11-1994

A Comparative Analysis of 30 Bonus-Malus Systems

Jean Lemaire
University of Pennsylvania

Hongmin Zi

Recommended Citation

Lemaire, J., & Zi, H. (1994). A Comparative Analysis of 30 Bonus-Malus Systems. *ASTIN Bulletin: The Journal of the IAA, 24* (2), 287-309. http://dx.doi.org/10.2143/AST.24.2.2005071

A Comparative Analysis of 30 Bonus-Malus Systems

Abstract
The automobile third party insurance merit-rating systems of 22 countries are simulated and compared, using as main tools the stationary average premium level, the variability of the policyholders' payments, their elasticity with respect to the claim frequency, and the magnitude of the hunger for bonus. Principal components analysis is used to define an "Index of Toughness" for all systems.

Disciplines
Business | Economics | Public Affairs, Public Policy and Public Administration

Volume 24, No. 2 November 1994

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.

EDITORIAL AND ANNOUNCEMENTS

EDITORIAL

ACTUARIES AND FINANCIAL ECONOMISTS

At the recent AFIR Colloquium in Orlando there were two meetings which gave me cause to think. I disagreed with the ideas being put forward, and I realised why I disagreed and why I prefer my own views to those being put forward. Let me explain.

The introductory lecture was given by Professor Stephen Ross. His theme was the effect of 'survivorship' on studies of manager performance. If you consider only fund managers that have been in business for the whole of a given period, you miss out those that have ceased business during the period. Since they may have ceased business because their performance was poor, the performance of the survivors is biased upwards. The same would be true of companies, though I do not recollect Ross saying this.

Ross then drew an analogy with rivers. Records of the level of the water in the Nile have been kept for millenia. The level appears to be statistically stationary. But, Ross argued, this was because the Nile was a survivor: river levels were more like random walks, but if the water level got too low the river dried up and if it got too high it became a lake (*laughter*). The Nile had been a survivor.

This analogy, though entertaining, does not support Ross's case, but it does support a different case. Over the few millenia we are taking about no major rivers have either dried up or become lakes. It is indeed in the nature of the water level in rivers to rise and fall, sometimes seasonally, and with more or less randomness, but still around some sort of central position. (Over geological time things may have been different.)

But so it is also with many investment time series. The dividend yield on ordinary shares seems to rise and fall about a central position, which varies from country to country but is typically in the 3% to 5% range. Interest rates too are stationary in the long run. The evidence for this also stretches back for millenia, as recorded by Sidney Homer (*A history of interest rates: 2000 B.C. to the present*, 1963).

The stationarity of dividend yields on shares is an example of the statistical time series concept known as cointegration, about which an increasing number of papers and books are appearing. The logarithms of share prices and of share dividends are cointegrated, so their difference, the logarithm of the yield, is stationary. The stationarity of dividend yields seems to have been first modelled by the Maturity Guarantees Working Party of the Institute of Actuaries and the Faculty of Actuaries in Britain in the late 1970s (see *Journal of the Institute of Actuaries*, **107**, 101-212, 1980), but it is now becoming recognised by other authors.

A technical difficulty is that a random walk, whether pure or modified (i.e. an I(1) series in the terminology of cointegration) has very similar short-run behaviour to a stationary autoregressive model (an I(0) series) with a rather slow mean reversion. Similarly, it is not possible to distinguish, in the short run, between a continuous Brownian motion and an Ornstein-Uhlenbeck process.

It is interesting that empirical time series investigation of stock market data on the eastern side of the Atlantic seems to have been more prolific than on the western side. This is not just because more actuaries have been involved in it, but it is at least partially so. Actuaries have longer memories and longer time horizons than some others in the investment market, and they should be able to make a helpful contribution to the world-wide discussion about investment modelling.

I am grateful to Stephen Ross for his analogy: but I submit that it supports my case better than it does his.

The second discussion in Orlando that I wish to refer to was one in which a number of papers on option pricing were discussed. One, by J. Ph. Jousseaume, made the useful distinction between what I shall call the 'actuarial value' of an option, and the 'arbitrage value'. For this purpose assume that the price of the underlying security performs a logarithmic Brownian motion, and consider the value of a European option. The distribution of the price of the security at the exercise data is lognormal, and it is easy to calculate the expected value of the option at this date. The formula involves the mean drift of the security price as well as the standard deviation. This expected value at maturity can then be discounted at some chosen interest rate to give a present value. This is the way actuaries traditionally calculate present values of contingent payments.

If we then follow the no-arbitrage line of argument, we substitute risk-neutral or martingale probabilities for the true probabilities. We replace the mean drift of the share price by the risk-free interest rate (with possibly also a term $\sigma^2/2$, depending on our definitions). We also use the risk-free interest rate for discounting, and we then obtain the Black-Scholes option pricing formula.

The no-arbitrage argument has gained great weight in recent years, and in the discussion Elias Shiu described it as 'the fundamental theorem of asset pricing'.

Why, then, do I think that the no-arbitrage argument is based on a fundamental misconception? First, the assets that most source investors, i.e. individuals, hold are not readily tradeable: these include private houses, pension rights and insurance policies. The financial intermediaries, those institutions whose liabilities are insurance policies and pension rights, equally cannot trade their liabilities. If these liabilities include options (which may be implicit or explicit) the intermediaries cannot necessarily set up a hedge portfolio with which to match these liabilities. They therefore have to consider the matching of assets and liabilities in a more traditional actuarial way, allowing for the possibility of mismatching, and allowing also for additional reserves to cover the risks of mismatching. Thus they need to value the options included in the liabilities using the actuarial value rather than the arbitrage value.

But even when trading is possible, as in many investment markets, perfect arbitrage is not possible. The problem is volume. In the derivation of the Black-Scholes diffential equation, it is asumed that the writer of a call option can set up a hedge portfolio by buying a suitable fraction of a share, *and that this does not affect the price of the security.*

It is plausible that the writing and hedging of a single option on an IBM share does not affect the price of IBM shares. But if the purchaser of the option asks for sufficiently many billion options, then the writer's hedging could only be done by putting forward a take-over bid for the company.

A fuller description of the process would include an additional term in the Black-Scholes differential equation, reflecting the change in the price of the security as a result of the hedging operation. This would make the equation more realistic, but a great deal harder to solve, because information about the sensitivity of share prices to volume demanded does not seem to have been readily modelled, although it has no doubt been investigated empirically.

The Maturity Guarantees Working Party that I have already referred to made essentially the same point when it was proposed to it that what is now called 'portfolio insurance' would allow life offices to offer policyholders an implicit put option on unit-linked contracts. The idea that life offices, in large volumes, could adopt a policy of selling shares when share prices were falling, without making them fall further, was considered by the Working Party to be unrealistic. Whether or not the stock market crash of October 1987 was caused by computer trading of portfolio insurance seems uncertain, but there is no doubt about the direction in which such trading would operate.

The question of volume is of relevance for the supervision of options markets. While many intermediaries may be able to balance their books appropriately, there must be source purchasers of options, and end writers who have not hedged; these may have cover in the form of matching securities or matching cash, but their ability to write options is limited by the cover they have available. The writer of uncovered options, like any insurance company writing risks that are not reinsured, requires additional solvency reserves.

This is not just an academic point. The solvency of the whole market in options trading would be threatened by the insolvency of any major payer. It is important that supervision is considered in an actuarial way, taking account of liquidity, mismatching and the solvency reserves that should be required.

This brings me to my third subject, one which was not inspired by any particular discussion at Orlando. The Capital Asset Pricing Model is an equilibrium model of the market, based on many assumptions, some of which can readily be relaxed without destroying the model, but others which seem to be both fundamental to it and mistaken. The three obstructions seem to me to be tradeability, *numeraires* and time horizons.

I have already noted that the personal assets of most individuals are not readily tradeable, nor are they readily sub-divisible; individuals therefore have very restricted investment opportunities. Life offices and pension funds consequently have substantial non-tradeable liabilities, which are not the same for all institutions. If they include their particular liabilities in a portfolio selection model, even if all the other assumptions of the CAPM are maintained, different institutions may well end up with different efficient frontiers (and not just different optimum portfolios along the same frontier).

Different investors and different institutions may also work using different numeraires. This may be because they work in different currencies, or because some work in 'real' and others in 'nominal' terms. The numeraire may be a matter of choice, or it may be directed by the liabilities. The life office with fixed money liabilities may well find a different efficient frontier from the pension fund whose liabilities are index-linked annuities, even if all other things are equal.

Time horizons may also be influenced by the liabilities. A mature pension fund may have a shorter horizon than a growing one, just as an older investor may have a shorter time horizon than a younger one. Fund managers who are judged on their quarterly performance may have even shorter time horizons. It may be possible to reconcile these different approaches by appealing to the uniformity of successive periods, but if my first proposition holds, that many investment series are stationary and autoregressive and thus are not independent from period to period, then a longer time period is not just a succession of identical shorter ones. The variance does not increase proportionately with time and the efficient frontier varies with the time horizon.

These three features lead me to conclude that any satisfactory equilibrium model needs to take account of the volume of investment in each sector, the volume committed to each type of non-traded liability, the volume using different numeraires, and the volume using different time horizons. The concepts are the same as those of 'market segmentation' in the bond market, different sectors of which may be dominated by investors with different tax positions or different 'preferred habitats' of duration or maturity date.

The equilibrium positions in such a model, taking account of volume, is much harder to find, and the elegant results derived from the CAPM may not be so readily forthcoming.

But actuaries should be among those who can recognise the individuality of each particular investment institution, and can adjust concepts of efficiency to match the requirements of each institution. We should not throw out the baby with the bathwater. Although I am criticising the naive CAPM applied on a global scale, the portfolio selection paradigm is undiminished, and may even be strengthened when it is realised that each institution has its own efficient frontier, and cannot necessarily rely on 'the market' finding the frontier for it.

In a debate at the Institute of Actuaries in London in March 1993 (see *Journal of the Institute of Actuaries*, **120**, 393-414, 1993), I proposed the motion that "*this house believes that the contribution of actuaries to investment could be enhanced by the work of financial economists*". I was ably seconded by Jim Tilley, and the opposing point of view was put forward by Terry Arthur and Robert Clarkson. I should like to propose to readers of *ASTIN Bulletin* and to members of AFIR that the reverse proposition is also true, that "*the work of financial economists in investment could be enhanced by the contribution of actuaries*". Let us see more papers to support this proposition in the pages of *ASTIN Bulletin*.

DAVID WILKIE

THE 4th AFIR INTERNATIONAL COLLOQUIUM

Disney World Orlando, the venue for the 4th AFIR Colloquium could not be more different from its predecessor, Rome. The organisers again made an inspired choice; the Colloquium being moved from the seat of one of the greatest concentrations of sights of artistic, architectural and historical interest in the world to the seat of one of the greatest concentrations of entertainment oriented, techical innovations.

The entertainment laid on by the hosts reflected the venue. The social highlight undoubtedly was the final evening visit to MGM studios. Around one thousand of the world's actuaries defied their traditional "grey suited" reputation by walking through a full scale New York Street scene protected from the spectacular thunderstorm by bright yellow Mickey Mouse ponchos. Crowds of screaming girls with autograph books were hired to greet us at the entrance and boost our egos; a 3-D Muppets Show, a trip on a Star Wars simulator and a spectacular fireworks display provided the memorable entertainement.

The format of the meeting, entertainment apart, was somewhat different from its predecessor. The AFIR Colloqiuim was combined with the Casualty Actuarial Society Special Interest seminar and Society of Actuaries' Spring Meeting. The result of this was a total of 82 sessions a number of which ran in parallel.

The Colloquium began with the main invited speaker Professor Stephen Ross of Yale University. He was there to defend the efficient market hypothesis and answer questions such as: do markets follow random walks? Do they have inherent cycles? Can we gain from fundamental investment analysis? Can we predict the success of good fund managers by looking at past performance?

Stephen Ross explained how the apparent ability of fund managers to outperform consistently could be explained by "survivorship bias". A similar phenomenon explained why market prices appeared to form cycles. There is a tendency for market analysts to look at historical data for surviving stock markets. A stock market which has risen to enormous heights (due to hyper-inflation) or collapsed to nothing (due to economic collapse) will not be amongst those which survive to have their course analysed. In a similar way, river levels appear to be cyclical: in fact they are not; any river the level of which has moved out of the fixed bounds has either flooded or dried up. It is no longer a river. Analysts tend not to take a group of rivers and chart their courses forwards in history: they take a group of surviving rivers and look backwards. Inevitably, the surviving rivers will have levels which have moved in cycles.

Many contributors debated the use of different risk measures or applied different risk measures to actuarial investment problems. The proponents of different risk measures fell into four "camps". Those who used traditional mean/variance approach to trading off risk and return; those who preferred downside measures of risk; those who preferred "shortfall controls" (probability of underperforming a particular benchmark); and those who preferred the use of utility theory as a risk management tool. The undercurrent of debate about risk measures went beyond the papers classified in that section. Some of the applied papers also contributed to the discussion on the measurement and control of investment risk.

There were various papers on stochastic investment modelling ranging from a simple ARCH approach to the use of transfer functions and time series models and one which involved elements of chaos theory. The application of stochastic investment models in simulation also featured prominently in the proceedings, especially in the areas of asset-liability management and measuring the risk of insurer insolvency.

A number of papers were submitted on the traditional stamping ground of immunization. One of the difficulties of immunization theory is the inability to deal with non-parallel yield curve movements. The problem was addressed in this section. Some demonstrated the wider fields potential of the profession by looking at credit risk, bank insolvency and money management.

The papers in the sections on Option Princing Techniques and a number of papers in the sections on Financial Instruments and the Analysis of Products with Investment Guarantees were connected with option pricing. So many of the applications of option pricing are in the actuarial field; it is clearly an area in which actuarial professions throughout the world need to take greater interest in both the theoretical and practical implications.

Three reports were also presented from task forces and study groups. Society of Actuaries' Task Forces presented reports on the application of cash flow techniques to pension plans and on losses from credit risk events. The Finnish Insurance Modelling Group reported on the suitability of possible stochastic investment models in the light of data available from twelve countries.

Finally, it was gratifying to have two contributions from Eastern Europe. Both discussed the Polish investment markets, one making pertinent comparisons with the Hungarian stock market. We hope that the proximity of the next two AFIRs to Eastern Europe will lead to more contributions from that area in both the technical and empirical subjects.

Each AFIR Colloquium in recent years seems to have been held in a place with its own special history and attractions (albeit a short history in the case of Orlando). We look forward to the next AFIR Colloquium in Brussels, the home of the European Union, in 1995.

PHILIP BOOTH
STEVEN HABERMAN
ALEN ONG

XXVth ASTIN COLLOQUIUM
CANNES, FRANCE, 11-15 september 1994.

The 25th Astin Colloquium was held at the Noga Hilton Hotel in Cannes. The Colloquium was attended by 163 participants with 43 accompanying persons from 22 countries. A welcoming cocktail took place on the Sunday evening at the terrasse Panorama of the Noga Hilton.

Björn AJNE, the Chairman of Astin, presided the opening ceremony on Monday morning. Lionel MOREAU, chairman of the organizing committee, and Pierre PETAUTON, chairman of the scientific committee, gave some precisions about the sessions format. Welcoming addresses were delivered by Mr J. LONGUET representing Mr M. MOUILLOT, mayor of Cannes, and by Mr BERTON, chairman of the Institut des Actuaires Français and also representing the chairman of the Institut des Sciences Financières et de l'Assurance de Lyon.

First invited lecture

Just after the opening ceremony, Mr D. KESSLER, chairman of the Fédération Française des Sociétés d'Assurance, gave the first invited lecture on the topic "L'Actuaire et l'évolution des assurances non-vie". Mr KESSLER detailed the factors explaining the results of the French non-life insurance companies over the last years and examined the new challenges emerging in the non-life insurance market.

First session: Great risks (topic nr 3)

This session, held during the remainder of monday morning, was chaired by C. STOOP and D. SKURNICK, Ch. LEVI reported on the four presented papers.

Z. BENABBOU and C. PARTRAT proposed an alternative to the truncation of the very big claims amounts commonly used for the tariff construction. This alternative consists in using a mixture of two distributions for modelling the claims amounts, the first distribution corresponding to ordinary claims, the second to exceptionnally big claims. W. HÜRLIMAN spoked about the hedging through a reinsurance agreement of the liabilities of an insurer under contracts with a claims dependent bonus provision. S. BERNEGGER, F. KRIETER, P. MEYER and A. BLOCH presented a risk model pemitting the derivation of the fluctuation loading for cat-portfolios with correlated layers and the breaking down of the overall loading between the different reinsurance contracts. J. P. CASANOVA and E. DUBREUIL explained the mechanism of catastrophe insurance futures and options sold on the Chicago Board of Trade, and the strategies available to an insurer for hedging his liabilities through these futures and options.

Second session: Financial stability of a non-life insurance company (topic nr 1)

This session held on Monday afternoon was chaired by D. BLANCHARD and T. CLARKE, R. DE LAROULLERE reported on the four presented papers. B. AJNE and E. JOHANSSON developed expressions for the expected value and the variance of the undiscounted and discounted claims reserves in a model allowing for stochastic variations in both the payment pattern and the interest rate. A. RENSHAW and R. VERRALL presented a generalised linear model for claims reserving producing the same results as the chain ladder technique. L. TRAN VAN LIEU insisted that any tariff decision has to be analysed with reference to the principles of modern corporate finance. J. JANSSEN presented a dynamic stochastic ALM model with assets and liabilities splitted into possibly correlated segments.

Third session: Financial stability of a non life insurance company (continued)

This session, hold on the remaining of Monday afternoon, was chaired by J.P. CASANOVA and M. GOOVAERTS, P. PETAUTON reported on the three presented papers.

W. HÜRLIMANN examined different stability criteria with respect to equity and technical premiums of an insurance company. L. CENTENO and J. ANDRADE E SILVA used the Bühlmann-Straub model to calculate the solvency ratio for some Portuguese non-life insurance companies. D.A. STANFORD and K.J. STROINSKI developed recursive algorithms for computing ruin probabilities at claim instants in the classical Poisson model with phase type distributed claim amounts. The method extends to the case of gamma distributed inter-claim times.

Second invited lecture

The second invited lecture was given by Professor Paul EMBRECHTS on Wednesday morning. Professor EMBRECHTS evidenced the necessity of filling the gap between theory and practice in actuarial science. In fact there must be a constant interaction between these two poles: new questions faced by the practitionners are the source of development of new techniques by the theoricians. Conversely, new techniques are to be brought to the practitionners facing the new questions. Professor EMBRECHTS illustrated this interactive process by some examples (application of Itô's calculus in the field of insurance futures; robust statistics based on a generalisation of Taylor's expansion applied in Credibility theory...).

Fourth session: risk selection and setting premium rates (topic nr 2)

This session, hold on the remaining of wednesday morning, was chaired by M.C. CHEYMOL and C. PATRICK, Ph. MARIE-JEANNE reported on the presented papers. F. BOULANGER developed a bonus-malus system for policies with several correlated covers.

F. BOULANGER and L. TRAN VAN LIEU considered the problem of determining an optimal premium rate taking into account the elasticity of the demand for insurance. J. LEMAIRE and H. ZI simulated and compared with respect of four different measures the bonus-malus systems in force in 22 countries for third party liability automobile insurance; they used factor analysis to define an "index of thoughness". E. LEVAY exemplified the use of historical data in modelling UK motor business. G. RAMACHANDRAN adressed the problem of determining premium rebates for deductibles which take into account fire protection measures adopted by the insured.

Fifth session: risk selection and setting premium rates (continued)

This session, hold on wednesday afternoon, was chaired by Ph. PICCARD and T. HOYLAND, J.L. BESSON reported on the five presented papers. F. BOULANGER advocated the use of the gamma distribution with random mean for estimating the claim size distribution. D. DANNENBURG compared different credibility estimators in the Hachemeister's regression credibility model. W. HÜRLIMANN established a result by SCHMITTER about the maximal stop-loss variance; he then examined the stop-loss retention minimising the total variance after reinsurance when the cedent and the reinsurer use the same variance principle. A. RENSHAW formulated the two components of the claims process, frequency and severity, by postulating two separate generalised linear models; this makes available other distributions than the classical Poisson distribution for frequencies and the gamma distribution for severities. O. HESSELAGER developed recursions for certain bivariate counting distributions and the corresponding compound distributions.

Sixth session: Speaker's Corner

This session, held on thursday morning, was heavily filled with twelve communications. Unfortunately only a few authors gave a written version of their talk. The authors and the titles of their talks are listed hereunder.

ASTIN General Assembly

The general assembly was held on Wednesday afternoon. It was confirmed that the next colloquium will take place in Louvain (Belgium) in September 1995, just after the Centenary International Congress of the I.A.A. The following ASTIN colloquium will be held in Copenhagen from 1 to 5 September 1996. The 1997 colloquium is expected to take place in August in Cairns, Australia, just before the Centenary of the Institute of Actuaries of Australia.

B. AJNE resigned from his duties of chairman. He will be replaced by J. STANARD. The viceChairman will be J. RANTALA.

An enquiry form about the format of the ASTIN colloquia was distributed and answered by the participants during the colloquium. Some first results from that enquiry were commented by J. STANARD. The answers will of course need a further detailed analysis.

Social programme

The social program consisted of a concert given on Monday evening in the Noga Hilton by the group "Les solistes de Cannes" and by the pianist David LEVY, and a full day tour on Tuesday: we visited a retrospective devoted to G. BRAQUE at the Foundation Maeght, the beautiful 11th century fortified town of Saint-Paul-de-Vence and the Fragonard perfume factory in Grasse. The colloquium dinner took place on Wednesday evening at the Hôtel Martinez on the Croisette. During the dinner, B. AJNE and P. PETAUTON made an assault of poetry; hereafter are their respective contributions. With this second experience, after a first one in Cambridge, the organizers of the forthcoming colloquia should consider to devote a session to actuarial poems (this should of course not inhibit the poetic accents usually emerging from most of the scientific papers).

There was un Colloque en Cannes
Petauton et Moreau made it fun
Many others be praised
Hard problems we faced
Non-life studies are not zero-one.

(B. AJNE)

Si j'étais un poète et si j'étais galant
J'aurais dû cette nuit complimenter les dames.
Croyez que j'eusse aimer posséder ce talent,
Pour célébrer ici la beauté de la femme.
Laissez-moi, je vous prie, essayer de rimer
Sur un sujet moins noble, un souvenir d'actuaire.
Déjà l'été dernier, notre collègue aimé,
L'éminent Bjorn Sundt, là-bas dans l'Angleterre
Nous avait dit un soir un aimable sonnet,
Lassé de corriger les formules inexactes.
C'était un parapluie qu'on nous avait donné
Qui fut l'inspiration et provoqua cet acte.
Pourtant l'objet là-bas n'était pas singulier
Et voilà qu'en Provence on pleure son absence.
Si le plus mal chaussé c'est bien le cordonnier,
L'assureur en ce jour manque de prévoyance.
Une ombrelle eut au moins évité les pépins.
Sommes-nous à l'abri dans cette salle d'attente?
Noyés sous un sprinkler et sans maillot de bain
Est une circonstance, il est vrai peu fréquente,
Mais dont on aurait dû évaluer le coût.
Mais craignant de lasser votre aimable patience,
Et prenant en pitié les anglophones à bout,
J'arrête en cet instant mon flot d'incohérences
Et mon propos de fin ne sera plus qu'un souhait
De mathématicien qui voit la triste époque.
Puisse le monde enfin, devenir à jamais
Beau comme un théorème à la fin d'un colloque.

(P. PETAUTON)

Warm thanks are due to L. MOREAU, P. PETAUTON and their organizing and scientific committees for this very successfull colloquium.

J.M. Reinhard, Bruxelles

List of papers

Topic 1

B. Ajne & E. Johansson
On the random structure of discounted loss reserves.

L. Centeno & J. Andrade e Silva
Applying credibility theory to solvency.

W. Hürlimann
On stable insurance business models.

A. Renshaw & R. Verall
The stochastic model underlying the chain-ladder technique.

L. Tran Van Lieu
La maximisation de la valeur comme objectif de rentabilité et outil de mesure d'une branche d'assurance dommage. Application au marché français de l'assurance automobile.

J. Janssen
A dynamic stochastic ALM model for insurance companies.

A. Egidio dos Reis
More about the time to recovery.

D. Stanford & K. Stroinski
Recursive algorithms for computing ruin probabilities at claim instants.

Topic 2

F. Boulanger
Estimation du coût des sinistres.

F. Boulanger
Systèmes de bonus malus multi-garanties.

F. Boulanger & L. Tran Van Lieu
Détermination du tarif commercial optimal dans un environnement concurrentiel.

D. Dannenburg
Estimators in the regression credibility model.

W. Hürlimann
From the inequalities of Bowers, Kremer and Schmitter to the total stop-loss risk.

E. Kremer
Robust credibility via robust Kalman filtering.

J. Lemaire & H. Zi
A comparative analysis of 30 bonus-malus systems.

E. Levay
Use of historical data in UK motor modelling.

G. Ramachandran
Rebates for deductibles and protection measures in industrial fire insurance.

A. RENSHAW
A note on some practical aspects of modelling the claims process in the presence of covariates.

O. HESSELAGER
Recursions for certain bivariate counting distributions and their compound distributions.

Topic 3

Z. BENABBOU & C. PARTRAT
Grands sinistres et lois mélanges.

W. HÜRLIMANN
Experience rating and reinsurance.

S. BERNEGGER, F. KRIETER, P. MEYER & A. BLOCH
Risk-theoretical fluctuation loading for the case of correlated layers.

J.P. CASANOVA & E. DUBREUIL
Le marché à terme des contrats d'assurance au C.B.O.T.

Speaker's corner

Ch. LEVI & J.F. BEAUTEMS
Couverture d'un portefeuille de traités américains par des contrats à terme "Cat. Nat." du CBOT.

W. HÜRLIMANN
An actuarial selection of orthogonal parameters to the mean.

E. PITACCO
LTC insurance: from the multistate model to practical implementations.

D.E.A. SANDERS
Reserving for catastrophes – the five S'.

J.P. VERLE
Etude de l'évolution des comptes techniques de l'assurance automobile en France.

Ch. GEOFFRAY & G. DUREUIL
Modélisation bivariée de fréquences de sinistres dépendantes.

M. LAPARRA, I. LION & C. PARTRAT
Design and analysis of market prices for the U.S. natural catastrophe excess reinsurance treaties.

L. SIGALOTTI
Equilibrium premium in a bonus-malus system.

L. PICECH
The merit rating factor in a multiplicative rate making system.

G. PARKER
A second order stochastic differential equation for the force of interest.

Ph. ARTZNER
Hedging of insurance contracts.

F. DUFRESNE
The efficiency of the Swiss bonus-malus system.

FURTHER RESULTS ON HESSELAGER'S RECURSIVE PROCEDURE FOR CALCULATION OF SOME COMPOUND DISTRIBUTIONS *

BY SHAUN WANG

University of Waterloo, Canada

AND MONICA SOBRERO

Università "La Sapienza", Italy

ABSTRACT

The recursive algorithm of HESSELAGER (1994) is extended to a more general class of counting distributions, which includes SUNDT's (1992) class as well as all the mixed Poisson distributions discussed by WILLMOT (1993).

KEYWORDS

Compound distributions; recursions; mixed Poissons; Sichel; Beta; Generalized Pareto; Inverse Gamma.

1. INTRODUCTION

In the collective risk model, compound distributions are used extensively in modeling the total claim for an insurance portfolio:

$$S = \sum_{i=1}^{N} X_i, \tag{1}$$

where the claim sizes X_i's are independent and identically distributed and independent of the claim frequency N.

If the claim frequency N has a probability function (p.f.) $\{p_0, p_1, \ldots\}$, and the claims sizes $\{X_1, X_2, \ldots\}$ have a common p.f.:

$$f_x = \Pr\{X = x\}, \qquad x = 0, 1, \ldots,$$

then the total claim S has a compound distribution with a p.f.

$$g_x = \sum_{n=0}^{\infty} p_n f_x^{*n}. \tag{2}$$

Since PANJER (1981), many resursive algorithms have been derived for a broad class of claim frequency distribtions (see WILLMOT and PANJER, 1987; SUNDT, 1992; WILLMOT, 1993; and others).

* The authors are grateful to the editor and the referees for their helpful comments.

HESSELAGER (1994) recently considered a class of claim frequency distributions satisfying:

$$(3) \qquad p_n = \frac{\sum_{i=0}^{k} a_i n^i}{\sum_{i=0}^{k} b_i n^i} p_{n-1}, \qquad n = 1, 2, \ldots$$

for some positive integer k, and derived recursions for the related compound distributions. For some counting distributions such as Generalized Waring, Hypergeometric and Polya-Eggenberger, Hesselager's method is more efficient than the ones provided by WILLMOT and PANJER (1987). However, Hesselager's class (3) does not include many other counting distributions such as Sichel, Poisson-Beta, Poisson Generalized Pareto, Poisson Inverse Gamma, Poisson Transformed Gamma and Poisson Transformed Beta (see WILLMOT, 1993).

In this paper, we extend Hesselager's recursive scheme to a broader family of counting distributions, which includes al the counting distributions which satisfy a finite order homogeneous recursion with polynomial coefficients. All the mixed Poisson distributions in WILLMOT (1993) are of this type (where the non-homogeneous terms can be eliminated).

2. RECURSIONS FOR THE EXTENDED CLASS OF COUNTING DISTRIBUTIONS

Assume that the claim frequency N has a p.f. $\{p_0, p_1 \ldots\}$ satisfying:

$$(4) \qquad \left(\sum_{i=0}^{k} b_i n^i\right) p_n = \sum_{j=1}^{s} \left\{\sum_{i=0}^{k} a_{j,i} (n-j)^i\right\} p_{n-j}, \qquad n = c, c+1, \ldots$$

where c is a positive integer and $p_n = 0$ for $n < 0$.

HESSELAGER (1994) introduced the following auxiliary functions

$$(5) \qquad g_{i,x} = \sum_{n=0}^{\infty} n^i p_n f_x^{*n}, \qquad x = 0, 1, \ldots; \qquad (i = 0, 1, \ldots)$$

(with $0^0 = 1$) and defined the vector

$$\vec{g}_x = (g_{0,x}, \ldots, g_{k,x})' .$$

Note that $g_{0,x}$ is the p.f. for the total claim distribution of S in (2).

Before we generalize Hesselager's result, we first introduce another auxiliary function:

$$(6) \qquad \Omega(x) = \sum_{n=1}^{c-1} \sum_{i=0}^{k} b_i n^i p_n f_x^{*n} - \sum_{j=1}^{s} \sum_{n=j}^{c-1} \sum_{i=0}^{k} a_{j,i} (n-j)^i p_{n-j} f_x^{*n},$$

where $\Sigma_{n=j}^{c-1}$ is zero if $j > c-1$. By letting $a_{0,i} = -b_i$ the expression for $\Omega(x)$ may be written as

$$\Omega(x) = -\sum_{j=0}^{s} \sum_{n=j}^{c-1} \sum_{i=0}^{k} a_{j,i} (n-j)^i p_{n-j} f_x^{*n} .$$

Note that $\Omega(x)$ does not depend upon the values of $\vec{g}_x$. In the special case of $k = s = c = 2$, we have

$$\Omega(x) = p_1 f_x \sum_{i=0}^{2} b_i - p_0 f_x a_{1,0}. \tag{7}$$

As in HESSELAGER (1994), let m be the smallest integer for which $f_m > 0$.

Theorem 1. For the claim frequencies in (4), the compound distribution $g_x = g_{0,x}$ can be evaluated by the following recursive method:

$$\vec{g}_x = T_x^{-1} \vec{t}_x, \qquad x \geq m \vee 1, \tag{8}$$

where

$$T_x = \begin{pmatrix} 1 & -m/x & 0 & \dots & 0 \\ 0 & 1 & -m/x & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (b_0 - \Sigma_{j=1}^{s} a_{j,0} f_0^j) & (b_1 - \Sigma_{j=1}^{s} a_{j,1} f_0^j) & \dots & \dots & (b_k - \Sigma_{j=1}^{s} a_{j,k} f_0^j) \end{pmatrix}, \tag{9}$$

and $\vec{t}_x = (t_{0,x}, \dots, t_{k,x})'$ is given by

$$t_{i,x} = \frac{1}{f_m} \sum_{y=1}^{x} f_{m+y} \left\{ \frac{m+y}{x} g_{i+1,x-y} + \frac{y-x}{x} g_{i,x-y} \right\}, \qquad i < k; \tag{10}$$

$$t_{k,x} = \sum_{y=1}^{x} \sum_{i=0}^{k} \sum_{j=1}^{s} a_{j,i} f_y^{*j} g_{i,x-y} + \Omega(x), \tag{11}$$

with starting values

$$g_{i,0} = \Sigma_{n=0}^{\infty} n^i p_n f_0^n, \qquad i = 0, \dots, k;$$

$$g_{i,x} = 0, \qquad i = 0, \dots, k; \qquad x = 1, \dots, m-1.$$

Proof: Let $F(z) = \Sigma_{x=0}^{\infty} f_x z^x$ be the probability generating function for the claim size distribution.

From the identity $\frac{d}{dz}[F(z)^n] = nF(z)^{n-1} F'(z)$, we have

$$0 = \sum_{y=0}^{x} \left[(n+1) \frac{y}{x} - 1 \right] f_y f_{x-y}^{*n}. \tag{12}$$

Multiplying (12) by $p_n n^i$ and summing over $n \geq 0$ yields

$$0 = \sum_{y=0}^{x} f_y \left\{ \frac{y}{x} g_{i+1,x-y} + \left(\frac{y}{x} - 1 \right) g_{i,x-y} \right\}. \tag{13}$$

By omitting the zero terms corresponding to $y = 0, \ldots, m-1$ and taking out the terms involving $\vec{g}_x$, we get for $i < k$,

(14) $$g_{i,x} - \frac{m}{x} g_{i+1,x} = t_{i,x},$$

where $t_{i,x}$ is given in (10).

Note that $f_x^{*n} = \Sigma_{y=0}^{x} f_y^{*j} f_{x-y}^{*(n-j)}$ for $j = 1, \ldots, s$.

Multiplying the left-hand side of (4) by f_x^{*n}, multiplying the j-th term of the right-hand side of (4) by $\Sigma_{y=0}^{x} f_y^{*j} f_{x-y}^{*(n-j)}$, and summing over $n \geq c$, we obtain the relation (for $x \geq 1$):

$$\sum_{i=0}^{k} b_i g_{i,x} - \sum_{n=1}^{c-1} \sum_{i=0}^{k} b_i n^i p_n f_x^{*n} = \sum_{y=0}^{x} \sum_{i=0}^{k} \sum_{j=1}^{s} a_{j,i} f_y^{*j} g_{i,x-y}$$
$$- \sum_{j=1}^{s} \sum_{n=j}^{c-1} \sum_{i=0}^{k} a_{j,i} (n-j)^i p_{n-j} f_x^{*n}.$$

By collecting the leading terms involving $g_{i,x}$, we get

(15) $$\sum_{i=0}^{k} \left(b_i - \sum_{j=1}^{s} a_{j,i} f_0^j \right) g_{i,x}$$
$$= \sum_{y=1}^{x} \sum_{i=0}^{k} \sum_{j=1}^{s} a_{j,i} f_y^{*j} g_{i,x-y} + \Omega(x), \qquad x \geq 1,$$

where $\Omega(x)$ is as defined in (6).

