risk premium in BM class j.
The situation involving general $h(.)$ is similar. However, in this case the composition of BM class j with respect to the prior expectation $\mu(\theta)$ will change over time. For example, there will be a tendency for the contracts with the lowest priors to migrate to the BM class with lightest claims experience. Thus, $\mu_j^{(t)}$, tracks the average prior in BM level j over time.

Despite this change, $r_j^{(t)}$, still denotes the factor by which experience revises the **average prior risk premium in BM level j**.

The relevance of this is as follows. The average prior $\mu_j^{(t)}$ is the average "standard premium rate" (i.e. the rate before recognition of experience) applicable to BM class j in the t-th year. Thus 100 $[r_j^{(t)} - 1]$ is the BM percentage justified by experience in class j.

Suppose that BM class K receives these standard rates. Then the factor which can be justified as relating BM class j to standard rates is $r_j^{(t)} / r_K^{(t)}$. These factors can be summarised in the vector

$$r_*^{(t)} = r^{(t)} / r_K^{(t)}, \tag{3.9}$$

where $r^{(t)}$ is the vector with components $r_j^{(t)}$.

The conclusion is that the maximum differentiation between premiums for different classes will be according to a factor

$$\max_j r_j / \min_j r_j, \tag{3.10}$$

with r_j defined by (3.6), i.e. a factor of

$$\max_j(\lambda_j / \mu_j) / \min_j(\lambda_j / \mu_j). \tag{3.11}$$

If the differentiation of priors μ_j over BM classes is left out of account, the differentiation of premiums wil be according to a factor of

$$\max_j \lambda_j / \min_j \lambda_j, \tag{3.12}$$

which will usually be substantially larger than (3.11).

4. NUMERICAL EXAMPLE

A specifically structured portfolio of risks, subject to a particular BM system, has been simulated and values of $\lambda_j^{(t)}$, $\mu_j^{(t)}$ recorded.
The portfolio consists of 10 groups of individuals structured as follows.

TABLE 4.1
PORTFOLIO STRUCTURE

Risk Group	*Mean cell average claim frequency* %	*Coefficient of variation of within-cell claim frequency* %	*Proportion of portfolio* %
1	6.5	75	4.0
2	8.9	65	18.9
3	11.4	60	15.8
4	13.7	55	20.1
5	16.1	50	12.0
6	20.1	45	11.6
7	24.9	40	10.3
8	29.7	40	4.5
9	36.0	40	2.1
10	50.5	40	0.6
TOTAAL	15.7		100

This structure was obtained by constructing a multiplicative model of claim frequency according to a number of covariates (but excluding BM), and then counting the numbers of policies in bands of modelled claim frequency, 5-7.5%, 7.5%-10%, etc.

The coefficient of variation of each band was chosen largely by informed guesswork, but subject to the criterion, again guesswork, that within-cell variance should increase in absolute terms with increasing frequency, but decrease in relative terms.

Individuals within a particular risk group are sampled from a certain gamma distribution with the parameters set out in Table 4.1, as will be described later. There are 9 BM classes, of which Class 6 is the standard. A higher class number indicates a higher premium. The rules for transition between the classes are as follows.

TABLE 4.2
BM TRANSITION RULES

Opening Class	Closing class after a year if 0 claims	1 claim	2 claims	3 or more claims
9	8	9	9	9
8	7	9	9	9
7	6	8	9	9
6	5	7	8	9
5	4	7	8	9
4	3	6	7	8
3	2	5	7	8
2	1	4	6	7
1	1	3	5	7

Appendix A gives the technical detail of the simulation
The claims experience of this portfolio is simulated over 30 years. At the beginning of year 1 all insureds are assumed to be in Class 6. The distribution appears to stabilise by about the end of Year 24. Consequently, the following results are averages over years 24 to 30.

TABLE 4.3
SIMULATION RESULTS

BM Class j	Average Proportion of portfolio %	True claim frequency λ_j %	Cell claim frequency μ_j %	Ratio: true/cell claim frequency r_j %
9	1	46	26	175
8	1	38	24	156
7	2	32	22	145
6	3	30	22	139
5	4	23	20	116
4	4	21	19	111
3	10	18	17	103
2	9	17	17	102
1	66	12	14	85

The table shows that, if base premiums reflect cell claim frequencies accurately, the BM scale should vary by a maximum factor of about 2 [cf (3.11)]. If the variation of the base premiums were left out of account, the BM scale would vary by a factor of nearly 4 [cf (3.12)]. The BM scale justified by the middle columns of Table 4.3 in the case K = 6 is as follows.

TABLE 4.4

PREMIUMS FOR BM CLASSES

BM Class j	*Premium as % of standard*	
	recognising covariates	*ignoring covariates*
9	126	152
8	112	125
7	104	107
6	100	100
5	83	77
4	80	71
3	74	59
2	73	57
1	61	40

If these premiums had been computed from the column of λ_j in Table 4.3, ignoring the effect of the covariates, quite different, and misleading, results would have been obtained, as shown in the final column of Table 4.4.

It is of interest to examine how results of this type vary as the BM system varies. Consider the case in which Table 4.2 is replaced by a simple set of rules which provide for:

- 1 step forward for each claim-free year;
- 4 steps back for each claim.

This is much more severe than Table 4.2 which is largely a 1-forward/2-back set of rules.

The new system replaces Table 4.4 by Table 4.5.

TABLE 4.5

PREMIUMS FOR BM CLASSES IN MORE SEVERE SYSTEM

BM Class j	*Premium as % of standard*	
	recognising covariates	*ignoring covariates*
9	123	150
8	113	125
7	104	109
6	100	100
5	92	83
4	91	81
3	86	75
2	83	72
1	73	55

If the ratio of the two columns in Table 4.5 is regarded as an "error ratio", measuring the error in ignoring covariates, the following comparison is noteworthy.

TABLE 4.6

ERROR RATIOS FOR DIFFERENT BM SYSTEMS

BM Class j	Error Ratio: Original BM system (Table 4.4) %	Error Ratio: Severe BM system (Table 4.5) %
9	121	122
8	112	111
7	103	105
6	100	100
5	93	91
4	89	89
3	79	88
2	78	86
1	66	75

Although the BM systems are very different, and so are the levels of bonus justified by them, there is a good deal of similarity between their error ratios.

APPENDIX A

TECHNICAL DETAIL OF SIMULATION

1 Individual claim frequency

Consider an individual in BM class j with $\mu_j^{(1)}$ given by Table 4.1. Let w_j be the associated coefficient of variation in Table 4.1. The value of Λ for this individual is assumed to be

$$\Lambda = \mu_j^{(1)}[1 + w_j^{(X/2-1)}], \tag{A.1}$$

where $X \sim \chi_2^2$.

Since χ_2^2 is gamma with mean 2 and standard deviation 2, (A.1) gives $E[\Lambda] = \mu_j^{(1)}$, s.d. $[\Lambda] = w_j\, \mu_j^{(1)}$, as required.

Values of X are simulated as:

$$X = X_1^2 + X_2^2, \tag{A.2}$$

where X_1, X_2 are independent, and

$$X_i \approx N(0,1),\ i = 1,2. \tag{A.3}$$

2 Claim inter-arrival times

For the individual discussed in Appendix A.1, it is assumed that the number of claims in a year is distributed Poisson (Λ). Hence inter-arrival times are exponentially distributed with mean $1/\Lambda$. These inter-arrival times have been simulated as:

$$T = -[\log(1-U)]/\Lambda, \tag{A.4}$$

where U is uniform [0,1].
Note that (A.4) is equivalent to:

$$U = 1 - \exp(-\Lambda T), \tag{A.5}$$

from which exponentiality of T is easily proved.

REFERENCES

LEMAIRE, J. (1985) *Automobile insurance: actuarial models.* Kluwer – Nijhoff Publishing, Boston.
LEMAIRE, J. (1995) *Bonus-malus systems in automobile insurance.* Kluwer – Academic Publishers, Boston.

APS REINSURANCE

By BRUNO KOLLER, NICOLE DETTWYLER

ABSTRACT

This paper presents a new reinsurance product, called 'Adaptive Pivot Smoothing' (APS). It is designed to reduce the variance of the risk reinsured without affecting the mean. Investment theories have provided the idea for the product.

KEYWORDS

Reinsurance, Financial Reinsurance, Smoothing.

1. THE NEED OF A NEW REINSURANCE PRODUCT

'Adaptive Pivot Smoothing', APS for short, is a new reinsurance product. Why do we need a new reinsurance product? Do existing products not already provide safe cover for all of the needs of insurance companies? We believe that traditional reinsurance treaties have three shortcomings in practice.

Firstly, traditional reinsurance products do not take account of the ideas of modern portfolio theory. Investors and insurance company managers alike aim to maximise the returns on their portfolios and, at the same time, minimise the volatility of the results. We have come up with a reinsurance product which does not alter the expectation of the claims distribution but which manages to reduce the variance to a level defined by the direct insurer. 'Pivot' in the acronym APS hints at this feature.

Secondly, there is more and more demand for reinsurance with potential risk of misuse. Health insurance policies spring to mind here most notably. Generally speaking, health insurance portfolios are quite homogeneous, without significant fluctuations in terms of loss load. In most cases reinsurance is not necessary. The actual problem facing health insurance companies is how to adapt premiums on time to the rapidly growing costs of health care. This is not always possible for political reasons. Consequently, insurers might be tempted to pass poor loss performance on to the reinsurer. If the reinsurance company, in turn, insists on increasing the premiums, the health insurance companies cancel their reinsurance policies. An experience rating component in our new product greatly reduces the potential for misuse of the cover. The term 'Adaptive' in APS reminds of this characteristic.

The third shortcoming often comes to light when a new insurance product is launched. In response to increased pressure from their competitors, insurance companies must develop covers and set rates with only very sketchy statistical material available sometimes. A reinsurance program would be advantageous

ASTIN BULLETIN, Vol. 27, No. 2, 1997, pp. 329-337

which a) offered temporary financing in cases where premiums are too low and b) skimmed off profits in cases where premiums are too high, setting these funds aside to be used later on. This gives the insurance company time to analyse business trends and to adapt premiums accordingly. Our new reinsurance product allows profits and losses to be carried forward to later years – the term 'Smoothing' in APS was chosen for this reason.

2. TRADITIONAL REINSURANCE

In technical terms, reinsurance means splitting a risk in two. One part of the risk (the retention) rests with the insurer, while the other part is ceded to the reinsurer. The split is determined by the payment function $h(S)$, which specifies how much the reinsurer is required to pay towards every claim S (S is a random variable). Thus, the insurance company carries the retention, $S - h(S)$, the reinsurer the risk $h(S)$.

The expected value of the retention distribution is

$$E[S - h(S)] = E[S] - E[h(S)],$$

and the expected value of the ceded distribution is

$$E[h(S)].$$

The variance and the standard deviation of the claims distribution of S are labelled $VAR\ [S]$ and $SDV\ [S]$ respectively. The variance of the retention distribution, $VAR\ [S\text{-}h\ (S)]$ is:

$$VAR[S - h(S)] = VAR[S] + VAR[-h(S)] + 2 \cdot COV[S, -h(S)]$$

$$= VAR[S] + VAR[h(S)] - 2 \cdot COV[S, h(S)].$$

Writing K for the correlation coefficient we get:

$$VAR[S - h(S)] = VAR[S] + VAR[h(S)] - 2K \cdot SDV[S] \cdot SDV[h(S)].$$

If we make the obvious assumption that the correlation between S and $h(S)$ is positive, the value K must be between 0 and 1, which enables us to make the following estimate:

$$(SDV[S] - SDV[h(S)])^2 \leq VAR[S - h(S)] \leq VAR[S] + VAR[h(S)].$$

If there is complete linear correlation ($K = 1$), the following equations apply:

$$VAR[S - h(S)] = (SDV[S] - SDV[h(S)])^2$$

$$SDV[S - h(s)] + SDV[h(S)] = SDV[S].$$

Thus, the reduction of the variance can be achieved most effectively by means of a linear payment function!

The situation of the cedent buying reinsurance can be quantified by

$$S - P[S] - h(S) + Q[h(S)];$$

here P represents the premium income and Q the reinsurance premium. If we assume that $Q[h(S)] > E[h(S)]$, consequently

$$E[S - P[S] - h(S) + Q[h(S)]] > E[S - P[S]].$$

In other words, reinsurance increases the mean burden on the cedent.

However, if reinsurance is unable to lower the mean burden at all, it is difficult to see how a payment function h can be used with $E[h(S)] > 0$. (Common to all traditional forms of reinsurance – proportional and non-proportional – is $E[h(S)] > 0$.) This merely amounts to an exchange in premiums and losses without giving rise to any economic benefit. It would be much wiser for the cedent to agree on a payment function with $E[h(S)] = 0$.

The reinsurer obviously cannot be expected to lower the cedent's burden. The benefit offered by reinsurance is that it brings about a reduction in volatility or, if volatility is measured in terms of variance, a precisely quantified reduction of the variance. Consequently, the cedent should be asking the reinsurer to reduce the variance of the claims distribution by x per cent; the service provided by the reinsurance company can then be assessed accurately.

To sum up, an insurance company would be best advised to take out a reinsurance policy in which $E[h(S)] = 0$ and $SDV[h(S)] = c \cdot SDV[S]$, where $0 < c < 1$.

This view of reinsurance is very much in accordance with modern portfolio theory. The investor is aiming to maximise expected returns, whilst at the same time minimise volatility. The risk manager at an insurance company endeavours to achieve precisely the same effect, stabilising profits on a high level with small fluctuations.

3. APS REINSURANCE

The above analysis clearly shows that the new product should display a linear payment function, whereby the ceded distribution should have an expected value of zero.

Assuming a linear payment function of

$$h(S) = c \cdot (S - E[S]), \text{where } 0 < c < 1,$$

we can derive the following relations:

$$E[h(S)] = 0$$

$$E[S - h(S)] = E[S]$$

$$VAR[h(S)] = c^2 \cdot VAR[S]$$

$$VAR[S - h(S)] = (1 - c)^2 \cdot VAR[S].$$

The point $(E[S], 0)$ where the linear payment function intersects the x-axis has been called the *'Pivot'* in order to emphasize that both losses and profits are affected by reinsurance. If the claims ratio is higher than the pivot, the reinsurer pays the cedent. However, if it is smaller, the cedent pays the reinsurer.

It is crucial to the success of this design that the payment function intersects the x-axis at the expected value of the claims distribution. Of course the 'true' expected value is hardly ever known. What is more, the expected value can shift with time. Therefore, we have to consider a mechanism which automatically adapts the payment function, or to be more precise the pivot, to current claims experience. This is a vital component of APS Reinsurance.

The adaptation mechanism must be defined in the treaty to ensure that the cedent and the reinsurer do not disagree on the 'correct' pivot. If the adaptation mechanism is to work properly, it must be unambiguous, simple and efficient.

It is important to understand that we are faced here with a forecast problem. We have to forecast the pivot for the coming financial year – not assess the pivot for the previous period. The cedent needs to know the exact terms of the reinsurance treaty in advance. Section 5 examines an adaptation mechanism, based on credibility; other methods are possible and may be even better, depending on the situation.

With our choice of the payment function, the expected value of claims being ceded is zero. Therefore the premium, calculated as 'expected claims plus loading', will be much lower than the premium for traditional reinsurance products. The choice of the premium calculation principle needs to be given very careful consideration. The 'expected value principle', for instance, is pointless here. The variance principle or the standard deviation principle are possible. However, we recommend that the principle of zero utility be used (cf. GERBER (1979), page 67) because also higher moments should be taken into account, given the dominance of the loading.

4. IMPLEMENTATION

APS Reinsurance is based on the *claims ratio r* of a certain portfolio. The claims ratio for financial year t is defined by

$$r_1(t) = -S(t)/P(t), \text{ where } S(t) \leq 0, \; P(t) > 0.$$

$S(t)$ represents the claims during financial year t, and $P(t)$ represents the premium income during the same period. When calculating $r_1(t)$, premium income, claims and claims reserve must be allocated to the correct year.

The *payment function h* becomes a function of claims ratio r_1:

$$h(r_1(t)) = a \cdot (r_1(t) - E[r_1(t)]), \text{ where } 0 < a < 1.$$

The parameter a is referred to as the *smoothing factor*. It defines the reduction in standard deviation brought about by the reinsurance.