From equations (14) and (15), we obtain the linear equations (8). □

Remark 1: As a special case, for $m = 0$ (i.e. $f_0 > 0$), we have the following recursions:

(16) $$g_{i,x} = \frac{1}{f_0} \sum_{y=1}^{x} f_y \left\{ \frac{y}{x} g_{i+1,x-y} + \frac{y-x}{x} g_{i,x-y} \right\}, \qquad i < k;$$

(17) $$g_{k,x} = \frac{1}{b_k - \Sigma_{j=1}^{s} a_{j,k} f_0^j} \times$$
$$\left\{ \sum_{y=1}^{x} \sum_{i=0}^{k} \sum_{j=1}^{s} a_{j,i} f_y^{*j} g_{i,x-y} + \sum_{i=0}^{k-1} \left(\sum_{j=1}^{s} a_{j,i} f_0^j - b_i \right) g_{i,x} + \Omega(x) \right\}.$$

Remark 2: HESSLAGER's (1994) recursive formula can be recovered from (8) as a particular case of $s = c = 1$.

3. EXAMPLES OF MIXED POISSON DISTRIBUTIONS

Mixed Poisson distributions are natural candidates for modeling the claim frequency for heterogenous risk portfolios. WILLMOT (1993) considered various mixed Poisson distributions and derived recursions for their probability functions. Some of the mixed Poisson distributions belong to the Pearson system (see ORD, 1972, p. 8) and WILLMOT (1993) derived recursions for their compound distributions. However, many mixed Poisson distributions, which are not in the Pearson system, belong to the generalized class (4). So our generalized recursive procedure can be used in the calculation of the compound distributions for these mixed Poisson frequencies.

Example 1: The Sichel distribution is obtained by mixing the Poisson mean over the Generalized Inverse Gaussian. WILLMOT (1993) derives a recursion for this mixed Poisson p.f. (for $n = 2, 3, \ldots$)

$$(1 + 2\beta)\, n(n-1) p_n = 2\beta (n-1)(n + \lambda - 1) p_{n-1} + \mu^2 p_{n-2}, \tag{18}$$

which corresponds to (4) with $k = s = c = 2$ and

$$\begin{array}{lll} b_0 = 0, & b_1 = -(1+2\beta), & b_2 = 1 + 2\beta; \\ a_{1,0} = 0, & a_{1,1} = 2\beta\lambda, & a_{1,2} = 2\beta; \\ a_{2,0} = \mu^2, & a_{2,1} = 0, & a_{2,2} = 0. \end{array}$$

Example 2: The Poisson Beta is obtained by mixing the Poisson mean over the Beta distribution. WILLMOT (1993) derives a recursion for this mixed Poisson p.f. (for $n = 2, 3, \ldots$)

$$n(n-1) p_n = (n-1)(n - 2 + \mu + \alpha + \beta) p_{n-1} - \mu (n - 2 + \alpha) p_{n-2}, \tag{19}$$

which corresponds to (4) with $k = s = c = 2$ and

$$\begin{array}{lll} b_0 = 0, & b_1 = -1, & b_2 = 1; \\ a_{1,0} = 0, & a_{1,1} = \mu + \alpha + \beta - 1, & a_{1,2} = 1; \\ a_{2,0} = -\mu\alpha, & a_{2,1} = -\mu, & a_{2,2} = 0. \end{array}$$

Example 3: The Poisson Generalized Pareto is obtained by mixing the Poisson mean over the Generalized Pareto. WILLMOT (1993) derives a recursion for this mixed Poisson p.f. (for $n = 2, 3 \ldots$)

$$n(n-1) p_n = (n-1)(n - 1 - \alpha - \mu) p_{n-1} + \mu (n - 2 + \beta) p_{n-2}, \tag{20}$$

which corresponds to (4) with $k = s = c = 2$ and

$$\begin{array}{lll} b_0 = 0, & b_1 = -1, & b_2 = 1; \\ a_{1,0} = 0, & a_{1,1} = -(\alpha + \mu), & a_{1,2} = 1; \\ a_{2,0} = \mu\beta, & a_{2,1} = \mu, & a_{2,2} = 0. \end{array}$$

Example 4: The Poisson Inverse Gamma is obtained by mixing the Poisson mean over the Inverse Gamma. WILLMOT (1993) derives a recursion for this mixed Poisson p.f. (for $n = 2, 3, \ldots$)

$$n(n-1) p_n = (n-1)(n - 1 - \alpha) p_{n-1} + \mu p_{n-2}, \tag{21}$$

which corresponds to (4) with $k = s = c = 2$ and

$$\begin{array}{lll} b_0 = 0, & b_1 = -1, & b_2 = 1; \\ a_{1,0} = 0, & a_{1,1} = -\alpha, & a_{1,2} = 1; \\ a_{2,0} = \mu, & a_{2,1} = 0, & a_{2,2} = 0. \end{array}$$

Remark: In all the above examples $\Omega(x)$ vanishes, which simplifies equation (8). However, T_x becomes a singular matrix at $x = m$, so the recursive evaluation by (8) should start with initial values $\{\vec{g}_0, \ldots, \vec{g}_m\}$. Other mixed Poisson distributions in the generalized class (4) are Poisson Transformed Gamma and Poisson Transformed Beta (see WILLMOT, 1993).

It is noted that SUNDT's (1992) class is a particular case of (4) when $k = 1$. However, when $k = 1$, Sundt's recursion involves only $g_{0,x}$, while recursion (8) requires both $g_{0,x}$ and $g_{1,x}$. In the same way, for mixed Poisson distributions in the Pearson system, Willmot's recursive method (see WILLMOT, 1993) is simpler than recursion (8) in evaluating their compound distributions. On the other hand, when $k \geq 2$, as in the earlier examples of this section where $k = s = c = 2$, the recursion (8) is more efficient than the ones given in WILLMOT and PANJER (1987).

The numerical aspects such as stability concerns of the recursion (8) need further study.

REFERENCES

HESSELAGER, O. (1994) A recursive procedure for calculation of some compound distributions. *ASTIN Bulletin*, to appear.

ORD, J. K. (1972) *Families Of Frequency Distributions*. Griffin, London.

PANJER, H.H. (1981) Recursive evaluation of a family of compund distributions. *ASTIN Bulletin* **12**, 22–26.

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

WILLMOT, G.E. (1993) On recursive evaluation of mixed Poisson probabilities and related quantities. *Scandinavian Actuarial Journal*, to appear.

WILLMOT, G.E. and PANJER, H.H. (1987) Difference equation approaches in evaluation of compound distributions. *Insurance: Mathematics and Economics* **6**, 43–56.

The mailing addresses of the authors are:

SHAUN WANG
Dept. of Statistics & Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

AND MONICA SOBRERO
Dipartimento di Scienze Attuariali e Matematica
Facoltà di Economia e Commercio
Università "La Sapienza", Roma, Italia

TWO STOCHASTIC APPROACHES FOR DISCOUNTING ACTUARIAL FUNCTIONS

BY GARY PARKER

Simon Fraser University

ABSTRACT

Two approaches used to model interest randomness are presented. They are the modeling of the force of interest accumulation function and the modeling of the force of interest. The expected value, standard deviation and coefficient of skewness of the present value of annuities-immediate are presented as illustrations. The implicit behavior of the force of interest under the two approaches is investigated by looking at a particular conditional expectation of the force of interest accumulation function.

KEYWORDS

Force of interest; Force of interest accumulation function; White Noise process; Wiener process; Ornstein-Uhlenbeck process; Present value function; Annuity-immediate.

1. INTRODUCTION

A wide variety of stochastic processes have been used to model interest randomness in the present value function and other actuarial functions. Not only are different processes used but they are also used in different ways. Two approaches that are used in existing literature are, firstly, the modeling of the force of interest accumulation function (see, for example, DEVOLDER (1986), BEEKMAN and FUELLING (1990, 1991, 1993), DE SCHEPPER et al. (1992a, 1992b), DE SCHEPPER and GOOVAERTS (1992)), and secondly, the modeling of the force of interest (see, for example, PANJER and BELLHOUSE (1980), DHAENE (1989), FREES (1990), PARKER (1992, 1993a, 1993b, 1994), NORBERG (1993)). The particular assumption that the forces of interest are independent and identically distributed (i.e. a White Noise process) will be seen to have an equivalent process for the force of interest accumulation function. IID interest notes have been used by WATERS (1978, 1990), DUFRESNE (1990) and PAPACHRISTOU and WATERS (1991) among others.

Although in the deterministic situation the two approaches are equivalent, they are truly different in the stochastic situation.

In this paper, we compare these two approaches for some simple Gaussian processes (see PARKER (1993c) for an earlier version presented at the XXIV ASTIN Colloquium). In Section 2, we define the random present value function and give an expression for its moments about the origin.

In Section 3, we present two stochastic processes, namely, the Wiener process and the Ornstein-Uhlenbeck process, for the force of interest accumulation function. The following section presents three stochastic processes, the White Noise, Wiener and Ornstein-Uhlenbeck processes, for modeling the force of interest.

In Section 5, we find the first three moments about the origin of the random present value of a n-year annuity-immediate of equal payments of 1. Some illustrations are presented in Section 6. Section 7 takes a closer look at an implicit difference between the two approaches. Finally, Section 8 summarizes the findings.

2. PRESENT VALUE FUNCTION

Let δ_s denote the force of interest at time s and let $y(t)$ denote the force of interest accumulation function at time t. We then have

$$y(t) = \int_0^t \delta_s \, ds . \tag{1}$$

The random present value at time 0 of a payment of 1 at time t is given by $e^{-y(t)}$.

Assuming that $y(t)$ is Gaussian, then the present value function is log-normally distributed with parameters $E[-y(t)]$ and $V[y(t)]$, and its mth moment about the origin is:

$$E[(e^{-y(t)})^m] = E[e^{-m \cdot y(t)}] = \exp\{-m \cdot E[y(t)] + .5 m^2 \cdot V[y(t)]\} \tag{2}$$

(see, for example, AITCHISON and BROWN (1963, p. 8)).

In the next section we will use two Gaussian stochastic processes to model the force of interest accumulation function. And, in the following section, Section 4, we will look at three Gaussian stochastic processes to model the force of interest.

3. MODELING THE FORCE OF INTEREST ACCUMULATION FUNCTION

A first approach to consider interest randomness is to model $y(t)$, the force of interest accumulation function. Here we present a Wiener process with deterministic drift δ and an Ornstein-Uhlenbeck process also with deterministic drift δ.

3.1. Wiener process

Let $y(t)$ be the sum of a deterministic drift of slope δ and a perturbation modeled by a Wiener process. That is

$$y(t) = \delta \cdot t + \sigma \cdot W_t , \tag{3}$$

where $\sigma \geq 0$ and W_t is the standardized Wiener process.

It can be shown that the expected value and autocovariance function of $y(t)$ are given by

$$E[y(t)] = \delta \cdot t , \tag{4}$$

and

$$\text{cov}\,[y(s), y(t)] = \sigma^2 \cdot \min(s, t). \tag{5}$$

(see ARNOLD (1974, Section 3.2)).

3.2. Ornstein-Uhlenbeck process

Let $y(t)$ be the sum of a deterministic drift of slope δ and a perturbation modeled by an Ornstein-Uhlenbeck process. That is

$$y(t) = \delta \cdot t + X(t), \tag{6}$$

where $X(t)$ is an Ornstein-Uhlenbeck process with parameters $\alpha \geq 0$ and $\sigma \geq 0$ and with an initial condition $X(0) = 0$. Therefore,

$$dX(t) = -\alpha \cdot X(t)\,dt + \sigma\,dW_t. \tag{7}$$

Using the results of ARNOLD (1974, p. 134), one can obtain the expected value and autocovariance function of $y(t)$ as defined in (6) and they are given by

$$E[y(t)] = \delta \cdot t, \tag{8}$$

and

$$\text{cov}\,[y(s), y(t)] = \frac{\sigma^2}{2\alpha} \cdot (e^{-\alpha(t-s)} - e^{-\alpha(t+s)}), \qquad s \leq t \tag{9}$$

or

$$\text{cov}\,[y(s), y(t)] = \rho^2 \cdot (e^{-\alpha(t-s)} - e^{-\alpha(t+s)}), \qquad s \leq t \tag{10}$$

where

$$\rho^2 = \frac{\sigma^2}{2\alpha}. \tag{11}$$

4. MODELING THE FORCE OF INTEREST

A second approach to model interest randomness is to model δ_s, the force of interest. Here we present a White Noise process, a Wiener process and an Ornstein-Uhlenbeck process. Note that the three processes will be defined so that they start at δ, not at the origin.

4.1. White Noise process

Let the force of interest be a White Noise process with mean δ and variance σ^2. That is, for $t > 0$,

$$\delta_t \sim N(\delta, \sigma^2). \tag{12}$$

The forces of interest are therefore modeled by Gaussian, independent and identically distributed random variables. Note that, in continuous time, White Noise is not a physical process but a mathematical abstraction (see KARLIN and TAYLOR (1981, p. 343)).

One may consider, in some sense, that the White Noise process is the derivative of the Wiener process (see, for example, ARNOLD (1974, p. 53) of KARLIN and TAYLOR (1981, p. 342)). (This indicates that assuming a stochastic process for $y(t)$ does not necessarily imply that a meaningful physical process for δ_t exist).

Then, $y(t)$, as defined in (1), is a Wiener process with expected value

$$E[y(t)] = \delta \cdot t, \tag{13}$$

and autocovariance function

$$\operatorname{cov}[y(s), y(t)] = \sigma^2 \cdot \min(s, t). \tag{14}$$

(see, for example, ARNOLD (1974, Section 3.2)).

Therefore, the model presented above is merely an alternative description of the Wiener process for the force of interest accumulation function presented in 3.1.

4.2. Wiener process

A second model for the force of interest is the Wiener process. Let the force of interest be defined as

$$\delta_t = \delta + \sigma \cdot W_t, \qquad \sigma \geq 0. \tag{15}$$

Adapting the results in Section 3.1 we find that the expected value and autocovariance function of this process are

$$E[\delta_t] = \delta, \tag{16}$$

and

$$\operatorname{cov}[\delta_s, \delta_t] = \sigma^2 \cdot \min(s, t). \tag{17}$$

Then, from the definition of $y(t)$ (see (1)), it follows that $y(t)$ is normally distributed with expected value

$$E[y(t)] = \delta \cdot t, \tag{18}$$

and autocovariance function

$$\operatorname{cov}[y(s), y(t)] = \int_0^s \int_0^t \operatorname{cov}[\delta_u, \delta_v] \, du \, dv, \tag{19}$$

which gives

$$\operatorname{cov}[y(s), y(t)] = \sigma^2 \cdot (s^2 t/2 - s^3/6), \qquad s \leq t. \tag{20}$$

4.3. Ornstein-Uhlenbeck process

As a third model for the force of interest we consider an Ornstein-Uhlenbeck process. Let the force of interest be defined by the following stochastic differential equation

$$d\delta_t = -\alpha(\delta_t - \delta) \cdot dt + \sigma \cdot dW_t \qquad \alpha \geq 0, \quad \sigma \geq 0, \tag{21}$$

with initial value $\delta_0 = \delta$ (see, for example, ARNOLD (1974, p. 134)).

Then, it can be shown that the expected value of δ_t is

$$E[\delta_t] = \delta, \tag{22}$$

and that its autocovariance function is

$$\operatorname{cov}[\delta_s, \delta_t] = \frac{\sigma^2}{2\alpha} \cdot (e^{-\alpha(t-s)} - e^{-\alpha(t+s)}), \qquad s \leq t. \tag{23}$$

Again, we will denote $\sigma^2/2\alpha$ by ρ^2.

The force of interest accumulation function, $y(t)$, is therefore a Gaussian process with expected value

$$E[y(t)] = \delta \cdot t, \tag{24}$$

and autocovariance function

$$\operatorname{cov}[y(s), y(t)] = \frac{\sigma^2}{\alpha^2} \min(s, t) + \tag{25}$$

$$+ \frac{\sigma^2}{2\alpha^3} [-2 + 2e^{-\alpha s} + 2e^{-\alpha t} - e^{-\alpha|t-s|} - e^{-\alpha(t+s)}].$$

(see, for example, PARKER (1994, Section 6)).

Note that the two models considered in Section 3 and the three models considered in this section have all been defined such that their expected values of the force of interest accumulation function are the same (i.e. $E[y(t)] = \delta \cdot t$). What varies over the models is the variance of $y(t)$ and the expected response in a given situation. This will be discussed further in Section 7.

5. ANNUITY-IMMEDIATE

We now consider a n-year annuity-immediate contract. Let $a_{\overline{n|}}$ be the present value of n equal payments of 1 made at the end of each of the next n years. Then, we have

$$a_{\overline{n|}} = \sum_{t=1}^{n} e^{-y(t)}. \tag{26}$$

We now consider the first three moments of $a_{\overline{n|}}$ using its assumed true probability distribution so that all moments have their usual interpretations. Note however that the expected value will be different than the market price of the annuity which requires that such price be in equilibrium for any purchasing strategy (see BÜHLMANN (1992)).

The expected value of $a_{\overline{n}|}$ may be obtained in the following way:

$$(27) \qquad E[a_{\overline{n}|}] = E\left[\sum_{t=1}^{n} e^{-y(t)}\right] = \sum_{t=1}^{n} E[e^{-y(t)}],$$

where from equation (2),

$$(28) \qquad E[e^{-y(t)}] = \exp\{-E[y(t)] + .5 \cdot V[y(t)]\}.$$

The particular values for $E[y(t)]$ and $V[y(t)]$ were given in Sections 3 and 4 for different modeling approaches and different stochastic processes.

The second moment about the origin of $a_{\overline{n}|}$ may be shown to be equal to

$$(29) \qquad E[(a_{\overline{n}|})^2] = \sum_{t=1}^{n} \sum_{s=1}^{n} E[e^{-y(t)-y(s)}].$$

Similarly, the third moment about the origin of $a_{\overline{n}|}$ is given by

$$(30) \qquad E[(a_{\overline{n}|})^3] = \sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{r=1}^{n} E[e^{-y(t)-y(s)-y(r)}].$$

In order to evaluate the expected values to be summed in (29) and (30), one simply notes that the exponential random variables involved are log-normally distributed. For example,

$$(31) \qquad e^{-y(t)-y(s)-y(r)} \sim \Lambda(\mu, \beta),$$

where

$$(32) \qquad \mu = -E[y(t)] - E[y(s)] - E[y(r)],$$

and

$$(33) \quad \beta = V[y(t)] + V[y(s)] + V[y(r)] + 2\operatorname{cov}[y(t), y(s)] + \\ + 2\operatorname{cov}[y(t), y(r)] + 2\operatorname{cov}[y(s), y(r)].$$

Therefore, from (2), we have:

$$(34) \qquad E[e^{-y(t)-y(s)-y(r)}] = \exp\{\mu + .5 \cdot \beta\}.$$

6. ILLUSTRATIONS

As a way to illustrate the different approaches and the different stochastic processes considered in this paper, we will evaluate their expected values, standard deviations and coefficients of skewness (see, for example, MOOD, GRAYBILL and BOES (1974, pp. 68, 76)) of $a_{\overline{n}|}$, for certain values of the parameters.

Some expected values are found in Table 1. Results are presented for values of the parameters δ set at .06 and .1 in each process. For the White Noise and Wiener processes, we let the parameter σ take the values .01 and .02. For the Ornstein-Uhlenbeck process, the parameter α is chosen to be .17 (this is the value obtained by BEEKMAN and FUELLING (1990, p. 186) from certain U.S. Treasury bill returns). We let the parameter ρ take the values .01 and .02 which correspond to σ equal

$.01 \cdot (.34)^{.5}$ and $.02 \cdot (.34)^{.5}$ respectively. This is consistent with some of the values used by BEEKMAN and FUELLING (1990, Tables 1 and 2).

It should be pointed out that an estimation procedure for finding the values of the different parameters from a data set of past interest rates would generally produce different values of the estimates of the parameters σ, α or ρ depending on the modeling approach used and on the stochastic process chosen. The estimators of the parameter δ, however, are likely to be roughly the same in all cases considered here. Using the same parameters under both approaches is believed to be appropriate to illustrate certain differences between these two approaches.

TABLE 1

EXPECTED VALUE OF $a_{\overline{n}|}$

Modeling the force of interest accumulation function

				n = 5	10	20	30	40
Wiener:		δ	σ					
		.06	.01	4.1920	7.2983	11.3057	13.5061	14.7143
		.06	.02	4.1938	7.3038	11.3202	13.5289	14.7435
		.10	.01	3.7418	6.0118	8.2246	9.0390	9.3387
		.10	.02	3.7433	6.0161	8.2337	9.0511	9.3524
O-U:	δ	α	ρ					
	.06	.17	.01	4.1915	7.2967	11.3013	13.4991	14.7052
	.06	.17	.02	4.1919	7.2975	11.3027	13.5008	14.7071
	.10	.17	.01	3.7413	6.0106	8.2218	9.0353	9.3346
	.10	.17	.02	3.7417	6.0113	8.2228	9.0364	9.3357

Modeling the force of interest

				n = 5	10	20	30	40
Wiener:		δ	σ					
		.06	.01	4.1943	7.3273	11.5925	14.4863	17.0285
		.06	.02	4.2030	7.4217	12.6140	19.5880	48.6888
		.10	.01	3.7437	6.0327	8.3788	9.4388	10.0567
		.10	.02	3.7510	6.1008	8.9232	11.3948	18.0414
O-U:	δ	α	ρ					
	.06	.17	.01	4.1920	7.3007	11.3221	13.5410	14.7658
	.06	.17	.02	4.1938	7.3135	11.3862	13.6702	14.9531
	.10	.17	.01	3.7417	6.0135	8.2336	9.0548	9.3586
	.10	.17	.02	3.7432	6.0229	8.2703	9.1151	9.4331

O-U: Ornstein-Uhlenbeck

From Table 1, one can see that the expected value of $a_{\overline{n}|}$ does not depend very much on the modeling approach used nor does it depend on the parameters of the process, except for the parameter δ, of course. The Wiener process, for n larger than say 20, when used to model the force of interest, is another exception.

Table 2 presents some standard deviations of $a_{\overline{n}|}$. It indicates that for a given stochastic process and a given modeling approach, the standard deviation is more or less proportional to the parameter σ (or ρ). It would appear that adjusting the parameters of a model cannot produce similar standard deviations to those of a different model for all n since the standard deviation exhibits significantly different patterns depending on the modeling approach and/or stochastic process selected.

TABLE 2

STANDARD DEVIATION OF $a_{\overline{n}|}$

Modeling the force of interest accumulation function

				n				
				5	10	20	30	40
Wiener:		δ	σ					
		.06	.01	.0605	.1342	.2623	.3503	.4053
		.06	.02	.1211	.2687	.5258	.7028	.8137
		.10	.01	.0530	.1058	.1734	.2037	.2160
		.10	.02	.1061	.2118	.3476	.4085	.4332
O-U:	δ	α	ρ					
	.06	.17	.01	.0258	.0457	.0645	.0705	.0724
	.06	.17	.02	.0517	.0913	.1291	.1411	.1448
	.10	.17	.01	.0228	.0368	.0463	.0479	.0482
	.10	.17	.02	.0456	.0736	.0926	.0959	.0964

Modeling the force of interest

				n				
				5	10	20	30	40
Wiener:		δ	σ					
		.06	.01	.1251	.5171	1.9640	4.2762	8.6273
		.06	.02	.2515	1.0710	5.1457	27.4239	1111.8356
		.10	.01	.1073	.3880	1.1483	1.9504	2.9114
		.10	.02	.2157	.8019	2.8968	10.1266	240.2379
O-U:	δ	α	ρ					
	.06	.17	.01	.0576	.1968	.5294	.7975	.9767
	.06	.17	.02	.1152	.3952	1.0736	1.6334	2.0169
	.10	.17	.01	.0495	.1495	.3263	.4202	.4610
	.10	.17	.02	.0991	.3001	.6604	.8563	.9433

O-U: Ornstein-Uhlenbeck

For example, we can compare the standard deviations of $a_{\overline{n}|}$ produced by the Ornstein-Uhlenbeck model with parameters $\delta = .06$, $\alpha = .17$ and $\rho = .02$ for the force of interest accumulation function, with those produced by the Ornstein-Uhlenbeck model with parameters $\delta = .06$, $\alpha = .17$ and $\rho = .01$ for the force of interest. Then the standard deviations presented for $n = 5$ are roughly the same (.0517 compared to .0576) while for $n = 40$, the latter (.9767) is almost 7 times larger than the former (.1448). Multiplying the value of ρ in the former by 7 would produce similar standard deviations for $n = 40$ but then the standard deviation in the former model would be about 7 times larger than in the latter model for $n = 5$.

Similar comparisons can be made between different processes under the same approach or different approaches.

This suggests that it is not possible to select different models that would be equivalent in the sense of producing similar standard deviations for all n.

The coefficient of skewness of $a_{\overline{n}|}$ for the same four models are contained in Table 3.

TABLE 3

COEFFICIENT OF SKEWNESS OF $a_{\overline{n}|}$

Modeling the force of interest accumulation function

				n = 5	10	20	30	40
Wiener:		δ	σ					
		.06	.01	.0481	.0640	.0841	.0963	.1040
		.06	.02	.0963	.1282	.1686	.1932	.2087
		.10	.01	.0530	.0616	.0772	.0844	.0876
		.10	.02	.0946	.1233	.1547	.1693	.1757
O-U:	δ	α	ρ					
	.06	.17	.01	.0197	.0202	.0185	.0171	.0165
	.06	.17	.02	.0394	.0404	.0370	.0343	.0330
	.10	.17	.01	.0194	.0198	.0183	.0176	.0175
	.10	.17	.02	.0389	.0395	.0366	.0353	.0349

Modeling the force of interest

				n = 5	10	20	30	40
Wiener:		δ	σ					
		.06	.01	.1338	.3488	.9732	2.1347	6.5145
		.06	.02	.2690	.7266	2.8689	56.9320	1.3×10^5
		.10	.01	.1311	.3336	.8718	1.7175	4.0382
		.10	.02	.2636	.6940	2.5013	41.5591	1.2×10^5
O-U:	δ	α	ρ					
	.06	.17	.01	.0585	.1205	.2157	.2773	.3166
	.06	.17	.02	.1172	.2421	.4379	.5693	.6564
	.10	.17	.01	.0573	.1154	.1961	.2383	.2580
	.10	.17	.02	.1148	.2318	.3977	.4874	.5311

O-U: Ornstein-Uhlenbeck

The coefficient of skewness also exhibits significantly different patterns depending on the model considered. This supports the observation made earlier that no two models can be seen as equivalent.

7. IMPLICIT BEHAVIOR OF THE FORCE OF INTEREST

Clearly, modeling the force of interest accumulation function has quite different implications on the random present value function and other actuarial functions than modeling the force of interest. Basically, when modeling the force of interest, it is δ_s that varies according to the chosen stochastic process. When modeling $y(t)$, then δ_s varies so that $y(t)$ follows the chosen stochastic process. Those differences have already been illustrated by the standard deviation and coefficient of skewness of $a_{\overline{n}|}$. Another useful way of illustrating the differences between the two approaches is to look at the conditional expected value of $y(t)$ given $y(s)$ and δ_s for $s < t$. This conditional expectation will provide some insight into the implicit behavior of each process.

7.1. Modeling the force of interest accumulation function

The conditional expected value of $y(t)$ given $y(s)$ and δ_s for $s < t$ when $y(t)$ follows an Ornstein-Uhlenbeck process may be obtained in the following way.

Using (6), we have

$$E[y(t) \mid y(s) = x, \delta_s = \varepsilon] = E[\delta \cdot t + X(t) \mid \delta \cdot s + X(s) = x, \delta_s = \varepsilon] \tag{35}$$

$$= \delta \cdot t + E[X(t) \mid X(s) = x - \delta \cdot s, \delta_s = \varepsilon], \tag{36}$$

since $X(t)|X(s)$ is independent of δ_s for $s < t$ from the Markovian property of $X(t)$, then

$$E[y(t)|y(s) = x, \delta_s = \varepsilon] = \delta \cdot t + E[X(t)|X(s) = x - \delta \cdot s], \tag{37}$$

which is [see, for example, BEEKMAN and FUELLING (1990, Section 2)]

$$E[y(t)|y(s) = x, \delta_s = \varepsilon] = \delta \cdot t + (x - \delta \cdot s) \cdot e^{-\alpha(t-s)}, \quad s < t. \tag{38}$$

One can proceed in a similar way to find the corresponding result when the force of interest accumulation function is modeled by a Wiener process.

7.2. Modeling the force of interest

The conditional expected value of $y(t)$ given $y(s)$ and δ_s for $s < t$ when δ_s follows an Ornstein-Uhlenbeck process may be obtained in the following way.

Using (1), we have

$$E[y(t) \mid y(s) = x, \delta_s = \varepsilon] = E\left[\int_0^t \delta_r \, dr \;\middle|\; \int_0^s \delta_r \, dr = x, \delta_s = \varepsilon\right] \tag{39}$$

$$= E\left[\int_0^s \delta_r \, dr + \int_s^t \delta_r \, dr \;\middle|\; \int_0^s \delta_r \, dr = x, \delta_s = \varepsilon\right] \tag{40}$$

and conditioning on $y(s) = x$, (40) becomes

$$(41)\quad E[y(t) \mid y(s) = x, \delta_s = \varepsilon] = x + E\left[\int_s^t \delta_r\, dr \;\Big|\; \int_0^s \delta_u\, du = x, \delta_s = \varepsilon\right]$$

$$(42)\quad = x + \int_s^t E\left[\delta_r \;\Big|\; \int_0^s \delta_u\, du = x, \delta_s = \varepsilon\right] dr.$$

From the Markovian property of the process, $\delta_r \mid \delta_s$ with $r > s$ is independent of all values of δ_u for $u < s$, we then have

$$(43)\quad E[y(t) \mid y(s) = x, \delta_s = \varepsilon] = x + \int_s^t E[\delta_r \mid \delta_s = \varepsilon]\, dr.$$

Finally, adapting the result for the conditional expectation of an Ornstein-Uhlenbeck process found in ARNOLD (1974, p. 134), we may write (43) as

$$(44)\quad E[y(t) \mid y(s) = x, \delta_s = \varepsilon] = x + \int_s^t \delta + (\varepsilon - \delta) \cdot e^{-\alpha(r-s)}\, dr.$$

$$(45)\quad = x + \delta\,(t - s) + (\varepsilon - \delta) \cdot \left(\frac{1 - e^{-\alpha(t-s)}}{\alpha}\right).$$

We can proceed similarly to find the corresponding conditional expectations when the force of interest is modeled by a White Noise or a Wiener process.

Table 4 summarizes these results and those obtained earlier in this paper.

TABLE 4

SUMMARY OF RESULTS ABOUT $y(t)$

Process	$E[y(t)]$	$V[y(t)]$	$E[y(t) \mid y(s) = x, \delta_s = \varepsilon]$
		Modeling the force of interest accumulation function	
Wiener	$\delta \cdot t$	$\sigma^2 \cdot t$	$x + \delta\,(t - s)$
O-U	$\delta \cdot t$	$\rho^2 \cdot (1 - e^{-2\alpha t})$	$\delta \cdot t + (x - \delta \cdot s) \cdot e^{-\alpha(t-s)}$
		Modeling the force of interest	
Wiener	$\delta \cdot t$	$\sigma^2 \cdot t^3/3$	$x + \varepsilon\,(t - s)$
O-U	$\delta \cdot t$	$\frac{2\rho^2 t}{\alpha} + \frac{\rho^2}{2\alpha}(-3 + 4e^{-\alpha t} - e^{-2\alpha t})$	$x + \delta\,(t - s) + (\varepsilon - \delta)\left(\frac{1 - e^{-\alpha(t-s)}}{\alpha}\right)$

We note from Table 4, as mentioned earlier, that the expected value of $y(t)$ is the same for all four models presented. Also, as noted earlier, the variances are quite different from one model to another. The salient feature of Table 4, however, is the fact that when modeling the force of interest accumulation function, the conditional expectation of $y(t)$ given $y(s)$ and δ_s does not depend on the values of δ_s. But when modeling the force of interest, this conditional expectation does depend on the value of δ_s.

In order to illustrate the possible implications of the conditional expected values of $y(t)$ presented in Table 4, we now consider the Consumer Price Index (CPI) for Canada for the 1960-1992 period (see Canadian Institute of Actuaries (1993, Table 1A)). Here, the CPI plays the role of the force of interest.

The results presented in Sections 2.2 and 6.4 of PANDIT and WU (1983) were used to estimate the parameters of the different models. The estimator for δ is .05335. The estimator of the parameter α when modeling the force of interest accumulation function is .01955, and when modeling the force of interest, it is .05389.

Using these values, the expected values of $y(t)$, $t > 10$, given $y(10) = .2771$ and $\delta_{10} = .0131$ were computed. The results are presented in Figure 1 where $t = 0$ corresponds to 1960. It is difficult to determine from this figure whether the fact that some models do not use the value of δ_{10} makes a significant difference.

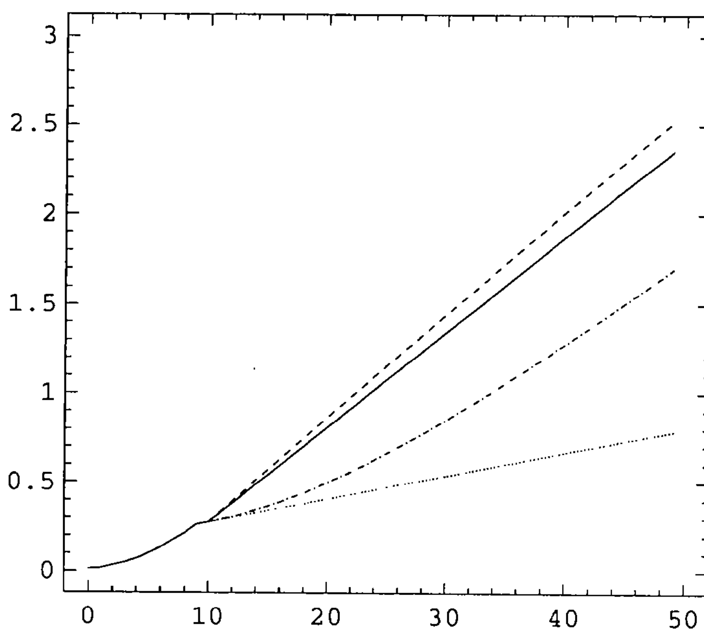


FIGURE 1. Conditional Expected Value of $y(t)$ given $y(10)$ and δ_{10}.

——	$\{y(s)\}_{s=0}^{10}$		with $y(10) = .2771$		$\delta_{10} = .0131$	
——	δ_t:	W-N	$\delta = .05335$	or	$y(t)$: Wiener	$\delta = .05335$
– –	$y(t)$:	O-U	$\delta = .05335$	$\alpha = .01955$		
· · ·	δ_t:	Wiener	$\delta = .05335$			
– · –	δ_t:	O-U	$\delta = .05335$	$\alpha = .05389$		

Figure 2 presents the expected values of δ_t, $t > 10$, given $y(10) = .2771$ and $\delta_{10} = .0131$. This last figure clearly indicates a possible implication resulting from modeling the force of interest accumulation function instead of the force of interest. That is, an expected value of the force of interest, in the immediate future, which can be significantly different from its current value.

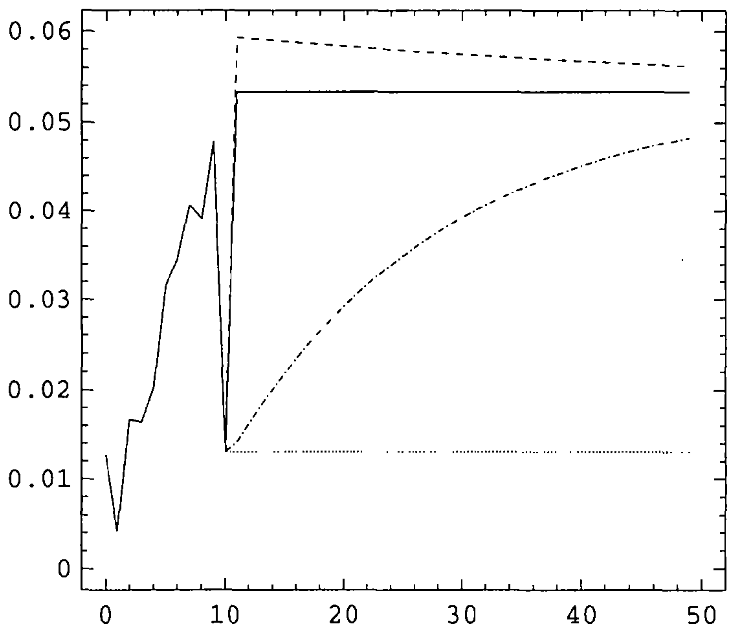


FIGURE 2. Conditional Expected Value of δ_t given $y(10)$ and δ_{10}.

———	$\{\delta_s\}_{s=0}^{10}$		with $y(10) = .2771$		$\delta_{10} = .0131$	
———	δ_t:	W-N	$\delta = .05335$	or	$y(t)$: Wiener	$\delta = .05335$
– –	$y(t)$:	O-U	$\delta = .05335$	$\alpha = .01955$		
· · ·	δ_t:	Wiener	$\delta = .05335$			
– · –	δ_t:	O-U	$\delta = .05335$	$\alpha = .05389$		

8. REMARKS AND SUMMARY

It should be noted that the numerical values presented in Tables 1 and 2 of this paper are not entirely comparable with those in BEEKMAN and FUELLING (1990, 1991). BEEKMAN and FUELLING (1990, 1991) study the continuous annuity, $\bar{a}_{\overline{n}|}$, and we chose to study the annuity-immediate, $a_{\overline{n}|}$. The choice of a discrete annuity was made in order to avoid errors involved in doing numerical integrations that would have been needed for the continuous annuity for some of the models considered.

In this paper, we have studied different models under two approaches to model the interest randomness. An annuity-immediate was used to present some illustrations.

As measured by the agreement of the expected values, standard deviations and coefficients of skewness, no two models can be seen as equivalent, even if one would try to select particular values of the parameters. The one exception to this is that a White Noise process for the force of interest is equivalent to a Wiener process for the force of interest accumulation function.

Further, when modeling the force of interest accumulation function, defined as $y(t)$, the conditional expected value of $y(t)$ given $y(s)$ and δ_s, $s < t$, does not depend on the value of the force of interest at time s. However, when modeling the force of interest, the expected value of $y(t)$ given $y(s)$ and δ_s, $s < t$, does depend on the value of the force of interest at time s.

Finally, another advantage to using one of the models presented for the force of interest is that they are special cases of one-factor interest rate term structure models. This means that the work that has already been done in finance could be used by actuaries interested in arbitrage-free pricing.

ACKNOWLEDGMENTS

I would like to thank Professor RAGNAR NORBERG and Professor HANS BÜHLMANN for their helpful comments on an earlier version of this paper. I am also grateful to the anonymous referees for their useful comments.

REFERENCES

AITCHISON, J. and BROWN, J.A.C. (1963) *The Lognormal Distribution*, 176 pp., Cambridge University Press.

ARNOLD, L. (1974) *Stochastic Differential Equations: Theory and Applications*, 228 pp., John Wiley & Sons, New York.

BEEKMAN, J.A. and FUELLING, C.P. (1990) Interest and Mortality Randomness in Some Annuities. *Insurance: Mathematics and Economics* **9**, 185–196.

BEEKMAN, J.A. and FUELLING, C.P. (1991) Extra Randomness in Certain Annuity Models. *Insurance: Mathematics and Economics* **10**, 275–287.

BEEKMAN, J.A. and FUELLING, C.P. (1993) One Approach to Dual Randomness in Life Insurance. To appear.

BÜHLMANN, H. (1992) Stochastic Discounting. *Insurance: Mathematics and Economics* **11**, 113–127.

CANADIAN INSTITUTE OF ACTUARIES (1993) *Rapport sur les Statistiques Économiques Canadiennes*, 1924–1992.

DE SCHEPPER, A., DE VYLDER, F., GOOVAERTS, M. and KAAS, R. (1992a) Interest Randomness in Annuities Certain. *Insurance: Mathematics and Economics* **11**, 271–281.

DE SCHEPPER, A., GOOVAERTS, M. and DELBAEN, F. (1992b) The Laplace Transform of Annuities Certain with Exponential Time Distribution. *Insurance: Mathematics and Economics* **11**, 291–294.

DE SCHEPPER, A. and GOOVAERTS, M. (1992) Some Further Results on Annuities Certain with Random Interest. *Insurance: Mathematics and Economics* **11**, 283–290.

DEVOLDER, P. (1986) Opérations Stochastiques de Capitalisation. *ASTIN Bulletin* **16S**, S5–S30.

DHAENE, J. (1989) Stochastic Interest Rates and Autoregressive Integrated Moving Average Processes. *ASTIN Bulletin* **19**, 131–138.

DUFRESNE, D. (1990) The Distribution of a Perpetuity, with Applications to Risk. Theory and Pension Funding. *Scandinavian Actuarial Journal*, 39–79.

FREES, E.W. (1990) Stochastic Life Contingencies with Solvency Considerations. *Transaction of the Society of Actuaries* **XLII**, 91–148.

KARLIN, S. and TAYLOR, H.M. (1981) *A Second Course in Stochastic Processes*, 542 pp., Academic Press, San Diego.

MELSA, J.L. and SAGE, A.P. (1973) *An Introduction to Probability and Stochastic Processes*. 403 pp., Prentice-Hall, New Jersey.

MOOD, A.M., GRAYBILL, F.A. and BOES, D.C. (1974) *Introduction to the Theory of Statistics*, 567 pp., McGraw-Hill.

NORBERG, R. (1993) A Solvency Study in Life Insurance. *Proceedings of the 3rd AFIR International Colloquium*, 821–830.

PANDIT, S.M. and WU, S.M. (1983) *Time Series and System Analysis with Applications*, 586 pp., John Wiley & Sons, New York.

PANJER, H.H. and BELLHOUSE, D.R. (1980) Stochastic Modelling of Interest Rates with Applications to Life Contingencies. *Journal of Risk and Insurance* **47**, 91–110.

PAPACHRISTOU, D.J. and WATERS, H.R. (1991) Some Remarks Concerning Stochastic Interest Rates in Relation to Long Term Insurance Policies. *Scandinavian Actuarial Journal*, 103–117.

PARKER, G. (1992) An Application of Stochastic Interest Rates Models in Life Assurance. Ph.D. thesis, Heriot-Watt University.

PARKER, G. (1993a) Distribution of the present value of future cash flows. *Proceedings of the 3rd AFIR International Colloquium*, 831–843.

PARKER, G. (1993b) Moments of the present value of a portfolio of policies. To appear in *Scandinavian Actuarial Journal*.

PARKER G., (1993c) Two Stochastic Approaches for Discounting Actuarial Functions. *Proceedings of the XXIV ASTIN Colloquium*, 367–389.

PARKER, G. (1994) Limiting distribution of the present value of a portfolio. *ASTIN Bulletin* **94-1**, 47–60.

WATERS, H. R. (1978) The Moments and Distributions of Actuarial Functions. *Journal of the Institute of Actuaries* **105**, **Part 1**, 61–75.

WATERS, H.R. (1990) The Recursive Calculation of the Moments of the Profit on a Sickness Insurance Policy. *Insurance: Mathematics and Economics* **9**, 101–113.

GARY PARKER
Dept. of Mathematics and Statistics, Simon Fraser University,
Burnaby, B.C. Canada V5A 1S6.

A MARKOV MODEL FOR LOSS RESERVING

By Ole Hesselager

University of Copenhagen, Denmark

Abstract

The claims generating process for a non-life insurance portfolio is modelled as a marked Poisson process, where the mark associated with an incurred claim describes the development of that claim until final settlement. An unsettled claim is at any point in time assigned to a state in some state-space, and the transitions between different states are assumed to be governed by a Markovian law. All claims payments are assumed to occur at the time of transition between states. We develop separate expressions for the IBNR and RBNS reserves, and the corresponding prediction errors.

Keywords

IBNR and RBNS reserves; marked point process; Markov chain; martingale.

1. Introduction

Hachemeister (1980) suggested to represent the information about an unsettled claim by modelling the development as the realization of a (discrete-time) Markov chain. The predicted future claims cost for a particular claim then depends on the current state of the claim, and the state space represents the possible types of information which the company may have (or want to consider) during the development process. In this paper we adopt the ideas of Hachemeister (1980) and describe the development of a claim from occurrence until final settlement as the realization of a time-inhomogeneous, continuous-time Markov chain. We extend the description by also modelling the claim occurrences — by a time-inhomogeneous Poisson process. This makes it possible to establish separate reserves for the pure IBNR (Incurred But Not Reported) claims and the RBNS (Reported But Not Settled) claims.

The reserving (or prediction) problem is conveniently formulated within the framework of marked (Poisson) point processes, which was advocated in the context of claims reserving by Arjas (1989), and further developed by Norberg (1993). In this context it is then assumed that the marks consist of the realization of a Markov chain together with the claims payments, which are assumed to occur at the times of transition between different states.

The present paper gives a time-continuous version of Hachemeister's (1980) model. Our way of modelling the claims payments, however, differs from that of

HACHEIMEISTER (1980), and hence also our formulas for the IBNR and RBNS reserves. In particular, the formulas given for the prediction errors, which are used to assess the quality of the IBNR and RBNS reserves, appear to be new.

2. THE MODEL

Consider a portfolio which has been observed during some time interval $[0, \tau]$, where τ represents the present moment. We denote by $K(t)$ the number of claims incurred during $[0, t]$, and by $0 < T_1 < T_2 < \dots$ the corresponding times of occurrence. With the ith claim we associate a mark Z_i, which describes the development of that claim until final settlement. The marks are constructed as

$$Z_i = Z^{(T_i)},$$

where $\{Z^{(t)}\}_{t \geq 0}$ is a family of random elements. For the claims generating process we assume that

(a) $\{K(t)\}_{t \geq 0}$ is a Poisson process with intensity $\{\mu(t)\}_{t \geq 0}$, and $\{Z^{(t)}\}_{t \geq 0}$ are mutually independent and independent of $\{K(t)\}_{t \geq 0}$.

A claim is at any point in time after occurrence assigned to one of at most countable many states, $\mathcal{S}$. Different states in the set $\mathcal{S}$ represent different types of information about the claim which the company may have. During the development process a claim may change state as new information becomes available, and (partial) payments may be made at the times of transition between states. We want the mark Z_i to carry the information about how the ith claim is classified in the course of time, and also payments being made on that claim. Thus, we let

$$Z^{(t)} = \{ \{S^{(t)}(u)\}_{u \geq 0}, \ \{Y^{(t)}_{mn,j}\}_{j=1,2,\dots} \ m \neq n, m, n \in \mathcal{S} \},$$

where $S^{(t)}(u) \in \mathcal{S}$ denotes the state at time $t + u$ of a claim incurred at time t, and $Y^{(t)}_{mn,j}$ denotes the payment made upon the jth transition from m to n.

For a claim incurred at time t, transitions from m to n occur at time epochs $t + U^{(t)}_{mn,1}, t + U^{(t)}_{mn,2}, \dots$, say. The payments $Y^{(t)}_{mn,j}$ are regarded as marks corresponding to the point process $0 < U^{(t)}_{mn,1} < U^{(t)}_{mn,2} < \dots$, and are constructed as

$$Y^{(t)}_{mn,j} = Y^{(t)}_{mn}(U^{(t)}_{mn,j}),$$

where $\{Y^{(t)}_{mn}(u)\}_{u \geq 0}$ is a family of random elements. For the development process we assume that

(b) $\{S^{(t)}(u)\}_{u \geq 0}$ is a time-inhomogeneous Markov chain with transition probabilities

$$p_{mn}(u, v) = \mathrm{P}(S^{(t)}(v) = n \mid S^{(t)}(u) = m),$$

and intensities

$$\lambda_{mn}(u) = \lim_{h \to 0+} p_{mn}(u, u + h)/h.$$

The amounts $\{Y^{(t)}_{mn}(u)\}_{u\geq 0}$ are mutually independent for all $m \neq n \in \mathcal{S}$ and $u \geq 0$, and are independent of $\{S^{(t)}(u)\}_{u\geq 0}$ with cumulative distribution function

$$F_{mn}(y \mid u) = \mathrm{P}\,(Y^{(t)}_{mn}(u) \leq y)\,.$$

Remark 2.1. According to assumption (b), the distribution of the mark $Z^{(t)}$ corresponding to a claim incurred at time t does not depend on time t. This assumption could be dropped without any consequences for the following — except that the intensities $\lambda_{mn}(u)$ and the distributions $F_{mn}(y \mid u)$ would then carry topscript t. A particular dependence on time t is that where $\lambda^{(t)}_{mn}(u) = \lambda_{mn}(t+u)$ depends on calendar time $t+u$ rather than waiting time u since occurrence of the claim. This may be a reasonable specification e.g. for transitions corresponding to the settlement of RBNS claims (see Examples 2 and 3 below). It is a different matter that the statistical estimation becomes more difficult in such cases. In fact, since the claims reserving problem is concerned with payments made at time epochs $t+u > \tau$, one will at time τ only be able to estimate the relevant intensities $\lambda_{mn}(t+u)$ if some parametric assumption is being made. □

Example 1. In the simplest possible model,

$$\boxed{0 \sim \text{IBNR}} \xrightarrow{\lambda_{0\Delta}(u)} \boxed{\Delta \sim \text{Settled}}$$

it is assumed that a claim is settled at the time of notification. Let W be the waiting time until notification, and let $G(u) = \mathrm{P}\,(W \leq u)$. With only two states, $\mathcal{S} = \{0, \Delta\}$, where $0 \sim$ IBNR and $\Delta \sim$ Settled, this model is trivially Markov, and the rate of settlement is $\lambda_{0\Delta}(u) = G'(u)/(1 - G(u))$. □

Example 2. Consider the model,

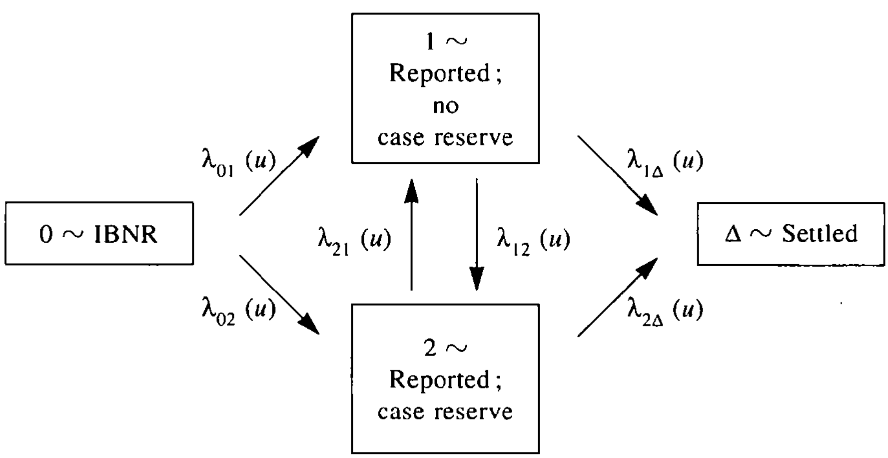
$$\boxed{0 \sim \text{IBNR}} \xrightarrow{\lambda_{01}(u)} \boxed{1 \sim \text{RBNS}} \xrightarrow{\lambda_{1\Delta}(u)} \boxed{\Delta \sim \text{Settled}}$$

where the reporting as well as the settlement of claims is subject to a delay. The assumption (b), that the intensities λ_{mn} depend on u, the time elapsed since occurrence, may seem inadequate as far as $\lambda_{1\Delta}$ is concerned. The management might want to assume that the rate of settlement for RBNS claims is determined by the amount of resources which are allocated to claims handling department, and that $\lambda_{1\Delta}$ should therefore depend on calendar time $t+u$. As pointed out in Remark 2.1 above, this is also possible within the current framework. One could take

$$\lambda^{(t)}_{01}(u) = \lambda_{01}(u)\,,$$
$$\lambda^{(t)}_{0\Delta}(u) = \lambda_{0\Delta}(t+u)\,,$$

in which case the rate of reporting depends on waiting time u since occurrence and the rate of settlement depends on calendar time $t + u$. □

Example 3. Consider the following example, inspired by HACHEMEISTER (1980). In some lines of business, with the possibility of having very large claims, it is customary that the claims handling department at the time of notification reviews the details concerning a claim and makes an estimate whether the ultimate claim amount is likely to exceed some prescribed limit, say DDK 200.000. If so, a case reserve (RBNS) is calculated for this claim. The company may later receive new information which causes it to revise the initial estimate. A claim which at the time of notification was judged to exceed the prescribed limit may then be re-classified as a "small" claim, and vice versa. Obviously the model could be refined by introducing more states, representing different intervals for the individual estimate (case reserve) for a claim.



□

For a claim incurred at time t we shall need the following quantities,

$I_m^{(t)}(u) = I(S^{(t)}(u) = m)$, the indicator of the event that the claim occupies state m at time $t + u$,

$N_{mn}^{(t)}(u)$, the number of direct transitions from m to n during $[t, t + u]$,

$\mathcal{H}_u^{(t)} = \sigma(\{S^{(t)}(v)\}_{0 \le v \le u}, \{Y_{mn,j}^{(t)}, j = 1, \dots, N_{mn}^{(t)}(u)\})$, the history generated during $[t, t + u]$ by the claim,

$y_{mn}(u) = \mathrm{E}\, Y_{mn}^{(t)}(u)$, the average claim amount paid at time $t + u$ if a transition from m to n occurs at that time,

$\sigma_{mn}^2(u) = \mathrm{Var}\, Y_{mn}^{(t)}(u)$, the variance on the claim amount paid at time $t + u$ if a transition from m to n occurs at that time.

We shall also make use of the fact that

$$dN_{mn}^{(t)}(u) = I_m^{(t)}(u-)\,\lambda_{mn}(u)\,du + dM_{mn}^{(t)}(u), \tag{2.1}$$

where $u-$ denotes the left-hand limit, and all $M_{mn}^{(t)}(u)$ for $m, n \in \mathcal{S}$ and $m \neq n$ are mutually orthogonal zero-mean martingales with respect to the internal history of the process $\{S^{(t)}(u)\}_{u \geq 0}$ (see e.g. ANDERSEN et al. (1985)). Because $\{S^{(t)}(u)\}_{u \geq 0}$ is stochastically independent of the claim amounts according to assumption (b), it is also true that $M_{mn}^{(t)}(u)$ is a zero-mean martingale with respect to the filtration $\{\mathcal{H}_u^{(t)}\}_{u \geq 0}$. Furthermore,

$$\operatorname{Var}[dM_{mn}^{(t)}(u) \mid \mathcal{H}_{u-}^{(t)}] = I_m^{(t)}(u-)\,\lambda_{mn}(u)\,du. \tag{2.2}$$

Let $X^{(t)}(u, v)$ denote the total payment made during $]t+u, t+v]$ in respect of a claim incurred at time t. We may write $X^{(t)}(u, v)$ as

$$X^{(t)}(u, v) = \sum_{m \neq n} \int_u^v Y_{mn}^{(t)}(\xi)\,dN_{mn}^{(t)}(\xi). \tag{2.3}$$

We make the convention that $0 \in \mathcal{S}$, and that this state represents IBNR claims. Also $\Delta \in \mathcal{S}$, and Δ is an absorbing state representing fully settled claims. With this convention the number of claims incurred during $[0, t]$, which at time τ are classified as IBNR and RBNS claims, respectively, can be written as

$$K_{IBNR}(t) = \int_0^t I_0^{(s)}(\tau - s)\,dK(s), \tag{2.4}$$

$$K_{RBNS}(t) = \int_0^t [1 - I_0^{(s)}(\tau - s) - I_\Delta^{(s)}(\tau - s)]\,dK(s), \tag{2.5}$$

and the corresponding outstanding (at time τ) claims payments are

$$X_{IBNR}(t) = \int_0^t X^{(s)}(\tau - s, \infty)\,dK_{IBNR}(s), \tag{2.6}$$

$$X_{RBNS}(t) = \int_0^t X^{(s)}(\tau - s, \infty)\,dK_{RBNS}(s). \tag{2.7}$$

In Section 4 we derive expressions for the IBNR and RBNS reserves, defined as the expected claims payments $X_{IBNR}(\tau)$ and $X_{RBNS}(\tau)$ given the available information at time τ, and the corresponding prediction errors. Before we proceed to do so, we shall in Section 3 derive the required moments of the future payments $X^{(s)}(\tau - s, \infty)$ in respect of a single claim.

3. FUTURE PAYMENTS ON A SINGLE CLAIM

Consider the claims payments $X^{(t)}(u, \infty)$ in respect of a single claim incurred at time t, as defined in (2.3). We shall derive expressions for the conditional moments of $X^{(t)}(u, \infty)$, given the individual history $\mathcal{H}_u^{(t)}$ of that claim. Since all quantities considered here are functions of the mark $Z^{(t)}$ corresponding to a claim incurred at time t, and the distribution of $Z^{(t)}$ does not depend on t according to assumption (b), we may in this section omit the superscript (t).

Consider the conditional distribution of $X(u, \infty)$ given $\mathcal{H}_u$. By the independence assumed in (b), the information about past claim amounts $Y_{mn}(v)$ for $v \leq u$ may be omitted from the history $\mathcal{H}_u$. From the Markov property it furthermore follows that the only information contained in $\mathcal{H}_u$ about the future development of $\{S(v)\}$ is the present state $S(u)$. Thus,

$$\text{(3.1)} \qquad \mathrm{E}\,[X(u, \infty) \mid \mathcal{H}_u] = \mathrm{E}\,[X(u, \infty) \mid S(u)] := V(u \mid S(u)),$$

$$\text{(3.2)} \qquad \mathrm{Var}\,[X(u, \infty) \mid \mathcal{H}_u] = \mathrm{Var}\,[X(u, \infty) \mid S(u)] := \Gamma(u \mid S(u)).$$

With $X(u, \infty)$ given by (2.3) we obtain by independence of $\{S(u)\}_{u \geq 0}$ and $\{Y_{mn}(u)\}_{u \geq 0}$, and by use of the decomposition (2.1), that

$$\begin{aligned}
\text{(3.3)} \quad V(u \mid j) &= \mathrm{E}\,[X(u, \infty) \mid S(u) = j] \\
&= \sum_{m \neq n} \int_u^\infty \mathrm{E}\,(Y_{mn}(\xi) \mid S(u) = j)\, \mathrm{E}\,(dN_{mn}(\xi) \mid S(u) = j) \\
&= \sum_{m \neq n} \int_u^\infty y_{mn}(\xi)\, \mathrm{E}\,[I_m(\xi-)\, \lambda_{mn}(\xi) + dM_{mn}(\xi) \mid S(u) = j] \\
&= \sum_{m \neq n} \int_u^\infty y_{mn}(\xi)\, p_{jm}(u, \xi)\, \lambda_{mn}(\xi)\, d\xi, \qquad j \in \mathcal{S},
\end{aligned}$$

where the latter equality in (3.3) follows by noting that

$$\mathrm{E}\,[dM_{mn}(\xi) \mid S(u) = j] = \mathrm{E}\,\{\mathrm{E}\,[dM_{mn}(\xi) \mid \mathcal{H}_u] \mid S(u) = j\} = 0$$

for $\xi \geq u$, because $\{M_{mn}(u)\}_{u \geq 0}$ is a martingale with respect to the history of that claim.

For the purpose of deriving formulas for the variance functions $\Gamma(u \mid j)$ in (3.2), we shall find it convenient to work will the loss corresponding to $]u, v]$, defined as

$$\text{(3.4)} \qquad L(u, v) = X(u, v) + V(v \mid S(v)) - V(u \mid S(u)).$$

The loss as defined in (3.4) plays a key role in connection with results of Hattendorff-type in life-insurance (e.g. PAPATRIANDAFYLOU & WATERS, 1984) due to the fact that $\{L(u, v)\}_{v \geq u}$ is a zero-mean martingale with respect to $\{\mathcal{H}_v\}_{v \geq u}$. This is most easily seen by writing

$$L(u, v) = \mathrm{E}\,[X(0, \infty) \mid \mathcal{H}_v] - \mathrm{E}\,[X(0, \infty) \mid \mathcal{H}_u], \quad v \geq u,$$

which for $u \leq \xi \leq v$ shows that

$$\begin{aligned}
\mathrm{E}\,(L(u, v) \mid \mathcal{H}_\xi) &= \mathrm{E}\,\{\mathrm{E}\,[X(0, \infty) \mid \mathcal{H}_v] \mid \mathcal{H}_\xi\} - \mathrm{E}\,[X(0, \infty) \mid \mathcal{H}_u] \\
&= \mathrm{E}\,[X(0, \infty) \mid \mathcal{H}_\xi] - \mathrm{E}\,[X(0, \infty) \mid \mathcal{H}_u] = L(u, \xi).
\end{aligned}$$

Because of (3.4), with $v = \infty$, and because the increments of a martingale are uncorrelated, we may then calculate the conditional variance (3.2) as

$$\text{Var}\,(X(u,\infty)|\,\mathcal{H}_u) = \text{Var}\,(L(u,\infty)|\,\mathcal{H}_u) = \int_u^\infty \text{Var}\,(L(d\xi)|\,\mathcal{H}_u), \tag{3.5}$$

where $L(d\xi)$ is a short-hand for the loss corresponding to an infinitesimal interval containing ξ. An expression for $\text{Var}\,(L(d\xi)|\,\mathcal{H}_u)$ may be obtained using calculations similar to those in NORBERG (1992). According to (3.4) it holds that

$$L(d\xi) = \sum_{m \neq n} Y_{mn}(\xi)\, dN_{mn}(\xi) + dV(\xi|\, S(\xi)). \tag{3.6}$$

By writing

$$V(\xi|\, S(\xi)) = \sum_m I_m(\xi)\, V(\xi|\, m),$$

we obtain that

$$dV(\xi|\, S(\xi)) = \sum_m \left\{ V(\xi|\, m)\, dI_m(\xi) + I_m(\xi) \left(\frac{d}{d\xi}\, V(\xi|\, m) \right) d\xi \right\}. \tag{3.7}$$

Since $I_m(\xi)$ increases by one if a transition into state m is made at time ξ and decreases by one if a transition out of state m is made at that time, we may write

$$dI_m(\xi) = \sum_{n:n \neq m} dN_{nm}(\xi) - \sum_{n:n \neq m} dN_{mn}(\xi). \tag{3.8}$$

The reserve $V(\xi|\, m)$ is a prospective reserve for a Markov model, as used in classical life-insurance mathematics. In the present case, the reserve (3.3) contains no interest or premium payments, and Thiele's differential equation then becomes,

$$\frac{d}{d\xi}\, V(\xi|\, m) = - \sum_{n:n \neq m} \lambda_{mn}(\xi)\, r_{mn}(\xi), \tag{3.9}$$

where

$$r_{mn}(\xi) = y_{mn}(\xi) + V(\xi|\, n) - V(\xi|\, m) \tag{3.10}$$

denotes the (expected) sum at risk at time ξ. Combining (3.7) with (3.8) and (3.9) yields

$$\begin{aligned} dV(\xi|\, S(\xi)) &= \sum_{m \neq n} V(\xi|\, m)\, (dN_{nm}(\xi) - dN_{mn}(\xi)) - I_m(\xi)\, \lambda_{mn}(\xi)\, r_{mn}(\xi)\, d\xi \\ &= \sum_{m \neq n} (V(\xi|\, n) - V(\xi|\, m))\, dN_{mn}(\xi) - I_m(\xi)\, \lambda_{mn}(\xi)\, r_{mn}(\xi)\, d\xi. \end{aligned}$$

Integrating (3.6) from u to v with $dV(\xi | S(\xi))$ given above, we then arrive at the expression

$$(3.11) \qquad L(u, v) = \sum_{m \neq n} \int_u^v [Y_{mn}(\xi) + V(\xi | n) - V(\xi | m)]\, dN_{mn}(\xi) -$$
$$- \sum_{m \neq n} \int_u^v I_m(\xi)\, \lambda_{mn}(\xi)\, r_{mn}(\xi)\, d\xi .$$

Since the latter integral in (3.1) is an ordinary Lebesgue integral, we may here replace $I_m(\xi)$ with its left-hand limit $I_m(\xi -)$, which allows the alternative expressiòn

$$(3.12) \qquad L(u, v) \doteq \sum_{m \neq n} \int_u^v \bar{Y}_{mn}(\xi)\, dN_{mn}(\xi) + \sum_{m \neq n} \int_u^v r_{mn}(\xi)\, dM_{mn}(\xi),$$

where $M_{mn}(\xi)$ is the martingale (2.1) and

$$\bar{Y}_{mn}(\xi) = Y_{mn}(\xi) - y_{mn}(\xi) .$$

For the purpose of calculating the conditional variance (3.5), the expression (3.12) is useful.

The terms $\bar{Y}_{mn}(\xi)\, dN_{mn}(\xi)$ are mutually uncorrelated given $\mathcal{H}_u$, as a consequence of assumption (b), and

$$\begin{aligned} \mathrm{Var}\, [\bar{Y}_{mn}(\xi)\, dN_{mn}(\xi) |\, \mathcal{H}_u] &= \mathrm{E}\, [\bar{Y}_{mn}(\xi)^2\, dN_{mn}(\xi)^2 |\, \mathcal{H}_u] \\ &= \sigma_{mn}^2(\xi)\, p_{S(u)m}(u, \xi)\, \lambda_{mn}(\xi)\, d\xi . \end{aligned}$$

The terms $r_{mn}(\xi)\, dM_{mn}(\xi)$ are mutally uncorrelated because the martingales $M_{mn}(\xi)$ are, and by use of (2.2),

$$\mathrm{Var}\, [r_{mn}(\xi)\, dM_{mn}(\xi) |\, \mathcal{H}_u] = r_{mn}(\xi)^2\, p_{S(u)m}(u, \xi)\, \lambda_{mn}(\xi)\, d\xi .$$

Finally, the terms $\bar{Y}_{mn}(\xi)\, dN_{mn}(\xi)$ and $r_{mn}(\xi)\, dM_{mn}(\xi)$ are uncorrelated as a consequence of assumption (b). From (3.5), (3.12) and the above expressions we then obtain that the variance functions $\Gamma(u | j)$ appearing in (3.2) can be expressed as

$$(3.13) \qquad \Gamma(u | j) = \sum_{m \neq n} \int_u^\infty p_{jm}(u, \xi)\, \lambda_{mn}(\xi)\, [\sigma_{mn}^2(\xi) + r_{mn}(\xi)^2]\, d\xi .$$

In the context of life insurance, variance formulas analogous to (3.13) were obtained by RAMLAU-HANSEN (1988) for a Markov model and by NORBERG (1992) in a more general counting process framework. However, in life insurance the size of the benefits is specified in the insurance contract, and these are consequently considered as deterministic. The variance $\sigma_{mn}^2(\xi)$ does therefore not appear in the formulas of RAMLAU-HANSEN (1988) and NORBERG (1992).

Remark 3.1. To obtain tables of $V(u|j)$ and $\Gamma(u|j)$ from (3.3) and (3.13), respectively, one has to calculate first the transition probabilities $p_{jm}(u,v)$ by solving Kolmogorov's differential equations. A computationally more convenient approach is to calculate $V(u|j)$ directly by solving Thiele's differential equation (3.9) with boundary conditions $V(\infty|j)=0$. In practice one will of course use a boundary condition $V(u_{\max}|j)=0$, where $u_{\max}$ is chosen such that all claims are fully settled within the first $u_{\max}$ time units after occurrence. Comparing (3.3) and (3.13) shows that the formula (3.13) can be obtained from (3.3) by replacing the average claim amount $y_{mn}(\xi)$ by $\sigma^2_{mn}(\xi)+r_{mn}(\xi)^2$. Taking the derivative with respect to u it then follows that $\Gamma(u|j)$ satisfies a Thiele's differential equation (3.9), except that $\sigma^2_{mn}(u)+r_{mn}(u)^2$ replaces $y_{mn}(u)$ also in this case. Thus,

$$(3.14)\quad \frac{d}{du}\,\Gamma(u|j) = -\sum_{m:m\neq j} \lambda_{jm}(u)\,[\sigma^2_{jm}(u)+r_{jm}(u)^2+\Gamma(u|m)-\Gamma(u|j)],$$

and $V(u|j)$ as well as $\Gamma(u|j)$ may be calculated without necessarily calculating the transition probabilities. □

4. CLAIMS RESERVES

By time τ we have registered all known (reported) claim occurrences during $[0,\tau]$, and for a reported claim incurred at time t, say, we have also registered the individual history $\mathcal{H}^{(t)}_{\tau-t}$ of that claim from the time of occurrence up to present time τ. Let $\mathcal{F}_\tau$ denote the collection of this information.

The IBNR and RBNS reserves at time τ are defined as

$$(4.1)\qquad V_{IBNR}(\tau) = \mathrm{E}\,(X_{IBNR}(\tau)|\,\mathcal{F}_\tau),$$

$$(4.2)\qquad V_{RBNS}(\tau) = \mathrm{E}\,(X_{RBNS}(\tau)|\,\mathcal{F}_\tau),$$

where $X_{IBNR}(t)$ and $X_{RBNS}(t)$ are defined in (2.6) and (2.7). The corresponding prediction errors are denoted by

$$(4.3)\qquad \Gamma_{IBNR}(\tau) = \mathrm{Var}\,(X_{IBNR}(\tau)|\,\mathcal{F}_\tau),$$

$$(4.4)\qquad \Gamma_{RBNS}(\tau) = \mathrm{Var}\,(X_{RBNS}(\tau)|\,\mathcal{F}_\tau).$$

Considering RBNS claims, the occurrences $\{K_{RBNS}(t)\}_{0\le t\le\tau}$ are known from $\mathcal{F}_\tau$, and from (2.7) we then obtain

$$(4.5)\qquad V_{RBNS}(\tau) = \int_0^\tau \mathrm{E}\,(X^{(t)}(\tau-t,\infty)|\,\mathcal{F}_\tau)\,dK_{RBNS}(t).$$

By assumption (a) the conditional expectation appearing in (4.5) depends only on the history $\mathcal{H}^{(t)}_{\tau-t}$ of that particular claim, and from (3.1) we then have,

$$(4.6)\qquad V_{RBNS}(\tau) = \int_0^\tau V(\tau-t|\,S^{(t)}(\tau-t))\,dK_{RBNS}(t),$$

where $V(u|j)$ is the reserve (3.3). By independence of the marks corresponding to different claims we also have,

$$\Gamma_{RBNS}(\tau) = \int_0^\tau \operatorname{Var}[X^{(t)}(\tau - t, \infty) |\ \mathcal{F}_\tau]\, dK_{RBNS}(t) \tag{4.7}$$

$$= \int_0^\tau \Gamma(\tau - t | S^{(t)}(\tau - t))\, dK_{RBNS}(t),$$

where $\Gamma(u|j)$ is defined in (3.13). Note that the integrals in (4.6), (3.13) simply represent summation over those claims which are RBNS at time τ. Thus, the RBNS reserve (4.6) is obtained by adding the reserves $V(u|j)$ corresponding to the current states and durations for the RBNS claims at time τ.