The claims ratio after reinsurance, r_2, becomes

$$r_2(t) = r_1(t) - h(r_1(t)) = (1 - a) \cdot r_1(t) + a \cdot E[r_1(t)]$$

r_2 lies between r_1 and $E[r_1(t)]$, thus damping the volatility of r_1.

At the end of the financial year t, the claims ratio $r_2(t)$ is calculated from $r_1(t)$ on the basis of the above equation. r_2 serves to calculate the amount of money to be paid, called the *smoothing benefit $L(t)$*:

$$L(t) = -r_2(t) \cdot P(t) - S(t).$$

It is important to understand that if L is positive (the observed claims ratio being higher than the expected value), the reinsurer pays benefit to the cedent. Otherwise, the cedent pays the reinsurer. In practice, the reinsurer generally incorporates (upper and lower) limits to the smoothing benefit into the terms of the contract.

The target of the smoothing procedure, i.e. the *pivot $b(t)$*, is the expected value of the random variable $r_1(t)$. Therefore, the equation for the claims ratio r_2 can also be written as:

$$r_2(t) = (1 - a) \cdot r_1(t) + a \cdot b(t) = b(t) + (1 - a) \cdot (r_1(t) - b(t)).$$

At the start of the reinsurance, the pivot is fixed on the basis of calculations or observations. Subsequently the pivot is automatically adapted on the basis of the most recent claims experience. One obvious choice for the *adaptation mechanism* is to use a credibility approach:

$$b(t + 1) = g(t) \cdot r_1(t) + (1 - g(t)) \cdot b(t)$$

$g(t)$ is called the *credibility weight*, or *credibility* for short. The advantage of this type of formula is that it incorporates all of the past experience, though with an exponential decrease in weighting. g is generally regarded independent of time. The above approach, in which weighting is time-related, is referred to as 'adaptive exponential smoothing' in time series analysis (cf. ABRAHAM and LEDOLTER (1983), page 377). In section 5 we propose a formula for $g(t)$ which has proved its worth in simulations of business procedures.

Thanks to the adaptation mechanism, *profits and losses* can be carried forward in APS Reinsurance. The smoothing benefit is carried over to the next year either in full or in part, and is offset against earlier payments. When the treaty is terminated, the accrued balance is paid back in full or in part over a predefined period of time. The adaptation mechanism guarantees that the (positive or negative) balance cannot grow without limits. Technically speaking, the profits and losses carried forward belong to the insured party and attract interest (positive or negative).

Depending on whether the balance is carried forward and/or settled at the end in full or only in part, APS Reinsurance becomes more or less a *financial reinsurance*.

In some cases it may be advisable to ask the cedent to pay a *deposit* at the start of the year. The deposit is offset against the smoothing benefit at the end of the year.

The ceded distribution can be calculated from the distribution of the aggregate claim amount and the payment function. If profits and losses are not being carried forward, the calculation of the *premium* is based on this distribution. If profits and losses are carried over and the final balance is not fully settled, we have to deal with a sum of random variables and the premium calculation is based on the folded distribution. Due account must be taken of the precise terms and conditions regarding termination of the treaty.

If the profits and losses are carried over in full and the final balance fully settled, there is no longer any technical risk. In this event, *commissions* should be requested instead of premiums. The commissions should be based on the difference between the premiums for risk r_1 and risk r_2 (cf. section 5), as this difference reflects the benefit which the cedent derives from the reinsurance.

5. EXAMPLE

In this section we will make assumptions about the claims distribution, specify the premium calculation principle and discuss a pivot adaptation mechanism.

We regard the claim S as being stochastic, whereas we view the cedent's income from premiums P as being deterministic. Let the claims be normally distributed. (Other distributions might be more realistic, but the normal distribution is better suited to illustrate the important points.) The standard deviation of the normal distribution is fixed, whereas the expected value is a function of time.

It follows that the claims ratio r_1 is also normally distributed. Thanks to the linear payment function, the ceded distribution of $r_1 - r_2$ and the retention distribution of r_2 are also normal with:

$$E[r_2] = E[(1-a) \cdot r_1 + a \cdot E[r_1]] = E[r_1]$$

$$E[r_1 - r_2] = E[r_1 - (1-a) \cdot r_1 - a \cdot E[r_1]] = 0$$

$$SDV[r_2] = SDV[(1-a) \cdot r_1 + a \cdot E[r_1]] = (1-a) \cdot SDV[r_1]$$

$$SDV[r_1 - r_2] = SDV[r_1 - (1-a) \cdot r_1 - a \cdot E[r_1]] = a \cdot SDV[r_1].$$

The premium calculation principle which we select is the exponential principle (cf. Gerber (1979) for instance) with risk aversion d. Instead of defining the utility of money, we regard it as a function of the claims ratio. Consequently, when we apply the exponential principle we obtain a premium rate instead of a premium amount. The premium rate is then calculated as

$$Q[r] = log\left(E\left[e^{dr}\right]\right)/d, \text{ where } d > 0.$$

Our approach is theoretically not quite correct. To a company sums of money are important, not ratios. Nevertheless, it provides us with a reinsurance premium which is proportional to the cedent's premium income the simplest approach in practice. It is advisable to adapt the risk aversion d depending on the size of the reinsured portfolio.

For a normal distribution the above formula becomes

$$Q[r] = E[r] + 0.5 \cdot d \cdot VAR[r].$$

Consequently, the reinsurer has to charge the following premium rate for the APS Reinsurance product:

$$Q[r_1 - r_2] = 0.5 \cdot d \cdot a^2 \cdot VAR[r_1]$$

and the *APSpremium* is calculated as $Q[r_1 - r_2] \cdot P$.
If we compare the APS premium rate with the rates for r_1 and r_2,

$$E[r_1] + 0.5 \cdot d \cdot VAR[r_1]$$

$$E[r_1] + 0.5 \cdot d \cdot (1 - a)^2 \cdot VAR[r_1],$$

we see that the value $Q[r_1 - r_2]$ is smaller than the difference between the two rates. In other words, the benefit which the cedent derives from the reinsurance product is higher than the reinsurance premium.

We discuss the following *pivot adaptation mechanism*:

$$b(t+1) = g(t) \cdot r_1(t) + (1 - g(t)) \cdot b(t).$$

It certainly makes sense that g is a function of r_1, b and $SDV[r_1]$: $g(t; r_1(t), b(t), SDV[r_1])$. The greater the difference between $r_1(t)$and $b(t)$, the higher the probability that the old estimate $b(t-1)$ is obsolete and that greater weighting should be allocated to the claims ratio currently observed: the credibility $g(t)$ should increase. Furthermore, the difference between the old pivot and the new observed value is all the more significant if the standard deviation is low: the credibility $g(t)$ should also increase in this case. The simplest quantity which meets these requirements is

$$|r_1(t) - b(t)|/SDV[r_1].$$

The above expression can have values between 0 and infinity, while $g(t)$ must be between 0 and 1. Therefore, we need a strictly increasing function which maps the positive real numbers on the interval [0, 1].

The above expression has a standard normal distribution; remember that b is the expected value of r_1. An obvious choice for the required transformation function is the probability of the interval

$$(-|r_1(t) - b(t)|/SDV[r_1], +|r_1(t) - b(t)|/SDV[r_1]).$$

The *credibility* is then calculated as

$$g(t) = 2 \cdot [SND(|r_1(t) - b(t)|/SDV[r_1]) - 0.5],$$

where SND stands for the standard normal distribution function.

6. APPLICATION OF APS REINSURANCE

APS Reinsurance is generally suitable for covering portfolios which entail a relatively *high frequency of claims*. Health insurance portfolios typically fulfill these requirements. If, on the other hand, claims are very rare like for instance in catastrophe reinsurance, the adaptation mechanism will not provide any suitable results.

APS Reinsurance is specially designed to suit sectors which are faced with *increasing claims costs*. The claims ratio typically rises over a period of a few years until the premiums are adapted. The claims ratio then drops and the cycle starts all over again. APS Reinsurance is ideal as a smoothing instrument in a case such as this.

The product can also play an important role when a new product is being set up. Statistics often prove to be unreliable at the start. With the aid of APS Reinsurance, results which are poor or which are (overly) good during those first few years can be carried over to later years, giving the cedent time to adapt the premiums.

Quota share reinsurance achieves the same goal as APS Reinsurance: it reduces the variance of the aggregate claims distribution to a certain percentage. However, the cedent is required to pay a considerable share of its premiums to the reinsurer. This is most of the time undesirable, so quota share reinsurance is no longer common.

Surplus Reinsurance also reduces the variance of the aggregate claims distribution. But the extent of the reduction is very difficult to calculate, whereas in APS Reinsurance the reduction percentage is part of the treaty. Stop loss reinsurance, which also serves to reduce the variance, has the disadvantage of a heavily loaded premium. This can be avoided in APS Reinsurance by implementing it as a financial reinsurance.

In spite of having a linear payment function, APS Reinsurance is not a proportional reinsurance. Proportional reinsurance is defined as being an agreement where the cedent and the reinsurer share premiums and losses in the same proportion. It is important to understand why this is not the case. APS Reinsurance does not carry losses – it reduces the volatility!

One of the advantages of APS is that it can be turned into a financial reinsurance. The dilemma with financial reinsurance is to provide enough risk exposure so that the supervising authorities class the treaty as an insurance and not as a banking transaction. If treated as a banking transaction, the smoothing effect disappears because of the profits and losses carried over show up in the balance sheet. There are a number of ways of including more risk, which have been discussed above (section 4). Another possibility not yet mentioned would be to limit the smoothing benefit and cover the excess by a conventional stop loss reinsurance; the stop loss reinsurance would then be part of the APS treaty.

REFERENCES

ABRAHAM, B. and LEDOLTER, J. (1983) *Statistical Methods for Forecasting*. John Wiley & Sons, New York.

GERBER, H.U. (1979) *An Introduction to Mathematical Risk Theory*. S.S. Huebner Foundation Monograph 8, University of Pennsylvania, Philadelphia.

BRUNO KOLLER
Dinkelbergstr. 21
CH-4125 Riehen

NICOLE DETTWYLER
Petersgraben 21
CH-4051 Basel

AN INTEGRATED DYNAMIC FINANCIAL ANALYSIS AND DECISION SUPPORT SYSTEM FOR A PROPERTY CATASTROPHE REINSURER[1]

By STEPHEN P. LOWE

Tillinghast – Towers Perrin

JAMES N. STANARD

RenaissanceRe Holdings, Ltd.

ABSTRACT

This paper describes the dynamic financial analysis model currently being used by a property catastrophe reinsurer to manage its business. The model is an integral part of the day-to-day operations at the Company, and is used as a decision making tool in the underwriting, investment, and capital management processes. The paper begins by describing the framework that the Company uses for risk management. This includes a classification of the risks facing the Company, which is used to define and prioritize their implementation in the model. Also included is a description of the conceptual approach the Company takes to evaluate the tradeoff between risk and return. The paper then goes on to describe the structure and operation of the dynamic financial analysis model and provides examples of its use at the Company, along with illustrative examples of the various types of output it produces.

KEYWORDS

Asset/Liability management; Capital adequacy; Dynamic financial analysis; Expected policyholder deficit; Modern portfolio theory; Property catastrophe reinsurance; Risk management; Simulation models; Underwriting cycles.

1. INTRODUCTION

The Company that is the subject of this paper is a major property catastrophe reinsurer, writing excess of loss coverage on a world-wide basis. It was formed in Bermuda in 1993 to provide additional capacity to the market, capitalizing on the

[1] An earlier version of this paper was prepared for the Casualty Actuarial Society's 1996 Call for Papers on Dynamic Financial Models of Property/Casualty Insurers. An updated version was presented at the XXVII ASTIN Colloquium in Copenhagen.

ASTIN BULLETIN, Vol. 27, No. 2, 1997, pp. 339-371

market dislocation following Hurricane Andrew. Since that time the Company has grown to be one of the largest specialist writers in the catastrophe reinsurance market.

Since its formation in 1993, a core strategic premise of the Company has been that an increased level of precision in the measurement and management of risk can be translated into a competitive advantage.

- Improved measurement of underlying exposure and modeling of losses allows underwriters to build a superior insurance portfolio, one that is less risky and/or more profitable than that of peers.
- Improved measurement of financial risk allows management to make more efficient use of capital, leading to superior returns on that capital.

The Company has developed systems and processes to support and implement this premise. Taken as a whole, these are used to facilitate ongoing dynamic financial analysis (DFA) of the enterprise. Perhaps most importantly, dynamic financial analysis activities are not restricted to technical staff operating apart from management. DFA has been integrated directly into the ongoing underwriting and financial management processes of the Company. Every senior manager is trained on the use of the system; thus, it is a practical and immediate resource for decision making.

The development of these capabilities has been a collaborative effort between the Company and an actuarial consulting firm (hence this co-authored paper). In addition to the authors, who co-led the development effort, many other people in both organizations contributed to the conceptualization, design, programming, and testing of the system.

Development of the system and its modeling capabilities is an ongoing activity; its design continues to evolve as experience with its use develops. Initially, the model was relatively simple, and focused only on measuring the principal risks facing the Company. As confidence in the model has grown, new features and additional risk components have been added. While this paper generally describes the model as it exists today, a few features are described that are under active development at the time of this writing, with the full expectation that they will be on line by the time of publication. A major goal of current development activity is to integrate the various components of the system more completely, strengthening the linkages between the risk elements in the process.

Finally, while the output exhibits presented in the paper are illustrative of those actually produced by the model, they are stylized versions of that output, and use figures that have been altered. The exhibits are included only to illustrate the varied uses of the model, and represent only a small sample of what has been produced. Many of the output exhibits, as well as the details of the system's implementation, are considered proprietary by the Company (key parts of the system are copyrighted). In preparing this paper it has been necessary to balance those interests against the goal of providing readers of the paper with useful insight into the structure, capabilities, and uses of the system.

The paper has four major sections:

- Section 1 provides an introduction and overview.
- Section 2 begins by describing the risk framework that was developed to guide the development of the model. The various types of risks facing an insurer are outlined and defined. The approach taken to evaluate the tradeoff between risk and return is then described.
- In Section 3, the structure of the dynamic financial analysis model is presented. This includes a system schematic and a description of the various inputs, variables, and calculation steps.
- Finally, in Section 4 the uses of the model are described and the output is illustrated.

Two appendices are included. The first provides a discussion of currency risk, which is present on both the asset and the liability side of the multinational insurer's balance sheet. The second provides a brief description of the expected policyholder deficit, a concept that is particularly relevant to the measurement of insurer risk and to the management of capital.

2. CONCEPTUAL FRAMEWORK

A necessary first step in the development of a dynamic financial model is establishing a conceptual framework to serve as a guide. The structure of the risks to be modeled must be defined in general, and then prioritized on the basis of the business profile of the company. Appropriate measures of risk must also be defined, and threshold values for the risk measures must be chosen.

2.1. Classification of Risk

The risks faced by an insurance enterprise have been classified in a variety of ways in the published literature on the subject. For example, see HARTMAN, et. al. (1992). There are three basic elements of risk, each of which must be considered in a dynamic financial analysis model. The three basic elements are:

1. Liability Risk: the risk that the cost of settling the insurance liabilities will be greater than expected (also referred to as obligation risk).
   - Claims on coverage already provided cost more to settle than anticipated.
   - Cost of claims generated on future coverage is greater than anticipated.
2. Asset Risk: the risk that the realizable value of assets will be less than anticipated.
   - The market value of invested assets declines.
   - Invested assets become non-performing.
   - Receivables from outward reinsurers become uncollectable.
   - Receivables from customers become uncollectable.
3. Business Risk: the general business risks faced by all enterprises.
   - Competitors will force market prices below costs to preserve their position/share.
   - Competitors will gain a competitive advantage, taking customers away.

- Regulators or legislators will interfere in the market in a harmful way.
- The company will be victimized by a crime.
- Operations will be adversely affected by a disaster at company premises.

The bullet points above are intended to be illustrative of the types of risks included in each element; these lists are not necessarily exhaustive.

As will be seen, the Company's dynamic financial analysis model is structured around this risk framework, explicitly incorporating each of these three major risk elements.

2.2. Liability Risk

Liability risk (or obligation risk) is viewed as the predominant risk element by most property/casualty insurers. As indicated, it includes existing claim obligations (whether known or not) on coverage provided in the past, as well as new claim obligations arising from future coverage provided on policies currently in force or written in the future. From the perspective of the actuary, liability risk includes what may loosely be referred to as *reserving* and *pricing risk*. It is the actuary's responsibility to estimate the cost of claims in each of the two contexts. Liability risk stems from the uncertainty of those estimates.