From (2.4) and (2.5) we note that $\{K_{IBNR}(t)\}_{0 \le t \le \tau}$ and $\{K_{RBNS}(t)\}_{0 \le t \le \tau}$ are obtained as a marker dependent partition of the Poisson process $\{K(t)\}$. From NORBERG (1993, Theorem 2) it then follows that the marked point processes corresponding to $\{K_{IBNR}(t)\}_{0 \le t \le \tau}$ and $\{K_{RBNS}(t)\}_{0 \le t \le \tau}$ are independent and (again) Poisson. The Poisson rate corresponding to IBNR claims is given by

$$\mu_{IBNR}(t) = \mu(t)\, \mathrm{P}\,(S^{(t)}(\tau - t) = 0) = \mu(t)\, p_{00}(\tau - t)$$

and the mark corresponding to an IBNR claim incurred at time t is distributed according to the conditional distribution of $Z^{(t)}$ given that $S^{(t)}(\tau - t) = 0$. Since the history $\mathcal{F}_\tau$ is generated by reported claims (only), it then also follows that $X_{IBNR}(t)$ is independent of $\mathcal{F}_\tau$, and from (4.1), (2.6) we obtain that

$$V_{IBNR}(\tau) = \mathrm{E}\, X_{IBNR}(\tau) \tag{4.8}$$

$$= \int_0^\tau \mu_{IBNR}(t)\, \mathrm{E}\,(X^{(t)}(\tau - t, \infty) |\, S^{(t)}(\tau - t) = 0)\, dt$$

$$= \int_0^\tau \mu_{IBNR}(t)\, V(\tau - t | 0)\, dt,$$

and

$$\Gamma_{IBNR}(\tau) = \operatorname{Var} X_{IBNR}(\tau) \tag{4.9}$$

$$= \int_0^\tau \mu_{IBNR}(t)\, \mathrm{E}\,(X^{(t)}(\tau - t, \infty)^2 |\, S^{(t)}(\tau - t) = 0)\, dt$$

$$= \int_0^\tau \mu_{IBNR}(t)\, [\Gamma(\tau - t | 0) + V(\tau - t | 0)^2]\, dt.$$

We have now derived formulas for the IBNR and RBNS reserves (4.6), (4.8), and the corresponding prediction errors (4.7), (4.9), expressed in terms of the reserve- and variance functions (3.3) and (3.13). The total reserve is (of course) the sum of IBNR and RBNS reserves. Because the marked point processes corresponding to $\{K_{IBNR}(t)\}$ and $\{K_{RBNS}(t)\}$ are stochastically independent, it also holds that the prediction error corresponding to the total reserve is obtained by adding the prediction errors corresponding to the IBNR and RBNS components.

Remark 4.1. If $V(u|j)$ and $\Gamma(u|j)$ are calculated directly by solving Thiele's differential equation as advocated in Remark 3.1, one also needs an expression for $p_{00}(0, u)$ in order to calculate (4.8) and (4.9). However, since state 0 is strongly transient, we have the expression

$$p_{00}(0, u) = \exp\left\{ - \int_0^u \sum_{m \neq 0} \lambda_{0m}(\xi)\, d\xi \right\}.$$

□

REFERENCES

ANDERSEN, P. K. and BORGAN, Ø. (1985) Counting process models for life history data: A review (with discussion). *Scan. Journ. Stat.*, **12**, 97–158.

ARJAS, E. (1989) The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin* **19**, 139–152.

HACHEMEISTER, C. A. (1980) A stochastic model for loss reserving. *Proceedings. ICA* **1980**, 185–194.

NORBERG, R. (1992) Hattendorff's theorem and Thiele's differential equation generalized. *Scand. Actuarial J.*, **1992**, 2–14.

NORBERG, R. (1993) Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin* **23**, 95–115.

PAPATRIANDAFYLOY, A. and WATERS, H. R. (1984) Martingales in life insurance. *Scand. Actuarial J.*, **1984**, 210–230.

RAMLAU-HAMSEN, H. (1988) Hattendorff's theorem: A Markov chain and counting process approach. *Scand. Actuarial J.*, **1988**, 143–156.

OLE HESSELAGER

Laboratory of Actuarial Mathematics, Universitetsparken 5,
University of Copenhagen, DK-2100 Copenhagen Ø.

MARTINGALE APPROACH TO PRICING PERPETUAL AMERICAN OPTIONS

By Hans U. Gerber

Université de Lausanne, Switzerland

and Elias S. W. Shiu

The University of Iowa, U.S.A.

Abstract

The method of Esscher transforms is a tool for valuing options on a stock, if the logarithm of the stock price is governed by a stochastic process with stationary and independent increments. The price of a derivative security is calculated as the expectation, with respect to the risk-neutral Esscher measure, of the discounted payoffs. Applying the optional sampling theorem we derive a simple, yet general formula for the price of a perpetual American put option on a stock whose downward movements are skip-free. Similarly, we obtain a formula for the price of a perpetual American call option on a stock whose upward movements are skip-free. Under the classical assumption that the stock price is a geometric Brownian motion, the general perpetual American contingent claim is analysed, and formulas for the perpetual down-and-out call option and Russian option are obtained. The martingale approach avoids the use of differential equations and provides additional insight. We also explain the relationship between Samuelson's high contact condition and the first order condition for optimality.

Keywords

Black-Scholes formula; option-pricing theory; equivalent martingale measure; Esscher transform; perpetual American call option; perpetual American put option; perpetual down-and-out American call option; perpetual American strangle; perpetual American straddle; Russian option; optional sampling theorem; optimal stopping; high contact condition; smooth pasting condition.

1. Introduction

The option-pricing theory of Black and Scholes (1973) is perhaps the most important development in the theory of financial economics in the past two decades. A fundamental insight in advancing the theory is the concept of *risk-neutral valuation* introduced by Cox and Ross (1976). Further elaboration on this idea was given by Harrison and Kreps (1979), Harrison and Pliska (1981) and others under the terminology of *equivalent martingale measure*. It is now understood that the absence of arbitrage is "essentially" equivalent to the existence of an equivalent martingale measure, and some authors (DYBVIG and ROSS, 1987; SCHACHERMAYER, 1992) call this the *Fundamental Theorem of Asset Pricing*.

Under the assumption that the logarithm of the stock price is governed by a stochastic process with stationary and independent increments, one may determine such an equivalent martingale measure by a time-honored technique in actuarial science — the Esscher transform (ESSCHER, 1932). An Esscher transform induces an equivalent probability measure on such a stock-price process. The risk-neutral Esscher parameter (which is unique) is determined so that the stock price, discounted by the risk-free interest rate less the dividend yield, becomes a martingale under the new probability measure. The price of a derivative security is the supremum of the expected discounted payoffs, where the expectation is taken with respect to this equivalent martingale measure and the discounting is calculated using the risk-free interest rate.

The pricing of American options with a finite expiration date has been a challenging problem in the field of financial economics. A main difficulty is the determination of the optimal exercise boundary. Some papers on American options in the past decade are BENSOUSSAN (1984), MACMILLAN (1986), BARONE-ADESI and WHALEY (1987), OMBERG (1987), KARATZAS (1988), JAILLET, LAMBERTON and LAPEYRE (1990), KIM (1990), JACKA (1991), CARR, JARROW and MYNENI (1992), MYNENI (1992), CHESNEY, ELLIOT and GIBSON (1993), LAMBERTON (1993), HULL and WHITE (1993), and TILLEY (1993). In this paper we study the pricing of American options without expiration date by means of the Esscher transform and the optional sampling (stopping) theorem. This is a more tractable problem because the optimal exercise boundary of a perpetual American option does not vary with respect to the time variable. We derive a simple, yet general formula for the price of a perpetual American put option on a stock whose downward movements are skip-free (jump-free). Similarly, we obtain a formula for the price of a perpetual American call option on a stock whose upward movements are skip-free. In the appendix, we present a family of stochastic processes for modeling such stock-price movements. This family includes the Wiener process, gamma process and inverse Gaussian process, and combinations of such processes.

Under the classical assumption that the stock price is a geometric Brownian motion, the general perpetual American contingent claim is analysed, and formulas for the perpetual down-and-out call option and Russian option are obtained. The martingale approach avoids the use of differential equations and provides additional insight. We also explain the relationship between Samuelson's high contact condition and the first order conditions for optimality.

2. THE RISK-NEUTRAL ESSCHER TRANSFORM

Let $S(t)$ be the price of a stock at time t, $t \geq 0$. We assume that the process, $\{X(t)\}_{t \geq 0}$, defined by

$$S(t) = S(0)\, e^{X(t)}, \qquad t \geq 0, \tag{2.1}$$

is one with stationary and independent increments. Let

$$F(x, t) = \Pr[X(t) \leq x], \qquad t \geq 0, \tag{2.2}$$

be the distribution of the random variable $X(t)$, and

$$M(z, t) = E[e^{zX(t)}], \qquad t \geq 0, \tag{2.3}$$

its moment generating function. Under a mild continuity condition (Breiman, 1968, Section 14.4),

$$M(z, t) = [M(z, 1)]^t, \qquad t \geq 0. \tag{2.4}$$

While the Esscher transform of a random variable is a well-established concept, in this paper we consider the Esscher transform of a stochastic process which satisfies (2.4). The Esscher transform (parameter h) of $\{X(t)\}_{t \geq 0}$ is again a process with stationary and independent increments; the modified distribution of $X(t)$ is now

$$\begin{aligned} F(x, t; h) &= \Pr[X(t) \leq x; h] \\ &= \frac{\int_{-\infty}^{x} e^{hy}\, dF(y, t)}{\int_{-\infty}^{\infty} e^{hy}\, dF(y, t)} \\ &= \frac{1}{M(h, t)} \int_{-\infty}^{x} e^{hy}\, dF(y, t). \end{aligned}$$

The corresponding moment generating function is

$$M(z, t; h) = \frac{M(z + h, t)}{M(h, t)}. \tag{2.5}$$

It follows from (2.4) that

$$\begin{aligned} M(z, t; h) &= \left[\frac{M(z + h, 1)}{M(h, 1)}\right]^t \\ &= [M(z, 1; h)]^t. \end{aligned} \tag{2.6}$$

Because the exponential function is positive, the old and new measures have the same null sets, i.e., they are equivalent probability measures. The appropriate parameter $h = h^*$ is determined according to the principle of risk-neutral valuation (Cox and Ross, 1976), or, using the terminology of Harrison and Kreps (1979) and Harrison and Pliska (1981), we seek $h = h^*$ to obtain an equivalent martingale measure.

In this paper we assume that the risk-free force of interest is constant, and it is denoted by δ. We also assume that the market is frictionless and trading is continuous. There are no taxes, no transaction costs, and no restriction on borrowing or short sales. All securities are perfectly divisible. Furthermore, we assume that the stock pays a continuous stream of dividends, at a rate proportional to its price, i.e., there is a nonnegative constant ρ such that the dividend paid between time t and $t + dt$ is $S(t)\rho dt$. The parameter $h = h^*$ is chosen so that the process $\{e^{-(\delta-\rho)t} S(t)\}_{t \geq 0}$ is a martingale with respect to the probability measure corresponding to h^*. In particular,

$$S(0) = E[e^{-(\delta-\rho)t} S(t); h^*]; \tag{2.7}$$

hence, by (2.1) and (2.6),

$$\begin{aligned} e^{(\delta-\rho)t} &= E[e^{X(t)}; h^*] \\ &= [M(1, 1; h^*)]^t, \end{aligned}$$

or

$$\ln [M(1, 1; h^*)] = \delta - \rho. \tag{2.8}$$

The Esscher measure corresponding to the parameter h^* is called the *risk-neutral Esscher measure*. The price of a derivative security, whose payments depend on $\{S(t)\}$, is calculated as a discounted expected value, where the expectation is taken with respect to the risk-neutral Esscher measure.

Under some regularity conditions, equation (2.8) has a unique solution. To see this, consider the function

$$g(h) = \ln [M(1, 1; h)] = \ln [M(1 + h, 1)] - \ln [M(h, 1)].$$

The formula

$$\frac{d}{dh} E[X(1); h] = \mathrm{Var}\,[X(1); h]$$

shows that $E[X(1); h]$ is an increasing function in h. Hence

$$g'(h) = E[X(1); 1 + h] - E[X(1); h]$$

is positive, showing that $g(h)$ is an increasing function. This proves the uniquencess of the solution of equation (2.8), which is

$$g(h) = \delta - \rho.$$

To discuss the existence of the solution, let M and m denote the right and left end point of the (essential) range of $X(1)$, respectively. (M may be $+\infty$ and m may be $-\infty$). We may assume

$$m + \rho < \delta < M + \rho,$$

or

$$m < \delta - \rho < M,$$

because otherwise arbitrage would be possible. Let (a, b) denote the interval of values of h for which $g(h)$ exists. Under some regularity conditions,

$$\lim_{h \downarrow a} g(h) = m, \qquad \lim_{h \uparrow b} g(h) = M,$$

in which case (2.8) does have a solution. It should be noted that, although the risk-neutral Esscher measure is unique, there may be other equivalent martingale measures; see DELBAEN and HAEZENDONCK (1989) for a study on equivalent martingale measures of compound Poisson processes.

The price of a derivative security is taken as the expectation of its discounted payoffs with respect to the risk-neutral Esscher measure. For example, consider a *European call option* on the stock with exercise price K and exercise date t, $t > 0$. Let $I(\cdot)$ denote the indicator function and $\kappa = \ln[K/S(0)]$. The price of the option (at time 0) is

$$\begin{aligned} (2.9) \qquad & e^{-\delta t} E[(S(t) - K)\, I(S(t) > K); h^*] \\ &= e^{-\delta t} E[S(t)\, I(S(t) > K); h^*] - e^{-\delta t} K E[I(S(t) > K); h^*]. \end{aligned}$$

The second expectation in the right-hand side of (2.9) is

$$\Pr[S(t) > K; h^*] = 1 - F(\kappa, t; h^*).$$

To evaluate the first expectation in the right-hand side of (2.9), note that, for each measurable function $g(\cdot)$,

$$\begin{aligned} (2.10) \qquad E[g(S(t)); h] &= \frac{E[g(S(t))\, e^{hX(t)}]}{E[e^{hX(t)}]} \\ &= \frac{E[g(S(t))\, S(t)^h]}{E[S(t)^h]}. \end{aligned}$$

With this formula, the following result can be proved.

Lemma: Let h and k be two real numbers. Assume that the Esscher transforms of parameters h and $h + k$ exist. Then, for each measurable function $\psi(\cdot)$,

$$(2.11) \qquad E[S(t)^k \psi(S(t)); h] = E[S(t)^k; h]\, E[\psi(S(t)); h + k].$$

Applying the Lemma [with $k = 1$, $\psi(x) = I(x > K)$ and $h = h^*$] and (2.7), we obtain

$$\begin{aligned} E[S(t)\, I(S(t) > K); h^*] &= E[S(t); h^*]\, E[I(S(t) > K); h^* + 1] \\ &= S(0)\, e^{(\delta - \rho)t} \Pr[(S(t) > K); h^* + 1]. \end{aligned}$$

Thus the price of the European call option is

$$(2.12) \qquad S(0)\, e^{-\rho t}[1 - F(\kappa, t; h^* + 1)] - K e^{-\delta t}[1 - F(\kappa, t; h^*)].$$

If $\{X(t)\}$ is a Wiener process with variance per unit time σ^2, then (2.12) (with (4.2) below) yields the expression

$$(2.13)\quad S(0)\,e^{-\rho t}\,\Phi\left(\frac{-\kappa+(\delta-\rho+\sigma^2/2)\,t}{\sigma\sqrt{t}}\right)-Ke^{-\delta t}\,\Phi\left(\frac{-\kappa+(\delta-\rho-\sigma^2/2)\,t}{\sigma\sqrt{t}}\right),$$

where $\Phi(\cdot)$ denotes the standardized normal distribution. For $\rho=0$ this is the celebrated *Black-Scholes formula.* Formula (2.13) is the same as formula (53) in SMITH (1976).

Remarks:

(1) We assume that the stock pays dividends at a constant proportional rate ρ. If all dividends are reinvested in the stock, then each share of the stock at time 0 grows to $e^{\rho t}$ shares at time t; this gives an interpretation for formula (2.7),

$$S(0)=E[e^{-\delta t}S(t)\,e^{\rho t};\,h^*].$$

On the other hand, we can also consider the situation where none of the dividends are reinvested in the stock, leading to the intuitive formula:

$$(2.14)\qquad S(0)=E\left[\int_0^t e^{-\delta u}S(u)\,\rho du+e^{-\delta t}S(t);\,h^*\right].$$

To prove (2.14), we interchange the order of expectation and integration on the right-hand side and apply the formula

$$E[e^{-\delta u}S(u);\,h^*]=e^{-\rho u}S(0);$$

thus

$$\begin{aligned}\text{R.H.S.}&=S(0)\left(\int_0^t e^{-\rho u}\rho du+e^{-\rho t}\right)\\&=S(0)\\&=\text{L.H.S.}\end{aligned}$$

(2) Formula (2.12) may be used to price currency exchange options, with $S(t)$ denoting the spot exchange rate at time t, δ the domestic force of interest and ρ the foreign force of interest. In this context, (2.13) is known as the *Garman-Kohlhagen formula.*

3. PRICING PERPETUAL AMERICAN OPTIONS

In this section, by applying the optimal sampling theorem, we derive pricing formulas for perpetual American put and call options on a stock. We make the assumptions about stock prices and dividends that were introduced in the previous section. In addition, when pricing a perpetual American put option, we assume that the downward movements of the stock price are skip-free. Similarly, when pricing a perpetual American call option, we assume that the upward movements of the stock price are skip-free. Under these convenient assumptions, attractive formulas can be obtained.

First, we consider a *perpetual American put option* with exercise price K. We temporarily assume that $K < S(0)$, so that an immediate exercise of the option can be excluded. The owner of this option exercises it according to some strategy: a stopping time T. Then, at time T, he will get

$$(K - S(T))_+ ,$$

where $x_+ = \text{Max}(x, 0)$. Thus the value (at time 0) associated with the strategy is

$$E[e^{-\delta T}(K - S(T))_+ ; h^*] . \tag{3.1}$$

To maximize this expression, we can limit ourselves to stationary strategies of the form

$$T_L = \inf\{t \mid S(t) \leq L\} , \tag{3.2}$$

where $L \leq K$; the option is exercised the first time when (if ever) the price of the stock falls below or equals the level L. The price of the option is the maximal value of

$$E[e^{-\delta T_L}(K - S(T_L))_+ ; h^*] . \tag{3.3}$$

With the assumption that the stock-price process, $\{S(t)\}_{t \geq 0}$, is skip-free downwards, the stock price is equal to L at the time when the option is exercised, i.e.,

$$L = S(T_L) = S(0)\, e^{X(T_L)} . \tag{3.4}$$

For simplicity, denote the current stock price $S(0)$ by S and expression (3.3) by $V(S, L)$. Since $L \leq K$,

$$V(S, L) = (K - L)\, E[e^{-\delta T_L}; h^*] . \tag{3.5}$$

The expectation in (3.5) is a Laplace transform of T_L and can be calculated by the following classical argument.

Consider the stochastic process $\{e^{-\delta t + \theta X(t)}\}_{t \geq 0}$. For $t \leq T_L$, it is a *bounded martingale* with respect to the risk-neutral Esscher measure if the coefficient θ is the negative solution of the equation

$$E[e^{-\delta t + \theta X(t)}; h^*] = 1 ,$$

or

$$M(\theta, 1; h^*) = e^{\delta} . \tag{3.6}$$

Equation (3.6) has two (real) solutions; one is negative and the other is greater than one. To see this, consider the function

$$\phi(\theta) = M(\theta, 1; h^*) = E[e^{\theta X(1)}; h^*] .$$

Since

$$\phi''(\theta) = E[X(1)^2 e^{\theta X(1)}; h^*] > 0,$$

the function $\phi(\theta)$ is convex. Consequently, equation (3.6),

$$\phi(\theta) = e^{\delta},$$

has at most two solutions. We note that

$$\phi(0) = 1 < e^{\delta}$$

and, because of (2.8),

$$\phi(1) = e^{\delta - \rho} < e^{\delta}.$$

Let us assume that

$$\Pr[X(1) < 0] > 0$$

and

$$\Pr[X(1) > 0] > 0,$$

from which it follows that $\phi(\theta) \to +\infty$ for $\theta \to -\infty$ and for $\theta \to +\infty$. Thus equation (3.6) has two solutions, $\theta_0 < 0$ and $\theta_1 > 1$.

By the optional sampling theorem, we have

$$E[e^{-\delta T_L + \theta_0 X(T_L)}; h^*] = 1,$$

which, because of (3.4), becomes

$$E[e^{-\delta T_L}; h^*] = \left(\frac{L}{S}\right)^{-\theta_0}. \tag{3.7}$$

Applying (3.7) to (3.5) yields, for $S \geq L$ and $K > L$,

$$V(S, L) = (K - L)\left(\frac{L}{S}\right)^{-\theta_0}. \tag{3.8}$$

For a given current stock price S, we seek the maximal value of (3.8) by varying the option-exercise boundary L. Let V_L denote the partial derivative of V with respect to L. Solving the equation

$$V_L(S, L) = 0$$

yields the optimal exercise boundary

$$L = \tilde{L} = \frac{-\theta_0}{1 - \theta_0} K. \tag{3.9}$$

Thus the maximal value is

$$V(S, \tilde{L}) = \frac{K}{1 - \theta_0}\left[\frac{-K\theta_0}{S(1 - \theta_0)}\right]^{-\theta_0}.$$

This is the price of the perpetual American put option provided that $S \geq \tilde{L}$. For $S < \tilde{L}$, the option is exercised immediately and the price is simply $K - S$. Hence the option price is

$$\begin{cases} \dfrac{K}{1-\theta_0}\left[\dfrac{-K\theta_0}{S(1-\theta_0)}\right]^{-\theta_0} & \text{if} \quad S \geq \tilde{L} \\ K - S & \text{if} \quad S < \tilde{L} \end{cases} \tag{3.10}$$

It may seem surprising that δ and ρ do not appear in (3.10). However, they were used to determine θ_0.

Next we study the pricing of a *perpetual American call option* with exercise price K, and we temporarily assume that $K > S$. For $M \geq K$, let

$$T_M = \inf\{t \mid S(t) \geq M\} \tag{3.11}$$

and

$$W(S, M) = E[e^{-\delta T_M}(S(T_M) - K)_+ ; h^*]. \tag{3.12}$$

With the assumption that the stock-price process, $\{S(t)\}_{t \geq 0}$, is skip-free upwards, the stock price is equal to M at the time when the option is exercised, i.e., $S(T_M) = M$. Since $M \geq K$, formula (3.12) becomes

$$W(S, M) = (M - K)\, E[e^{-\delta T_M} ; h^*]. \tag{3.13}$$

The expectation in (3.13) is evaluated in the same way as above, except that we now use θ_1, the positive root of (3.6), to make sure that $\{\exp[-\delta t + \theta_1 X(t)]\}$ is a bounded martingale (with respect to the risk-neutral Esscher measure) for $t \leq T_M$. The resulting formula is

$$E[e^{-\delta T_M}; h^*] = \left(\frac{S}{M}\right)^{\theta_1}. \tag{3.14}$$

For given current stock price S, the maximal value of

$$W(S, M) = (M - K)\left(\frac{S}{M}\right)^{\theta_1} \tag{3.15}$$

is attained at

$$M = \tilde{M} = \frac{\theta_1}{\theta_1 - 1} K, \tag{3.16}$$

and

$$W(S, \tilde{M}) = \frac{K}{\theta_1 - 1}\left[\frac{S(\theta_1 - 1)}{K\theta_1}\right]^{\theta_1}. \tag{3.17}$$

This gives the price of the perpetual American call option provided $S \leq \tilde{M}$. For $S > \tilde{M}$, the option is exercised immediately and the price is simply $S - K$. Thus the option price is

$$(3.18) \qquad \begin{cases} \dfrac{K}{\theta_1 - 1}\left[\dfrac{S(\theta_1 - 1)}{K\theta_1}\right]^{\theta_1} & \text{if} \quad S \le \tilde{M} \\ S - K & \text{if} \quad S > \tilde{M}. \end{cases}$$

Remarks: As the dividend yield ρ tends to 0, the coefficient θ_1 tends to 1, the optimal exercise boundary $\tilde{M}$ tends to ∞, and the price of the perpetual American call option tends to S, the current stock price. These limiting results can be verified by direct calculations: for $\rho = 0$, $\theta_1 = 1$, (3.15) reduces to

$$(3.19) \qquad W(S, M) = \left(1 - \frac{K}{M}\right)S, \qquad M \ge K.$$

Since this is a strictly increasing function of M, its supremum is not attained for a finite value of M, and the maximal value (the value of the option) is S. Thus, if $\rho = 0$, the perpetual American call option will never be exercised, but nevertheless it has a positive value. To avoid this anomaly (to which INGERSOLL (1987, p. 373) refers as the problem of "infinities") we might modify the payoff of the call option as

$$[(S(T) - K)_+]^{\alpha}, \qquad 0 < \alpha < 1.$$

Then

$$W(S, M) = (M - K)^{\alpha}\frac{S}{M},$$

which is maximal for

$$\tilde{M} = \frac{K}{1 - \alpha}.$$

3.1. The high contact condition

Each of (3.10) and (3.18), as a function of the current stock price S, has a continuous first derivative, because

$$(3.1.1) \qquad \begin{aligned} V(\tilde{L}, \tilde{L}) &= K - \tilde{L}, \\ V_S(\tilde{L}, \tilde{L}) &= -1, \\ W(\tilde{M}, \tilde{M}) &= \tilde{M} - K, \end{aligned}$$

and

$$W_S(\tilde{M}, \tilde{M}) = 1.$$

Formulas (3.1.1) and (3.1.2) are special cases of the so-called *high contact condition* (SAMUELSON, 1965); in the literature about optimal stopping problems (SHIRAYAYEV, 1978, p. 160) the term is *smooth pasting condition*. SHEPP and

SHIRYAEV (1983) use the term the *principle of smooth fit* and attribute it to Kolmogorov. (Shirayayev is the same person as Shiryaev).

MERTON (1973, p. 171, footnote 60; 1990, p. 296, footnote 47) has derived the high contact condition as a first order condition necessary for optimality. (Merton's proof is reformulated on page 189 of BREKKE and ØKSENDAL (1991)). Under some weak conditions, the converse is also true — a solution proposal to an optimal stopping problem satisfying the high contact condition is in fact an optimal solution to the problem; a recent paper on this is BREKKE and ØKSENDAL (1991). It is easy to check that condition (3.1.1) does determine the optimal exercise boundary $\tilde{L}$, while (3.1.2) determines $\tilde{M}$.

We now derive a formula explaining how the high contact condition (3.1.1) and optimality for $V(S, \cdot)$ are related. Let

$$\lambda(S, L) = E[e^{-\delta T_L}; h^*]. \tag{3.1.3}$$

From (3.7) or simply by interpretation, it follows that, for $0 < x < S - L$,

$$\lambda(S, L) = \lambda(S, L + x)\, \lambda(L + x, L) \tag{3.1.4}$$

(cf. Lemma 7.1 on page 243 of KARLIN and TAYLOR (1981)). Differentiating (3.1.4) with respect to x and setting $x = 0$ yields

$$0 = \lambda_L(S, L) + \lambda(S, L)\, \lambda_S(L, L). \tag{3.1.5}$$

Now, let

$$\pi(x) = (K - x)_+ \tag{3.1.6}$$

denote the payoff function. Then (3.5) becomes

$$V(S, L) = \pi(L)\, \lambda(S, L). \tag{3.1.7}$$

Differentiating (3.1.7) with respect to L and applying (3.1.5) yields

$$\begin{aligned} V_L(S, L) &= \pi'(L)\, \lambda(S, L) + \pi(L)\, \lambda_L(S, L) \\ &= \pi'(L)\, \lambda(S, L) - \pi(L)\, \lambda(S, L)\, \lambda_S(L, L) \\ &= \lambda(S, L)\, [\pi'(L) - V_S(L, L)]. \end{aligned} \tag{3.1.8}$$

(Formula (3.1.8) can also be derived using (3.8)). Since $\lambda(S, L)$ is positive, $V_L(S, L) = 0$ if and only if

$$V_S(L, L) = \pi'(L). \tag{3.1.9}$$

Equation (3.1.8) shows explicitly that the optimal exercise boundary $\tilde{L}$ does not depend on the current stock price S. We note that (3.1.7), (3.1.8) and (3.1.9) are valid for payoff functions $\pi(\cdot)$ more general than (3.1.6).

Similarly, one can derive the formula

$$W_M(S, M) = \mu(S, M)\, [\pi'(M) - W_S(M, M)], \tag{3.1.10}$$

where

$$\mu(S, M) = E[e^{-\delta T_M}; h^*].$$

4. LOGARITHM OF THE STOCK PRICE AS A WIENER PROCESS

The stochastic process with stationary and independent increments and sample paths which are both skip-free upwards and downwards (i.e., continuous) is the Wiener process. In this section we assume that $\{X(t)\}_{t \geq 0}$ is a Wiener process; this is the classical geometric Brownian motion model for stock-price movements (SAMUELSON, 1965; BLACK and SCHOLES, 1973). Let μ and σ^2 denote, respectively, the mean and variance of $\{X(t)\}$ per unit time. In terms of a stochastic differential equation, the assumption is

$$\frac{dS(t)}{S(t)} = \left(\mu + \frac{\sigma^2}{2}\right)dt + \sigma dW(t), \qquad t \geq 0,$$

where $\{W(t)\}_{t \geq 0}$ denotes the standardized Wiener process.

Since

$$M(z, t) = \exp\,[(\mu z + ½\sigma^2 z^2)t],$$

it follows from (2.5) that

$$\ln\,[M(z, t; h)] = [(\mu + h\sigma^2)z + ½\sigma^2 z^2]\,t.$$

This shows that the transformed process has modified mean per unit time $\mu + h\sigma^2$ and unchanged variance per unit time σ^2. From (2.8) we get

$$(\mu + h^* \sigma^2) + ½\sigma^2 = \delta - \rho. \tag{4.1}$$

Thus to evaluate a derivative security, we use a Wiener process with mean per unit time

$$\mu + h^* \sigma^2 = \delta - \rho - ½\sigma^2. \tag{4.2}$$

From (3.6) we obtain

$$(\delta - \rho - ½\sigma^2)\,\theta + ½\sigma^2\theta^2 = \delta,$$

or

$$\sigma^2\theta^2 + (2\delta - 2\rho - \sigma^2)\,\theta - 2\delta = 0. \tag{4.3}$$

The roots of this quadratic equation are

$$\theta_0 = \frac{-(2\delta - 2\rho - \sigma^2) - \sqrt{(2\delta - 2\rho - \sigma^2)^2 + 8\sigma^2\delta}}{2\sigma^2} \tag{4.4}$$

and

$$\theta_1 = \frac{-(2\delta - 2\rho - \sigma^2) + \sqrt{(2\delta - 2\rho - \sigma^2)^2 + 8\sigma^2\delta}}{2\sigma^2}. \tag{4.5}$$

Formula (4.5) should be attributed to MCKEAN (1965, Section 3) who studied the pricing of perpetual warrants; at that date of course he did not solve the problem in terms of the risk-neutral measure. With zero dividend yield ($\rho = 0$), formula (4.4) becomes

$$\theta_0 = \frac{-2\delta}{\sigma^2}, \tag{4.6}$$

which was first given by MERTON (1973, Section 8; 1990, Section 8.8), who evaluated the perpetual American put option by adopting MCKEAN's (1965) technique. Discussions on pricing perpetual American options can also be found in the books by KARLIN and TAYLOR (1975, p. 365), INGERSOLL (1987, p. 375) and LAMBERTON and LAPEYRE (1991, p. 82), and in the recent articles by KARATZAS (1988, p. 59, e.g. 6.7), KIM (1990) and JACKA (1991, Proposition 2.3). (In formula (9) of KIM (1990), the denominator $1-\beta$ should be $\beta - 1$).

In the finance literature, the formulas for pricing perpetual American options are usually derived as follows. Let D denote the value of a derivative security. It follows from the hedging argument first given by BLACK and SHOLES (1973) that D satisfies the partial differential equation

$$\tfrac{1}{2}\sigma^2 S^2 D_{SS} + (\delta - \rho) S D_S - \delta D + D_t = 0, \tag{4.7}$$

subject to the appropriate boundary conditions. In the case of a perpetual option, we have $D_t = 0$ and (4.7) becomes a homogeneous, linear, second-order ordinary differential equation in S,

$$\tfrac{1}{2}\sigma^2 S^2 D_{SS} + (\delta - \rho) S D_S - \delta D = 0. \tag{4.8}$$

The function $D = S^\theta$ is a solution of (4.8) if the number θ satisfies the quadratic equation,

$$\tfrac{1}{2}\sigma^2 \theta(\theta - 1) + (\delta - \rho)\theta - \delta = 0, \tag{4.9}$$

which is the same as (4.3). Then any solution of (4.8) is of the form

$$D = c_0 S^{\theta_0} + c_1 S^{\theta_1}, \tag{4.10}$$

where c_0 and c_1 are independent of S.

In this paper we use the martingale approach and avoid differential equations. Additional insight for (4.10) is provided in the following; see (4.1.16) below.

4.1. Perpetual contingent claims

In this section we consider the pricing of perpetual contingent claims with U-shaped payoff functions such as

$$\pi(x) = a_1 (K_1 - x)_+ + a_2 (x - K_2)_+ . \tag{4.1.1}$$

For $a_1 = a_2 = 1$, the contingent claim may be called a *perpetual American strangle* if $K_1 < K_2$, and called a *perpetual American straddle* if $K_1 = K_2$. The assumption on $\{X(t)\}$ remains that it is a Wiener process.

Let $S = S(0)$ denote the current stock price. We consider exercise strategies arising from stopping times of the form

$$T_{L,M} = \inf\{t \mid S(t) = L \text{ or } S(t) = M\},$$

where $0 \le L \le S \le M$. The value of the contingent claim according to such a strategy is

$$V(S, L, M) = E[\pi(S(T_{L,M}))\, e^{-\delta T_{L,M}}; h^*]. \tag{4.1.2}$$

Put

$$\lambda(S, L, M) = E[I(S(T_{L,M}) = L)\, e^{-\delta T_{L,M}}; h^*] \tag{4.1.3}$$

and

$$\mu(S, L, M) = E[I(S(T_{L,M}) = M)\, e^{-\delta T_{L,M}}; h^*]. \tag{4.1.4}$$

Then

$$V(S, L, M) = \pi(L)\,\lambda(S, L, M) + \pi(M)\,\mu(S, L, M). \tag{4.1.5}$$

For $\theta = \theta_0$ and $\theta = \theta_1$ (the roots of equation (4.3)), the process $\{e^{-\delta t + \theta X(t)}\}$ is a bounded martingale (with respect to the risk-neutral measure) for $t \le T_{L,M}$. Applying the optional sampling theorem to these two martingales yields the equations

$$\lambda(S, L, M)\left(\frac{L}{S}\right)^{\theta_0} + \mu(S, L, M)\left(\frac{M}{S}\right)^{\theta_0} = 1 \tag{4.1.6}$$

and

$$\lambda(S, L, M)\left(\frac{L}{S}\right)^{\theta_1} + \mu(S, L, M)\left(\frac{M}{S}\right)^{\theta_1} = 1, \tag{4.1.7}$$

respectively, from which we obtain

$$\lambda(S, L, M) = \frac{M^{\theta_1} S^{\theta_0} - M^{\theta_0} S^{\theta_1}}{M^{\theta_1} L^{\theta_0} - M^{\theta_0} L^{\theta_1}} \tag{4.1.8}$$

and

$$\mu(S, L, M) = \frac{S^{\theta_1} L^{\theta_0} - S^{\theta_0} L^{\theta_1}}{M^{\theta_1} L^{\theta_0} - M^{\theta_0} L^{\theta_1}}. \tag{4.1.9}$$

Note that

$$\lim_{M \to \infty} \lambda(S, L, M) = \left(\frac{S}{L}\right)^{\theta_0} = \left(\frac{L}{S}\right)^{-\theta_0}, \tag{4.1.10}$$

which confirms (3.7), and

$$\lim_{L \downarrow 0} \mu(S, L, M) = \left(\frac{S}{M}\right)^{\theta_1}, \tag{4.1.11}$$

which is (3.14).

The remaining problem is to optimize $V(S, L, M)$, considered as a function of the exercise boundaries L and M. The first order conditions are

$$V_L(S, \tilde{L}, \tilde{M}) = 0$$

and

$$V_M(S, \tilde{L}, \tilde{M}) = 0.$$

These conditions do not depend on S (as long as S is between L and M). At first this seems surprising, but it follows immediately from the formulas

$$V_L(S, L, M) = \lambda(S, L, M)\,[\pi'(L) - V_S(L, L, M)] \tag{4.1.12}$$

and

$$V_M(S, L, M) = \mu(S, L, M)\,[\pi'(M) - V_S(M, L, M)], \tag{4.1.13}$$

which generalize (3.1.8) and (3.1.10), respectively. Thus the first order conditions become

$$V_S(\tilde{L}, \tilde{L}, \tilde{M}) = \pi'(\tilde{L}) \tag{4.1.14}$$

and

$$V_S(\tilde{M}, \tilde{L}, \tilde{M}) = \pi'(\tilde{M}), \tag{4.1.15}$$

which are the high contact conditions. The optimal exercise boundaries $\tilde{L}$ and $\tilde{M}$ are determined by solving (4.1.14) and (4.1.15) simultaneously. For $\tilde{L} \le S \le \tilde{M}$, the price of the perpetual contingent claim is

$$\begin{aligned} V(S, \tilde{L}, \tilde{M}) &= \pi(\tilde{L})\,\lambda(S, \tilde{L}, \tilde{M}) + \pi(\tilde{M})\,\mu(S, \tilde{L}, \tilde{M}) \\ &= (S^{\theta_0} \quad S^{\theta_1}) \begin{pmatrix} \tilde{L}^{\theta_0} & \tilde{L}^{\theta_1} \\ \tilde{M}^{\theta_0} & \tilde{M}^{\theta_1} \end{pmatrix}^{-1} \begin{pmatrix} \pi(\tilde{L}) \\ \pi(\tilde{M}) \end{pmatrix}. \end{aligned} \tag{4.1.16}$$

To prove (4.1.12), consider the identities

$$\lambda(S, L, M) = \lambda(S, L + x, M)\,\lambda(L + x, L, M)$$

and

$$\mu(S, L, M) = \mu(S, L + x, M) + \lambda(S, L + x, M)\,\mu(L + x, L, M)$$

where $0 < x < S - L$. Differentiating these equations with respect to x and setting $x = 0$ yields

$$0 = \lambda_L(S, L, M) + \lambda(S, L, M)\,\lambda_S(L, L, M) \tag{4.1.17}$$

and

(4.1.18) $$0 = \mu_L(S, L, M) + \lambda(S, L, M)\,\lambda_S(L, L, M),$$

respectively. Differentiating (4.1.5) with respect L and applying (4.1.17) and (4.1.18), we have

$$\begin{aligned} V_L(S, L, M) &= \pi'(L)\,\lambda(S, L, M) + \pi(L)\,\lambda_L(S, L, M) + \pi(M)\,\mu_L(S, L, M) \\ &= \lambda(S, L, M)\,[\pi'(L) - \pi(L)\,\lambda_S(L, L, M) - \pi(M)\,\mu_S(L, L, M)] \\ &= \lambda(S, L, M)\,[\pi'(L) - V_S(L, L, M)], \end{aligned}$$

which is (4.1.12). The proof of (4.1.13) is similar.

Remarks: For general payoff functions, there might be several disjoint optimal non-exercise intervals. For a matrix derivation of the results in this Section, see Section 5 of GERBER and SHIU (1994). There are closed-form formulas for *deferred* perpetual American call and put options; see GERBER and SHIU (1993b).

4.2. Perpetual down-and-out option

In this section we consider the pricing of a perpetual "down-and-out" American call option with exercise price K. The option contract becomes null and unexercisable, if the stock price declines to the knock-out price L, $L < K$. When this occurs, a rebate or refund of amount R is given. For $M \geq K$, it follows from (4.1.5) that the value of the strategy to exercise the call option when the stock price increases to M for the first time is

(4.2.1) $$V(S, L, M) = R\lambda(S, L, M) + (M - K)\,\mu(S, L, M), \qquad L \leq S \leq M.$$

Note that the lower exercise boundary L is fixed, and the problem is to maximize V as a function of the upper exercise boundary M.

We now consider the special case where the stock pays no dividends (hence $\theta_1 = 1$ and $\theta_0 = -2\delta/\sigma^2$). We shall show that the maximal value of (4.2.1) is obtained for $M \to \infty$ and that it is

(4.2.2) $$\begin{aligned} V(S, L, \infty) &= S + (R - L)\left(\frac{L}{S}\right)^{-\theta_0} \\ &= S + (R - L)\left(\frac{L}{S}\right)^{2\delta/\sigma^2}. \end{aligned}$$

This result can also be found in MERTON (1973, (57); 1990, (8.57)) and INGERSOLL (1987, p. 372, (39)).

For the proof we first observe that $\lambda(S, L, M)$ is an increasing function of M and hence the first term on the right-hand side of (4.2.1) is bounded by

$$R\lambda(S, L, \infty) = R\left(\frac{L}{S}\right)^{-\theta_0}.$$

The second term on the right-hand side of (4.2.1) may be estimated as follows:

$$\begin{aligned}(M-K)\,\mu(S,L,M) &= (M-K)\,\frac{SL^{\theta_0}-S^{\theta_0}L}{ML^{\theta_0}-M^{\theta_0}L}\\ &= \frac{M-K}{M-\left(\frac{L}{M}\right)^{-\theta_0}L}\left[S-L\left(\frac{L}{S}\right)^{-\theta_0}\right]\\ &< S-L\left(\frac{L}{S}\right)^{-\theta_0}\\ &= \lim_{M\to\infty}\ (M-K)\,\mu(S,L,M).\end{aligned}$$

4.3. The Russian option

Let M be a number such that $M \geq S$. Let

(4.3.1) $$M(t) = \max\,\{M, \max\,[S(u)\,|\,0 \leq u \leq t]\},$$

which can be interpreted as the historical maximum of the stock prices at time t. Note that the pair $\{S(t), M(t);\, t \geq 0\}$ is a homogeneous Markov process. The term "Russian option" was coined by SHEPP and SHIRAYAEV (1993) to describe a perpetual American option whose payoff is $M(t)$, if it is exercised at time t, $t \geq 0$. That is, the holder of a Russian option has the privilege of receiving the historical maximum of the stock prices up till when he chooses to exercise the option. The price at time 0 of the option is the supremum, over all stopping times $T \geq 0$, of

(4.3.2) $$E[e^{-\delta T}M(T);\, h^*].$$

SHEPP and SHIRYAEV (1993) show that there is a number $\tilde{k}$, which depends only on δ, ρ and σ, such that (if $S(0) > \tilde{k}M$) the optimal strategy is to exercise the option at the first time t when

(4.3.3) $$S(t) = \tilde{k}M(t).$$

Here we shall show how $\tilde{k}$ can be determined in a very transparent fashion. Let k be a number, $0 < k < 1$. For a current stock price $S = S(0)$ with $kM \leq S$, we consider the strategy to exercise the option at the stopping time

(4.3.4) $$T_k = \inf\,\{t\,|\,S(t) = kM(t)\}.$$

The value of this strategy is denoted by $R(S, M; k)$; we note that

$$R(S, M; k) = M\,R(S/M, 1; k).$$

From this and the definitions (4.1.3) and (4.1.4) it follows that

(4.3.5) $$\begin{aligned}R(S, M; k) &= M\lambda(S, kM, M) + R(M, M; k)\,\mu(S, kM, M)\\ &= M\,[\lambda(S, kM, M) + R(1, 1; k)\,\mu(S, kM, M)].\end{aligned}$$

Applying (4.1.8) and (4.1.9), we obtain

$$R(S, M;\ k) = M[\lambda(S/M, k, 1) + R(1, 1; k)\mu(S/M, k, 1)] \tag{4.3.6}$$

$$= \frac{M}{k^{\theta_0} - k^{\theta_1}} \{[(S/M)^{\theta_0} - (S/M)^{\theta_1}] + R(1, 1; k)\ [k^{\theta_0}(S/M)^{\theta_1} - k^{\theta_1}(S/M)^{\theta_0}]\},$$

where $R(1, 1; k)$ needs to be determined by the boundary condition at $S = M$. This condition can be derived by the following heuristic argument. If the current stock price S is very close to M, we can be "almost sure" that the stock price will attain the level M (and hence that the maximum will be increased) before the option is exercised. Thus, if S is close to M, $R(S, M; k)$ does not depend on the exact value of M,

$$R_M(M, M; k) = 0. \tag{4.3.7}$$

From this and (4.3.6) we get the condition

$$\frac{1}{k^{\theta_0} - k^{\theta_1}} \{[(1-\theta_0) - (1-\theta_1)] + R(1, 1; k)\,[k^{\theta_0}(1-\theta_1) - k^{\theta_1}(1-\theta_0)]\} = 0,$$

which yields

$$R(1, 1; k) = \frac{\theta_1 - \theta_0}{(1-\theta_0)k^{\theta_1} - (1-\theta_1)k^{\theta_0}}. \tag{4.3.8}$$

We substitute this expression into formula (4.3.6) and obtain after simplification the result that

$$R(S, M; k) = M\frac{(1-\theta_0)\,(S/M)^{\theta_1} + (\theta_1 - 1)\,(S/M)^{\theta_0}}{(1-\theta_0)k^{\theta_1} + (\theta_1 - 1)k^{\theta_0}}. \tag{4.3.9}$$

Now it is clear that the optimal value of k is the one that minimizes the denominator, whose derivative is

$$(1-\theta_0)\,\theta_1 k^{\theta_1 - 1} + (\theta_1 - 1)\,\theta_0 k^{\theta_0 - 1}.$$

Hence the optimal value is

$$\tilde{k} = \left(\frac{\theta_0(1-\theta_1)}{\theta_1(1-\theta_0)}\right)^{1/(\theta_1-\theta_0)}, \tag{4.3.10}$$

and the price of the Russian option is

$$\begin{cases} R(S, M; \tilde{k}) & \text{if} \quad \tilde{k}M \le S \le M \\ M & \text{if} \quad S \le \tilde{k}M \end{cases}. \tag{4.3.11}$$

Formulas (4.3.10) and (4.3.11) are equivalent to (2.3) and (2.4) of SHEPP and SHIRYAEV (1993).

ACKNOWLEDGEMENTS

The authors thank FRANÇOIS DUFRESNE for pointing out the paper by SHEPP and SHIRYAEV (1993) and the anonymous referees for their comments. ELIAS SHIU gratefully acknowledges the support from the Principal Financial Group.

REFERENCES

BARONE-ADESI, G. and WHALEY, R. E. (1987) Efficient analytic approximation of American option values. *Journal of Finance* **42**, 301–320.

BHATTACHARYA, S. and CONSTANTINIDES, G. (ed.) (1989) *Theory of Valuation: Frontier of Modern Financial Theory* **Volume** 1. ROWMAN and LITTLEFIELD, Totowa, N.J.

BLACK, F. and SCHOLES, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637–659.

BENSOUSSAN, A. (1984) On the theory of option pricing. *Acta Applicandae Mathematicae* **2**, 139–158.

BREIMAN, L. (1968) *Probability*. Addison-Wesley, Menlo Park, California, Reprinted by the Society of Industrial and Applied Mathematics, Philadelphia, 1992.

BREKKE, K. A. and ØKSENDAL, B. (1991) The high contact principle as a sufficiency condition for optimal stopping. In LUND and ØKSENDAL (1991), 187–208.

CARR, P., JARROW, R. and MYNENI, R. (1992) Alternative characterizations of American put options. *Mathematical Finance* **2**, 87–106.

CHESNEY, M., ELLIOTT, R. J. and GIBSON, R. (1993) Analytical solutions for the pricing of American bond and yield options. *Mathematical Finance* **3**, 277–294.

COOTNER, P. H. (ed.) (1967) *The Random Character of Stock Market Prices*, revised edition. M.I.T. Press, Cambridge, Massachusetts.

COX, J. C. and ROSS, S. A. (1976) The valuation of options for alternative stochastic processes. *Journal of Financial Economics* **3**, 145–166.

DELBAEN, F. and HAEZENDONCK, J. (1989) A martingale approach to premium calculation principles in an arbitrage free market. *Insurance: Mathematics and Economics* **8**, 269–277.

DUFRESNE, F., GERBER H. U. and SHIU, E. S. W. (1991) Risk theory with the gamma process. *ASTIN Bulletin* **21**, 177–192.

DYBVIG, P. H. and ROSS, S. A. (1987) Arbitrage. In *The New Palgrave: A Dictionary of Economics* **Vol. 1**, edited by J. EATWELL, M. MILGATE and P. NEWMAN, Macmillan, London, 100–106. Reprinted in *The New Palgrave: Finance*, edited by J. EATWELL, M. MILGATE and P. NEWMAN, W. W. NORTON, New York (1989), 57–71.

ESSCHER, F. (1932) On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift* **15**, 175–195.

GERBER, H. U. and SHIU, E. S. W. (1993a) Option pricing by Esscher transforms. *Proceedings of the 24th ASTIN Colloquium at Cambridge University* **Volume 2**, 305–344. Also to appear in the *Transactions of the Society of Actuaries* **46**, (1994).

GERBER, H. U. and SHIU, E. S. W. (1993b) Discussion on TILLEY (1993). *Transactions of the Society of Actuaries* **45**.

GERBER, H. U. and SHIU, E. S. W. (1994) From perpetual strangles to Russian options. *Insurance: Mathematics and Economics*.

GOLDMAN, M. B., SOSIN, H. B. and GATTO, M. A. (1979) Path dependent options: "Buy at the low, sell at the high". *Journal of Finance* **34**, 1111–1127.

HARRISON, J. M. and KREPS, D. M. (1979) Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* **20**, 381–408.

HARRISON, J. M. and PLISKA, S. (1981) Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* **11**, 215–260.

HESTON, S. L. (1993) Invisible parameters in option prices. *Journal of Finance* **48**, 933–947.

HULL, J. and WHITE, A. (1993) Efficient procedures for valuing European and American path-dependent options. *Journal of Derivatives* **1**, 21–31.

INGERSOLL, J. E., Jr. (1987) *Theory of Financial Decision Making*. ROWAN and LITTLEFIELD, Totowa, N.J.

JACKA, S. D. (1991) Optimal stopping and the American put. *Mathematical Finance* **1** (2), 1–14.

JAILLET, P., LAMBERTON, D. and LAPEYRE, B. (1990) Variational inequalities and the pricing of American options. *Acta Applicandae Mathematicae* **21**, 263–289.
KARATZAS, I. (1988) On the pricing of American options. *Applied Mathematics and Optimization* **17**, 37–60.
KARLIN, S. and TAYLOR, H. (1975) *A First Course in Stochastic Processes* (2nd edn). Academic Press, New York.
KARLIN, S. and TAYLOR, H. (1981) *A Second Course in Stochastic Processes*. Academic Press, New York.
KIM, I.J. (1990) The analytic valuation of American options. *Review of Financial Studies* **3**, 545–572.
LAMBERTON, D. (1993) Convergence of the critical price in the approximation of American options. *Mathematical Finance* **3**, 179–190.
LAMBERTON, D. and LAPEYRE, B. (1991) *Introduction au calcul stochastique appliqué à la finance*. Mathématiques & Applications #9, Société de Mathématiques Appliquées et Industrielles; distributed by ELLIPSES, Paris, France.
LUND, D. and ØKSENDAL, B. (ed.) (1991) *Stochastic Models and Option Values: Applications to Resources, Environment and Investment Problems*. North-Holland, Amsterdam.
MACMILLAN, L.W. (1986) Analytic approximation for the American put. *Advances in Futures and Options Research* **1** (A), 119–139.
MCKEAN, H.P. (1965) A free boundary problem for the heat equation arising from a problem in mathematial economics: An appendix to SAMUELSON (1965). *Industrial Management Review* **6** (2), 32–39. Reprinted in COOTNER (1967), 525–532.
MERTON, R.C. (1973) Theory of rational option pricing. *Bell Journal of Economics and Management Science* **4**, 141–183. Reprinted in BHATTACHARYA and CONSTANTINIDES (1989), 229–271, and as Chapter 8 in MERTON (1990), 225–308.
MERTON, R.C. (1990) *Continuous-Time Finance*. Basil Blackwell, Oxford.
MYNENI, R. (1992) The pricing of the American option. *Annals of Applied Probability* **2**, 1–23.
OMBERG, E. (1987) The valuation of American put options with exponential exercise polices. *Advances in Futures and Options Research* **2**, 117–142.
SAMUELSON, P.A. (1965) Rational theory of warrant pricing. *Industrial Management Review* **6** (2), 13–32. Reprinted in COOTNER (1967), 506–525.
SCHACHERMAYER, W. (1992) A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete time. *Insurance: Mathematics and Economics* **11**, 249–257.
SHEPP, L. and SHIRYAEV, A.N. (1993) The Russian option: Reduced regret. *Annals of Applied Probability* **3**, 631–640.
SHIRAYAYEV, A.N. (1978) *Optimal Stopping Rules*. Springer-Verlag, New York.
SMITH, C.W., Jr. (1976) Option pricing: A review. *Journal of Financial Economics* **3**, 3–51.
TILLEY, J.A. (1993) Valuing American options in a path simulation model. *Transactions of the Society of Actuaries* **45**.

APPENDIX

A1. Semi-continuous sample paths

In the rest of this paper, we consider the assumption that the sample paths of $\{S(t)\}$, or equivalently, those of $\{X(t)\}$, are skip-free downwards. (This assumption was used in deriving (3.10)). Then the following decomposition holds:

$$X(t) = Y(t) + \nu^2 W(t) - ct, \qquad t \geq 0. \tag{A.1.1}$$

Here, $\{Y(t)\}$ is either a compound Poisson process with positive increments or the limit of such processes; $\{W(t)\}$ is an independent standardized Wiener process (with zero drift and unit variance per unit time); the last term, ct, represents a deterministic drift. The cumulant generating function of the random variable $X(t)$ is of the form

$$\ln[M(z, t)] = t\left\{\int_0^\infty (e^{zx} - 1)\,[-dQ(x)] + \nu^2 z^2/2 - cz\right\}, \tag{A.1.2}$$

where $Q(x)$ is some nonnegative and nonincreasing function with $Q(\infty) = 0$. Note that, for each positive number ε, the integral

$$\int_{\varepsilon}^{\infty} (e^{zx} - 1)\,[-dQ(x)],$$

as a function in z, is the cumulant generating function of a compound Poisson distribution with Poisson parameter

$$\lambda(\varepsilon) = Q(\varepsilon),$$

and jump amount distribution

$$P(x; \varepsilon) = \frac{Q(\varepsilon) - Q(x)}{Q(\varepsilon)}, \qquad x \geq \varepsilon.$$

For notational simplicity, we assume that

$$-dQ(x) = q(x)\,dx$$

for some nonnegative function $q(x)$. Let μ and σ^2 denote, respectively, the mean and variance of $\{X(t)\}$ per unit time. Then

(A.1.3) $$\mu t = E[X(t)] = \left[\int_0^{\infty} x\,q(x)\,dx - c\right] t,$$

(A.1.4) $$\sigma^2 t = \mathrm{Var}\,[X(t)] = \left[\int_0^{\infty} x^2\,q(x)\,dx + v^2\right] t,$$

and

(A.1.5) $$E[(X(t) - \mu t)^3] = t \int_0^{\infty} x^3\,q(x)\,dx.$$

In general, for $n \geq 3$, the *n-th* cumulant of $X(t)$ is given by

$$t \int_0^{\infty} x^n\,q(x)\,dx.$$

If follows from (2.5) and (A.1.2) that

(A.1.6) $$\ln[M(z, t; h)] = \ln[M(z + h, t)] - \ln[M(h, t)]$$
$$= t\left\{\int_0^{\infty} (e^{zx} - 1)\,e^{hx}\,q(x)\,dx + v^2 z^2/2 - (c - v^2 h)z\right\}.$$

Thus the Esscher transform (parameter h) of a process defined by (A.1.1) is of the same type, with the following modifications:

(A.1.7) $$q(x) \to e^{hx} q(x),$$

(A.1.8) $$\nu^2 \to \nu^2 \quad \text{(unchanged)},$$

(A.1.9) $$c \to c - \nu^2 h.$$

Furthermore, it follows from (A.1.6) that (2.8) and (3.6) can be written as

(A.1.10) $$\int_0^\infty (e^x - 1)\, e^{h^* x} q(x)\, dx + \nu^2 h^* = c + \delta - \rho - \frac{\nu^2}{2}$$

and

(A.1.11) $$\int_0^\infty (e^{\theta x} - 1)\, e^{h^* x} q(x)\, dx + \frac{\nu^2 \theta^2}{2} - (c - \nu^2 h^*)\theta = \delta,$$

respectively.

A2. A particular family

For the model defined by (2.1) and (A.1.1), we now assume that $\nu = 0$, i.e.,

$$S(t) = S(0)\, e^{Y(t) - ct},$$

and that

(A.2.1) $$q(x) = a x^{\alpha - 1} e^{-bx}, \qquad x > 0,$$

where $a > 0$, $\alpha > -1$, and $b > 0$ are three parameters. In the context of risk theory, DUFRESNE, GERBER and SHIU (1991) have considered such a $q(x)$ function.

According to (A.1.7), for $h < b$, the Esscher transform of a process defined by (A.2.1) is a member of the same family, with b replaced by

(A.2.2) $$b(h) = b - h.$$

The moment generating function of $Y(t)$ is

(A.2.3) $$\exp\left[t \int_0^\infty (e^{zx} - 1)\, q(x)\, dx\right] = \exp\left[at \int_0^\infty (e^{zx} - 1)\, x^{\alpha - 1} e^{-bx}\, dx\right]$$

$$= \begin{cases} \left(\dfrac{b}{b - z}\right)^{at} & \text{if } \alpha = 0 \\[2ex] \exp\left(\dfrac{a\Gamma(\alpha)}{b^\alpha}\left[\left(\dfrac{b}{b - z}\right)^\alpha - 1\right] t\right) & \text{if } \alpha \neq 0 \end{cases}$$

Thus, for $\alpha = 0$, $\{Y(t)\}_{t \geq 0}$ is a gamma process; for $\alpha > 0$, it is a compound Poisson process with Poisson parameter

$$\lambda(a, \alpha, b) = \frac{a\Gamma(\alpha)}{b^{\alpha}},$$

and gamma jump density

$$p(x; \alpha, b) = \frac{b^{\alpha}}{\Gamma(\alpha)} x^{\alpha - 1} e^{-bx}, \qquad x > 0.$$

For $-1 < \alpha < 0$, the most prominent case is $\alpha = -\frac{1}{2}$, where $\{Y(t)\}_{t \geq 0}$ is an inverse Gaussian process and the density function of $Y(t)$ is

$$\frac{at}{x^{3/2}} \exp\left[\frac{-(\sqrt{b}x - \sqrt{\pi}at)^2}{x}\right], \qquad x > 0.$$

The condition for

$$b^* = b(h^*) = b - h^*$$

becomes

$$\text{(A.2.4)} \qquad \frac{b^*}{b^* - 1} = e^{\frac{c + \delta - \rho}{a}} \qquad \text{if} \qquad \alpha = 0$$

and

$$\text{(A.2.5)} \qquad \frac{1}{(b^* - 1)^{\alpha}} - \frac{1}{b^{*\alpha}} = \frac{c + \delta - \rho}{a\Gamma(\alpha)} \qquad \text{if} \qquad \alpha \neq 0.$$

Solving (A.2.4) yields

$$\text{(A.2.6)} \qquad b^* = \frac{1}{1 - e^{-(c + \delta - \rho)/a}},$$

which, with $\rho = 0$, is formula (3.1.7) in GERBER and SHIU (1993a). In general, equation (A.2.5) does not yield a closed-form solution for b^*. However, if $\alpha = 1$ (exponential jump amounts), one finds

$$\text{(A.2.7)} \qquad b^* = \frac{1 + \sqrt{1 + \dfrac{4a}{c + \delta - \rho}}}{2}.$$

A discussion of the case where $\alpha = -\frac{1}{2}$ can be found in GERBER and SHIU (1993a).

For each fixed α, we might determine the parameters, a, b and c, by the method of moments. Thus we assume that we know μ, σ and the third central moment of $X(1)$, which we write as $\gamma\sigma^3$ (γ being the coefficient of skewness). Matching the first three moments (by means of formulas (A.1.3), (A.1.4) and (A.1.5)) yields the equations:

$$\mu = \int_0^\infty x\, q(x)\, dx - c = \frac{a\Gamma(\alpha+1)}{b^{\alpha+1}} - c\,,$$

$$\sigma^2 = \int_0^\infty x^2\, q(x)\, dx = \frac{a\Gamma(\alpha+2)}{b^{\alpha+2}}\,,$$

and

$$\gamma\sigma^3 = \int_0^\infty x^3\, q(x)\, dx = \frac{a\Gamma(\alpha+3)}{b^{\alpha+3}}\,.$$

From these equations we obtain

$$b = \frac{\alpha+2}{\gamma\sigma}$$

(to be replaced by b^* for the evaluation of a derivative security),

$$a = \frac{(\alpha+2)^{\alpha+2}}{\Gamma(\alpha+2)\,\gamma^{\alpha+2}\,\sigma^{\alpha}}\,, \tag{A.2.8}$$

and

$$c = \frac{\alpha+2}{\alpha+1}\;\frac{\sigma}{\gamma} - \mu\,. \tag{A.2.9}$$

These formulas generalize (and explain!) the formulas in Sections V.2 and V.3 of GERBER and SHIU (1993a). We note that HESTON (1993) has independently introduced the gamma process for modeling stock-price movements; his formula (10a) is the same as formula (4.1.7) of GERBER and SHIU (1993a).

A3. Formulas for the negative root

With the assumptions $\nu = 0$ and

$$q(x) = a x^{\alpha-1} e^{-bx}, \qquad x > 0,$$

equation (A.1.11) becomes

$$a\int_0^\infty (e^{\theta x} - 1)\, e^{h^* x} x^{\alpha-1} e^{-bx}\, dx - c\theta = \delta\,,$$

or

$$\int_0^\infty (e^{\theta x} - 1)\, x^{\alpha-1} e^{-b^* x}\, dx = \frac{\delta + c\theta}{a}\,. \tag{A.3.1}$$

The value of the integral in the left-hand side of (A.3.1.) can be read off from (A.2.3).

If $\alpha = 0$, then (A.3.1) becomes

$$\frac{b^*}{b^* - \theta} = e^{\frac{\delta + c\theta}{a}}. \tag{A.3.2}$$

Substituting b^* in (A.3.2) with formula (A.2.6) yields

$$e^{-c\theta/a} + \theta\,[e^{\delta/a} - e^{-(c-\rho)/a}] = e^{\delta/a}. \tag{A.3.3}$$

By (A.2.8) and (A.2.9),

$$\frac{1}{a} = \frac{\gamma^2}{4}$$

and

$$\frac{c}{a} = \frac{\gamma}{2}\left(\sigma - \frac{\mu\gamma}{2}\right).$$

For example, assume that $\delta = 0.1$, $\rho = 0$, $\mu = 0.1$, $\sigma = 0.2$ and $\gamma = 1$. Then (A.3.3) becomes

$$e^{-3\theta/40} + \theta\,[e^{1/40} - e^{-3/40}] = e^{1/40},$$

from which we obtain

$$\theta_0 = -7.559609675.$$

Note that, in the Wiener process model (with $\delta = 0.1$ and $\sigma = 0.2$), $\theta_0 = -5$ by formula (4.6).

If $\alpha \neq 0$ and $\alpha > -1$, then (A.3.1) becomes

$$\frac{1}{(b^* - \theta)^{\alpha}} - \frac{1}{b^{*\alpha}} = \frac{\delta + c\theta}{a\Gamma(\alpha)}, \tag{A.3.4}$$

where b^* is defined by (A.2.5). In the special case where $\alpha = 1$, (A.3.4) simplifies as

$$\frac{1}{b^* - \theta} = \frac{1}{b^*} = \frac{\delta + c\theta}{a}, \tag{A.3.5}$$

which is a quadratic equation in θ, where b^* is given by (A.2.7),

$$a = \frac{27}{2\gamma^3\sigma}$$

and

$$c = \frac{3}{2}\frac{\sigma}{\gamma} - \mu.$$

Now, consider the zero dividend case ($\rho = 0$), then the positive root of (A.3.5) is $\theta_1 = 1$, and the negative root is

$$\theta_0 = -\delta b^*/c\,;$$

using the same numerical values as above, $\delta = 0.1$, $\mu = 0.1$, $\sigma = 0.2$ and $\gamma = 1$; we obtain

$$b^* = \frac{1 + \sqrt{901}}{2}$$

and

$$\theta_0 = -7.75416551\,.$$

HANS U. GERBER
Ecole des hautes études commerciales,
Université de Lausanne,
CH-1015 Lausanne, Switzerland.

ELIAS S. W. SHIU
Department of Statistics and Actuarial Science,
The University of Iowa,
Iowa City, Iowa 52242-1419, U.S.A.

ROBUST CREDIBILITY VIA ROBUST KALMAN FILTERING

BY ERHARD KREMER

Hamburg

ABSTRACT

Credibility theory is closely related to Kalman filtering. As a consequence, methods proposed for robustifying the Kalman filter can often be specialised to obtain robust credibility rating procedures. The application of one such method to several classical credibility models is shown in this paper.

1. INTRODUCTION

Credibility theory is a very old branch of risk theory and non-life insurance mathematics. Eearly results are by MOWBRAY (1914) and WHITNEY (1918). A theoretically elegant approach was given by BÜHLMANN (1967) and BÜHLMANN & STRAUB (1970).

The classical models presented by those authors can be generalised to regression models, hierarchical models and evolutionary models. Generalisations have been studies intensively in the actuarial literature over the past twenty years. Some key references are HACHEMEISTER (1975), TAYLOR (1979), SUNDT (1980, 1983), NORBERG (1980, 1986), KREMER (1988a, 1988b).

In later years several authors have investigated ways of robustifying credibility rating methods. The aim of robustification is to limit the influence of extremely large claim amounts on the estimated premium. The reader is referred to, e.g. GISLER (1980), BÜHLMANN et al. (1982), KREMER (1991), KÜNSCH (1992), GISLER & REINHARD (1993).

MEHRA (1973) pointed out that credibility estimation can be achieved by the Kalman filtering technique. DE JONG & ZEHNWIRTH (1983) explored the correspondence for the classical credibility models, and ZEHNWIRTH (1985) explored its implications for evolutionary models.

Robust versions of the Kalman filter have been studies for some time, for example by MASRELIEZ & MARTIN (1977), MEINHOLD & SINGPURWALLA (1989) and CIPRA & ROMERA (1991). Due to the close relation between Kalman filtering and credibility theory, it is obvious that corresponding robust versions of credibility rating techniques can be derived. In the present paper we specialise the method proposed by CIPRA & ROMERA (1991) to the three most important credibility models. The resulting robust credibility techniques turn out to be quite tractable.

2. PRELIMINARIES

Suppose that all probabilistic statements are based on a probability space (Ω, Δ, P) and consider a risk during periods with indices $i = 1, 2, 3, \ldots, n + l$. Assume that the claims behavior of a risk over all periods can be described by a parameter θ. Suppose that the value of this parameter is unknown and interpret it as a realisation of a random variable:

$$\theta : (\Omega, \Delta, P) \rightarrow (\Theta, \tau),$$

with the parameter space Θ and the σ-algebra τ on Θ. Let the observed claims amounts (or loss ratios) of the risk be represented by the nonnegative random variables:

$$X_i, \qquad \text{with} \qquad i = 1, 2, 3, \ldots, n + 1,$$

defined on (Ω, Δ, P). It is assumed that all X_i lie in the Hilbert space L_2 of measurables, square integrable functions f (identified with the equivalence class of all g which are P-a.e. equal to f) defined on Δ with scalar product:

$$\langle f_1, f_2 \rangle = E(f_1 \cdot f_2)$$

and norm:

$$\|f\| = E(f^2)^{1/2}.$$

In the given insurance context the conditional expectation

$$m_i = E(X_i \mid \theta)$$

is called the **net premium** (or **net loss ratio**) in period no. i. Then the credibility estimator is nothing else but the linear-affine prediction of m_{n+1} from $X_1, X_2, \ldots, X_n$. Defining the subspace A_n of all linear-affine combinations

$$f_n = a_0 + \sum_{i=1}^{n} a_i \cdot X_i,$$

the **credibility estimator** is defined as the projection of m_{n+1} on A_n, i.e. as the random variable $m_{n+1} \in A_n$ with

$$\|m_{n+1} - \hat{m}_{n+1}\| \leq \|m_{n+1} - f_n\|$$

for all $f_n \in A_n$.

In general the credibility estimator can be determined by solving certain normal equations. Under more special model assumptions one can derive explicit formulas for the credibility estimator. Three special models are given in the Section 4. Especially in the regression model one can calculate explicit credibility estimators, see e.g. NORBERG (1980).

It is well-known that the usual credibility estimators are not robust against extremely large claim amounts. This has led to attempts to contruct robust versions of the classical credibility estimators. One method to robustify the credibility estimator was given by Gisler already in 1980.

Alternative robustifications, which also cover the general regression model, are given by KREMER (1991).

3. ROBUST KALMAN FILTERING

The KALMAN filter is a well-known instrument for recursive prediction in dynamic linear systems. A dynamic linear system is defined by the two stochastic linear recursions:

$$X_i = H_i \cdot b_i + v_i$$
$$b_i = F_i \cdot b_{i-1} + w_i,$$

where X_i is a p-dimensional stochastic vector of observations, H_i a known (pxq)-dimensional design matrix, b_i a q-dimensional stochastic parameter vector and F_i a known (qxq)-dimensional transition matrix. The v_i, w_i are random disturbances with:

$$\begin{array}{lll} E(v_i) = 0, & E(w_i) = 0, & \\ E(v_i \cdot v_j^T) = 0, & E(w_i \cdot w_j^T) = 0, & i \neq j \\ E(v_i \cdot v_i^T) = R_i, & E(w_i \cdot w_i^T) = Q_i & \\ E(v_i \cdot w_j^T) = 0, & i \neq j, & \end{array}$$

where R_i, Q_i are known covariance matrices. The Kalman filter algorithm gives handy recursions for the optimal affine-linear predictor of b_i from $X_1, X_2, \ldots, X_{i-1}$. For more details see Section 3 in DE JONG et al. (1983).

Like other standard methods the Kalman filter is not robust to outliers. As a consequence, several authors have proposed robust versions of the usual Kalman filter algorithm. Recently, a handy robutistification was proposed by CIPRA & ROMERA (1991). We give a brief summary of their result.