In the definition of liability risk, cost is expressed in terms of present value. Liability risk includes the timing of the claim cash flows, as well as their nominal amounts. It also includes the expenses of settling the claims, as well as the claim payments themselves.

Uncertainty of liabilities includes both process risk, which arises from the random nature of claim events, and parameter risk, which arises from the inability to know the claim frequency and severity distributions from which the events are drawn. These distributions cannot be known in advance, because they are dependent on future social and economic conditions that cannot be predicted with certainty.

For most lines of insurance, a company can write sufficient volumes of business to diversify away process risk. In these cases parameter risk will be the dominant component of liability risk, with process risk considered de minimis. However, in property catastrophe reinsurance process risk is not diversifiable by volume. Even on a world-wide market basis the covered events are too few to achieve a stable annual result. (We will have to wait for the market to expand to include a few other worlds beyond earth to achieve diversification by volume.) For this line, both process and parameter risk must be accommodated in a dynamic financial analysis model.

Finally, a complicating factor for an international insurer is the issue of currency. Insurance contracts are typically issued with claims to be settled in a specific currency, typically the local currency of the contract. However, from the perspective of the owner, claim costs are ultimately measured by their impact on equity as measured in the owner's currency. Thus the cost of liabilities includes the cost of converting them from the local contract currency to the owner's currency, and liability risk includes movements in exchange rates that affect conversion costs.

2.3. Asset Risk

By definition, assets are capable of generating an expected positive cash flow. The positive cash flow may be contractual (e.g., a bond), or may stem from the potential sale value in the market (e.g., home office real estate). Asset risk deals with the uncertainty associated with the realization of the cash flow. This uncertainty stems from two fundamental sources. One is the risk of non-performance of the obligor, such as the default of a bond or the insolvency of a reinsurer. The other is a change in conditions that affects the value or performance of the asset. Examples of the latter would include a recession causing a decline in the stock market, or a rise in mortgage interest rates that lowers the rate of refinancing on a Collateralized Mortgage Obligation.

The inclusion of reinsurance recoverables with asset risk aligns the risk classification structure with contemporary GAAP thinking, and not with traditional U.S. statutory accounting where the financial presentation suggests that obligation risk be measured on a net basis.

As is the case with liabilities, much of the risk associated with individual assets is diversifiable. Thus the movement of individual stock prices or the default of individual bonds is not usually relevant to asset risk, unless the individual holding is material. Instead the primary focus is on the non-diversifiable components of risk associated with each asset class.

Asset risk also has a currency dimension. To the extent that assets are held in currencies different from that of the owner, changes in exchange rates contribute to asset risk. The influence of currency on asset and liability risks is discussed more fully in APPENDIX A.

2.4. Business Risk

General business risk has been given relatively little attention in the actuarial literature. This is unfortunate, because it is a significant source of risk in insurance. Business risk contributes significantly to underwriting risk in ways that cannot be described by simple random processes. Severe underwriting losses at the bottom of the U.S. property/casualty underwriting cycle are neither random nor unforeseen events. They aren't caused by claim costs being higher than expected (i.e., by liability risk), but rather by market price levels being set below the level of expected costs. During a down-cycle many companies are aware that their prices are too low and that underwriting results will be poor.

A variety of forces acts on price levels in the insurance marketplace, most notably the level of overall capacity in relation to demand. Prices will fall when capacity exceeds demand, and will rebound only when capacity is withdrawn. The operation of these forces depends on the structure of the market and external conditions at the time. External economic conditions can play a reinforcing role, particularly such factors as the level of interest rates.

Competitive position is also important to the business risk of individual companies operating within the market. One example would be the cost of distribution. Companies with a high-cost distribution system should not expect to achieve adequate returns, unless that distribution system offers enough value to them or their customers to warrant its excess cost. In a competitive market, the companies with the lower distribution costs will simply set the market price at a level that produces sub-par returns for their high-cost competitors.

Competitive advantage is not only about distribution costs. It includes the effectiveness of the company's marketing, underwriting, claim, and capital management functions. While the overall industry results over the last few years have generally been lackluster, many individual companies have produced attractive returns during this period by superior execution in one or more of the above areas. Conversely, the disappearance of several national multiline companies over the same period can be attributed to their inability to perform successfully in these areas. Competitive risks are both significant and real in this industry.

Business risks arising from market competition are not at all unique to insurance. One only has to look as far as the U.S. airline industry to witness the same risks playing themselves out in a non-insurance context. There, too, an excess of capacity in relation to demand has forced a blood-letting as competitors vied to retain market share. Airline managements knew that fares were inadequate, but market forces were beyond their control.

From a dynamic financial analysis perspective, the authors believe that underwriting risk must be broken down into business risk and liability risk components, with each component modeled separately. While the two types of risk are not entirely unrelated, the drivers of each are different. Modeling them as a single risk (i.e., modeling underwriting risk via loss ratios) is therefore an inherently weak approach.

2.5. Measuring Risk and Return

Application of dynamic financial analysis requires that financial constraints be defined. For example, while the results of an analysis might indicate that there is an x% probability of impairment, defined as the loss of y% or more of capital, those results alone do not tell management what actions to take. To translate analysis results into action, management (or the board of directors) must decide whether or not the indicated level of impairment probability is too high. Also, while impairment probability might be an appropriate constraint, it is probably not the only constraint relevant to the enterprise. In fact, a variety of constraints are relevant, depending on the question the analysis is designed to answer.

Dynamic financial analysis also requires the definition of financial performance objectives. If a reinsurance program were offered to the company that reduced its probability of impairment from x_1% to x_2%, management can only judge the benefit of that reduction in relation to the cost of the reinsurance. This issue becomes particularly relevant when there are several alternative reinsurance programs, each with different ruin reductions and different costs. The issue is further complicated when the cost of a particular program is variable, or when its effects are spread into several future accounting periods in a multi-year deal.

In developing its dynamic financial analysis model, the Company has adopted the Asset/Liability Efficient Frontier (ALEF^{SM}) as a basic framework for resolving these issues in a logically consistent manner.[1] (Additional discussion of ALEF can be found in BUFF (1990) and DOLL, et. al. (1994).)

The efficient frontier concept is taken from modern portfolio theory, and is attributed to MARKOWITZ (1959). In its most basic formulation, the investor is presented with several alternative classes of assets in which he can invest. For each class of asset, the investor knows the expected return, the risk associated with that return (as measured by its standard deviation), and the correlation of returns with all other classes of asset. His problem is to choose a portfolio by specifying the mix of assets by class. Markowitz's contribution was to recognize that not all asset mixes are optimal: alternative mixes can be found for which either a higher return can be achieved for the same level of risk, or the same return can be

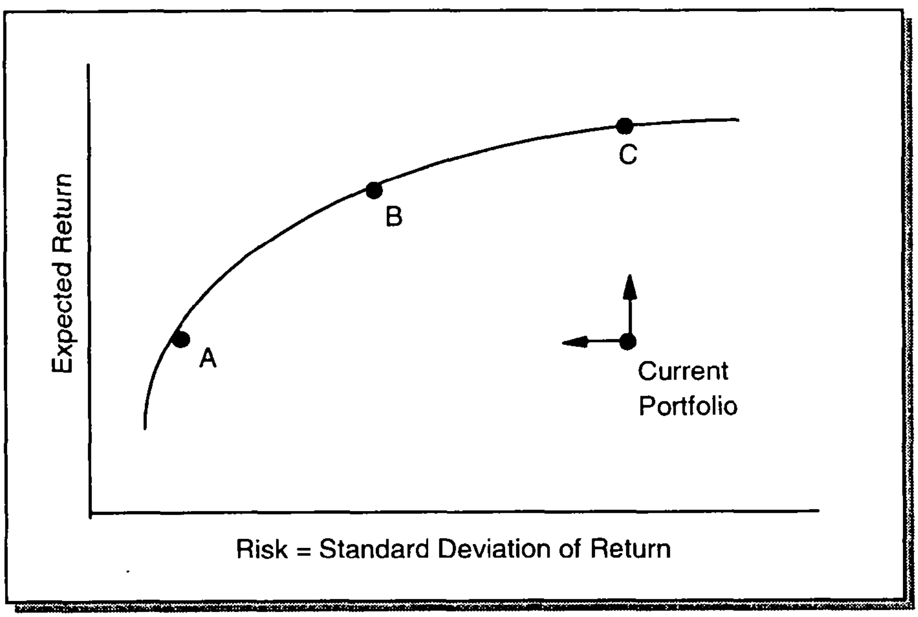


FIGURE 1: In the classical efficient frontier of Modern Portfolio Theory, asset mixes A, B, and C are efficient; the asset mix of the current portfolio is not.

[1] ALEF is a registered service mark of Tillinghast – Towers Perrin.

achieved for a lower risk. There is, however, a frontier to the set of possible asset mixes consisting of those portfolios that are efficient in the sense that one cannot improve upon them. Figure 1 illustrates these concepts.

The investment portfolios on the efficient frontier are all good choices; choosing among them is a matter of the investor's risk/return preferences.

ALEF is a generalization of the efficient frontier to the optimization of more general business strategies. The definition of both of the two axes in the chart above are generalized. In the ALEF approach the X-axis is labeled generically as "level of risk" and the Y-axis is labeled generically as "expected performance".

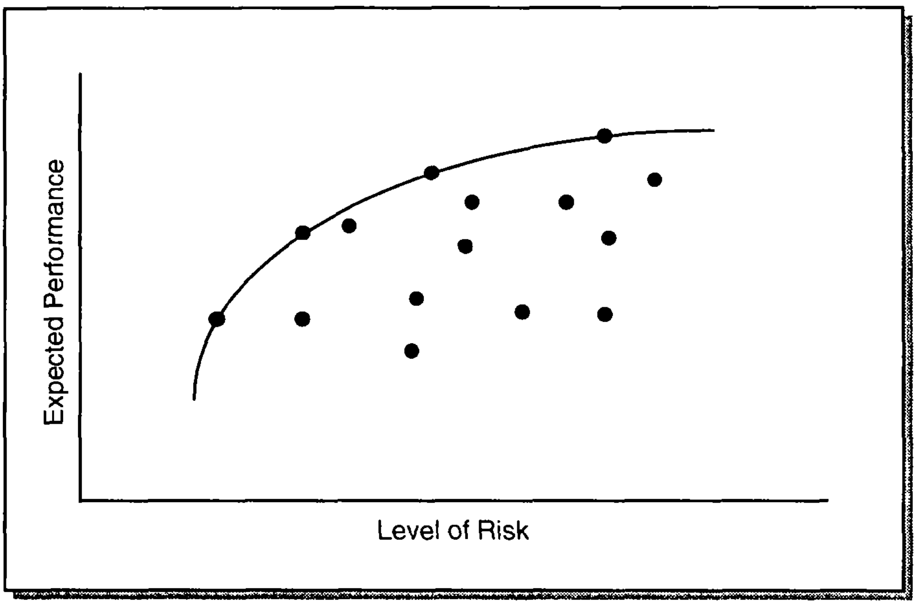


FIGURE 2: Using the Asset/Liability Efficient Frontier, Strategies can be evaluated in a generalized risk/reward framework.

The user must define each of these terms. Similarly, the strategies to be analyzed are generalized from asset mix to any set of decision variables relevant to the enterprise. Once the problem is specified in these terms, the dynamic financial analysis model can be used to find the efficient frontier from the available choices.

In contrast to the classical efficient frontier objective, in which performance is measured exclusively by single-period economic returns, the ALEF performance objective can be any financial or economic measure that management believes is most important, or any combination of such measures. Generally, the measure should be consistent with the maximization of shareholder value, but it can be reflective of any specific component (such as reported profits, change in statutory surplus, or revenue growth). In the case of multiple measures, management must specify the relative weight assigned to each so that they can be combined into a single index. (The function combining the measures need not be linear.) The measures can be based on economic or accounting values, since both are relevant to the operation of the enterprise. The measures can be expressed in terms of absolute dollars, returns in relation to capital employed, or relative performance when compared to peers. Finally, the measures can reflect any chosen time horizon.

The only overriding requirement of the performance measure used is that it must be consistent. Management must always want to choose the strategy that maximizes the measure's expected value, all other things being equal.

As a measure of risk, standard deviation has been the subject of ample criticism. Much of this criticism stems from the fact that standard deviation focuses only on the dispersion of the outcomes, without any special recognition of the greater disutility of the adverse outcomes. While most people equate risk with uncertainty of outcomes, they also equate risk with the likelihood and severity of adverse outcomes. In the ALEF framework, risk can be any measure of adverse outcomes that management feels is most relevant. Examples would include:

- Probability of ruin over the next ten years
- Probability of combined ratio above 110% next year
- Expected policyholder deficit[1] on current business
- Probability of suffering a net decline in surplus of 20% or more in three years
- Probability of failing an RBC test at any point in the next five years
- Probability of a ratings downgrade by A.M. Best
- Probability of a combined ratio two points or more worse than the industry average
- Probability of revenues being 25% or more below plan.

As was the case with the measure of performance, several different measures of risk can be combined to produce an overall index of risk, with weights reflective of their relative importance. Figure 2 illustrates the generalized ALEF framework.

ALEF is a powerful and flexible tool for managing an insurance company. It can be customized to mirror the business philosophy of the company, both the financial objectives to be maximized and the risks to be controlled.

The Company uses the ALEF framework in conjunction with its dynamic financial analysis model to evaluate a variety of strategic issues. The Company has developed a vector of multiple risk constraints that collectively capture its appetite for risk. This vector is used consistently in each analysis. While the types of strategic issues analyzed are discussed in subsequent sections of the paper, the Company considers its risk constraint vector to be confidential.

3. DESCRIPTION OF THE MODEL

A conceptual schematic of the Company's dynamic financial analysis model is presented in Figure 3. The model consists of the following basic components:

[1] Appendix B provides a description of the expected policyholder deficit and discusses its application in this context.

- A liability scenario generator, which produces distributions of aggregate underwriting results for the insurance portfolio.
- An asset scenario generator which, when combined with the liability generator, produces a distribution of operating results for the combined insurance/investment portfolio.
- A multi-period financial model, which extends the distributions over a longer time horizon.

Each of these components produces dynamic output that is used to manage different aspects of the business.

As can be seen from the schematic, the model is not a single system, but a linked set of programs and databases that can be used in a variety of combinations to facilitate the needs of any given analysis. A key attribute of this structure is flexibility. While the core calculation engines are written in high-order programming languages to achieve efficiency, many of the inputs and outputs of each component are held in spreadsheets to facilitate their manipulation "on the fly" by the user. The spreadsheets also facilitate the creation of graphical output for analysis of results.

3.1. Liability Scenario Generator

Because the Company's core business is property catastrophe reinsurance, a heavy emphasis is placed on detailed modeling of the volatile claim experience inherent to that line. The models are used extensively in the underwriting of individual contracts. In the context of this paper, however, the focus of presentation is on their use as an input to the enterprise-level DFA model. The advantage of this tightly integrated approach is that the effect of any one underwriting decision on the key corporate DFA objective functions can be easily determined by the underwriter, and therefore taken into account at the point of decision in the underwriting process.

For each peril in each region of the world a set of catastrophic events has been developed. The events vary according to their location, size, and intensity as well as to the ensuing insured damage they would generate. Relative probabilities are also assigned to each event in the set, based on the likelihood of that particular combination of event parameters occurring at once. The probabilities for each set of events sum to one. In conjunction with the insured losses associated with each event, they represent a sample severity distribution for the particular peril. Similarly, for each peril in each region a frequency distribution is specified, reflecting the likelihood of a given number of events happening within a year. For example, a frequency distribution is specified for the number of landfall hurricanes hitting the U.S. over the course of a season.

Within the system, the frequency and severity distributions for each peril are convoluted to produce annual aggregate catastrophe losses. In the current configuration, 40,000 scenarios of annual losses are created, which are deemed sufficient for analysis purposes. (The sampling process is stratified, not Monte Carlo, so that the tails of the resulting aggregate distribution are considerably more robust.)

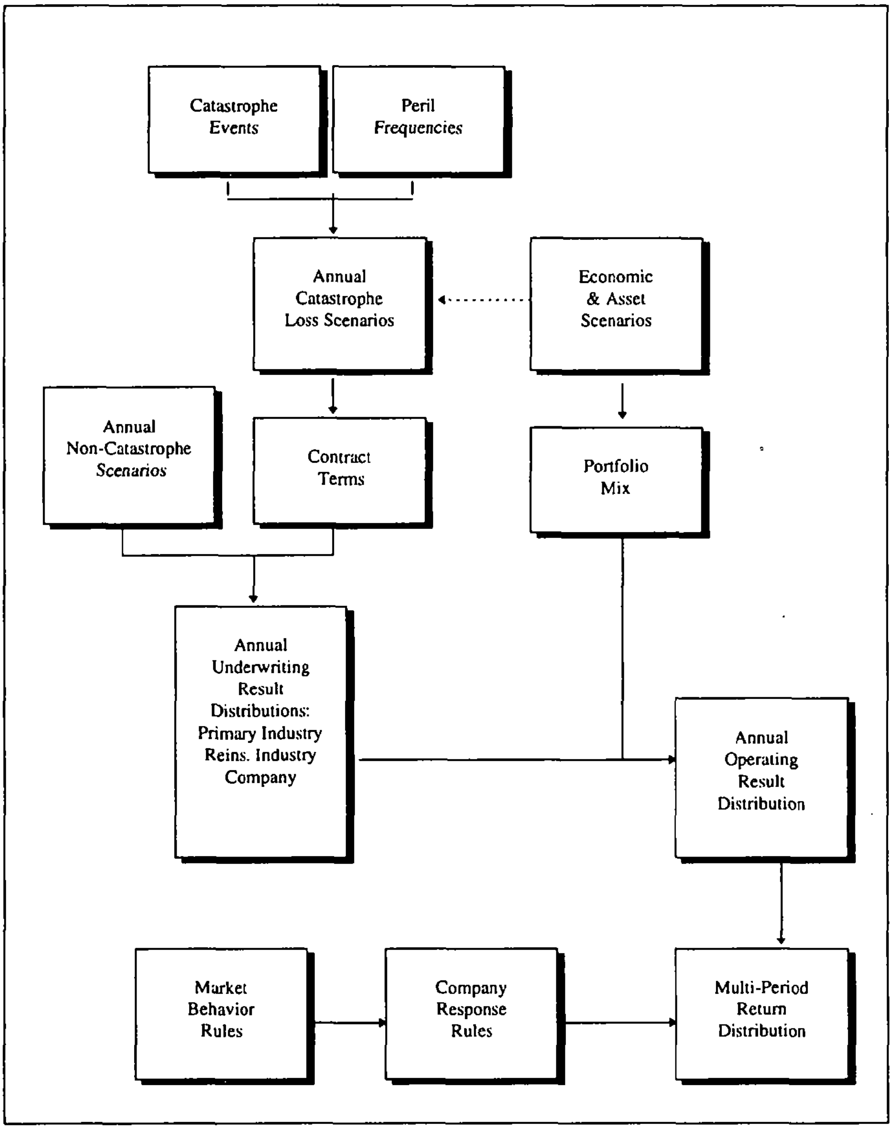


FIGURE 3: Conceptual Schematic of the Dynamic Financial Analysis Model

At this juncture in the system, the losses in each scenario are those of the primary ceding company. The primary losses are then run through the applicable reinsurance contract terms to obtain the corresponding losses to the reinsurance contract. A database containing the actual terms of all catastrophe reinsurance contracts in the portfolio is maintained, so that world-wide aggregate underwriting results for the entire portfolio for each scenario can be obtained and analyzed. The system is on-line, so that portfolio results can be obtained at any time. A complete portfolio run takes about two hours to process through the system on a Silicon Graphics workstation. Analysis of the marginal impact of adding a contract to the portfolio takes less than five minutes. In addition to ongoing ad hoc portfolio analysis, portfolio results are produced and analyzed in detail in a formal underwriting meeting each quarter, after the latest cycle of contracts have been written.

The Company writes small amounts of other types of reinsurance from time to time, which are incorporated into the system using a less formal modeling approach. A spreadsheet containing the estimated underwriting distributions applicable to this business is maintained, and is incorporated into the overall results as a "last step" in the overall process. This assures that the complete underwriting portfolio is modeled within the system.

The principal output of this component of the system is a distribution of underwriting results for the Company. The distribution reflects all elements in the underwriting result that vary directly with losses: reinstatement premiums, losses, brokerage, and federal excise taxes/premium taxes.

These elements are calculated on a contract-by-contract basis, reflecting the actual applicable terms and conditions. Other elements such as operating expenses may be added as a last step in the process.

In addition to Company underwriting experience, supplemental industry-wide information is produced showing the corresponding losses for the primary industry and the estimated portion of those losses that would be ceded to the property catastrophe reinsurance industry.

Since the Company's functional currency is the U.S. dollar, all transactions relating to contracts involving other currencies are converted to their U.S. equivalent. Within the system, exchange rates can be varied to test the impact of adverse movements on underwriting results.

Each of the underlying catastrophe events has an associated day of the year. Thus, each underwriting scenario generated by the model has a pattern of losses throughout the year. At the present time, the models do not consider the variability in the timing from event occurrence to claim payment. Such risk is considered fairly immaterial. Neither is there any consideration of "reserving risk", in the sense that actual payments might be greater than estimated in the financial statements.

Parameter risk is not explicitly included within the modeling process itself. Instead, the parameters are sensitivity-tested in a variety of ways and the results are used to introduce conservatism into the final parameter assumptions. These sensitivity tests take two forms:

- First, output can be generated using event files created by different vendors. In addition to developing its own event files for various perils and regions, the Company has developed relationships with many of the primary catastrophe modeling consultants, including Applied Insurance Research, RMS, Dames & Moore, EQECAT, and Tillinghast – Towers Perrin. Event files have been constructed and incorporated into the system using the catastrophe models developed by these firms. Comparing the results generated by these different event files, reflecting the different approaches and assumptions of each firm, provides a measure of the impact of varying the underlying event parameters, and helps to assure that the results obtained are not dependent on the specific catastrophe model used.
- Second, sensitivity testing is performed by altering the underlying frequency and severity distributions. Results are routinely tested using higher peril frequencies. This is particularly relevant in light of the research being done by global climatologists (such as that published by GRAY (1990) and popularized in the media), and the record level of hurricane activity experienced in 1995. The generated peril severity distributions have also been adjusted to consider various factors such as the demand-driven inflation that occurred after hurricane Andrew.

Finally, results can be produced for the entire portfolio of reinsurance contracts or any defined subset. This facilitates analysis of sources of risk, and also can be used to analyze the value of potential retrocessions. Hypothetical portfolios can be run to test alternative underwriting strategies as well.

3.2. Asset Scenario Generator

The Company uses the Global CAP: Link system to obtain scenarios for various economic and investment variables for several different currencies. On request, a CAP:Link output file containing 1,000 scenarios is provided to the Company, with each scenario reflecting a future path of interest rates, inflation rates, currency exchange rates, and rates of return by asset class for each of five major currencies. Each scenario is a plausible path of the annual movement of the variables; taken together the scenarios describe the range of variation in each of the variables.

The CAP: Link system uses a stochastic diffusion model to generate economic and capital market scenarios on a global basis. Scenarios are generated on the basis of a cascading set of stochastic differential equations, structured so that the proper relationship between the modeled variables is maintained over time. These include serial correlation effects, reinvestment risks, and path volatility characteristics. The top of the cascade is a yield curve scenario generator, based on a variant of the two-factor yield model proposed by BRENNAN and SCHWARTZ (1982). These yield results are then passed down to generators for other variables such as inflation and stock returns, which are conditionally related in the cascade. The developers of the CAP:Link system believe that it is superior to other popular approaches such as lognormal models, time series models based on ARIMA or

Box-Jenkens, or models based on Vector AutoRegression. A more detailed description of the stochastic diffusion model, and a discussion of its performance relative to other models can be found in MULVEY and THORLACIUS (Forthcoming in 1997).

The asset scenarios from CAP:Link are convoluted with the liability scenarios. Each individual annual scenario consists of:

- Economic conditions: annual inflation rates by currency and exchange rate movements for the year
- Capital market conditions: interest rates and annual rates of return by asset class and currency
- Catastrophic conditions: a set of catastrophic events and primary and reinsurance industry losses ensuing from those events.

The Company underwriting result distribution is combined with investment results reflecting the cash flows and investment returns for each scenario, so that an annual operating result distribution for the Company can be obtained. Note that both the liabilities and the assets are dynamically adjusted for changes in exchange rates. The operating result distribution can be produced either for the current mix of investments, or for any hypothetical alternative mix (as well as for different insurance portfolios). This facilitates the testing of alternative investment portfolio strategies, including the mix of investments by currency.

At the time of writing, the catastrophe losses at the detailed scenario level are not dynamically linked directly to the economic scenarios (hence the dotted line in the schematic diagram). This is an enhancement that is currently under development. Once it is completed the losses will vary according to the inflation rates in each scenario.

3.3. Multi-Period Model

Up to this point, the description of the model has focused on the short-term, annual time horizon. The liability and asset legs of the model focus on annualized results in the context of the current business environment. The multi-period model extends the analysis to a longer-term horizon (currently five years) and introduces key elements of business risk into the analysis. Underwriting results in future periods will be influenced by loss experience (liability risk) and market price levels (business risk).

The first step in this process is to encapsulate the behavior of the market in a set of rules. The critical question is how market price levels will move over the five-year time horizon, and what factors will affect that movement. In this area the Company has an advantage over the large multiline insurers, for whom this would be a vast and daunting question. Such insurers would need to specify the market behavior and drivers for each product line they offer in each market, as well as the interrelationships across the different product lines and markets. In the Company's case only one product line and market, property catastrophe reinsurance, must be addressed.

The fundamental behavior of prices in the property catastrophe reinsurance market can be stated succinctly.

- If results are good, prices will decline from their current level.
- Prices will continue to decline until results are poor, at which point they will rise.
- The rate of decline is related to how good the results are; the rate of increase is related to how poor results are.
- Rises in prices include nominal increases in rates-on-line, and also implicit increases through higher retentions and other coverage reductions.

Since the market has exhibited this general behavior over an extended period, it is reasonable to assume the behavior will continue. The difficult part of the problem is translating the qualitative behavior rules into quantitative terms. While the historical responsiveness of prices to results can serve as a guide, changes in the market's structure that influence its behavior must also be considered. For example, one could argue that the new capital provided to the reinsurers in Bermuda may be less forgiving, and will be withdrawn more rapidly, if and when results are bad. Similarly, the growing use of catastrophe models by the reinsurers in underwriting may inject a greater degree of discipline, reducing the rate of price decline in the face of favorable results.

The approach taken by the Company is to relate catastrophe reinsurance price levels in each subsequent year to the industry-wide catastrophe experience in several preceding years. A market price index has been constructed, the movement of which is dependent on emerging industry experience. The market price index is based on information from several sources: the actual price movements observed by the Company since its formation; historical price movements over a longer time period, derived from information from several sources; discussions with brokers and other experts in the market; and judgment.

The responsiveness of price levels to experience over several years involves significant parameter risk. The Company has performed in-depth sensitivity testing of this element of the model to gain insight into how alternative assumptions influence results.

The starting point in the multi-period simulation is the current distribution of annual underwriting results. Using a Monte Carlo approach, a first-year scenario with the associated underwriting result for the Company is chosen from that distribution. On the basis of the corresponding industry-wide result, the movement in the price level index for year two is determined. The annual underwriting result distribution is then modified to reflect the effect of the change in price level to obtain a distribution for the second year. A second-year result is then chosen from the modified underwriting result distribution. This stochastic process continues until five years of results have been generated.

In addition to the market behavior rules, company response rules reflecting the actions of Company management must also be defined. These actions fall in three areas.

- Market share actions must be defined, reflecting the Company's willingness to write business at the prevailing price level. Depending on the perceived adequacy of prices, the Company will either seek to grow, hold steady, decrease, or severely reduce its market share. This decision feeds back into Company results as follows: the price level on the Company's portfolio relative to the market price level improves/degrades as the Company's market share declines/grows, due to more/less selective underwriting.

- Capitalization actions must be specified, reflecting the changing needs of the Company over time. For example, at some threshold level a portion of excess capital is returned to shareholders. Similarly, if actual capital falls below specified requirements, market share is forced down to the level allowed by the requirements. Both normal and extraordinary dividend policies must be defined.
- Debt/capital levels over the five-year period must be specified, and debt actions in relation to operating losses must be defined.

The multi-period model starts with an opening balance sheet, simulates the underwriting result for the first-year, translates that result into a first-year operating result, determines the market behavior for the next year, and implements the company responses. This process continues iteratively until the full five years have been generated. Typically, 20,000 trials are run to produce a distribution of five-year returns to shareholders, which is based on the stream of dividends and the final equity at the end of the fifth year. In addition to return measures, appropriate risk measures are also generated. The model can be run using different company response strategies; the risk and return associated with each strategy can be compared by placing it in an ALEF context.

The multi-period model successfully captures the liability and business risk elements which, taken together, comprise underwriting risk for a property catastrophe reinsurer. Other types of business risk, such as regulatory interference or fraud, are not directly incorporated into the model.

4. MODEL USES AND SAMPLE OUTPUT

One of the key advantages of a highly integrated system such as the one described is that many different types of decisions can be tested against a consistent risk/return "yardstick", which is based on a common set of underlying probability distribution assumptions. These include:

- Ongoing evaluation of the adequacy of capital to support the current risks undertaken
- Evaluation of the value of retrocessional coverage offerings
- Analysis of alternative capital structures
- Development of asset mix investment policy
- Analysis of currency risk
- Studies of alternative market and underwriting strategies
- Individual underwriting decisions reflecting the marginal effect of a given contract on risk and return constraints.

Exhibit 1 is an example of output from the liability scenario generator. It shows graphically the right-hand tail of an underwriting result distribution for a

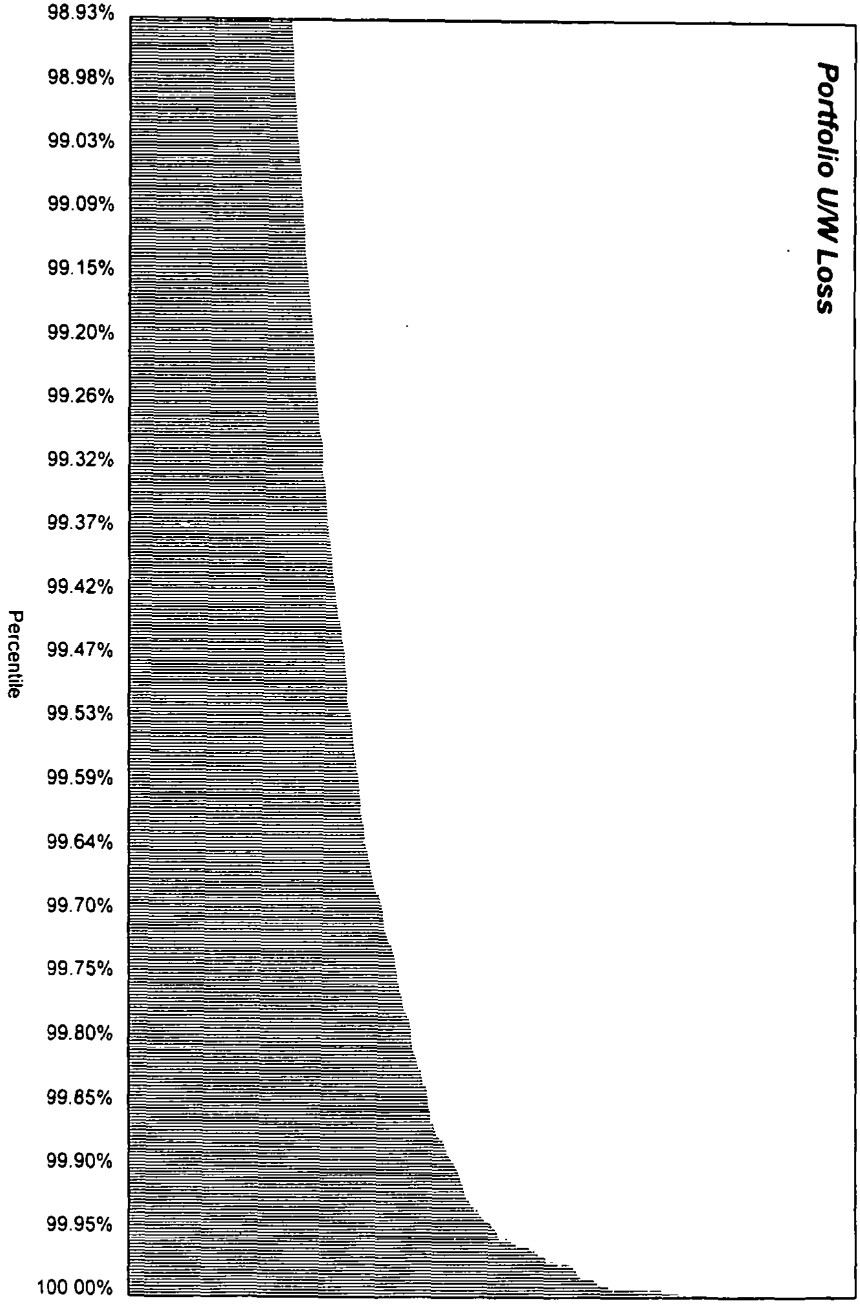

Portfolio U/W Loss
Percentile
98.93%
98.98%
99.03%
99.09%
99.15%
99.20%
99.26%
99.32%
99.37%
99.42%
99.47%
99.53%
99.59%
99.64%
99.70%
99.75%
99.80%
99.85%
99.90%
99.95%
100 00%