Denote by $\hat{b}_i$, the one-step ahead prediction of b_i, based on the observations $X_1, \ldots, X_{i-1}$, and define its error covariance matrix as

$$C_i = E(b_i - \hat{b}_i)(b_i - \hat{b}_i)^T.$$

For a given covariance matrix M, we define as $M^{-1/2}$ any matric which satisfies $M^{-1/2} \cdot M \cdot (M^{-1/2})^T = I$.

With this convention we now introduce the matrices

$$A_i^{q \times q} = C_i^{-1/2},$$
$$D_i^{p \times q} = R_i^{-1/2} \cdot H_i,$$

and the random vectors

$$s_i^{p \times 1} = R_i^{-1/2} \cdot X_i,$$
$$p_i^{q \times 1} = A_i \cdot \hat{b}_i.$$

Note that, conditional on b_i, the stochastic vector s_i has mean $D_i \cdot b_i$ and covariance matrix I. Further note that

$$E(p_i - A_i \cdot b_i)(p_i - A_i \cdot b_i)^T = I.$$

In the spirit of ***M*-Estimation**, CIPRA & ROMERA (1991) propose to determine the updated estimate $\hat{b}_i^i$ of b_i, given all observations $X_1, \ldots, X_i$, as the solution of a minimisation problem, namely

$$\text{(3.1)} \qquad \text{minimise}\left\{\sum_{k=1}^{q} \rho_{1k}(p_{ki} - a_{ki}^T \cdot \hat{b}_i^i) + \sum_{j=1}^{p} \rho_{2j}(s_{ji} - d_{ji}^T \cdot \hat{b}_i^i)\right\}$$

Here $\rho_{11}, \ldots, \rho_{1q} \geq 0$ and $\rho_{21}, \ldots, \rho_{2q} \geq 0$ are arbitrary robustifying functions, the k-th row of A_i is denoted by a_{ki}^T, and the j-th row of D_i is denoted by d_{ji}^T. Denote the derivatives of the functions ρ by Ψ.

It is seen that the resulting estimator $\hat{b}_i^i$ will be the result of a compromise between the desire to minimise deviation from the one-step ahead prediction (note $p_i - A_i \cdot \hat{b}_i = 0$) and the desire to have $\hat{b}_i^i$ reflect the information in the new data as represented by s_i.

Having solved (3.1) to obtain $\hat{b}_i^i$, one obtains the one-step ahead prediction to use in the next recursion by

$$\hat{b}_{i+1} = F_{i+1} \cdot \hat{b}_i^i.$$

For $\hat{b}_i^i$ one has the normal equations (see (2.8) in CIPRA & ROMERA (1991)):

$$\text{(3.2)} \qquad \sum_{k=1}^{q} a_{ki}^T \cdot \Psi_{1k}(p_{ki} - a_{ki}^T \cdot \hat{b}_i^i) + \sum_{j=1}^{p} d_{ji}^T \cdot \Psi_{2j}(s_{ji} - d_{ji}^T \cdot \hat{b}_i^i) = 0.$$

By approximating $\hat{b}_i^i$ by $\hat{b}_i$ (3.2) gives a certain approximate normal equation

$$\text{(3.3)} \qquad \sum_{k=1}^{q} w_{1ki} \cdot a_{ki}^T \cdot (p_{ki} - a_{ki}^T \cdot \hat{b}_i^i) + \sum_{j=1}^{p} w_{2ji} \cdot d_{ji}^T \cdot (s_{ji} - d_{ji}^T \cdot \hat{b}_i^i) = 0$$

where the weights w_{1ki}, w_{2ji} are defined as:

$$w_{1ki} = \lim_{x \to 0}\left(\frac{\Psi_{1k}(x)}{x}\right)$$

$$w_{2ji} = \frac{\Psi_{2j}(s_{ji} - d_{ji}^T \cdot \hat{b}_i)}{s_{ji} - d_{ji}^T \cdot \hat{b}_i}.$$

So far for the general robustification of the Kalman-filter. Turn now to more practicable, special cases.

It is reasonable to assume that the disturbances w_i do not produce outliers. This results in the choice:

$$\Psi_{1k}(x) = x.$$

Furthermore one is willing to take:

$$\Psi_{2j}(\cdot) = \Psi(\cdot)$$

independent of j, e.g.:

$$\Psi_{2j}(\cdot) = \Psi_H(\cdot)$$

In the practical case $p = 1$ one can choose the so called (one-sided) **Huber-function**:

$$\Psi_H(z) = z, \quad \text{for} \quad z \leq c$$
$$= c, \quad \text{for} \quad z > c,$$

where c is a given positive constant. CIPRA & ROMERA (1991) propose to take (3.2) in case of the Huber-function Ψ_H for Ψ and the approximation (3.3) in case of general Ψ. For general Ψ the formula (3.3) gives the following recursion in case that $p = 1$:

$$\hat{b}_i^i = \hat{b}_i + \left(\frac{C_i \cdot H_i^T}{H_i \cdot C_i \cdot H_i^T + R_i/k_i}\right) \cdot (X_i - H_i \cdot \hat{b}_i) \tag{3.4}$$

where:

$$C_i = F_i \cdot C_{i-1}^{i-1} \cdot F_i^T + Q_i \tag{3.5}$$

$$C_i^i = C_i - \frac{C_i \cdot H_i^T \cdot H_i \cdot C_i}{H_i \cdot C_i \cdot H_i^T + R_i/k_i} \tag{3.6}$$

$$k_i = \frac{\Psi(R_i^{-1/2} \cdot (X_i - H_i \cdot \hat{b}_i))}{R_i^{-1/2} \cdot (X_i - H_i \cdot \hat{b}_i)}. \tag{3.7}$$

Note that the 'ordinary' Kalman filter is just (3.4)-(3.6) with $k_i = 1$.
For the special $\Psi = \Psi_H$ one gets from (3.2) in case $p = 1$ the recursion:

$$\hat{b}_i^i = \hat{b}_i + C_i \cdot H_i^T \cdot R_i^{-1/2} \cdot \Psi_H\left(\frac{R_i^{1/2} \cdot (X_i - H_i \cdot \hat{b}_i)}{H_i \cdot C_i \cdot H_i^T + R_i}\right) \tag{3.8}$$

Cipra & Romera propose to update C_i like in the original (nonrobust) Kalman-filter, i.e.:

$$C_i = F_i \cdot C_{i-1}^{i-1} \cdot F_i^T + Q_i \tag{3.9}$$

$$C_i^i = C_i - \frac{C_i \cdot H_i^T \cdot H_i \cdot C_i}{H_i \cdot C_i \cdot H_i^T + R_i}. \tag{3.10}$$

Obviously the resulting recursions are quite tractable.

4. ROBUST CREDIBILITY

We consider three well-known credibility models. For each model the robustifications (3.4)-(3.10) are presented, giving recursions for a robust credibility estimator.

4.1. Bühlmann-Straub (1970) model

Suppose that $X_1, \ldots, X_{n+1}$ are conditionally independent given θ. There exist measurable functions

$$\begin{aligned} \mu \ &: (\Theta, \tau) \rightarrow (IR, IB) \\ \sigma^2 &: (\Theta, \tau) \rightarrow (IR_+, IB_+) \end{aligned}$$

such that:

$$\begin{aligned} E(X_i \mid \theta) &= \mu(\theta) \\ \mathrm{Var}\,(X_i \mid \theta) &= \sigma^2(\theta)/V_i, \end{aligned}$$

where V_i, $i \geq 1$ are known volume measures. Explicit formulas for the credibility estimator $\hat{m}_{n+1}$ can be found in the original article, equivalent recursions for it based on the classical Kalman filter in DE JONG et al. (1983). Obviously a dynamic linear model like in Section 3 is given with:

$$\begin{aligned} p &= 1, \qquad H_i = 1, \qquad R_i = \sigma_0^2/V_i \\ q &= 1, \qquad F_i = 1, \qquad Q_i = 0, \end{aligned}$$

where:

$$\sigma_0^2 = E(\sigma^2(\theta)).$$

One has:

$$\begin{aligned} b_i &= m_i = \mu(\theta) \\ \hat{b}_i &= \hat{b}_{i-1}^{i-1} = \hat{m}_i, \qquad C_i = C_{i-1}^{i-1} \end{aligned}$$

implying from (3.4), (3.6), (3.7) in case of general Ψ:

$$\hat{m}_{i+1} = \hat{m}_i + \left(\frac{C_i \cdot V_i \cdot k_i}{C_i \cdot V_i \cdot k_i + \sigma_0^2}\right) \cdot (X_i - \hat{m}_i)$$

$$C_{i+1} = C_i - \frac{C_i^2 \cdot V_i \cdot k_i}{C_i \cdot V_i \cdot k_i + \sigma_0^2}$$

$$k_i = \frac{\Psi(V_i^{1/2} \cdot (X_i - m_i)/\sigma_0)}{V_i^{1/2} \cdot (X_i - m_i)/\sigma_0}$$

and from (3.8), (3.10) in case of special $\Psi = \Psi_H$:

$$\hat{m}_{i+1} = \hat{m}_i + C_i \cdot \left(\frac{V_i^{1/2}}{\sigma_0}\right) \cdot \Psi_H\left(\frac{\sigma_0 \cdot V_i^{1/2} \cdot (X_i - \hat{m}_i)}{C_i \cdot V_i + \sigma_0^2}\right)$$

$$C_{i+1} = (V_i/\sigma_0^2 + C_i^{-1})^{-1}.$$

Obviously one has quite handy recursions for the (robust) credibility estimator $\hat{m}_i$.

4.2. Hachemeister's (1975) regression model

Conditionally given θ the $X_1, \ldots, X_{n+1}$ are independent. Suppose that there exist functions:

$$b \ : (\Theta, \tau) \to (IR^q, IB^q)$$
$$\sigma^2 : (\Theta, \tau) \to (IR_+, IB_+)$$

such that:

$$E(X_i \mid \theta) = a_i^T \cdot b(\theta)$$
$$\mathrm{Var}\,(X_i \mid \theta) = \sigma^2(\theta)/V_i,$$

where a_i is a known q-dimensional vector and V_i, $i \geq 1$ are given volume measures. The credibility estimator $\hat{m}_{n+1}$ of m_{n+1} is given by:

$$\hat{m}_{n+1} = a_{n+1}^T \cdot \hat{b}_{n+1},$$

where $\hat{b}_{n+1}$ is the (vector-valued) credibility estimator of $b(\theta)$ based on $X_1, \ldots, X_n$.

Explicit formulas for $\hat{b}_{n+1}$ can be found in the original paper of HACHEMEISTER (1975), equivalent recursions based on the classical Kalman-filter in DE JONG & ZEHNWIRTH (1983).

Obviously the model fits into the framework of Section 3.

Choose simply there:

$$p = 1, \qquad H_i = a_i^T, \qquad R_i = \sigma_0^2/V_i$$
$$F_i = 1, \qquad Q_i = 1,$$

where:

$$\sigma_0^2 = E(\sigma^2(\theta)).$$

One has:

$$b_i = b(\theta)$$
$$b_i = b_{i-1}^{i-1}, \qquad C_i = C_{i-1}^{i-1}.$$

The recursion (3.4), (3.6), (3.7) give in case of general Ψ:

$$\hat{b}_{i+1} = \hat{b}_i + \left(\frac{C_i \cdot a_i \cdot V_i \cdot k_i}{a_i^T \cdot C_i \cdot a_i \cdot V_i \cdot k_i + \sigma_0^2}\right) \cdot (X_i - a_i^T \cdot \hat{b}_i)$$

$$C_{i+1} = C_i - \left(\frac{C_i \cdot a_i \cdot a_i^T \cdot C_i \cdot V_i \cdot k_i}{a_i^T \cdot C_i \cdot a_i \cdot V_i \cdot k_i + \sigma_0^2}\right)$$

$$k_i = \left(\frac{\Psi(V_i^{1/2} \cdot (X_i - a_i^T \cdot \hat{b}_i)/\sigma_0)}{V_i^{1/2} \cdot (X_i - a_i^T \cdot \hat{b}_i)/\sigma_0}\right)$$

and the recursions (3.8)-(3.10) in case of special $\Psi = \Psi_H$:

$$\hat{b}_{i+1} = \hat{b}_i + C_i \cdot a_i \cdot \left(\frac{V_i^{1/2}}{\sigma_0}\right) \cdot \Psi_H\left(\frac{\sigma_0 \cdot V_i^{1/2} \cdot (X_i - a_i^T \cdot \hat{b}_i)}{a_i^T \cdot C_i \cdot a_i \cdot V_i + \sigma_0^2}\right)$$

$$C_{i+1} = C_i - \left(\frac{C_i \cdot a_i \cdot a_i^T \cdot C_i \cdot V_i}{a_i^T \cdot C_i \cdot a_i \cdot V_i + \sigma_0^2}\right).$$

Obviously also these recursions for a (robust) credibility estimator are quite practicable.

4.3. Gerber & Jones' (1975) evolutionary model

Suppose that $X_1, \ldots, X_{n+1}$ are conditionally independent given θ. Furthermore assume that:

$$m_i = m_{i-1} + w_i,$$

where the random disturbances satisfy:

$$E(w_i) = 0, \qquad E(w_i^2) = w$$

$$E(w_i \cdot w_j) = 0, \qquad i \neq j, \qquad E(w_i \cdot m_0) = 0.$$

Finally let:

$$\sigma^2 : (\Theta, \tau) \to (IR_+, IB_+)$$

such that:

$$\operatorname{Var}(X_i \mid \theta) = \sigma^2(\theta)/V_i,$$

where V_i, $i \geq 1$ are given volume measures. Recursions for the credibility estimator are given e.g. in the paper SUNDT (1981).

This model is a special case of the dynamic linear model of Section 3. Choose simply there:

$$p = 1, \qquad H_i = 1, \qquad R_i = \sigma_0^2/V_i$$

$$F_i = 1, \qquad Q_i = w,$$

with:

$$\sigma_0^2 = E(\sigma^2(\theta)).$$

The recursions (3.4), (3.6), (3.7) give in case of general Ψ:

$$\hat{m}_{i+1} = \hat{m}_i + \left(\frac{C_i \cdot V_i \cdot k_i}{C_i \cdot V_i \cdot k_i + \sigma_0^2}\right) \cdot (X_i - \hat{m}_i)$$

$$C_{i+1} = C_i - \frac{C_i^2 \cdot V_i \cdot k_i}{C_i \cdot V_i \cdot k_i + \sigma_0^2} + w$$

$$k_i = \frac{\Psi(V_i^{1/2} \cdot (X_i - \hat{m}_i)/\sigma_0)}{V_i^{1/2} \cdot (X_i - \hat{m}_i)/\sigma_0}$$

and the recursions (3.8), (3.10) in case of special $\Psi = \Psi_H$:

$$\hat{m}_{i+1} = \hat{m}_i + (C_i + w) \cdot \left(\frac{V_i^{1/2}}{\sigma_0}\right) \cdot \Psi_H\left(\frac{\sigma_0 \cdot V_i^{1/2} \cdot (X_i - \hat{m}_i)}{C_i \cdot V_i + w \cdot V_i + \sigma_0^2}\right)$$

$$C_{i+1} = C_i - \frac{C_i^2 \cdot V_i}{C_i \cdot V_i + \sigma_0^2} + w.$$

5. A SIMULATION STUDY

For the model of Section 4.1 data was simulated with the choice $V_i = 1$ for all i. The conditional distribution of X_i given $\theta = \vartheta$ was assumed to be given according

(5.1) $$P(X_i = k \mid \theta = \vartheta) = (1 - \pi) \cdot p_\vartheta(k) + \pi \cdot p_{\vartheta_0}(k)$$

$$\text{(for } k = 0, 1, 2, 3, \ldots)$$

with the Poisson-probabilities:

$$p_\vartheta(k) = \left(\frac{\vartheta^k}{k!}\right) \cdot \exp(-\vartheta)$$

and a probability π. For $\vartheta_0 > \vartheta$ π can be interpreted as the probability of an outlier occurring according to the probabilities $p_{\vartheta_0}(k)$. The risk parameter ϑ was simulated according to the Gamma-model with density on $(0, \infty)$:

(5.2) $$f^\theta(\vartheta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \vartheta^{\alpha - 1} \cdot \exp(-\beta \cdot \vartheta)$$

where α and β are the nonnegative parameters. One gets for $\pi = 0$ as credibility estimator:

(5.3) $$\hat{m}_{n+1} = \frac{\alpha + \sum_{i=1}^{n} X_i}{\beta + n}$$

what shall be compared with a robust variant calculated recursively according to the formulas:

$$\hat{m}_{n+1} = \hat{m}_n + (C_n/\sigma_0) \cdot \Psi_H \cdot \left(\frac{X_n - \hat{m}_n}{(C_n/\sigma_0) + \sigma_0}\right) \tag{5.4}$$

$$C_{n+1} = \frac{\sigma_0^2}{1 + (\sigma_0^2/C_n)}$$

where $\sigma_0^2 = (\alpha/\beta)$ and Ψ_H is the Huber-function with $c = 1.645$. The recursions start with:

$$\hat{m}_1 = (\alpha/\beta), \qquad C_1 = (\alpha/\beta^2).$$

The aim of the study was to compare the results of (5.3), (5.4) with the 'true' value ϑ and to see which one gives the smaller mean squared error:

$$\frac{1}{n - n_0 + 1} \cdot \sum_{i=n_0}^{n} (\hat{m}_{i+1} - \vartheta)^2, \tag{5.5}$$

where n_0 is an adequate number smaller than n. The claims data had to be simulated with a sufficiently large ϑ_0 and an adequate small π. The author chosed $\pi = 0.05$ and for ϑ_0 the values 20, 25, 30, whereas he took $\alpha = 100$, $\beta = 10$, giving for $E(\theta)$ the value 10. With these parameter choices he simulated 100 risk parameters ϑ_j, $j = 1, \ldots, 100$ according to the model (5.2) and for each $\vartheta = \vartheta_j$ independently $n = 9$ values X_i according to (5.1) (with $\pi = 0.05$). In (5.5) he took $n_0 = 6$. He got for the overall mean squared error

$$MSE = \frac{1}{500} \cdot \sum_{j=1}^{100} \sum_{i=6}^{10} (\hat{m}_{i+1}^{(j)} - \vartheta_j)^2$$

the results of the following table:

	$\vartheta_0 = 20$	25	30
with			
(5.3)	0.956	1.231	1.593
(5.4)	0.806	0.807	0.807

showing the strong superiority of (5.4) for situations where bigger outliers can occur with small probability but one wants to rate the normal risk (i.e. case $\pi = 0$). For further illustration the simulation results shall be given for two typical cases. In the first row of the following tables the simulated X_i are given, in the second the $\hat{m}_i$ of (5.3) and in the third the $\hat{m}_i$ of (5.4).

$\boldsymbol{\vartheta_0 = 20}$:

$\vartheta_j = 11.30$

9	13	11	22	13	15	14	14	16	—
10.00	9.91	10.17	10.23	11.07	11.20	11.44	11.59	11.72	11.95
10.00	9.91	10.17	10.23	10.63	10.79	11.05	11.23	11.38	11.62

$\vartheta_j = 8.42$

21	8	12	9	4	8	9	19	8	—
10.00	11.00	10.75	10.85	10.71	10.27	10.13	10.06	10.56	10.42
10.00	10.52	10.31	10.44	10.34	9.91	9.80	9.75	10.05	9.95

$\boldsymbol{\vartheta_0 = 25}$**:**

$\vartheta_j = 9.60$

7	19	11	11	11	33	12	11	11	—
10.00	9.73	10.50	10.54	10.57	10.60	12.00	12.00	11.94	11.89
10.00	9.73	10.20	10.26	10.31	10.36	10.70	10.78	10.80	10.81

$\vartheta_j = 10.71$

12	8	24	12	15	15	10	13	11	—
10.00	10.18	10.00	11.08	11.14	11.40	11.63	11.53	11.61	11.58
10.00	10.18	10.00	10.43	10.55	10.84	11.10	11.03	11.15	11.14

$\boldsymbol{\vartheta_0 = 30}$**:**

$\vartheta_j = 10.77$

11	7	13	6	7	28	7	40	11	—
10.00	10.09	9.83	9.79	9.60	9.60	10.75	10.53	12.17	12.11
10.00	10.09	9.83	10.08	9.79	9.60	9.95	9.77	10.08	10.13

$\vartheta_j = 8.42$

31	8	12	9	4	8	9	29	8	—
10.00	11.91	11.58	11.62	11.43	10.93	10.75	10.65	11.67	11.47
10.00	10.52	10.31	10.44	10.34	9.91	9.80	9.75	10.05	9.95

6. FINAL REMARKS

By applying robustifications of the Kalman-filter to credibility models one can derive fairly practicable recursions for a (robust) credibility estimator. For the Bühlmann-Straub and Hachemeister models one gets an alternative to an already existing approach to robust credibility (see KREMER (1991), KÜNSCH (1992)). For practical application of the above robustified recursions one needs (robust) estimators for the unknown model parameters. Desirable would be such estimators in a recursive form. Obviously here is something left for further research.

REFERENCES

BÜHLMANN, H. (1967) Experience rating and credibility. *ASTIN Bulletin.*

BÜHLMANN, H. and STRAUB, E. (1970) Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.*

BUHLMANN, H., GISLER, A. and JEWELL, W.S. (1982) Excess claims and data trimming in the context if credibility rating procedures. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.*

CIPRA, T. and ROMERA, R. (1991) Robust Kalman filter and its application in time series analysis. *Kybernetika.*

DE JONG, P. and ZEHNWIRTH, B. (1983) Credibility theory and the Kalman filter. *Insurance: Mathematics and Economics.*

GERBER, H.U. and JONES, D.A. (1975) Credibility formulas of the updating type. *Transactions of the society of actuaries.*

GISLER, A. (1980) Optimum trimming of data in the credibility model. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.*

GISLER, A. and REINHARD, P. (1993) Robust Credibility. *ASTIN Bulletin.*

HACHEMEISTER, Ch. (1975) Credibility for regression models with application to trend. In *Credibility: Theory and applications* (ed. P.M. KAHN), New York.

KREMER, E. (1988a) Box-Jenkins Credibility. *Blätter der deutschen Gesellschat für Versicherungsmathematiker.*

KREMER, E. (1988b): General exponential smoothing credibility. *Bulletin of the Belgian Actuarial Association.*

KREMER, E. (1991) Large claims in credibility. *Blätter der deutschen Gesellschaft für Versicherungsmathematik.*

KÜNSCH, H.R. (1992) Robust methods for credibility. *ASTIN Bulletin.*

MASRELIEZ, C.J. and MARTIN, R.D. (1977) Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Automatic Control AC 22.*

MEHRA, R.K. (1975) Credibility theory and Kalman filtering with extensions. *International Institute for applied system analysis.* Laxenburg, Austria.

MEINHOLD, R.J. and SINGPURWALLA, N.D. (1989) Robustification of Kalman filter models. *Journal of the American Statistical Association 84.*

MOWBRAY, A.H. (1914) How extensive a payroll exposure is necessary to give a dependable pure premium. *Proceedings of the Casualty Actuarial society.*

NORBERG, R. (1980) Empirical Bayes credibility. *Scandinavian Actuarial Journal.*

NORBERG, R. (1986) Hierarchical credibility: Analysis of a random effect linear model with nested classification. *Scandinavian Actuarial Journal.*

SUNDT, B. (1980) A multi-level hierarchical credibility model. *Scandinavian Actuarial Journal.*

SUNDT, B. (1981) Recursive credibility estimation. *Scandinavian Actuarial Journal.*

SUNDT, B. (1983) Finite credibility formulae in evolutionary models. *Scandinavian Actuarial Journal.*

TAYLOR G.C. (1979) Credibility analysis of a general hierarchical model. *Scandinavian Actuarial Journal.*

WHITNEY, A.W. (1918) The theory of experience rating. *Proceedings of the Casualty Actuarial society.*

ZEHNWIRTH, B. (1975) A linear filtering approach to recursive credibility estimation. *ASTIN Bulletin.*

ERHARD KREMER
Institut für Mathematische Stochastik, Bundesstraße 55,
20146 Hamburg, FRG.

RECURSIVE METHODS FOR COMPUTING FINITE-TIME RUIN PROBABILITIES FOR PHASE-DISTRIBUTED CLAIM SIZES

By D. A. Stanford and K. J. Stroiński

Department of Statistical and Actuarial Sciences
University of Western Ontario, London, Ontario, Canada
N6A 5B7

Abstract

Finite time ruin methods typically rely on diffusion approximations or discretization. We propose a new method by looking at the surplus process embedded at claim instants and develop a recursive scheme for calculating ruin probabilities. It is assumed that claim sizes follow a phase-type distribution. The proposed method is exact. The application of the method reveals where in the future the relative vulnerability to the company lies.

Keywords

Finite-time ruin probability; Recursive methods; Phase-type distributions; Risk theory

1. Introduction

In this article we consider a classic model describing the evolution over time of the surplus of an insurance company. The value of the surplus process at time t consists of the initial surplus plus premiums received, minus the value of claims that have occurred by time t. There is a vast literature describing the situation where the surplus process becomes negative for the first time.

The classic results concerning ruin probabilities were obtained by Arfwedson (1950), Beekman (1966) and Cramér (1955) and generalized by Thorin (1968). The classic problem was addressed also by Prabhu (1961 and 1965) and Takács (1967). In the case of a compound Poisson claims process and a fixed rate of premium income c the non-ruin probability $\phi(a, t)$ over the finite horizon t given an initial surplus a is a solution of the integro-differential equation

$$c \frac{\partial \phi(a, t)}{\partial a} = \frac{\partial \phi(a, t)}{\partial t} + \phi(a, t) - \int_0^a \phi(a-y, t)\, dB(y)$$

where $B(\cdot)$ is the claim size distribution (see Gerber (1979)).

The typical solution to this equation is stated in terms of Laplace transforms. Numerical difficulties involving the inversion of the Laplace transform solution were pointed out by JANSSEN and DELFOSSE (1982), TAYLOR (1978) and others. For the probability of eventual ruin, exact results were obtained for claim sizes given by combinations of exponential distributions and combinations of gamma distributions; see GERBER et al. (1987). Those results were generalized in DUFRESNE and GERBER (1988) for the family of combinations of shifted exponential distributions. Recursive calculations of ruin probabilities were developed by DEVYLDER and GOOVAERTS (1988) and SHIU (1988).

For more general cases, DICKSON and WATERS [(1991) and (1992)] present a discrete-time approximation for both the probability and the severity of ruin in finite time. A pertinent review of ruin theory results by TAYLOR (1985) contains references to many other texts on the subject.

It was recently shown by ASMUSSEN and ROLSKI (1991) that the exact results for the probability of eventual ruin can be obtained for distributions belonging to be so-called "phase-type" family. (A similar observation was also made by JANSSEN (1982).) In particular, if the claim size distribution is of the phase type, in the compound Poisson model, then the distribution of the maximal aggregate loss is a known phase-type distribution whose parameters are easily calculated. The probability of ruin is obtained as the tail of this distribution. Recently results involving phase-type distributions applied to the ruin problem were obtained by ASMUSSEN and BLADT (1992). However, most finite, continuous time ruin algorithms employ either diffusion approximations (GARRIDO (1988)) or discretize the surplus process, often employing PANJER's (1981) recursion formula.

The model we study can be described as follows: The initial surplus is a, and claims occur according to a Poisson process. Premiums are earned at a constant rate, and claim amounts are assumed to be non-negative, i.i.d. random variables with common distribution function $B(x)$. Claim amounts are assumed to be independent of the claim number process.

Since ruin occurs upon payment of claims, our approach is to observe the process embedded at claim instants only. The methods used lead to exact recursive formulae for the probability of ruin on a specific claim number. This in turn allows us to see where the relative vulnerability of the company lies over the duration of the process. We develop algorithms for the methods presented here assuming that the claim size distribution is of a phase-type.

Our interest in the probability of ruin on a specific claim has been motivated in part by the similarity between the surplus process and the workload process in a single-server queue. Similarities between ruin theory and queueing theory have been extensively explored in WILLMOT (1990).

In the next subsection necessary information about phase-type distributions is presented. In section 2 the general algorithm for phase-type distributions is described, and specific cases are considered. Numerical examples are presented in section 3 followed by conclusions in section 4.

1.1. Phase-type Distributions

Phase-type distributions have become an extremely popular tool for applied probabilists wishing to generalize beyond the exponential while retaining some of its key properties. The phase-type family includes the exponential, mixture of exponentials, Erlangian and Coxian distributions as special cases. Among the appealing characteristics of phase-type distributions are the following closure properties: a) finite n-fold convolutions of phase-types are again of phase-type; b) a random modification (i.e. forward recurrence time) of a phase-type is again of phase-type; c) geometric mixtures of n-fold convolutions of phase-types are again of phase-type. This latest property is of particular interest vis-à-vis the ruin problem: couched in terms of its $M/G/1$ waiting time analogue, NEUTS (1981) has shown that the distribution of the maximum aggregate loss is phase-type with easily determined parameters whenever the claim size distribution is phase-type as well. This fact was later used by JANSSEN (1982) and ASMUSSEN and ROLSKI (1991), who developed formulas for the ultimate probability of ruin for phase-type claim size distributions.

Phase-type distributions were first introduced by NEUTS in 1975, but the most popular standard reference for them has become NEUTS (1981). A shortened treatment can be stated as follows. Consider a Markov process with transient states $\{1, 2, ..., m\}$ and absorbing state $(m+1)$, whose infinitesimal generator Q has the form

$$Q = \begin{bmatrix} T & t_0 \\ 0 & 0 \end{bmatrix}.$$

The diagonal entries T_{ii} are necessarily negative, other entries are non-negative, and $\boldsymbol{t}_0 = -T\boldsymbol{e}$ represents the rates at which transitions occur from the individual transient states to the absorbing state.

Let the process start in state i with probability α_i, $i = 1, \ldots, m+1$, and let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$. (In many practical problems, $\alpha_{m+1} = 0$.) Now let $B(x)$ denote the distribution of the time to absorption, X, into state $m+1$. The distribution $B(\cdot)$ thus described is said to be of "phase-type with representation $(\boldsymbol{\alpha}, T)$", and

$$B(x) = 1 - \boldsymbol{\alpha} \exp(Tx)\boldsymbol{e};\ x \geq 0.$$

Assuming $\alpha_{m+1} = 0$ its density is $b(x) = \boldsymbol{\alpha} \exp(Tx)\boldsymbol{t}_0$, $x \geq 0$, its Laplace-Stieltjes transform is $\Phi_B(s) = E\{e^{-sX}\} = \boldsymbol{\alpha}(sI - T)^{-1}\boldsymbol{t}_0$, and its nth noncentral moment is given by

$$E\{X^n\} = n!\boldsymbol{\alpha}(-T)^{-n})\boldsymbol{e}$$

Many of these properties are quoted in the development which follows. The interested reader is directed to NEUTS (1981) for a rigorous treatment of phase-type distributions.

2. RECURSIONS FOR PHASE-DISTRIBUTED CLAIM SIZES

The method we are about to describe works with the incomplete density for the reserve remaining after the nth claim occurs. (It is of course incomplete because ruin already have occurred.) Define

$$p_n(y) = \frac{d}{dy} Pr\ \{\text{non-ruin up to } n\text{th claim, and remaining reserve} \leq y\}$$

and define the Laplace transforms

$$\mathcal{L}_n(s) = \int_0^\infty e^{-sy} p_n(y)\,dy \tag{2.1}$$

(Note in particular that $\mathcal{L}_n(0) = Pr$ {non-ruin up to nth claim}.)

Next, define the "increment" between two consecutive claims as the difference between the revenue earned and the claim amount. Let $g(y)$ be its density (defined on $(-\infty, \infty)$ because the increment can assume both positive and negative values), and let

$$G(s) = \int_{-\infty}^{\infty} e^{-sy} g(y)\,dy. \tag{2.2}$$

Let the claim size have distribution function $B(y)$, density $b(y)$, and Laplace-Stieltjes transform

$$\Phi_B(s) = \int_0^\infty e^{-sy}\,dB(y). \tag{2.3}$$

Since the number of claims is given by a Poisson process, inter-claim times (and hence revenue amounts earned between claims) are exponentially distributed. Furthermore, premiums are collected at a constant rate, so it follows that the revenue collected between consecutive claims is also exponentially distributed, and we denote the mean revenue by $(1/\lambda)$. Since the increment is the difference between the revenue and the claim size, these definitions lead to the following results after straightforward manipulations:

$$g(y) = \begin{cases} \Phi_B(\lambda)\,\lambda e^{-\lambda y}; & y \geq 0 \\ \displaystyle\int_0^\infty \lambda e^{-\lambda t} b(t-y)\,dt; & y < 0 \end{cases} \tag{2.4}$$

$$G(s) = (\lambda/\lambda + s)\,\Phi_B(-s) \tag{2.5}$$

Theorem: Let the claim size be given by a phase-type distribution with representation $(\boldsymbol{\alpha}, T)$, and let $\boldsymbol{t}_0 = -T\boldsymbol{e}$. Then

$$\mathcal{L}_n(s) = \mathcal{L}_{n-1}(s)\,G(s) + \boldsymbol{\nu} \int_{x=0}^{\infty} p_{n-1}(x) \exp(Tx)\,dx\,(sI + T)^{-1} \boldsymbol{t}_0 \tag{2.6}$$

where $\boldsymbol{\nu} = \lambda \boldsymbol{\alpha} (\lambda I - T)^{-1}$.

Proof: Having described the increment as above, the reserve after the nth claim is the sum of the reserve after the $(n-1)$th claim and the ensuing increment. Therefore

$$p_n(y) = \int_0^\infty p_{n-1}(x)\,g(y-x)\,dx. \tag{2.7}$$

The equivalent expression involving Laplace transforms is

$$\mathcal{L}_n(s) = \int_{y=0}^{\infty} e^{-sy} p_n(y)\,dy = \int_{y=0}^{\infty} \int_{x=0}^{\infty} e^{-sy} p_{n-1}(x) g(y-x)\,dxdy.$$

Due to absolute integrability, we can reverse the order of integration to obtain

$$\mathcal{L}_n(s) = \int_{x=0}^{\infty} e^{-sx} p_{n-1}(x) \int_{y=0}^{\infty} e^{-s(y-x)} g(y-x)\,dydx$$

$$= \mathcal{L}_{n-1}(s) G(s) - \int_{x=0}^{\infty} e^{-sx} p_{n-1}(x) \int_{y=x}^{\infty} e^{sy} g(-y)\,dydx \qquad (2.8)$$

Formula (2.8) applies for *all* non-negative generally distributed claim sizes. To complete the proof of the theorem, we must rely on properties of phase-type distributions. Recall that $b(y) = \alpha \ \exp(Ty)t_0$. Therefore for $x<0$, we find

$$g(x) = \int_{t=0}^{\infty} \lambda e^{-\lambda t} b(t-x)\,dt$$

$$= \boldsymbol{\alpha} \int_{t=0}^{\infty} \lambda e^{-\lambda t} \exp(T(t-x))\,dt\boldsymbol{t}_0$$

$$= \lambda\boldsymbol{\alpha}(\lambda I - T)^{-1} \exp(-Tx)\boldsymbol{t}_0 = \boldsymbol{\nu} \exp(-Tx)\boldsymbol{t}_0$$

Consequently we can evaluate the inner integral of the 2nd term of (2.8) as follows:

$$\int_{y=x}^{\infty} e^{sy} g(-y)\,dy = \boldsymbol{\nu} \int_{y=x}^{\infty} e^{sy} \exp(Ty)\,dy\boldsymbol{t}_0$$

$$= \boldsymbol{\nu} \int_{y=x}^{\infty} \exp((sI+T)y)\,dy\boldsymbol{t}_0$$

$$= -e^{sx}\boldsymbol{\nu} \exp(Tx)(sI+T)^{-1}\boldsymbol{t}_0 \qquad (2.9)$$

When (2.9) is substituted into (2.8), equation (2.6) is obtained.

Remark: By evaluating (2.6) at $s=0$, one finds

$Pr\{$non-ruin up to the nth claim$\} =$
$Pr\{$non-ruin up to the $(n-1)$th claim$\}$

$$+ \ \boldsymbol{\nu} \int_{x=0}^{\infty} p_{n-1}(x) \exp(Tx)\,dxT^{-1}\boldsymbol{t}_0$$

from which one concludes that

$$P(n) \equiv Pr\,\{\text{ruin on the } n\text{th claim}\} = \nu \int_{x=0}^{\infty} p_{n-1}(x)\exp(Tx)\,dxe. \tag{2.10}$$

In what follows, we develop computational algorithms from (2.6) and (2.10) for the specific cases of i) exponential, ii) mixtures of exponential claim size distributions, and iii) Erlang-N distributions.

2.1 Recursive Algorithm for Exponential Claim Sizes

If one assumes that claim sizes are exponentially distributed with mean $(1/\mu)$, we can simplify the recursions for $\mathcal{L}_n(s)$ and $P(n)$ greatly. All of the matrix quantities reduce to scalar results. In particular:

$$G(s) = (\lambda/\lambda+s)(\mu/\mu-s),\ T=[-\mu],\ \boldsymbol{t}_0=\mu,$$

$$\nu=\lambda/(\lambda+\mu),$$

$$(sI+T)^{-1} = 1/(s-\mu), \text{ and } \exp(Tx) = e^{-\mu x}.$$

Therefore (2.6) becomes

$$\mathcal{L}_n(s) = (\mu/\mu-s)\,[(\lambda/\lambda+s)\,\mathcal{L}_{n-1}(s) - (\lambda/\lambda+\mu)\,\mathcal{L}_{n-1}(\mu)] \tag{2.11}$$

and

$$P(n) = (\lambda/\lambda+\mu)\,\mathcal{L}_{n-1}(\mu). \tag{2.12}$$

For the remainder of this section, let $\Phi \equiv \Phi_B(\lambda) = (\mu/\mu+\lambda)$. We seek an algorithm to determine the form of $\mathcal{L}_n(s)$ in terms of $\mathcal{L}_{n-1}(s)$, and the corresponding expression for $P(n)$. If the initial reserve is a, then $\mathcal{L}_0(s)=e^{-as}$. Thus from (2.12)

$$P(1) = (1-\Phi)\,\mathcal{L}_0(\mu) = (1-\Phi)\,e^{-a\mu}. \tag{2.13}$$

Using (2.11) we can find $\mathcal{L}_1(s)$ as follows

$$\begin{aligned}
\mathcal{L}_1(s) &= \lambda\mu/(\mu-s)\,[e^{-as}/(\lambda+s) - e^{-a\mu}/(\lambda+\mu)] \\
&= \lambda\mu e^{-a\mu}\,[(\lambda+\mu)\,e^{a(\mu-s)} - (\lambda+s)]/\{(\mu-s)(\lambda+s)(\lambda+\mu)\} \\
&= \lambda\mu e^{-a\mu}\,[(\mu-s) + (\lambda+\mu)\sum_{k=1}^{\infty}(a(\mu-s))^k/k!]/[(\mu-s)(\lambda+s)(\lambda+\mu)] \\
&= \lambda/(\lambda+s)\,e^{-a\mu}\left[\Phi + (a\mu)\sum_{k=1}^{\infty}(a(\mu-s))^{k-1}/k!\right].
\end{aligned} \tag{2.14}$$

Again using (2.12) for $n=2$ we find

$$P(2) = (1-\Phi)\,\mathcal{L}_1(\mu) = (1-\Phi)^2 e^{-a\mu}\,[\Phi + a\mu] \tag{2.15}$$

and (2.11) gives rise to the following expression:

$$\mathcal{L}_2(s) = e^{-a\mu}\left[(\Phi + a\mu)\{\Phi(\lambda/\lambda+s)^2 + \Phi(1-\Phi)(\lambda/\lambda+s)\}\right.$$
$$\left. + (a\mu)^2(\lambda/\lambda+s)^2 \sum_{k=2}^{\infty} (a(\mu-s))^{k-2}/k!\right] \tag{2.16}$$

We now establish the general form of $\mathcal{L}_n(s)$.

Theorem: For $n \geq 1$, the Laplace transform of the distribution of the reserve following the nth claim is given by

$$\mathcal{L}_n(s) = e^{-a\mu}\left[\sum_{j=1}^{n} c_j^{(n)}(\lambda/\lambda+s)^j(1-\Phi)^{n-j} + (a\mu)^n(\lambda/\lambda+s)^n \sum_{k=n}^{\infty} \frac{(a(\mu-s))^{k-n}}{k!}\right] \tag{2.17}$$

Furthermore, the coefficients $c_j^{(n)}$ at the nth stage are related to those at the $(n-1)$th stage via

$$c_j^{(n)} = \Phi\left\{\sum_{k=\max(1,j-1)}^{n-1} c_k^{(n-1)} + \frac{(a\mu)^{n-1}}{(n-1)!}\right\}; \; j = 1, \ldots, n \tag{2.18}$$

Proof: The proof proceeds via induction. For $n = 1$, (2.14) show that the form of (2.17) is correct, with $c_1^{(1)} = \Phi$. For $n = 2$, (2.16) reveals that $c_1^{(2)} = c_2^{(2)} = \Phi^2 + (a\mu)\Phi$.

Now assume that (2.17) is valid up to index $n = N - 1$. Thus

$$\mathcal{L}_{N-1}(s) = e^{-a\mu}\left[\sum_{j=1}^{N-1} c_j^{(N-1)}(\lambda/\lambda+s)^j(1-\Phi)^{N-1-j}\right.$$
$$\left. + (a\mu)^{N-1}(\lambda/\lambda+s)^{N-1} \sum_{k=N-1}^{\infty} \frac{(a(\mu-s))^{k-N+1}}{k!}\right] \tag{2.19}$$

Substituting (2.19) into (2.11) for $n = N$ we find

$$\mathcal{L}_N(s) = \mu/(\mu-s)\, e^{-a\mu}\left[\sum_{j=1}^{N-1} c_j^{(N-1)}(1-\Phi)^{N-1-j}((\lambda/\lambda+s)^{j+1} - (\lambda/\lambda+\mu)^{j+1})\right.$$
$$\left. + (a\mu)^{N-1}\left\{\frac{(\lambda/\lambda+s)^N}{(N-1)!} - \frac{(\lambda/\lambda+\mu)^N}{(N-1)!}\right\} + (\lambda/\lambda+s)^N \sum_{k=N}^{\infty} \frac{(a(\mu-s))^{k-N+1}}{k!}\right]$$
$$= e^{-a\mu}\left[\sum_{j=1}^{N-1} c_j^{(N-1)}(1-\Phi)^{N-1-j}\left(\frac{\mu}{\lambda+\mu}\right)\left(\frac{\lambda}{\lambda+s}\right)\sum_{l=0}^{j}(\lambda/\lambda+s)^l(\lambda/\lambda+\mu)^{j-l}\right.$$
$$+ \frac{(a\mu)^{N-1}}{(N-1)!}\left(\frac{\mu}{\lambda+\mu}\right)\left(\frac{\lambda}{\lambda+s}\right)\sum_{l=0}^{N-1}(\lambda/\lambda+s)^l(\lambda/\lambda+\mu)^{N-1-l}$$
$$\left. + (a\mu)^N(\lambda/\lambda+s)^N \sum_{k=N}^{\infty} \frac{(a(\mu-s))^{k-N}}{k!}\right] \tag{2.20}$$

The infinite sum in (2.20) already satisfies the form required by (2.17) for $n = N$. It remains for us to rearrange the other terms in (2.20) and establish the recursion posed by (2.18). Note that these can be written as

$$e^{-a\mu}\left[\sum_{l=0}^{N-1} \Phi \sum_{j=\max(1,l)}^{N-1} c_j^{(N-1)}(1-\Phi)^{N-1-j}(\lambda/\lambda+s)^{l+1}(\lambda/\lambda+\mu)^{j-l}\right.$$

$$\left.+\sum_{l=0}^{N-1} \Phi \frac{(a\mu)^{N-1}}{(N-1)!}(\lambda/\lambda+s)^{l+1}(\lambda/\lambda+\mu)^{N-1-l}\right]$$

$$= e^{-a\mu}\left[\sum_{l=1}^{N}(\lambda/\lambda+s)^l(1-\Phi)^{N-l}\,\Phi\left\{\sum_{k=\max(1,l-1)}^{N-1} c_k^{(N-1)}+\frac{(a\mu)^{N-1}}{(N-1)!}\right\}\right] \tag{2.21}$$

A comparison of equivalent powers of $(\lambda/\lambda+s)$ between (2.21) and the first sum of (2.17) establishes (2.18) and completes the proof by induction.

When (2.17) is substituted into (2.12) one obtains

$$P(n) = (1-\Phi)^n e^{-a\mu}\left\{\sum_{j=1}^{n-1} c_j^{(n-1)}+\frac{(a\mu)^{n-1}}{(n-1)!}\right\}.$$

However, in light of (2.18) this can be restated as follows.

Corollary:

$$P(n) = (1-\Phi)^n e^{-a\mu} c_1^{(n)}/\Phi, \qquad n = 1, 2, \ldots \tag{2.22}$$

The algorithm for determining $P(n)$ thus consists of (2.18) and (2.22), starting from $c_1^{(1)} = \Phi$.

2.2 Recursive Algorithm for Mixtures of Exponentials

Mixtures of a finite number of exponential distributions have been used frequently as a generalization of the single-exponential case, see for instance GERBER (1979).

The phase-type formulation of a mixture of K exponentials has the following form: $\boldsymbol{\alpha} = [p_1, p_2, \ldots, p_K]$; $T = -\operatorname{diag}[\mu_1, \mu_2, \ldots, \mu_K]$, and $\boldsymbol{t}_0 = [\mu_1, \mu_2, \ldots, \mu_K]'$. Similarly $\Phi_B(s)$

$$= [p_1, p_2, \ldots, p_K]\operatorname{diag}[(s+\mu_1)^{-1}, (s+\mu_2)^{-1}, \ldots, (s+\mu_K)^{-1}][\mu_1, \mu_2, \ldots, \mu_K]'$$

$$= \sum_{i=1}^{K} p_i\mu_i/(s+\mu_i).$$

The other computations for (2.6) are equally straightforward, and the resulting recursion for $\mathcal{L}_n(s)$ is

$$\mathcal{L}_n(s) = \sum_{i=1}^{K} p_i(\mu_i/\mu_i - s)[\mathcal{L}_{n-1}(s)(\lambda/\lambda+s) - \mathcal{L}_{n-1}(\mu_i)(\lambda/\lambda+\mu_i)] \tag{2.23}$$

and the probability of ruin on the nth claim is

$$P(n) = \sum_{i=1}^{K} p_i(\lambda/\lambda + \mu_i)\mathcal{L}_{n-1}(\mu_i) \tag{2.24}$$

Again, we seek explicit expressions for $\mathcal{L}_n(s)$ and $P(n)$ which can be calculated recursively. Defining $\Phi_i = (\mu_i/\lambda + \mu_i)$, we find (since $\mathcal{L}_0(s) = e^{-as}$):

$$P(1) = \sum_{i=1}^{K} p_i(1 - \Phi_i)e^{-a\mu_i}, \tag{2.25}$$

$$\mathcal{L}_0(s) = (\lambda/\lambda + s) \sum_{i=1}^{K} p_i e^{-a\mu_i}\left[\Phi_i + (a\mu_i) \sum_{k=1}^{\infty} \frac{(a(\mu_i - s))^{k-1}}{k!}\right] \tag{2.26}$$

from which $P(2)$ can be readily calculated using (2.24). Determination of $\mathcal{L}_n(s)$ for $n \geq 2$ in the general case becomes increasingly complicated due to the need to evaluate $\mathcal{L}_{n-1}(\mu_i)$ at all K rates. Tractable results are, however, available for two important sub-cases. For both (2.24) can then be used to find the corresponding probabilities.

Case A: No Initial Reserve ($a = 0$): In this case

$$\mathcal{L}_1(s) = (\lambda/\lambda + s) \sum_{i=1}^{K} p_i \Phi_i$$

and similar methods to those of the previous section can be used to establish the following recursion:

$$\mathcal{L}_n(s) = \sum_{j=1}^{n} c_j^{(n)}(\lambda/\lambda + s)^j \qquad n = 1, 2, \ldots \tag{2.27}$$

where

$$c_j^{(n)} = \sum_{k=\max(1, j-1)}^{(n-1)} c_k^{(n-1)} \left\{\sum_{i=1}^{K} p_i \Phi_i (1 - \Phi_i)^{k+1-j}\right\} \qquad j = 1, 2, \ldots, n \tag{2.28}$$

and where

$$c_1^{(1)} = \sum_{i=1}^{K} p_i \Phi_i.$$

Case B: Mixture of 2 Exponentials

Theorem: For the case of a mixture of 2 exponentials the following relationship applies for $\mathcal{L}_n(s)$, $n \geq 1$:

$$\mathcal{L}_n(s) = \sum_{j=1}^{n} c_j^{(n)}(\lambda/\lambda + s)^j + (\lambda/\lambda + s)^n A^{(n)}(s) \tag{2.29}$$

where

$$A^{(n)}(s) = \sum_{k=n}^{\infty} \frac{a^{k-n}}{k!} \sum_{m=1}^{2} p_m \left\{ D_m^{(n)} (\mu_m - s)^{k-n} \right.$$
$$\left. + \sum_{j=0}^{n-2} f_{jm}^{(n)} \sum_{l=0}^{k-n} \binom{k-n-l+j}{j} (\mu_m - s)^l (\mu_m - \mu_{3-m})^{k-n-l} \right\},$$

where the coefficients $c_j^{(n)}$, $D_m^{(n)}$ and $f_{jm}^{(n)}$, $n \geq 2$, satisfy the recursions

$$c_j^{(n)} = \sum_{i=1}^{2} p_i \Phi_i \left[\sum_{k=\max(1,j-1)}^{n-1} c_k^{(n-1)} (\lambda/\lambda + \mu_i)^{k+1-j} + (\lambda/\lambda + \mu_i)^{n-j} A_i^{(n-1)} \right], \tag{2.30}$$

$$D_m^{(n)} = D_m^{(n-1)} r_m, \tag{2.31}$$

$$f_{jm}^{(n)} = \begin{cases} r_m f_{0m}^{(n-1)} + r_{3-m} D_m^{(n-1)} & j = 0 \\ r_m f_{jm}^{(n-1)} + r_{3-m} f_{j-1,m}^{(n-1)} & j = 1, 2, \ldots, n-3 \\ r_{3-m} f_{n-3,m}^{(n-1)} & j = n-2 \end{cases} \tag{2.32}$$

where $r_m = p_m a \mu_m$ and

$$A_i^{(n-1)} \equiv A^{(n-1)}(\mu_i). \tag{2.33}$$

The coefficients are initialized as follows:

$$c_1^{(1)} = \sum_{i=1}^{2} p_i e^{-a\mu_i} \Phi_i; \qquad D_i^{(1)} = e^{-a\mu_i}(a\mu_i), \qquad i = 1, 2$$

and

$$f_{0m}^{(2)} = r_{3-m} D_m^{(1)}; \qquad m = 1, 2$$

Proof: Follows after tedious but straightforward substitution using (2.23) and (2.26).

2.3 Recursive Algorithm for Erlang-*N* Claims

The Erlang-*N* distribution is the *N*-fold convolution of the exponential distribution, and thus forms a subset of the family of Gamma distributions. It is often used to model distributions which are less variable relative to the mean than the exponential. Let the mean per exponential stage be $1/\theta$. The Erlang-*N* can be viewed within the phase-type framework as a process successively moving through states 1 through m prior to absorption into state $m+1$. Thus $\boldsymbol{\alpha} = [1, 0, \ldots, 0]$;

$$T = \begin{bmatrix} -\theta & \theta & & 0 \\ & -\theta & \theta & \\ 0 & & \ddots & -\theta \end{bmatrix} \quad ; \text{ and } \boldsymbol{t}_0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \theta \end{bmatrix}.$$

Theorem: The recursion for $\mathcal{L}_n(s)$ when the claim size distribution is Erlang-N as described above is as follows:

$$\mathcal{L}_n(s) = \Bigg\{[\mathcal{L}_{n-1}(s)(\lambda/\lambda+s),\ 0,\ \ldots,\ 0]$$

$$-\left[p\mathcal{L}_{n-1}(\theta),\ p(-\theta)\mathcal{L}'_{n-1}(\theta)+pq\mathcal{L}_{n-1}(\theta),\ \ldots,\ \sum_{l=0}^{N-1} pq^l \frac{(-\theta)^{N-1-l}}{(N-1-l)!}\cdot\mathcal{L}_{n-1}^{(N-1-l)}(\theta)\right]\Bigg\}$$

$$\times\begin{bmatrix} (\theta/\theta-s)^N \\ (\theta/\theta-s)^{N-1} \\ \vdots \\ (\theta/\theta-s) \end{bmatrix} \tag{2.34}$$

where $p=(\lambda/\lambda+\theta)$ and $q=1-p$.

Proof: One can show after elementary calculations that

$$(-sI-T)^{-1}t_0 = [(\theta/\theta-s)^N,\ (\theta/\theta-s)^{N-1},\ \ldots,\ (\theta/\theta-s)]'$$

$$\nu = \lambda\alpha(\lambda I-T)^{-1} = [\lambda/\lambda+\theta),\ \lambda\theta/(\lambda+\theta)^2,\ \ldots,\ \lambda\theta^{N-1}/(\lambda+\theta)^N];$$

and

$$\exp(Tx) = \begin{bmatrix} f_0(\theta x) & f_1(\theta x) & f_2(\theta x) & \ldots & f_{N-1}(\theta x) \\ 0 & f_0(\theta x) & f_1(\theta x) & \ldots & f_{N-2}(\theta x) \\ 0 & 0 & f_0(\theta x) & \ldots & f_{N-3}(\theta x) \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & & f_0(\theta x) \end{bmatrix}$$

where $f_k(\theta x) = (-1)^k \sum_{l=k}^{\infty} \frac{(-\theta x)^l}{k!\,(l-k)!} = \frac{(\theta x)^k}{k!} e^{-\theta x}$. Therefore

$$\int_{x=0}^{\infty} p_{n-1}(x)\exp(Tx)\,dx$$

$$= \begin{bmatrix} \mathcal{L}_{n-1}(\theta) & (-\theta)\mathcal{L}'_{n-1}(\theta) & \theta^2\mathcal{L}''_{n-1}(\theta)/2 & \ldots & (-\theta)^{N-1}\mathcal{L}_{n-1}^{(N-1)}(\theta)/(N-1)! \\ 0 & \mathcal{L}_{n-1}(\theta) & (-\theta)\mathcal{L}'_{n-1}(\theta) & & (-\theta)^{N-2}\mathcal{L}_{n-1}^{(N-2)}(\theta)/(N-2)! \\ 0 & 0 & \mathcal{L}_{n-1}(\theta) & . & (-\theta)^{N-3}\mathcal{L}_{n-1}^{(N-3)}(\theta)/(N-3)! \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & \mathcal{L}_{n-1}(\theta) \end{bmatrix}.$$

Substitution of these expressions into (2.6) in light of (2.5) gives rise to the theorem's result. Substitution of $s = 0$ in (2.34) leads to the following result after elementary manipulations.

Corollary: In the case of Erlang-N claim-size distributions, the probability of ruin on the nth claim is given by

$$P(n) = \sum_{i=0}^{N-1} \mathcal{L}_{n-1}^{(i)}(\theta)\{(-\theta)^i[1 - q^{N-i}]/i!\}. \tag{2.35}$$

In the case of Erlang-2 claims, this reduces to

$$P(n) = \mathcal{L}_{n-1}(\theta)(1 - q^2) - p\theta\mathcal{L}'_{n-1}(\theta). \tag{2.36}$$

Using the same methods as before, one can again develop recursions for $\mathcal{L}_n(s)$. We state here as an example the recursion for the case of Erlang-2 claim sizes:

$$\mathcal{L}_n(s) = e^{-a\theta}\left[\sum_{j=1}^{n} c_j^{(n)}(\lambda/\lambda + s)^j + (a\theta)^{2n}(\lambda/\lambda + s)^n \sum_{k=2n}^{\infty} \frac{(a(\theta - s))^{k-2n}}{k!}\right] \tag{2.37}$$

where the coefficients $c_j^{(n)}$ satisfy the recursion:

$$c_j^{(n)} = \left[\frac{(a\theta)^{2n-2}q^2}{(2n-2)!}(n+1-j) + \frac{(a\theta)^{2n-1}}{(2n-1)!}q\right]p^{n-j}$$

$$+ q^2 \sum_{k=\max(2,j)}^{n} c_{k-1}^{(n-1)}p^{k-j}(k+1-j). \quad j = 1, \ldots, n;\ n = 1, 2, \ldots \tag{2.38}$$

The recursion starts with $c_1^{(1)} = q^2 + a\theta q$.

3. Numerical Examples

A series of numerical examples have been carried out to demonstrate the effect of the security loading, the initial reserve, and the claim size distribution on the probability of ruin. The results are displayed in Figures 1 through 10.

Figures 1 and 2 display the cumulative probability of ruin for exponentially distributed claims. In Figure 1, it is assumed that there is no initial reserve. Although the probabilities of this and all later graphs are in fact valid only for integer claim numbers, continuous trajectories have been fitted to these points to show the overall trend. Figure 1 shows that these trajectories approach their ultimate limit of $1/(1 + \theta)$ (where θ is the relative security loading). This limit is a well-known result in risk theory; see for example Bowers et al. (1986) p. 359, eq. (12.5.2). The speed of convergence to this limit increases with θ. This is not surprising—if the security loading is large, one would expect either to be ruined very soon, or else to have built up enough surplus to weather further fluctuations. For small θ, a smaller reserve accumulates, so the period of vulnerability lasts longer.

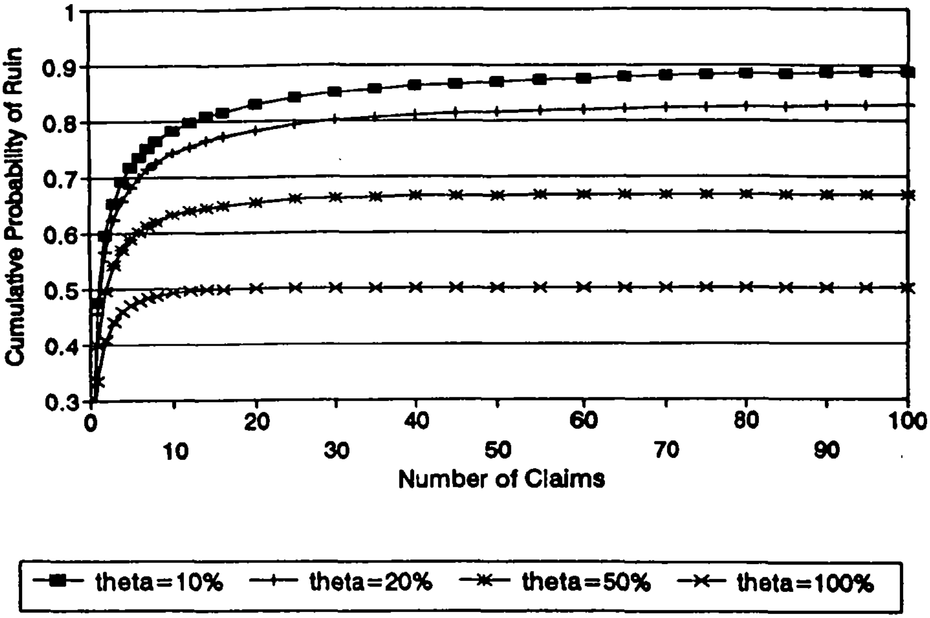


FIGURE 1. Effect of security loading (no initial reserve).

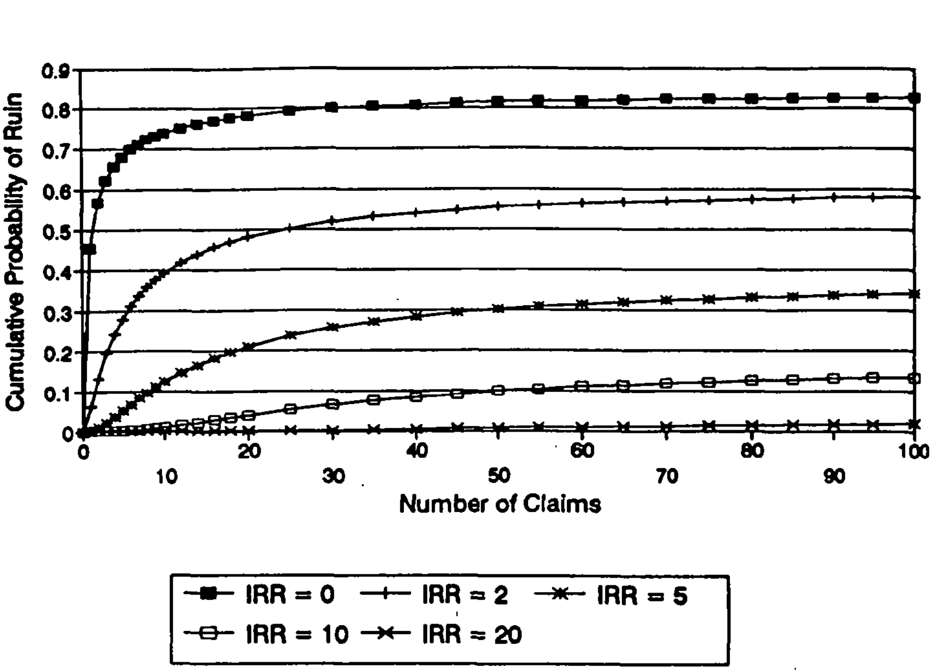


FIGURE 2. Effect of initial reserve (security loading = 20%).

Figure 2 displays the cumulative probability of ruin as a function of the time and the initial reserve level. Here the acronym IRR refers to the initial reserve ratio, defined are the initial reserve divided by the expected amount of a single claim. The security loading is 20%. The figure shows that even small reserves have a major benefit as opposed to having no reserve at all. For larger reserves, the commencement of the period of vulnerability is delayed due to the extremely small chance of a large number of claims during a single time unit.

The probability of ruin on a given claim number reveals the relative vulnerability of the company. This vulnerability is explored in Figures 3 though 6, where exponential claim sizes have been assumed. Figures 3 and 4 show that with a non-zero initial surplus, the ruin probabilities for the process with a smaller relative security loading θ are spread over a much wider range than those for larger θ.

Figures 5 and 6 show that for larger initial surpluses, the ruin probabilities are again spread over a wider range for smaller θ's. Further, the larger initial surplus reduces the possibility of early ruin, however, the relative vulnerability is more spread out.

To demonstrate the effect of more variable claim size distributions, Figures 7 and 8 display the probabilities of ruin on the nth claim for a series of mixtures of two exponentials with balanced means. This balanced means assumption ($p_1/\mu_1 = p_2/\mu_2$) is common when fitting a mixture of 2 exponentials to only the first two moments $E\{B\}$ and $\mathrm{Var}\{B\}$. Define $c^2 = \mathrm{Var}\{B\}/E\{B\}^2$; that is, c^2 is the squared coefficient of variation (SCV) of the claim size distribution. Then the parameters p_i, μ_i, $i = 1, 2$, are found from the following equations:

$$R = \sqrt{(c^2-1)/(c^2+1)}, \tag{3.1}$$

$$p_1, p_2 = (1 \pm R)/2, \tag{3.2}$$

$$\mu_i = 2p_i/E\{B\}. \tag{3.3}$$

Figures 7 and 8 consider the case where there is no initial reserve and $\theta = 100\%$. The SCV takes on values of 1.0 (corresponding to the ordinary exponential), 4.0 and 9.0. The figures show that around the 40th claim, the probability of being ruined on a given claim is almost 100 times more likely when SCV = 9.0 than with the ordinary exponential. In the short-run, however, there is more chance of being ruined by a less variable claim size distribution. This seemingly contradictory result can be understood according to the following reasoning: when SCV = 9, we get $R = 0.905$, $p_1 = .952$ and $p_2 = .048$. Thus, for every large claim in the long-run, there are roughly 20 smaller ones, but the rare large-sized claim is roughly 20 times larger on average. Therefore, in the very short-run, there is little chance of ruin, but eventually, the larger claims start to occur, and their impact is so much greater.

Figure 9 compares the ruin probabilities for exponential and Erlang-2 claim sizes for two values of IRR, and assuming a relative security loading of 20%. As the figure demonstrates, there is substantially less likelihood of ruin for Erlang-2 claims. This is due to the reduced variability of the claim sizes. (The Erlang-2 case has a c^2 equal to half that of the exponential.)

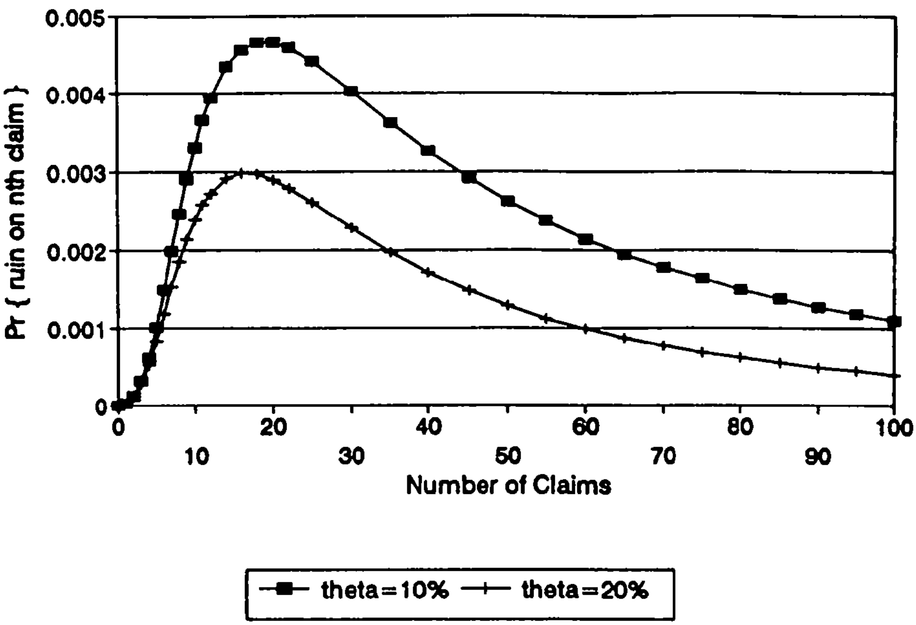


FIGURE 3. Effect of security loading (initial reserve ratio = 10).

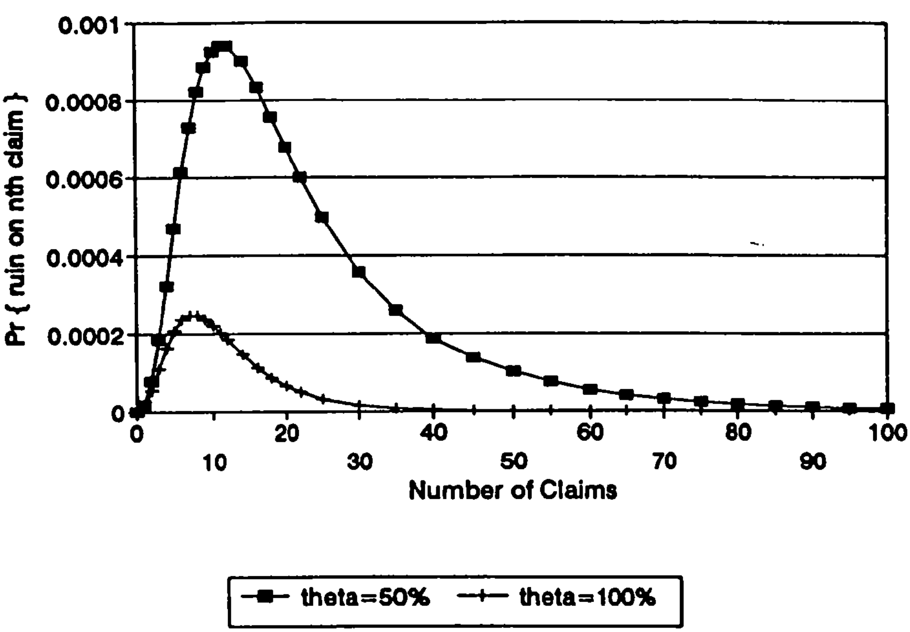


FIGURE 4. Effect of security loading (initial reserve ratio = 10).

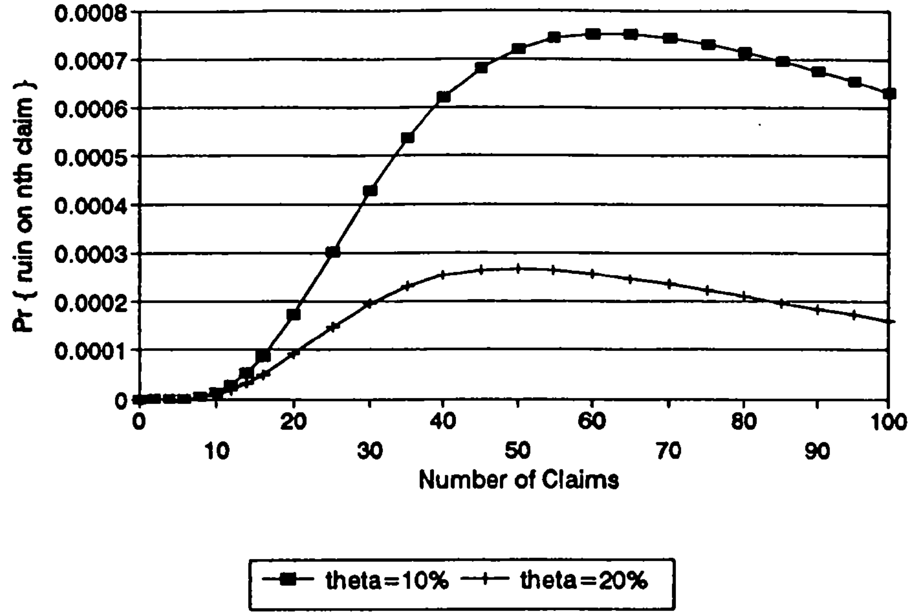


FIGURE 5. Effect of security loading (initial reserve ratio = 20).

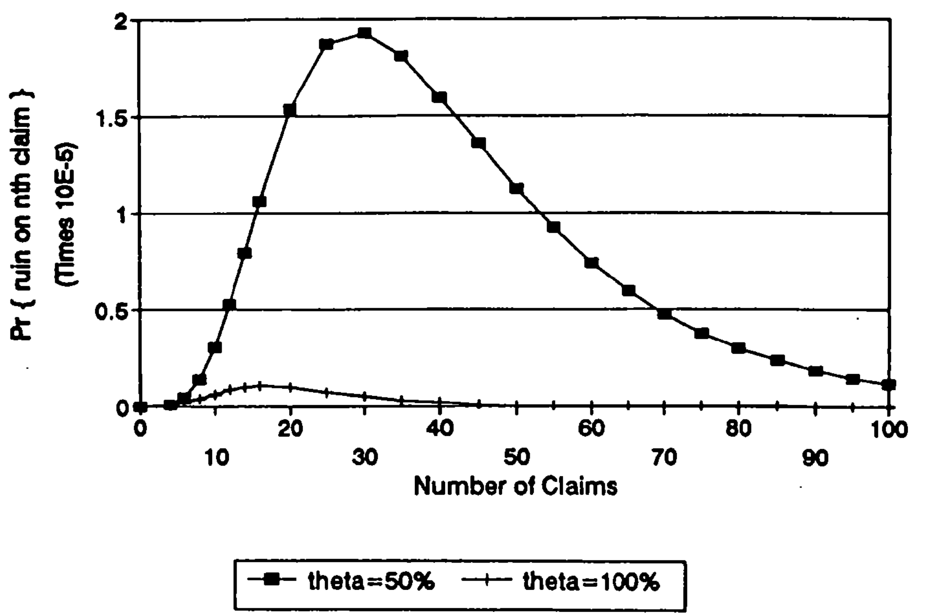


FIGURE 6. Effect of security loading (initial reserve ratio = 20).