EXHIBIT 1

Worldwide Portfolio as of xx,xx,xxxx
Based on Peril/Assumption Set 23

Description	Return Interval	Peril	Industry Loss	Portfolio Gross Loss
			(billion)	*(million)*
Northeast/NY	1/xxx	H	xxx	xxx
Northeast/MA	1/xxx	H	xxx	xxx
[illegible]	1/xxx	H	xxx	xxx
	1/xxx	H	xxx	xxx
	1/xxx	H	xxx	xxx
	[illegible]	H	xxx	xxx
		E	xxx	xxx
			xxx	xxx
			[illegible]	xxx
				xxx
Gulf/[illegible]				
Gulf/LA				
California				
South CA				
North CA	1/xxx			
South CA	1/xxx			
North CA	1/xxx	E		
New Madrid/TN	1/xxx	E	[illegible]	
New Madrid/MO	1/xxx	E	xxx	
Hawaii	1/xxx	H	xxx	xxx
Northwest/WA	1/xxx	E	xxx	xxx
N Europe/UK	1/xxx	W	xxx	xxx
N Europe/UK	1/xxx	W	xxx	xxx
N Europe/GER	1/xxx	W	xxx	xxx
N Europe/GER	1/xxx	W	xxx	xxx
Japan Wind	1/xxx	H	xxx	xxx
Japan Wind	1/xxx	H	xxx	xxx
Japan Quake	1/xxx	E	xxx	xxx
Japan Quake	1/xxx	E	xxx	xxx
Australia	1/xxx	H	xxx	xxx
New Zealand	1/xxx	E	xxx	xxx
Caribbean	1/xxx	H	xxx	xxx
Northridge/CA	1/xxx	E	xxx	xxx
Loma Prieta/CA	1/xxx	E	xxx	xxx
Great NE Hurricane/NY	1/xxx	H	xxx	xxx
Andrew/FL	1/xxx	H	xxx	xxx
Hugo/SC	1/xxx	H	xxx	xxx
90A - Daria/Europe	1/xxx	W	xxx	xxx
90G - Vivian/Europe	1/xxx	W	xxx	xxx
90D - Herta/Europe	1/xxx	W	xxx	xxx
87J - StormV/Europe	1/xxx	W	xxx	xxx
76B - Capella/Europe	1/xxx	W	xxx	xxx
Mirielle/Japan	1/xxx	H	xxx	xxx
New Castle/Australia	1/xxx	E	xxx	xxx

EXHIBIT 2

US Primary and Retro Business
Inforce Date - xx,xx,xxxx

Occurrence Distribution

Zone of Event	Occurrence Probabilities					
	90.0%	95.0%	99.0%	99.6%	99.8%	99.9%
Eastern New England	xxx	xxx	xxx	xxx	xxx	xxx
Eastern LI & CT	xxx	xxx	xxx	xxx	xxx	xxx
...thern NJ	xxx	xxx	xxx	xxx	xxx	xxx
	xxx	xxx	xxx	xxx	xxx	xxx
	xxx	xxx	xxx	xxx	xxx	xxx
	xxx	xxx	xxx	xxx	xxx	xxx
		xxx	xxx	xxx	xxx	xxx
			xxx	xxx	xxx	xxx
				xxx	xxx	xxx
					xxx	xxx
						xxx
Northern Pla...						
Northwest						
Northern California						
Southern California						
Hawaii	xxx					
National Occurrence	xxx	xxx	xxx			
Wind Storms	xxx	xxx	xxx	xxx	xxx	xxx
Earthquake	xxx	xxx	xxx	xxx	xxx	xxx
National Aggregate	xxx	xxx	xxx	xxx	xxx	xxx

Share of Market Loss

Zone of Event	*Primary Industry* Market Share					
	90.0%	95.0%	99.0%	99.6%	99.8%	99.9%
Eastern New England	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%
...rn LI & CT	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%
...NJ	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%
	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%
	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%
	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%
		x.x%	x.x%	x.x%	x.x%	x.x%
			x.x%	x.x%	x.x%	x.x%
				x.x%	x.x%	x.x%
					x.x%	x.x%
						x.x%
Torna...						
Northern Plains & C...						
Northwest						
Northern California						
Southern California						
Hawaii	x.x%					
National Occurrence	x.x%	x.x%	x.x%			
Hurricane	x.x%	x.x%	x.x%	x.x%	x.x...	
Wind Storms	x.x%	x.x%	x.x%	x.x%	x.x%	x.x...
Earthquake	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%

EXHIBIT 3

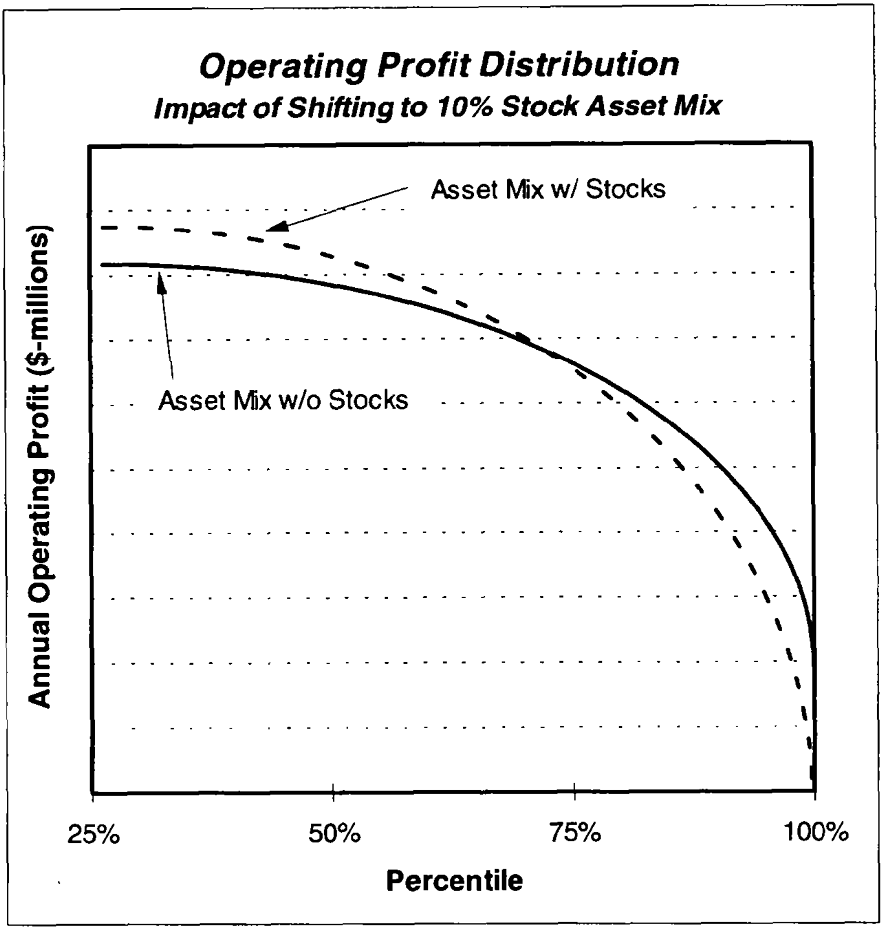


EXHIBIT 4

portfolio. As indicated previously, this information (along with accompanying risk and return statistics) can be generated for any vendor/peril scenario, and any segment of the portfolio of reinsurance contracts.

Exhibit 1 is a relatively simple graphic, but when it is coupled with the risk/return measures it is a powerful management tool. For example, distributions can be generated with and without a retrocessional cover that is being considered. Comparison of the two allows management to evaluate the marginal impact of the cover on underwriting risk and return, and ultimately to assess the value of the cover. Alternatively, reinsurance accounts that have a particularly detrimental impact on the distribution can be isolated for potential re-underwriting at renewal.

Management can also track changes in the distribution over time, as a measure of underwriting performance.

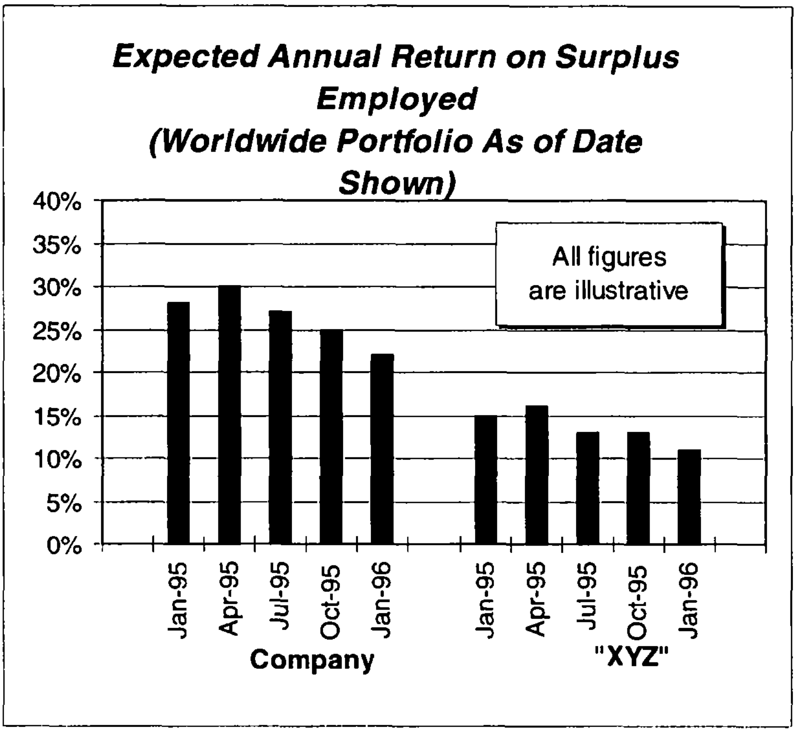


EXHIBIT 5

A variety of routinely produced diagnostic exhibits allows management to gain insight into the sources of adverse underwriting scenarios: perils, regions, reinsurance layers, etc. Comparative information on primary and reinsurance industry losses is also included. Exhibits 2 and 3 are illustrative of these types of exhibits.

Exhibit 2 displays industry and portfolio experience on a standard defined-event set. The defined events reflect a range of different likelihoods for various perils and regions. (The "break" in the exhibit indicates that it is longer than actually shown; only the beginning and end of the exhibit is shown in the illustration.) Exhibit 3 displays percentiles of severity distributions for the portfolio by (illustrative) geographic zone, and the Company' s share of the industry loss at that percentile.

In addition to underwriting profit distributions, operating profit distributions reflecting investment as well as underwriting risk are produced by the model, such as those shown on Exhibit 4. These can be used to translate underwriting risk into operating profit terms, or to test the effect of introducing various levels of asset risk via changes to the mix of investments.

Multiperiod Financial Planning Model
Expected Operating Performance by Strategy
Baseline Market Behavior Assumption

Operating Leverage: 50% | Debt/Capital Ratio: 0% | Company Response To Market: Modest

	Year 1	Year 2	Year 3	Year 4	Year 5
Written Premium	191	207	209	210	170
Net Operating Profit	120	129	126	124	98
Dividends	16	60	73	176	
Surplus	381	486	555	608	556

Operating Leverage: 50% | Debt/Capital Ratio: 30% | Company Response To Market: Modest

	Year 1	Year 2	Year 3	Year 4	Year 5
Written Premium	190	207	211	214	180
Net Operating Profit	110	114	110	107	84
Dividends	35	56	64	129	
Surplus	381	489	572	638	608

Operating Leverage: 65% | Debt/Capital Ratio: 30% | Company Response To Market: Modest

	Year 1	Year 2	Year 3	Year 4	Year 5
Written Premium	267	290	295	302	270
Net Operating Profit	153	158	152	148	123
Dividends	70	94	87	131	
Surplus	381	501	594	687	714

Operating Leverage: 80% | Debt/Capital Ratio: 30% | Company Response To Market: Modest

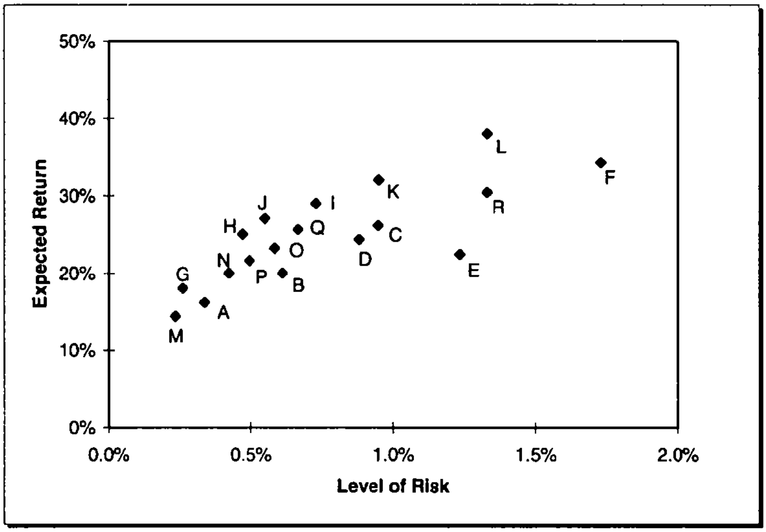
	Year 1	Year 2	Year 3	Year 4	Year 5
Written Premium	305	332	332	333	270
Net Operating Profit	177	181	171	158	138
Dividends	89	113	111	131	
Surplus	381	509	607	708	743

EXHIBIT 6

Many of the risk measures suggested in Section 2 can be translated into *boundary constraints*, reflecting their maximum level of acceptability. For example, one possible risk measure is the probability of suffering a surplus decline of 20% or more. If that were a chosen risk measure, management would presumably seek to minimize that probability for a given level of return, and would only be willing to accept an increase in that probability in exchange for a higher return. Management might also impose a boundary constraint that in no event will management allow that probability to exceed 3%.

One can invert the boundary constraint relationship to obtain an implied surplus requirement. For example, if the current annual operating profit distribution for a hypothetical company indicates that there is a 3% chance of

Multiperiod Financial Planning Model
Asset Liability Efficient Frontier
Baseline Market Behavior



Strategy	Operating Leverage	Debt/ Capital	Dividend Policy	Response to Market
A	85%	20%	Standard	Level
B	100%	20%	Standard	Level
C	115%	20%	Standard	Level
D	85%	40%	Standard	Level
E	100%	40%	Standard	Level
F	115%	40%	Standard	Level
G	85%	20%	Standard	Modest
H	100%	20%	Standard	Modest
I	115%	20%	Standard	Modest
J	85%	40%	Standard	Modest
K	100%	40%	Standard	Modest
L	115%	40%	Standard	Modest
M	85%	20%	Standard	Aggressive
N	100%	20%	Standard	Aggressive
O	115%	20%	Standard	Aggressive
P	85%	40%	Standard	Aggressive
Q	100%	40%	Standard	Aggressive
R	115%	40%	Standard	Aggressive

EXHIBIT 7

suffering an operating loss of $70 million or greater, then the minimum required surplus for the company is $350 million. At that level of surplus, it will be just inside the boundary constraint.

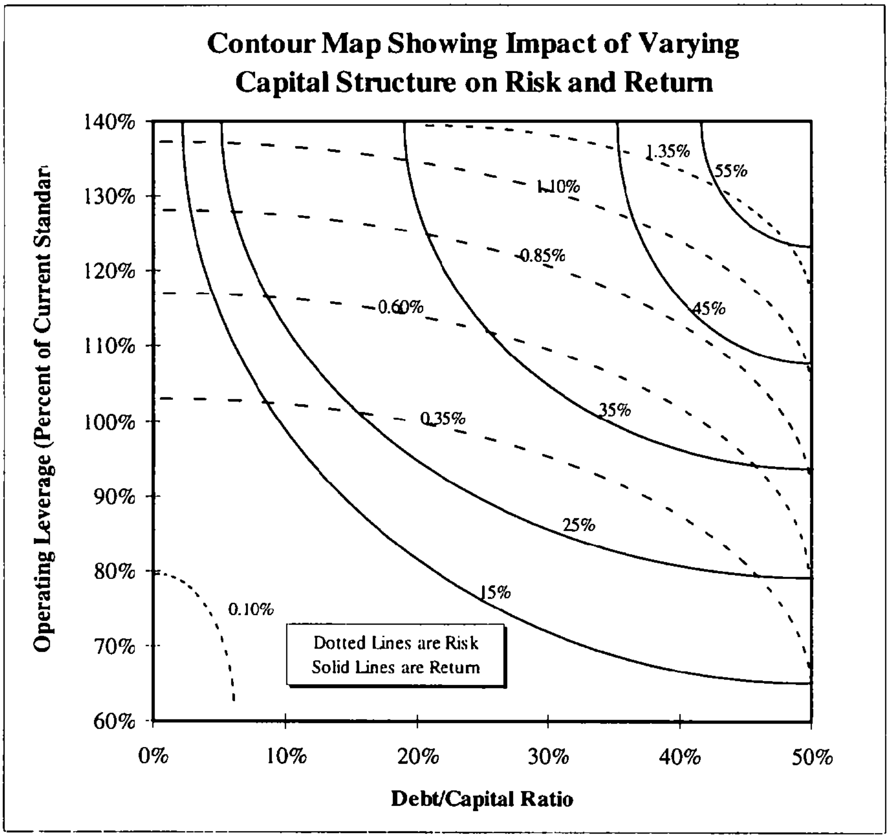


EXHIBIT 8

The Company has established several such boundary constraints, and uses them to measure surplus employed on an ongoing basis, on the basis of the operating profit distributions generated by the model each quarter. This approach is not only directly useful in the capital management of the company, but also facilitates the measurement of expected returns on surplus employed. Exhibit 5 illustrates this type of information. In addition to Company results, the model generates the results for an "index fund" of a cross-section of the entire excess property catastrophe market (for certain regions) labeled as 'XYZ', so that comparative performance can also be measured.

A variety of exhibits can be generated from the multi-period model, since it can be used to test so many different strategy variables: operating leverage, debt/capital ratios, dividend strategies, and responses to changing market conditions. Exhibits 6, 7 and 8 are illustrative of the types of output generated by this analytic tool. Exhibit 6 shows the Company's expected results as generated by the model for four sample strategies. In actual practice, basic exhibits like these have been generated for hundreds of alternative strategies and assumptions sets.

Exhibit 7 is an illustration of an asset/liability efficient frontier for 18 different strategies, which are listed on the lower half of the exhibit. In this example, the Company is considering raising or lowering its operating leverage by 15% from current levels, varying its debt/capital ratio from 20% to 40%, and altering its response to changes in market price levels from "modest" to either "flat" or "aggressive". While the exhibit is a highly stylized version of such an analysis, it is indicative of the approach actually taken.

Finally, Exhibit 8 is a supporting exhibit to Exhibit 7, showing the trade-off between risk and return associated with the operating leverage and debt/capital variables. Risk and return measures from the multi-period planning model have been used to construct a contour map for a range of values of each variable. The contour map shows how risk and return rise and fall in each region of variable values. (The actual contour lines are more involved than shown, with multiple inflection points.) To find an efficient frontier point, one follows a particular return line, looking for the region where the line also achieves minimum risk. Exhibits such as these are used as diagnostics in the efficient frontier analysis.

In addition to using different decision variables, the model is run with varying assumptions to test how the resulting frontiers and contour maps are affected.

5. CONCLUSION

To make the dynamic financial analysis system described in this paper useful in the decision making process, a significant continuing investment is required in:

- Maintaining the underlying databases current and error-free
- Including all types of business and perils to which the company is exposed
- Training all professional staff in the details of the model
- Designing the system so that the DFA results are produced quickly, with easily understandable output reports
- Selecting employees and establishing a culture where decision making in this framework is considered natural and practical.

The substantial investment in building and maintaining the system has clearly been justified – but only because of its usefulness in many of the practical decisions facing the company.

A final challenge is for employees using this admittedly complex system to develop good judgment as to how much weight to give its results in their actual decision making. This requires a thorough understanding of the theory and the practical details of the system, and an appreciation of the limitations and assumptions underlying the results. A good sense of how to weigh system results with unmodeled factors is the essence of the amorphous term "underwriting judgment".

ACKNOWLEDGMENTS

The authors would like to acknowledge specifically the significant contributions made by Jayant Khadilkar, Richard Rafferty, William Riker, and Cary Sparrow towards the development of the Company's DFA system. These individuals toiled many late night hours fighting the "devil in the details" of the system.

The authors also want to thank Dr. Hans Bühlmann for pointing out the early references to expected policyholder deficit in the German actuarial literature, and to Mark Scully for assistance in translation.

APPENDIX A – CURRENCY RISK

The Company is an international reinsurer writing contracts covering exposures in many different countries. Since it operates in multiple currencies, the resulting revenues, assets, and liabilities are affected by currency movements. It is instructive to observe the interplay of currency movements on asset and liability risks.

The measurement of performance and risk in the ALEF framework must take the perspective of the owner. While assets and liabilities may be held in a variety of currencies, ultimate returns and settlement costs must be measured in terms of their impact on equity, as measured in the owner's currency. For this reason, the Company's DFA model expresses all results in terms of U.S. dollars, reflecting gains and losses at the time of conversion as part of the cost or benefit. Specifically:

- The cost of future claim liabilities includes the cost/benefit of converting them to U.S. dollars at future exchange rates.
- The benefit of future reinstatement premiums includes the cost/benefit of converting them to U.S. dollars at future exchange rates.
- The total return on non-U.S. investments includes the gains/losses due to currency movements during the period.

Thus, currency risk is treated as an embedded element of asset and liability risks, and not as a separate risk element.

If potential investors have a principal currency other than U.S. dollars, they may be interested in measuring risks and returns from the vantage point of another currency. Since the Company's stock is traded only on a U.S. exchange, the U.S. dollar perspective to risk and return in the model appears reasonable. Investors from outside the U.S. must overlay the risk/return associated with holding a U.S.-denominated asset to the risk/return as measured by the model.

International reinsurance contracts can pose particularly complex currency risk issues, for the following reasons.

- *The underlying exposures may be in one or several currencies.* The primary insurer will be paying claims in the local currency. It is even possible that the primary insurer could be paying claims from a single event in more than one currency (for example, French francs and Danish kroner).
- *The reinsurance contract terms (i.e., retentions, limits, etc.) may be in one or more currencies, possibly different from the currency of the underlying claims.* In such an instance the contract may specify that underlying claims be converted, using a specified currency exchange rate or the rate prevailing at the time of the event.

- *The reinsurance contract may also specify settlement by the reinsurer in a particular currency.*

The ultimate cost (in U.S. dollars) of claims on such reinsurance contracts is dependent on the interaction of the underlying claims with the prevailing exchange rates and the contract terms.

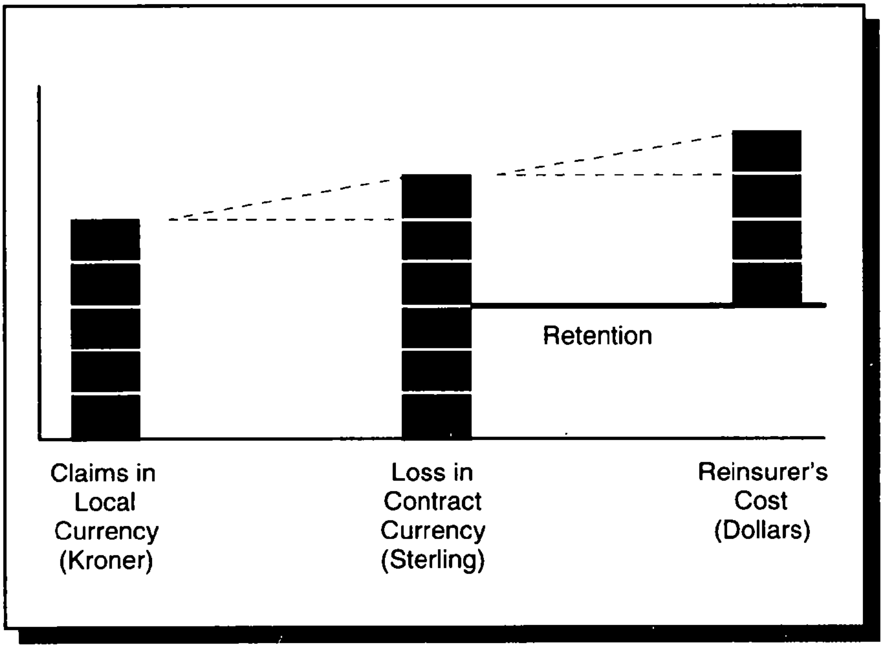


FIGURE 4: Conversion of underlying loss in local currency to u.s. Dollar cost to reinsurer.

To illustrate, consider the simple (and admittedly unrealistic) example illustrated in Figure 4 below. The contract involves underlying exposures in Danish kroner, with contract losses settled in U.K. sterling, the cost of which must ultimately be expressed in terms of U.S. dollars. Given the underlying losses shown in the left-hand bar, the reinsurer incurs the cost shown in the right-hand bar. To measure its loss, the underlying losses (5 units in kroner) must be converted from their original cost in kroner to sterling at the prevailing exchange rate (6/5 in the example); the retention of 3 units (expressed in sterling) must be applied; and the resulting loss to the layer must be converted from sterling to dollars at the prevailing exchange rate (4/3 in the example). Thus an underlying loss of 5 units in kroner creates a cost of 4 units in dollars to the reinsurer.

To illustrate the interplay of currency risks on the contract, consider an alternative scenario involving adverse movements in all currencies. This alternative scenario is presented in figure 5. First, a higher-than-anticipated Danish inflation rate causes the underlying loss in kroner to be greater (the left-hand bar is now 6 units, rather than 5). Next, adverse movement in the kroner-to-sterling exchange rate causes the loss to be even greater when measured in the contract currency (the exchange rate has moved from 6/5 to 8/6). The increase is leveraged by the fixed retention (3 units of sterling in either scenario). Finally, adverse movement in the sterling-to-dollar exchange rate causes the U.S. dollar loss to the layer to be greater still (the exchange rate has moved from 4/3 to 7/5).

As a result of adverse movements in inflation and exchange, the reinsurer's cost has grown from 4 units of dollars to 7 units of dollars.

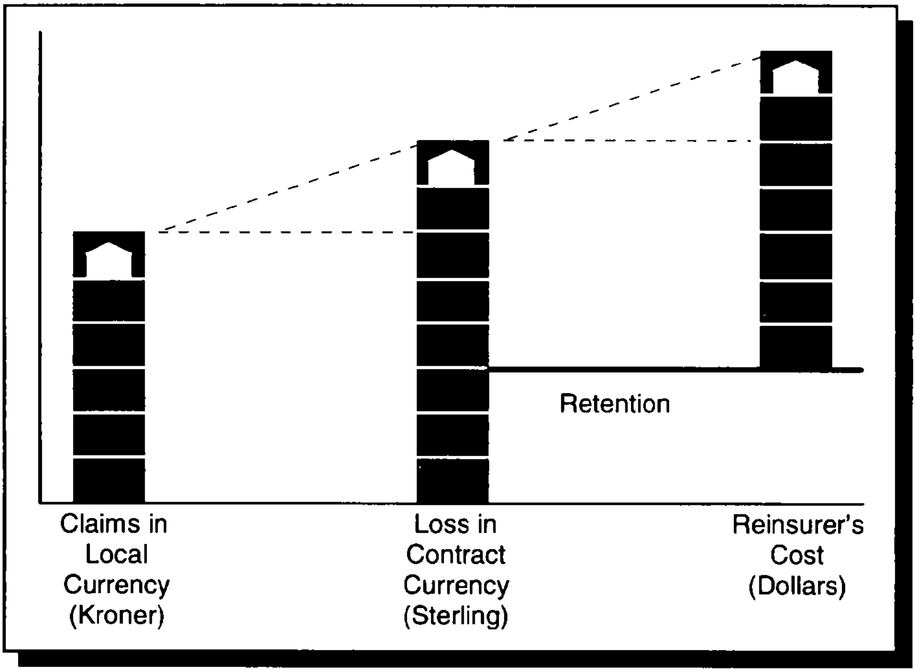


FIGURE 5 Alternative scenario showing the impact of currency movements.

The example neatly divides the currency portion of liability risk into three components: inflation risk, which affects the magnitude of the underlying losses in their original currency; contract exchange rate risk, which affects the conversion of losses from original to contract currency; and settlement exchange rate risk, which affects the conversion of losses from contract currency to dollar terms. All three components need to be incorporated in the pricing and underwriting of reinsurance contracts.

The example is contrived and also unrealistic in its assumed exchange rates. In addition, the adverse scenario is particularly unrealistic in that purchase power parity would imply that an increase in kroner inflation would generally be expected to be associated with a favorable movement in the kroner-to-sterling exchange rate (the kroners would have less purchasing power, so it would take relatively fewer pounds sterling to buy them). Of course, the direction of movements in the illustration could certainly happen in the short run.

A particularly strong feature of Global CAP:Link is its ability to model inflation and exchange rates between multiple currencies in a logical and consistent manner, so that unreasonable purchase power parity relationships are not created. Each scenario generated by the system is plausible, with variability in parity occurring over short intervals and a greater tendency towards parity over longer intervals. By integrating the CAP:Link scenarios into the generation of both the liability scenarios and the asset return scenarios, the overall risks posed by operations in multiple currencies can be managed.

While the illustration above describes the risks as working from left to right on the chart, in reality they work in the opposite direction. The Company must exchange U.S. assets for sterling to settle the claim. Some readers may question whether settlement exchange risk is real, and not created by currency mismatch. In other words, if the Company is holding some of its assets in sterling, it can settle the claims on sterling-based contracts without suffering any gains or losses due to exchange rate fluctuations by simply paying the claims out of its sterling funds. Thus, it might be argued that settlement exchange risk only exists to the extent that insufficient sterling assets to pay the claims are available.

However, the above line of reasoning confuses the existence of risk with its immunization. For liabilities that are fixed and certain, the Company can immunize itself against overall currency risk by holding a matched set of assets equal to the liabilities in the same currency. In such a case, any change in the exchange rate will cause the decline in asset value to be offset by an equal decline in liability value (both measured relative to U.S. dollars); conversely, an increase in asset value will be offset by an equal increase in liability value. Thus, although settlement exchange risk and asset currency risk are both present, they are negatively correlated, facilitating the immunization.

If liabilities in each currency were fixed, known amounts, the minimum risk position would appear to be to hold funds in each currency sufficient to settle the liabilities. (This pre-supposes that no arbitrage opportunities exist and that the investment risks and expected returns are the same in each currency.) But when liabilities are uncertain as to amount, timing, and currency, it is not quite so clear how to minimize currency risk. This is where effective modeling can be an invaluable tool.

APPENDIX B – RISK MEASURES AND THE EXPECTED POLICYHOLDER DEFICIT

In the ALEF framework risk can be any measure of adverse outcomes that management believes is most relevant to the enterprise. One such measure is the expected policyholder deficit (EPD), a term developed as part of the U.S. risk-based capital initiative and attributed to BUTSIC (1994). Since some readers may not have been exposed to the concept, a brief description is included herein.

All insurers face the possibility that, at some point in the future, their obligations may exceed their assets. The magnitude of this risk is a function of the asset, liability, and business risks faced by the insurer, and the level of capital held to support those risks. Insolvency risk has traditionally been measured in terms of the probability of ruin. However, from the perspective of the policyholder this measure is insufficient because it fails to take into account the severity of the insolvency.

TABLE 1, taken from BUTSIC, illustrates this point. In this simple example, two insurers have identical balance sheets. Each insurer has assets of $13,000, liabilities of $10,000, and capital of $3,000. Although the assets of each company are certain, the liabilities (unpaid claims) are uncertain, subject to the probability distributions shown.

TABLE 1

CALCULATING THE EXPECTED POLICYHOLDER DEFICIT

	Asset Amount	Probability of Outcome	Liability Amount	Capital Amount	Claim Payment	Deficit
Insurer A						
Scenario 1	13,000	0.2	6,900		6,900	
Scenario 2	13,000	0.6	10,000		10,000	
Scenario 3	13,000	0.2	13,100		13,000	100
Expectation	13,000		10,000	3,000	9,980	20
Insurer B						
Scenario 1	13,000	0.2	2,000		2,000	
Scenario 2	13,000	0.6	10,000		10,000	
Scenario 3	13,000	0.2	18,000		13,000	5,000
Expectation	13,000		10,000	3,000	9,000	1,000

In this simplest of examples, there are no expenses or taxes, no time value to money, and no other business transactions to consider. For each company, the ultimate outcome will be one of the three scenarios shown. Due to the corporate form of the enterprise (assumed to be a non-assessable stock corporation), the payments to policyholders are limited to the available assets. Each insurer is subject to an equal probability of ruin, with a 20% chance that obligations will exceed resources, claim payments will be limited, and the insurer will be forced to go out of business. Both insurers exhibit the same balance sheet leverage.

However, the claim payment column in the chart clearly indicates that the policyholders of Insurer B are significantly worse off than those of Insurer A. While policyholders of Insurer A receive only a minor reduction in claim payments in one of the three possible scenarios for their liabilities, policyholders of Insurer B may suffer a substantial underpayment, receiving only 13/18 of their indicated claim payment. Overall, policyholders of Insurer A expect *a priori* to recover all but $20 of the expected claim payments, while policyholders of Insurer B expect to recover only $9,000 of the expected $10,000 liability.

The expected policyholder deficit is defined as the expected value of the difference between the amount of the claim obligation and the actual claim payment. For Insurer A, the EPD is $20, or 0.2% of expected obligations. For Insurer B, the EPD is $1,000, or 10.0% of expected obligations.

While the ruin probabilities and reported financial leverage ratios of Insurer A and Insurer B are the same, the value of coverage afforded by each is clearly different. Insurer A offers considerably greater real value, from the perspective of the policyholder; expected recoveries are a substantially greater proportion of expected losses than is the case with Insurer B. In comparing the security offered, Insurer A's *EPD ratio* of 0.2% is stronger than Insurer B's 10.0%. To offer the same level of security, Insurer B would need to increase its *capital ratio* from the current 30% of expected losses to 79% of expected losses (i.e., raise its assets to 17,900, so that it could pay all but $100 of the losses in Scenario 3).

The expected policyholder deficit concept can easily be adapted to consider asset risks as well as liability risks, by expanding the scenarios to include changes in asset values as well as liability values. For each scenario, the realized value of the assets is compared to the settlement value of the liabilities to determine whether or not there is a deficit.

From a financial standpoint, the EPD is the value of the put option held by the shareholders of a corporated enterprise. In the event that aggregate obligations exceed total assets, the shareholders can put the obligations to the regulators in exchange for the assets. When customers purchase insurance from a particular company, they implicitly give this option to the company.

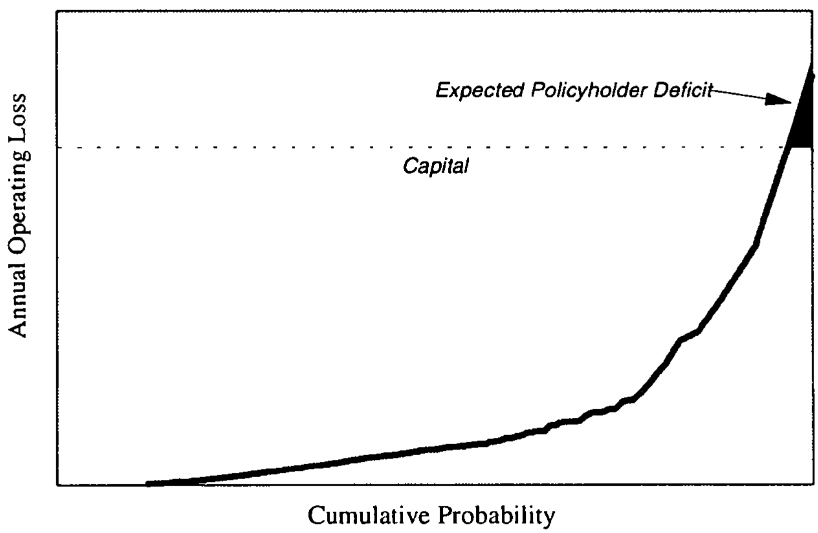


FIGURE 6: The distribution of annual aggregate operating losses for the enterprise, and the level of capital, determine the expected policyholder deficit

BUTSIC and others have argued that capital requirements for different lines of business should be set by equating EPD ratios (as opposed to ruin probabilities or leverage ratios). For each line of business the capital requirements should be set so that the expected deficit is the same percentage of expected losses. This approach is most relevant in a pricing context, when the cost of employed capital is being considered. Since all policyholders give up the same option, setting returns on capital that is apportioned in this manner assures equity among policyholders.

The expected policyholder deficit concept can be extended to consider all types of risks, to the extent that they can be incorporated into the dynamic financial model. The key model output is the distribution of aggregate operating losses for the enterprise. Different strategies can then be evaluated in terms of their impact on the EPD ratio. Alternatively, for a given target EPD ratio, different strategies can be evaluated in terms of their impact on the capital required to achieve that ratio.

It turns out that the concept of expected policyholder deficit is not at all new. In a paper published in 1868, HATTENDORF refers to "mittleres Risiko", the mean risk, as defined by WITTSTEIN and KANNER (1867). HATTENDORF discusses the concept in the context of mortality risk in life insurance; excerpts of that discussion are loosely translated below.

> Because it is not possible to calculate an absolutely correct mortality table for an infinite number of observations, and because in reality the number of insureds with the same age is always finite, an insurer must accept that results will deviate from the expected level. Such a deviation can be favorable for the insurer, but it can also require greater payments than expected. And so, the company takes a risk in that it promises the payment of all insured sums *under any conditions*, but its remuneration from the insured is based on the *expected* case.

> If one defines risk as the financial loss which one accepts, then it is clear that a narrower definition of the concept is required. One can speak of the largest and the smallest risk. The smallest is clearly equal to zero. The largest is the entire insured sum on all policies, less the available funds in reserves and premiums. Far more important is the mean risk. *By this one means the sum of all possible operating losses, each multiplied by its probability.* This definition is well-defined, permits no uncertainty, and with it one can compute the mean risk for a given insurance portfolio.

The HATTENDORF paper develops a methodology for estimating WITTSTEIN and KANNER'S mean risk for a portfolio of life insurance contracts. While WITTSTEIN and KANNER had proposed the concept, they had not developed a practical means of manually calculating mean risk for large numbers of contracts.

REFERENCES

BUFF, J. (1990) Extending the Efficient Frontier. *Imphasis* **1990/4**, Tillinghast - Towers Perrin, New York.

BUTSIC, R. (1994) Solvency Measurement for Property-Liability Risk-Based Capital Applications. *The Journal of Risk and Insurance* **Volume 61, Number 4**, 656-690.

BRENNAN, M. and SCHWARTZ, E. (1982) An Equilibrium Model of Bond Pricing and a Test of Market Efficiency. *Journal of Financial and Quantitative Analysis* **17**, 75-100.

DAYKIN, C.D. et. al. (1987) Assessing the Solvency and Financial Strength of a General Insurance Company. *Journal of the Institute of Actuaries* **September, 114**, 227-325

DAYKIN, C.D. and HEY, G.B. (1990) Managing Uncertainty in a General Insurance Company. *Journal of the Institute of Actuaries* **September, 117**, 173-277.

DOLL, D. and SONLIN, S. (1994) Optimizing ALEF Studies. *Emphasis* **1994/4**, Tillinghast – Towers Perrin, New York.

GRAY, W.M. (1990) Strong association between West African rainfall and U.S. landfall of intense hurricanes. *Science* **249, 1251-1256.**

HARTMAN, D.G. et. al. (1992) Property-Casualty Risk-Based Capital Requirement - A Conceptual Framework. *The Forum* **Spring 1992**, Casualty Actuarial Society, New York, 211-280.

HATTENDORF, K. (1868) Ueber die Berechnung der Reserven und des Risico bei der Lebensversicherung. In *Rundschau der Versicherungen* **Volume XVIII**, Leipzig

HODES, D.M. et. al. (1996) The Financial Modeling of Property/Casualty Insurance Companies. *The Forum* **Spring 1996**, Casualty Actuarial Society, New York, 3-88.

MARKOWITZ, H. (1959) Portfolio Selection: Efficient Diversification of Investments. John Wiley & Sons, New York.

MULVEY, J.M. and THORLACIUS, A.E. (Forthcoming in 1997) The Towers Perrin Global Capital Market Scenario Generation System: CAP:Link. In *Worldwide Asset and Liability Modeling* (ed. W.T. Ziembarand and J.M. Mulvey), Cambridge University Press.

WITTSTEIN and KANNER (1867) Determination of the Mean Risk in Life Insurance. *Deutsche Versicherunge-Zeitung*.

STEPHEN P. LOWE
Chief Actuary
Tillinghast – Towers Perrin
175 Powder Forest Drive
Weatogue, Connecticut 06089-9658, USA
lowes@towers.com

JAMES N. STANARD
Chairman, President and Chief Executive Officer
RenaissanceRe Holdings, Ltd.
Renaissance House, East Broadway
P.O. Box HM 2527
Hamilton HMGX, Bermuda
jns@renre.bm

BOOK REVIEWS

D. R. DANNENBURG, R. KAAS, M. J. GOVAERTS (1996): *Practical Actuarial Credibility Models.* IAE (Institute of Actuarial Science and Econometrics of the University of Amsterdam), 157 pages.

The courageous foreword of this book caught my interest. The authors made three points about existing credibility literature:

1. Credibility is currently taught in a needlessly complicated way.
2. Credibility is emphatically set in a Bayesian framework.
3. Often advanced mathematics is applied, even in situations where more elementary methods could be used.

These statements raise high expectations about the book, but I was not disappointed. In an accessible language the authors provide an extensive and rather complete guide to credibility. The theories are well documented with verbal motivation, practical considerations and numerical examples. Many readers will appreciate the fact that the book does not neglect the ties to statistics, such as analysis of variance and regression. The book has the following eight chapters:

1. Introduction,
2. The Buhlmann-Straub model,
3. Jewell's hierarchial model,
4. General properties of credibility estimators,
5. Credibility and Bayes estimators,
6. The two-way cross classification model,
7. Credibility models applied to IBNR problems,
8. The Hachemeister model.

By writing this text, the authors have rendered a service to the international actuarial community: the book can be recommended to practitioners as well as to advanced actuarial students and their teachers.

Information how to order the book can be obtained from Rob Kaas, either at his IAA Yearbook address, or at RobKaas@fee.uva.nl by e-mail.

HANS U. GERBER

D. G. HART, R. A. BUCHANAN, B. A. HOWE (1996): *The Actuarial Practice of General Insurance.* Institute of Actuaries of Australia, Sydney. 591 pp. ISBN 0-85813-055-6.

This book is the general insurance textbook used by students of the Institute of Actuaries of Australia. In broad terms, its coverage can be broken down into the following three areas:

- The nature and operation of general insurance (chapters 1-5).
- Actuarial techniques for general insurance (chapters 6-16).
- Actuarial practise of general insurance (chapters 17-40).

The first part provides a thorough, but very readable introduction into the nature of general insurance. Starting with a description of the risks covered by different classes of insurance and a brief historical overview, it then moves into the area of insurance law and insurance regulation. Having dealt with these issues at considerable length, the authors then discuss the actual operation of an insurance company, including underwriting and claims management, coinsurance and reinsurance, and financial reporting. The last two chapters of the first part cover the data requirements of a general insurer and, briefly at this stage, the role of the actuary in general insurance.

By necessity, the first part of this book is steeped in Australian legislation, which can be a bit daunting to a reader unfamiliar with the federal structure of Australia. Of somewhat greater concern is the fact that the authors attempt to give an up-to-date presentation of several state-based statutory insurance schemes such as Workers' Compensation. As such schemes are constantly being modified, frequent revisions of this book will be inevitable. The reviewer's hope is that a future edition will be written with a view to being less ephemeral. For the current reader, however, this part of the book provides a wealth of information on how to operate an insurance company in any country.

The second part provides an overview of important actuarial and statistical techniques for the general insurance actuary. The following areas are covered: claim frequency and size distributions leading to aggregate claim distributions; basic risk theory leading to methods for optimising capital, profit and reinsurance; experience rating and risk classification; forecasting and run-off techniques; premium rating.

This part is very comprehensive and provides many useful quantitative methods. All the methods presented are clearly motivated and described with numerical examples. Due consideration is given to situations with less-than-perfect data, such as incomplete run-off triangles (the assumption of perfect data is a frustrating aspect of many theoretical textbooks).

Naturally it has been impossible for the authors to cover all the available methods, or to provide more than an outline of the theoretical derivation and statistical properties of the methods presented: for this, the interested reader must consult other sources.

The third part of the book covers different aspects of the actuary's role and tasks in general insurance. Areas of actuarial involvement covered include:

- Rating and design.
- Reserving.
- Financial control.
- Reinsurance.
- Appraisal.
- Compensation schemes.
- Risk management.
- Statutory supervision.

Within each area, the authors provide a wide-ranging discussion of the relevant practical, professional and legal issues that need to be addressed by the general insurance actuary in the performance of his or her duties.

The authors of a book as comprehensive as this one must make a number of compromises. Thus one could point out a number of topics that are treated in greater depth by dedicated theoretical texts. However, within the limitation of almost 600 pages, the authors have succeeded remarkably well in balancing the general discussion of issues, with a presentation of quantitative techniques that has an acceptable degree of mathematical rigour.

In the opinion of this reviewer, *The Actuarial Practice of General Insurance* is an excellent introduction to actuarial work in general insurance.

WALTHER NEUHAUS

ICIAM 99 IN EDINBURGH

The Fourth International Congress on Industrial and Applied Mathematics (ICIAM) will be held in Edinburgh from 5th to 9th of July 1999. More than 2,000 delegates are expected to attend. Previous ICIAM congresses were held in Paris 1987, Washington 1991, and Hamburg 1995 and this event is now firmly established as the premier international conference in applied mathematics. The last mathematical congress of comparable importance and size to be held in the U.K. was the International Congress of Mathematicians held in 1958. This, too, was held in Edinburgh. The success of that congress bodes well for ICIAM 99.

The congress will focus worldwide attention on the importance of mathematical and computational methods in the solution of real-world problems. The main features of the programme will be:

- 25 general lectures by leading international experts on current developments of industrial, computational and applied mathematics. Mathematical methods for the qualitative and quantitative analysis of models will be presented and important practical applications will be discussed extensively. Particular themes will include:
  - Financial Mathematics, Insurance, Investment and Banking
  - Mathematical Modelling in Industry
  - Mathematics of Medicine
  - Geophysical and Oil Sciences
  - Large Scale Computation
  - Environmental and Climate Science
  - Cryptography, Coding and Computer Security
- 300 mini-symposia and organised discussion sessions to provide integrated presentations and discussion by international panels on the latest mathematical and computational techniques. Research on industrial, commercial and environmental applications will be discussed as well as other issues including applied mathematics education, public perception of mathematics, and the organisation of applied mathematics societies.
- Special all-day sessions will be run in conjunction with learned societies and other organisations discussing new research in the mathematical and computational sciences and outlining novel applications.
- End of Conference Session with a panel overviews and perspectives, drawing conclusions from lectures and mini-symposia, and looking forward to the challenges and problems of the next century.

The U.K. actuarial profession will be organising mini-symposia for the Congress.

Organisation of the Congress

The Joint Patrons of the Congress are H.R.H. The Prince Philip, Duke of Edinburgh, K.G., K.T. and The Right Hon. The Lord Mackay of Clashfern.

ASTIN BULLETIN, Vol. 27, No. 2, 1997, pp. 377-378

The scientific programme and, in particular, the selection of invited speakers lies in the hands of an international Scientific Programme Committee chaired by Professor J. R. C. Hunt, Horonary Professor at University of Cambridge and, until recently, Chief Executive of the U.K. Meteorological Office. The committee has 30 members from 16 different countries.

Other U.K. organisations involved include the British Computer Society, Royal Statistical Society, Institute of Mathematics and its Applications, International Centre for Mathematical Sciences, London Mathematical Society, Operational Research Society, Royal Society, Royal Society of Edinburgh, Engineering and Physical Sciences Research Council.

Information about ICIAM 99

Further information on the Congress can be found on the World Wide Web under the address

http://www.maths.ed.ac.uk/conferences/iciam99/

where the current information is constantly brought up to date and you can preregister for the meeting following the easy instructions.

If you do not have access to the World Wide Web, further information can be obtained by writing/telephoning/faxing/emailing

ICIAM '99 Congress Secretariat
c/o Meeting Makers
50 George Street
Glasgow G1 1QE
U.K.
Telephone: +44 (0)141 553 1930
Fax: +44 (0)141 552 0511
email: geninfo.iciam@meetingmakers.co.uk

THE 30TH INTERNATIONAL ASTIN COLLOQUIUM

Tokyo, Japan, 22-25 August, 1999

The 30th International ASTIN Colloquium will be held in Tokyo, Japan from the 22nd through the 25th of August 1999. The Colloquium will cover the following areas related to most non-life actuaries.

- Sound management of non-life insurance companies
- Rating system and method
- Statistical analysis of insurance
- Miscellaneous

1999 marks the centennial anniversary of the Institute of Actuaries of Japan (IAJ). The AFIR Colloquium and IAJ's Centennial Celebration Convention are scheduled to follow the 30th ASTIN Colloquium in Tokyo. The final day of the 30th ASTIN Colloquium, August 25th, will be "Joint Day" with AFIR. IAJ considers it a privilege that Japan will be the first Asian country to host the ASTIN Colloquium.

While a detailed schedule of the Colloquium is still being developed, we hope many ASTIN members from overseas will be able to attend and present research papers. Active participation in this event will be most appreciated.

Tokyo is not only one of the largest cities in the world, but it is also a city that retains Japanese tradition and culture. Excursions would include areas around Tokyo that will allow you to visit many famous places and historic sites. In addition, a two-hour bullet train ("Shinkansen") ride can take you to Kyoto or Nara. In these ancient cities, you can fully appreciate historical Japan. We would like to invite you to attend the Colloquium in Tokyo and to take advantage of the opportunity to enjoy Japanese tradition and culture.

IAJ has formed the ASTIN Scientific Committee (Chairman Dr. Yoshizoe, Professor of Aoyama Gakuin University) and the Organizing Committee, for preparation and management of the Colloquium. Committee members are eargerly anticipating the Colloquium.

Further information and details will follow shortly.

REPORT ON THE INTERNATIONAL ASTIN/AFIR COLLOQUIA 1997

The 28th ASTIN International Colloquium and the 7th International AFIR Colloquium were held in the same week from August 11th to August 15th 1997, in the Cairns International Hotel, North Queensland, Australia. This first ever joint meeting culminated in a "joint day" on the Wednesday, recognising the fact that ASTIN and AFIR delegates have an increasing number of interests in common and that both actuarial and financial skills are necessary to cope with the new challenges of today. The CNN announcement of the merger of the Winterthur Insurance Company and the Crédit Suisse Group, news of which spread rapidly among delegates, emphasised this point.

The joint meeting was also an excellent opportunity to celebrate the 100th anniversary of the Institute of Actuaries of Australia. The first gathering of Australian actuaries to discuss the formation of an institute took place on August 12th 1897.

This historic meeting (and possibly also the chance to visit the Great Barrier Reef and the tropical rainforest, two of Australia's World Heritage Areas) attracted a large audience: 193 people registered for ASTIN and 209 registered for AFIR; both numbers include 126 people who took the opportunity to register for both colloquia. The delegates came from 31 different countries with groups of at least ten delegates from Australia, Belgium, Canada, Denmark, Germany, Japan, Netherlands, Norway, Sweden, Switzerland, United Kingdom and the USA.

Eight internationally recognised keynote speakers gave lectures during the week; 21 speakers presented papers during the two ASTIN days, nine during the afternoon of the joint day; 31 speakers presented their research, partly in concurrent sessions, during the two AFIR days. The combined proceedings of both colloquia have an approximate length of two thousand pages. Obviously, we can only mention the general themes and some highlights of the colloquia, which reflect our personal tastes and understanding.

The ASTIN colloquium opened with a session on the topical subject of catastrophe risk. Bruce Harper held an entertaining invited address on the modelling of wind hazards and insurance risks in Australia. This talk was impressive in showing how, in the absence of reliable data on the losses caused by tropical storms, insurance events could be simulated using physical models for storms and for the damage caused by storms with different characteristics.

In the second invited address Prof. Paul Embrechts, ETH Zürich, outlined the possibilities offered by extreme value theory in the modelling of catastrophic losses in the situation where some data are available. The basic message of extreme value theory is that there are natural probability distributions and models for extremely large observations in the same way that there are natural models for average values (such as the well known Gaussian normal distribution). Further contributed talks on extreme values by Alexander McNeil and Dietmar Pfeifer indicated that this interesting branch of probability theory is now coming to the attention of practising actuaries.

A second major topic on the first day was classification of risks, and this session showed the broad palette of statistical techniques now being used in insurance research. Talks ranged from the application of cluster analysis in the formation of tariff classes using neural network-based implementations to the evaluation of occupational risks using methods from survival analysis. Greg Taylor, who later in the week became the recipient of the first ever gold medal of the Institute of Actuaries of Australia, submitted two papers in this section, one on the use of Whitaker spatial smoothing to obtain good estimates of risks which vary geographically, the second on the setting up of a bonus-malus scale of premiums in the presence of additional rating factors.

The final session of the day consisted of papers on the subject of premium rating in which one identifiable theme was the use of Markov state models. Papers presented by Jose Garrido and by Ermanno Pitacco addressed the use of such models in disability and health insurance tarification. The day ended with delegates enjoying dinners at one of two exotic locations. One party sampled traditional Queensland fare at the Riverstone Homestead, a historical sugar plantation house; a second group dined at the luxury Paradise Palms Golf Course.

In view of the copiousness and excellence of the food and wine it was all the more remarkable that attendance had not declined on the second morning when delegates reconvened to hear talks in two sessions entitled statistics and reserving. In the former session David Dickson described an alternative to the classical compound Poisson risk process; he derived results for the probability and the extent of ruin when claims occur as more general renewal processes. In the latter session two papers looked at different aspects of the calculation of development factors, or link ratios, in the loss development problem. The first paper by Glen Barnett and Ben Zehnwirth focused on the use of diagnostics in the selection of competing models for loss development; the second paper by Erhard Kremer looked at a robust version of the classical chain-ladder model. After these short but intensive scientific sessions delegates spent the afternoon on an enjoyable excursion with the Kuranda historical railway from the coast up to Kuranda, passing the Barron Falls, and down again with Skyrail above the canopy of the tropical rainforest. This provided a most unusual but congenial backdrop for the important conference activity of catching up and networking with colleagues.

The joint ASTIN/AFIR day was mainly devoted to the securitization of insurance risk, the issue which best represents the convergence of ASTIN and AFIR interests. The first keynote lecture by James A. Tilley addressed the securitization of catastrophic property risks, the second by Prof. Neil Doherty, University of Pennsylvania, was about financial innovation in the management of catastrophe risk.

For ASTIN delegates the first talk followed nicely from the discussion of catastrophe risk on day one. The interest in securitization arises because of increased exposure to catastrophes and the empirical observation that frequency and severity of large losses are on the increase. Insurance companies alone may not have the capacity to handle the mega-catastrophes of the future.

A possible way to pass the insurance risk to investors (in return for a corresponding risk premium) are catastrophe bonds. Investors can use these bonds to diversify their portfolios because natural disasters, for example, have a very small correlation with financial market risk. Catastrophe bonds can be classified as:
— pure catastrophe bonds, when the principal and the coupons are at risk;
— principal-protected catastrophe bonds, when only the coupons are at risk;
— deferred catastrophe bonds, when no payment as such is at risk, but the date of the payments can be deferred, leaving the issuer of the bond an interest gain in case of a catastrophe.

James Tilley described some products which have so far been developed, such as the California Earthquake Authority risk bonds (an example of the second type of bond above), and looked at reasons why the market for securitized products has generally developed slowly. Among these reasons are the favourable catastrophe experience since 1994, the rehabilitation of Lloyds and the weaknesses of proposed securitization structures. However, he suggested there was still a potential need for cost efficient products with flexible annual renewal possibilities and more of the simplicity which makes traditional reinsurance arrangements appealing.

In the afternoon presentations of the joint day, several points raised in the keynote lectures were studied more deeply. The correspondence of catastrophe bonds and defaultable bonds was discussed, both leading to an incomplete market setting which causes difficulties in the pricing methodology. An approach, advocated by Prof. Martin Schweizer, is to decompose the risk of say a catastrophe bond into a hedgeable part and a residual part, which is treated by standard actuarial methods to obtain a price. This leads to a bid-ask price spread for the catastrophe bond. A specific principal-protected catastrophe bond, the Winterthur Insurance convertible bond with WinCat coupons "hail", was considered by Uwe Schmock and several methods for the estimation of the coupon values were presented; model risk for the statistical analysis and the corresponding pricing of the bond were investigated. Further talks addressed selected topics of asset liability modelling, risk-based capital allocation, risk-adjusted performance management and reserving for future claims taking stochastic interest rates into account.

The joint day ended with a lavish ASTIN/AFIR colloquia banquet in the Great Hall of the Cairns Convention Centre, where a gold medal of the Institute of Actuaries of Australia was awarded to Greg Taylor, the chairman of the ASTIN scientific committee.

The first pure AFIR day started with Prof. Phelim Boyle's keynote lecture on quasi-Monte Carlo methods for numerical integration. He showed that deterministic low discrepancy sequences outperform crude Monte Carlo methods in low dimensions, but that this superiority diminishes for high dimensions or discontinuous integrands. Randomisation of low discrepancy sequences and reduction of the effective dimension of the integration problem can come in handy in these cases.

After morning tea, the AFIR prize winning papers were presented and the certificates awarded. The first prize was given to Glen R. Harris, AMP Society, Australia, for his paper on "Regime switching vector autoregressions: a Bayesian Markov chain Monte Carlo approach". The second prize was divided between D.J.F. Nonnenmacher and Jochen Russ, University of Ulm, Germany, for their paper "Equity-linked life insurance in Germany: quantifying the risk of additional policy reserves" and Ken Seng Tan and Prof. Phelim Boyle, University of Waterloo, Canada, for their paper "Applications of scrambled low discrepancy sequences to exotic options".

Prof. Stanley Pliska, University of Illinois, gave his keynote lecture on a model for risk-sensitive dynamic asset allocation. For measuring the performance of the model, he presented various infinite-horizon criteria: expected growth rate of the portfolio, expected utility growth rate of the portfolio and a risk-sensitive growth rate criterion. A major aim of this model was to combine the statistical work for parameter estimation with the forecast for asset management. Applied to a historic data set, the corresponding management strategy showed an impressive performance.

Several contributed talks also presented and compared asset/liability management strategies for various settings such as continuous-time pension fund models or portfolios of defaultable assets. Some further themes ranged from inflation modelling to an axiomatic classification of usurious loans and tax-efficient, option-based compensation packages for employees.

The last day of the colloquium started with the keynote lecture by Prof. J. David Cummins, University of Pennsylvania, about the use of financial derivatives in corporate risk management, participation and volume decisions in the insurance industry. According to this talk, the main reasons for the use of derivatives in the insurance industry are to avoid:

— financial distress costs like bankruptcy costs, additional regulatory restrictions and reputational losses affecting relationships with key employees, suppliers and customers;
— duration problems in the asset/liability management, including the liquidity risk of private placements or real estate;
— foreign exchange risk, and
— losses due to the convexity of income tax schedules.

In the contributed talks on Friday, partly in concurrent sessions, Peter Antal applied ideas mentioned on the joint day to the pricing of regular options, arguing that option prices should contain a risk premium in any case because a dynamic hedge as in the Black-Scholes model is not possible in practice. Godfrey Perrott presented policyholder considerations for the demutualisation of a company. David Wilkie showed that different prices of risk and different portfolios cause a failure of the capital asset pricing model in a multi-currency world. Robert Clarkson critically discussed the financial risk in the Markowitz and Black-Scholes worlds, explaining that investors would not accept a certain level of risk (bankruptcy for example), no matter how high the offered risk premium is; a comparison with the risk of death in some dangerous sports made his point clear. Further talks concerned the fitting of the term structure, pricing rate of return guarantees, the use of genetic algorithms, the Italian pension plan or the social security system in Indonesia, for example.

The final keynote lecture was given by Dato'Abdul Khalid bin Ibrahim, currently the Group Chief Executive of Kumpulan Guthrie Berhad. He gave an overview of the Asian capital markets with special focus on the development of the capital markets in Malaysia and Singapore, and he encouraged the audience to invest in these markets. The colloquium closed with the annual general meeting of the AFIR section.

We close with a glimpse of future events. The 8th International AFIR Colloquium will take place in Cambridge, United Kingdom, 15th to 17th of September 1998. Write to David Golder, AFIR 1998 Colloquium Secretariat, Institute of Actuaries, Staple Inn Hall, London, WC1V 7QJ, United Kingdom, for information. The next General Insurance Convention & ASTIN Colloquium will take place in Glasgow, Scotland, 7th to 10th of October 1998. For information write to Linda Pritchard at the above address. The next joint International ASTIN and AFIR Colloquia are scheduled to take place in Tokyo, Japan, from 22nd to 25th of August 1999, followed by the Centenary Convention 29th to 31st of August.

ALEXANDER MCNEIL and UWE SCHMOCK
Department of Mathematics
ETHZ, Zürich
Switzerland

ACTUARIAL VACANCY

DRAKE UNIVERSITY ACTUARIAL SCIENCE POSITION

Position: A tenure track position in actuarial science in the College of Business and Public Administration, to begin August, 1998. Rank and salary based on qualifications.

Duties: Ph.D. in actuarial science or a related area along with Associateship in the CAS or SOA is preferred. Applicants with a Fellowship and Master's degree will be considered.

Applications: Submit a curriculum vitae and arrange for three letters of reference to be sent to Professor Stuart Klugman, FSA; CBPA; Drake University; Des Moines, IA 50311. Applications will be accepted until the position is filled.

Drake University is an equal opportunity/affirmative action employer and actively seeks applications from women and minority group members who are qualified for this position.

Stuart Klugman, FSA
Principal Financial Group Professor of Actuarial Science
Drake University
2507 University Avenue
Des Moines, IA 50311 USA

Phone: 515-271-4097
e-mail: Stuart.Klugman@drake.edu

ACKNOWLEDGEMENT TO REFEREES

"The referees of ASTIN Bulletin, through thorough and conscientious reviews, make this journal possible. Their task is difficult and time-consuming with no material reward. On behalf of the ASTIN Bulletin readership, the Editors and the Editorial board offer them their sincere thanks."

GUIDELINES TO AUTHORS

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

Paul Embrechts,
Department of Mathematics, ETH-Zentrum,
CH-8092 Zurich, Switzerland.

D. Harry Reid,
Eagle Star Insurance Company Ltd,
The Grange, Bishop's Cleeve
Cheltenham Glos GL52 4XX, United Kingdom.

or to one of the Co-Editors:

René Schnieper,
Zurich Insurance Company,
P.O. Box, CH-8022 Zurich, Switzerland.

Andrew Cairns
Department of Actuarial Mathematics
and Statistics
Heriot-Watt University
Edinburgh EH 14 4AS, United Kingdom.

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 ···hor(s), year, title, journal (in italics, cf. point 9), volume (in boldface, cf. point 9), and pages. book references give author(s), year, title (in italics), publisher, and city.

ımples:

BARLOW, R.E. and PROSCHAN, F. (1975) *Mathematical Theory of Reliability and Life Testing*. Holt, Rinehart, and Winston, New York.

JEWELL, W.S. (1975a) Model variations in credibility theory. In *Credibility: Theory and Applications* (ed. P.M. KAHN), pp. 193–244, Academic Press, New York.

JEWELL, W.S. (1975b) Regularity conditions for exact credibility. *ASTIN Bulletin* **8**, 336–341.

References in the text are given by the author's name followed by the year of publication (and possibly a letter) in parentheses.

8. The address of at least one of the authors should be typed following the references.

Continued overleaf.

COMMITTEE OF ASTIN

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.