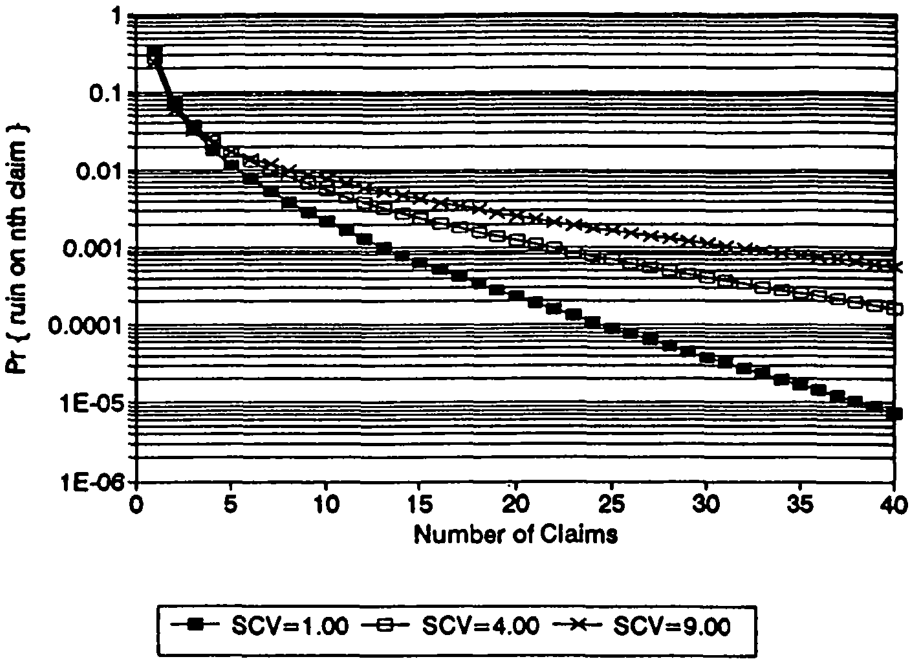


FIGURE 7. Effect of claim size variability.

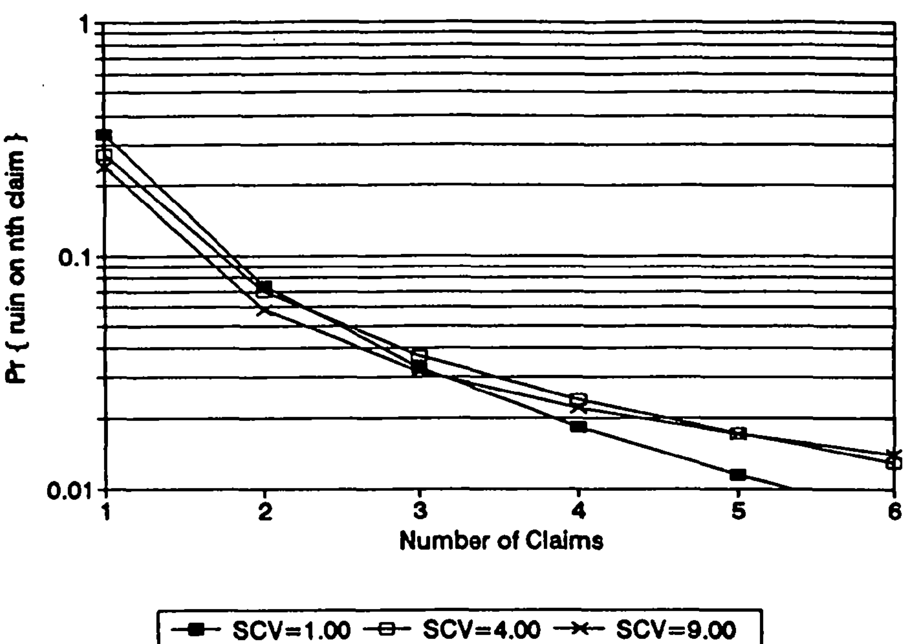


FIGURE 8. Effect of claim size variability.

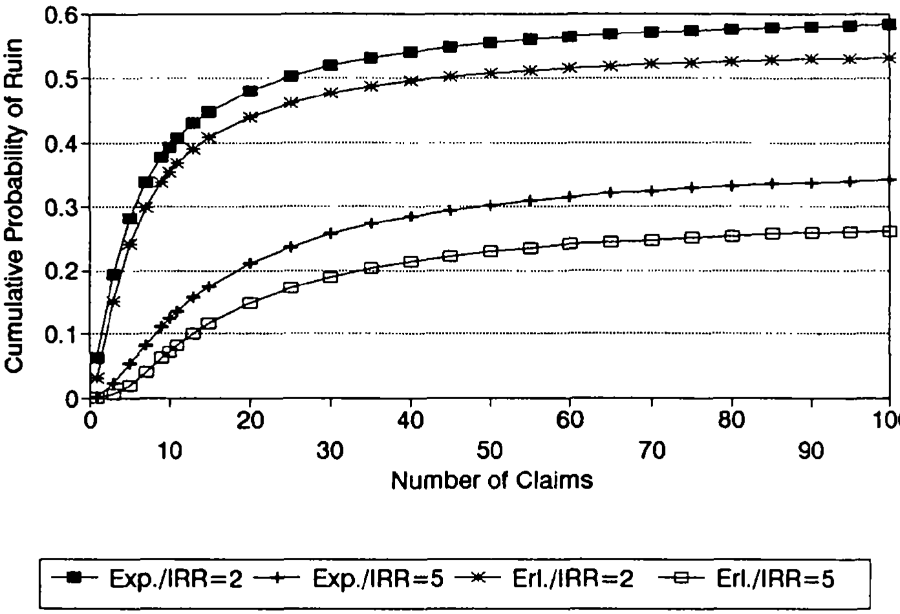


FIGURE 9. Effect of claim size distribution (security loading = 20%).

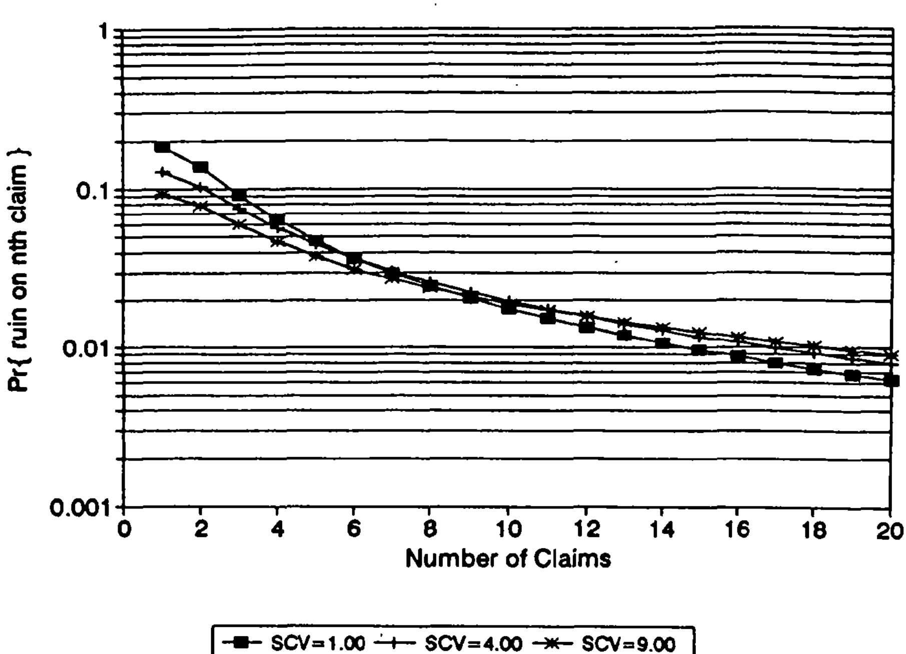


FIGURE 10. Effect of claim size variability.

Figure 10 provides a comparison of the ruin probabilities for the same three claim sizes distributions as Figures 7 and 8, assuming a non-zero initial reserve. The IRR equals 1 in this case, and the relative security loading is 0%. Although ruin is therefore certain, we chose this example to demonstrate the ruin behaviour over finite time. Again, we see that in the short run, the exponential case has the highest likelihood of ruin, but in the long run the situation is reversed.

4. Conclusions and Future Work

The current paper has presented recursive methods for determining the probability of ruin at claim instants. Among the advantages of this approach are the fact that it is exact, and that it reveals where in the future that the relative vulnerability to the company lies. We hope to extend this method to include non-Poisson claims processes in further work.

5. Acknowledgements

We would like to thank our joint graduate student Chit Nam Chua for programming the algorithms and preparing the graphs. We would also like to thank the referees for their helpful comments. This work has been financially supported by Canada's Natural Sciences and Engineering Research Council operating grant number OGP0041187.

References

Arfwedson, G. (1950) Some problems in the collective theory of risk. *Skand. Aktuartidskr.* 1-38.

Asmussen, S. and Bladt, M. (1992) Phase-type distributions and risk processes with state-dependent premiums. *Technical Report* **R92-2022**, The University of Aalborg.

Asmussen, S. and Rolski, T. (1991) Computational methods in risk theory: a matrix-algorithmic approach. *Technical Report* **R91-13**, The University of Aalborg.

Beekman, J.A. (1966) Research on the collective risk stochastic process. *Skand. Aktuartidskr.* 65-77.

Bowers, N.L., Gerber, H.V., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1986) *Actuarial Mathematics.* Society of Actuaries, Itasca, Illinois.

Cramér, M. (1955) Collective Risk Theory. *Jubilee Volume of Försäkringsaktiebogalet Skandia.*

De Vylder, F. and Goovaerts, M.J. (1988) Recursive calculations of finite-time ruin probabilities. *Insurance: Mathematics and Economics* 7, 1-8.

Dickson, D.C.M. and Waters, H.R. (1991) Recursive calculation of survival probabilities. *ASTIN Bulletin* **21**, 199-221.

Dickson, D.C.M. and Waters, H.R. (1992) The probability of severity of ruin in finite and infinite time. *ASTIN Bulletin* **22**, 177-190.

Dufresne, F. and Gerber, H.U. (1988) The probability and severity of ruin for combinations of exponential claim amount distributions and their Translations. *Insurance: Mathematics and Economics* 7, 75-80.

Garrido, J. (1988) Diffusion premiums for claim severities subject to inflation. *Insurance: Mathematics and Economics* 7, 123-129.

Gerber, H.U. (1979) *An Introduction to Mathematical Risk Theory.* Monograph No. 8, S.S. Huebuer Foundation, Distributed by R. Irwin, Homewood, IL.

Gerber, H.U., Goovaerts, M.J. and Kaas, R. (1987) On the probability and severity of ruin. *ASTIN Bulletin* **17**, 151-163.

Janssen, J. (1980) Some transient results on the *M/SM/*1 special semi-Markov model in risk and queueing theories. *ASTIN Bulletin* **11**, 41-51.

Janssen, J. (1982) On the interaction between risk and queueing theories. *Blätter* **15**, 383-395.

Janssen, J. and Delfosse, P. (1982) Some numerical aspects in transient risk theory. *ASTIN Bulletin* **13**, 99-113.

NEUTS, M.F. (1981) *Matrix-geometric Solutions in Stochastic Models.* John Hopkins University Press, Baltimore.

PANJER, H.H. (1981) Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* **12**, 22-26.

PRABHU, N.U. (1961) On the ruin problem of collective risk theory. *Annals of Math. Statistics* **32**, 757-764.

PRABHU, N.U. (1965) Queues and Inventories, John Wiley & Sons, New York.

SHIU, E.S.W. (1988) Calculation of the probability of eventual ruin by Beekman's convolution series. *Insurance: Mathematics and Economics* **7**, 41-47.

TAKÁCS, L. (1967) *Combinatorial Methods in the Theory of Stochastic Processes*, John Wiley & Sons, New York.

TAYLOR, G.C. (1978) Representation and explicit calculation of finite time ruin probabilities. *Scan. Actuarial Journal* **78** (1), 1-18.

TAYLOR, G.C. (1985) A heuristic review of some ruin theory results. *ASTIN Bulletin* **15**, 73-88.

THORIN, O. (1968) An identity in the collective risk theory with some applications. *Skand. Aktuartidskr.* 26-44.

WILLMOT, G. (1990) A queueing theoretic approach to the analysis of the claims payment process. *Transactions of the Society of Actuaries*, XLII, 447-497.

ON THE COMPOUND GENERALIZED POISSON DISTRIBUTIONS

BY R. S. AMBAGASPITIYA AND N. BALAKRISHNAN

University of Calgary — McMaster University

ABSTRACT

GOOVAERTS and KAAS (1991) present a recursive scheme, involving Panjer's recursion, to compute the compound generalized Poisson distribution (CGPD). In the present paper, we study the CGPD in detail. First, we express the generating functions in terms of Lambert's W function. An integral equation is derived for the pdf of CGPD, when the claim severities are absolutely continuous, from the basic principles. Also we derive the asymptotic formula for CGPD when the distribution of claim severity satisfies certain conditions. Then we present a recursive formula somewhat different and easier to implement than the recursive scheme of GOOVAERTS and KAAS (1991), when the distribution of claim severity follows an arithmetic distribution, which can be used to evaluate the CGPD. We illustrate the usage of this formula with a numerical example.

KEYWORDS

Compound generalized Poisson distributions; moments; integral equations; recursive equation; tail behaviour.

1. INTRODUCTION

Modelling the claim frequency data is one of the most important areas in risk theory. Traditionally, the Poisson distribution, when the mean number of claims is equal to its variance, and the negative binomial distribution, when the variance of the number of claims exceeds its mean, have been used because of their convenient mathematical properties. Several authors including GOSSIAUX and LEMAIRE (1981), SEAL (1982) and WILLMOT (1987) have considered alternatives to Poisson and negative binomial distributions for this purpose. CONSUL (1990) has compared the Generalized Poisson distribution (GPD) suggested by CONSUL and JAIN (1973) with several well known distributions and concluded that GPD is a plausible model for claim frequency data. GOOVAERTS and KAAS (1991) presented a recursive scheme to compute the total claim distribution under the assumptions that the claims are independently and identically distributed integer random variables with the GPD claim frequency.

In this paper, we discuss the compound generalized Poisson distribution (CGPD) in detail and derive a somewhat easy to programmable recursive relation than one given by GOOVAERTS and KAAS (1991). In Section 2, we present a brief summary of the properties of the generalized Poisson distribution. In Section 3, we express the generating functions of CGPD in terms of Lambert's W function and illustrate the derivation of moments. In Section 4, we present an integral equation similar to Volterra's integral equation of second kind for the density function of CGPD when the distribution of claim severity is absolutely continuous. In addition, we discuss the tail behaviour of CGPD when the claim severity is non arithmetic. In Section 5, we present a recursive formula for the probability function of CGPD when the distribution of claim severity is arithmetic. We illustrate the usage of this formula through an example.

2. GENERALIZED POISSON DISTRIBUTION (GPD)

CONSUL and JAIN (1973) proposed a new generalization of the discrete Poisson distribution which was modified by CONSUL and SHOUKRI (1985) to: A discrete random variable N is said to have a generalized Poisson distribution (GPD) if its probability mass function is given by

$$(2.1)\quad \Pr(N=n)=p_n(\lambda,\theta)=\begin{cases}\lambda(\lambda+n\theta)^{n-1}\dfrac{\exp(-\lambda-n\theta)}{n!} & \text{for } n=0,1,2\ldots\\ 0 & \text{for } n>m \text{ when } \theta<0\end{cases}$$

and zero otherwise, where $\lambda > 0$, $\max(-1, -\lambda/m) \le \theta < 1$ and $m(\ge 4)$ is the largest positive integer for which $\lambda + \theta m > 0$ when θ is negative. This generalization of the Poisson probability model in the sense that is probability generating function (pgf) is given by the Lagrange expansion of any pgf under a suitable transformation (CONSUL and SHENTON (1972)). The GPD reduces to the Poisson distribution when $\theta = 0$ and it possesses the twin properties of over-dispersion and under-dispersion according as $\theta > 0$ or $\theta < 0$. The GPD gets truncated for negative values of θ but the truncation error is always less than 0.07%. A recent book by CONSUL (1989) discusses various properties, inference and numerous applications of this model in biology, ecology, and other disciplines. For simplicity, from here on we assume the parameter $\theta > 0$. AMBAGASPITIYA and BALAKRISHNAN (1993) has recently expressed the moment generating function $M_N(t)$ and the probability generating function of the GPD in terms of Lambert's W function when $\theta > 0$ as follows:

$$(2.2)\qquad M_N(t) = \exp\left\{-\frac{\lambda}{\theta}[W(-\theta\exp(-\theta+t))+\theta]\right\}$$

$$(2.3)\qquad P_N(z) = \exp\left\{-\frac{\lambda}{\theta}[W(-\theta z\exp(-\theta))+\theta]\right\}$$

where W is the Lambert's W function defined as

$$W(x)\exp(W(x)) = x.$$

For more details about Lambert's W function see CORLESS et al. (1994).

2.1. Central moments of GPD

We can obtain the central moments of GPD by differentiating (2.2) with respect to t as illustrated by AMBAGASPITIYA and BALAKRISHNAN (1993), or from the basic principles as described by CONSUL (1989), or by using the method suggested by GOOVAERTS and KAAS (1991). The resulting expressions for first four central moments are as follows:

$$\begin{aligned} \mu_1 &= \lambda M \\ \mu_2 &= \lambda M^3 \\ \mu_3 &= \lambda(3M-2)M^4 \\ \mu_4 &= 3\lambda^2 M^6 + \lambda(15M^2 - 20M + 6)M^5 \end{aligned} \tag{2.4}$$

where $M = (1-\theta)^{-1}$.

2.2. Maximum likelihood estimators of λ and θ

Let a random sample of n items be taken from the GPD model and let $x_1, x_2, \ldots, x_n$ be their corresponding values. If the sample values are classified into class frequencies and n_i denotes the frequency of the ith class ($n_i = \# \{x_j : 1 \le j \le n, x_j = i\}$), the ML estimate $\hat{\theta}$ as described in CONSUL and SHOUKRI (1984) is given by the unique root of θ given by the equation

$$\sum_{i=0}^{k} n_i \frac{i(i-1)}{\bar{x} + (i-\bar{x})\theta} - n\bar{x} = 0 \tag{2.5}$$

where k (≤ 2) is the number of classes, $n = \sum_{i=1}^{k} n_i$ and $\bar{x}$ is the sample mean. Note that (2.5) does not give a value for θ when $k = 0$ or 1. The ML estimate $\hat{\lambda}$ is then given by

$$\hat{\lambda} = \bar{x}(1-\hat{\theta}) \tag{2.6}$$

2.3. Tail behaviour of GPD

Lemma 2.1: For fixed λ, θ and $n \to \infty$

$$\Pr(N=n) \approx \frac{\lambda}{\theta\sqrt{2\pi}} \exp\left(-\lambda + \frac{\lambda}{\theta}\right) n^{-3/2} \cdot (\theta \exp(1-\theta))^n \tag{2.7}$$

Proof:

For large n, using the Stirling approximation to $n!$ we can write the pmf in (2.1) as

$$\Pr(N=n) \approx \frac{\lambda(\lambda+n\theta)^{n-1} \exp(-\lambda - n\theta)}{\sqrt{2\pi}\, n^{n+1/2} \exp\left(-n + \dfrac{\theta_1}{12n}\right)} \tag{2.8}$$

where $\theta_1 = \theta_1(n)$ satisfies $0 < \theta_1 < 1$. After some rearrangement, we have

$$\text{(2.9)} \quad \Pr(N = n) \approx \left\{\frac{\lambda}{\theta\sqrt{2\pi}}\left[1 + \frac{\lambda}{\theta n}\right]^{n-1} \exp\left(-\lambda - \frac{\theta_1}{12n}\right)\right\} n^{-3/2}\,(\theta \exp(1-\theta))^n;$$

Note that the term inside the { } tends to the required constant as $n \to \infty$ and hence the proof.

■

3. COMPOUND GENERALIZED POISSON DISTRIBUTION (CGPD)

Let N denote the number of claims produced by a portfolio of policies in a given time period. Let X_i denote the amount of the ith claim. Then

$$\text{(3.1)} \qquad S = X_1 + X_2 + \ldots + X_N$$

represents the aggregate claims generated by the portfolio for the period under study. In order to make the model tractable, two fundamental assumptions are made in risk theory and they are

1. $X_1, X_2, \ldots$ are identically distributed random variables with the distribution function $F(x)$.
2. The random variables $N, X_1, X_2, \ldots$ are mutually independent.

When a GPD is chosen for N, the distribution of S is called a compund generalized Poisson distribution. In terms of the convolution operation, we can write the distribution function of S as:

$$F_S(x) = \sum_{n=0}^{\infty} F^{*n}(x)\,\lambda(\lambda + n\theta)^{n-1}\,\frac{\exp(-\lambda - n\theta)}{n!}.$$

The moment generating function of S is given by

$$\text{(3.2)} \qquad M_S(t) = M_N(\log M_X(t)),$$

where $M_N(t)$ is the moment generating function (mgf) of the GPD and $M_X(t)$ is the mgf of the claim amount distribution. By using the expression given in (2.2), we can write the mgf of S as

$$\text{(3.3)} \qquad M_S(t) = \exp\left\{-\frac{\lambda}{\theta}\left[W(-\theta\exp(-\theta)\,M_X(t)) + \theta\right]\right\}.$$

Similarly, the probability generating function (pgf) of S, when the distribution of claim severity is arithmetic, can be written as

$$\text{(3.4)} \qquad P_S(z) = \exp\left\{-\frac{\lambda}{\theta}\left[W(-\theta\exp(-\theta)\,P_X(z)) + \theta\right]\right\},$$

where $P_X(z)$ is the pgf of claim amount distribution.

3.1. Central moments of S

The moments of S can be obtained by directly differentiating the mgf of S given in (3.3). For this differentiation, one may use the following identity, involving Lamberts W functions:

$$\frac{dW(x)}{dx} = \frac{W(x)}{x(1+W(x))}.$$

After some lengthy algebra, we obtain the following expressions for the first three central moment of S:

$$E(S) = \lambda p_1 M$$

$$\operatorname{Var}(S) = \lambda p_2 M^3 + \lambda (p_2 - p_1^2) M$$

$$E((S-E(S))^3) = \lambda (3M-2) p_1^3 M^4 + 3\lambda p_1 (p_2 - p_1^2) M^3 + (p_3 - 3p_2 p_1 + 2p_1^3)\lambda M$$

where $M = (1-\theta)^{-1}$ and p_i, $i = 1, 2, 3$ are the ith non-central moments of claim severity.

4. PROPERTIES OF CGPD: ABSOLUTELY CONTINUOUS SEVERITIES

Theorem 4.1: If the claim sizes are absolutely continuous with pdf $f(x)$ for $x > 0$, then the pdf $g(\lambda, \theta; x)$ of CGPD satisfy the integral equation

$$g(\lambda, \theta; x) = p_1(\lambda, \theta) f(x) + \frac{\lambda}{\lambda+\theta} \int_0^x \left(\theta + \lambda \frac{y}{x}\right) g(\lambda+\theta, \theta; x-y) f(y)\, dy \tag{4.1}$$

where $p_1(\lambda, \theta) = \Pr(N = 1)$ in the GPD with parameters λ and θ.

Proof:

Consider

$$g(\lambda, \theta; x) = \sum_{i=1}^{\infty} p_i(\lambda, \theta) f^{*i}(x) \tag{4.2}$$

$$= p_1(\lambda, \theta) f(x) + \sum_{i=2}^{\infty} p_i(\lambda, \theta) f^{*i}(x) \tag{4.3}$$

By using the following identity of GPD,

$$p_i(\lambda, \theta) = \frac{\lambda}{\lambda+\theta} \left(\theta p_{i-1}(\lambda+\theta, \theta) + \frac{\lambda}{i} p_{i-1}(\lambda+\theta, \theta)\right) \qquad i = 1, 2, \ldots \tag{4.4}$$

we have

$$\sum_{i=2}^{\infty} p_i(\lambda, \theta) f^{*i}(x) = \frac{\lambda}{\lambda+\theta} \left(\theta \sum_{i=2}^{\infty} p_{i-1}(\lambda+\theta, \theta) f^{*i}(x) + \right. \tag{4.5}$$

$$\left. + \lambda \sum_{i=2}^{\infty} \frac{p_{i-1}(\lambda+\theta, \theta)}{i} f^{*i}(x)\right)$$

Using the identities

$$f^{*i}(x)=\int_0^x f^{*(i-1)}(x-y)f(y)\,dy \tag{4.6}$$

and

$$\frac{f^{*i}(x)}{i}=\int_0^x \frac{y}{x}f^{*(i-1)}(x-y)f(y)\,dy \tag{4.7}$$

we have

$$\sum_{i=2}^{\infty} p_i(\lambda,\theta)f^{*i}(x)=\frac{\lambda}{\lambda+\theta}\left(\theta\sum_{i=2}^{\infty} p_{i-1}(\lambda+\theta,\theta)\int_0^x f^{*(i-1)}(x-y)f(y)\,dy\right.$$
$$\left.+\lambda\sum_{i=2}^{\infty} p_{i-1}(\lambda+\theta,\theta)\int_0^x \frac{y}{x}f^{*(i-1)}(x-y)f(y)\,dy\right) \tag{4.8}$$

By interchanging the order of summation and the integration and realizing the fact

$$\sum_{i=2}^{\infty} p_{i-1}(\lambda+\theta,\theta)f^{*(i-1)}(x-y)=\sum_{i=1}^{\infty} p_i(\lambda+\theta,\theta)f^{*i}(x-y) \tag{4.9}$$

$$=g(\lambda+\theta,\theta;x-y) \tag{4.10}$$

we have

$$\sum_{i=2}^{\infty} p_i(\lambda,\theta)f^{*i}(x)=\frac{\lambda}{\lambda+\theta}\int_0^x\left(\theta+\lambda\frac{y}{x}\right)g(\lambda+\theta,x-y)f(y)\,dy \tag{4.11}$$

Substitution of (4.11) in (4.3) yields the required result and hence the theorem.

■

One has to solve the integral equation (4.1) numerically. Although, there are many algorithms and implementations available to solve Volterra integral equations of the second kind, one has to modify them to solve (4.1). We are currently investigating the problem of finding the best algorithm and we hope to report this finding in a future article.

4.1. Tail behaviour of CGPD

Theorem 4.2: If there exists a number $\kappa>0$ satisfying

$$\frac{\exp(\theta)}{e\theta}=\mathcal{L}_X(-\kappa) \tag{4.12}$$

for X non-arithmetic and if $-\mathcal{L}_X'(-\kappa) < \infty$, then

$$1 - F_S(x) \sim Cx^{-3/2} \exp(-\kappa x), \tag{4.13}$$

where C is given by

$$C = \frac{\lambda}{\theta\sqrt{2\pi}} \exp\left(-\lambda + \frac{\lambda}{\theta}\right)\sqrt{\kappa\left[-\frac{\exp(\theta-1)}{\theta}\mathcal{L}_\lambda'(-\kappa)\right]}$$

Proof:

The Proof of this theorem directly follows from the Lemma 2.1 and from the theorem of EMBRECHTS, MAEJIMA, and TEUGELS (1982).

■

5. PROPERTIES OF CGPD: ARITHMETIC SEVERITIES

Theorem 5.1: If the claim sizes are random variables on the positive integers with probability mass function $f(x) = \Pr(X = x)$, $x = 0, 1, 2, \ldots$, then the probability mass function $g(\lambda, \theta; x)$ of CGPD satisfies the recurrence equation

$$g(\lambda, \theta; x) = \frac{\lambda}{\lambda+\theta} \sum_{y=1}^{x} \left(\theta + \lambda\frac{y}{x}\right) g(\lambda+\theta, \theta; x-y) f(y). \tag{5.1}$$

Proof:

This theorem can be proved following the same line of reasoning as Theorem 4.1 or the standard proof of Panjer's recursion (see Theorem 6.6.1 and Corollary 6.6.1 in PANJER and WILLMOT (1992)).

■

A result analogue to Theorem 4.2 can be established for discrete severity case using Lemma 2.1 and the theorem given in WILMOT (1989).

5.1. Recursive evaluation

The recursive formula given in (5.1) is easily programmable and also simple to use for manual calculations. For the latter, one may use the following schematic approach:

$g(\lambda, \theta, 0)$	$g(\lambda+\theta, \theta, 0)$	$g(\lambda+2\theta, \theta, 0)$	$g(\lambda+3\theta, \theta, 0)$	$g(\lambda+4\theta, \theta, 0)$
$g(\lambda, \theta, 1)$	$g(\lambda+\theta, \theta, 1)$	$g(\lambda+2\theta, \theta, 1)$	$g(\lambda+3\theta, \theta, 1)$	
$g(\lambda, \theta, 2)$	$g(\lambda+\theta, \theta, 2)$	$g(\lambda+2\theta, \theta, 2)$		
$g(\lambda, \theta, 3)$	$g(\lambda+\theta, \theta, 3)$			
$g(\lambda, \theta, 4)$				

The first row of the above scheme is obtained by using the fact that $g(\lambda + i\theta, 0) = p_0(\lambda + i\theta, \theta) = \exp(-\lambda - i\theta)$ for $i = 0, 1 \ldots$. To calculate the probability mass function given in the (i, j)th location, one has to use the elements in $(l, j+1)$ where $l = 0, 1, \ldots, i-1$. Since the scheme is of an upper diagonal form, we can carry out the computations for each row starting from right to left. For example, if one wishes to compute $g(\lambda, \theta, 4)$ one may start from $g(\lambda + 4\theta, \theta, 0)$ and move along the diagonal from right to left, i.e. calculate $g(\lambda + (4-i)\theta, \theta, i)$, $i = 0, 1, 2, 3, 4$ in that order.

Example:

Suppose that S has a CGPD with $\lambda = 0.8$, $\theta = 0.5$ and the distribution of individual claim amounts is as follows:

x	$\Pr(X = x)$
1	0.25
2	0.45
3	0.30

Then, by using the recursive method described above, the pmf of S has been tabulated for $s = 0(1)59$ and these values are presented in Table 1.

TABLE 1

THE PROBABILITY MASS FUNCTION OF S

s	$\Pr(S = s)$	s	$\Pr(S = s)$	s	$\Pr(S = s)$
0	.44933	20	.00269	40	.00017
1	.05451	21	.00231	41	.00015
2	.10555	22	.00198	42	.00013
3	.09329	23	.00171	43	.00012
4	.04809	24	.00148	44	.00010
5	.04813	25	.00128	45	.00009
6	.03595	26	.00111	46	.00008
7	.02737	27	.00096	47	.00007
8	.02320	28	.00083	48	.00006
9	.01835	29	.00073	49	.00006
10	.01505	30	.00063	50	.00005
11	.01248	31	.00055	51	.00004
12	.01029	32	.00048	52	.00004
13	.00860	33	.00042	53	.00003
14	.00720	34	.00037	54	.00003
15	.00605	35	.00032	55	.00003
16	.00512	36	.00028	56	.00002
17	.00434	37	.00025	57	.00002
18	.00369	38	.00022	58	.00002
19	.00315	39	.00019	59	.00002

ACKNOWLEDGMENTS

The authors would like to thank both referees for their valuable comments when revising this paper. This work was supported by the Natural Siences and Engineering Research Council of Canada.

REFERENCES

AMBAGASPITIYA, R.S. and BALAKRISHNAN, N. (1993) Some remarks on Lagragian distributions, Submitted for publication.

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

CONSUL, P.C. (1990) A model for distributions of injuries in auto-accidents. *Itteilungen der Schweiz Vereinigung der Versicherungsmathematiker* **Heft 1**, 161–168.

CONSUL, P.C. and JAIN, G.C. (1973) A generalization of 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 Journal of Applied Mathematics* **23**, 239–248.

CONSUL, P.C. and SHOUKRI, M.M. (1984) Maximum likelihood estimation for the generalized Poisson distribution. *Communications in Statistics- Theory and Methods* **10**, 977–991.

CONSUL, P.C. and SHOUKRI, M.M. (1985) The generalized Poisson distribution when the sample mean is larger than the sample variance. *Communications in Statistics — Simulation and Computation* **14**, 1533–1547.

CORLESS, R.M., GONNET, G.H., HARE, D.E.G. and JEFFREY, D.J. (1994) The Lambert *W* function. To appear in Advances in Computational Mathematics.

EMBRECHTS, P., MAEJIMA, M. and TEUGELS, J. (1982) Asymptotic behaviour of compound distributions. *ASTIN Bulletin* **15**, 45–48.

GOOVAERTS, M.J. and KAAS, R. (1991) Evaluating compound generalized Poisson distributions recursively. *ASTIN Bulletin* **21**, 193–197.

GOSSIAUX, A. and LEMAIRE, J. (1981) Methodes d'adjustement de distribution de sinistres. *Bulletin of the Association of Swiss Actuaries* **81**, 87–95.

PANJER, H.H. and WILLMOT, G.E. (1992) *Insurance Risk Models.* Society of Actuaries. SEAL, H. (1982) Mixed Poisson-an ideal distribution of claim numbers? *Bulletin of the Association of Swiss Actuaries* **82**, 293–295.

WILLMOT, G.E. (1987) The Poisson-Inverse Gaussian distribution as an alternative to the Negative Binomial. *Sandinavian Actuarial Journal*, 113–127.

WILLMOT, G.E. (1989) Limiting tail behaviour of some discrete compound distributions. *Insurance: Mathematics and Economics* **8**, 175–185.

R.S. AMBAGASPITIYA
Department of Mathematics and Statistis, University of Calgary,
Calgary, Alberta, Canada, T2N 1N4.

AND N. BALASKRISHNAN
Department of Mathematics and Statistics, McMaster University,
Hamilton, Ontario, Canada, L8S 4K1.

MODELLING THE CLAIMS PROCESS IN THE PRESENCE OF COVARIATES

BY ARTHUR E. RENSHAW

Department of Actuarial Science & Statistics
The City University, London

ABSTRACT

An overview of the potential of Generalized Linear Models as a means of modelling the salient features of the claims process in the presence of rating factors is presented. Specific attention is focused on the rich variety of modelling distributions which can be implemented in this context.

KEYWORDS

Claims Process; Rating Factors; Generalized Linear Models; Quasi-Likelihood; Extended Quasi-Likelihood.

1. INTRODUCTION

The claims process in non-life insurance comprises two components, claim frequency and claim serverity, in which the product of the underlying expected claim rate and expected claim severity defines the pure or risk premium. Specifically, considerable attention is given to the probabalistic modelling of various aspects of a single batch of claims, often focusing on the aggregate claims accruing in a time period of fixed duration, typically one year, under a variety of assumptions imposed on the claim frequency and claim severity mechanisms.

In this paper, attention is refocused on the considerable potential of generalized linear models (GLMs) as a comprehensive modelling tool for the study of the claims process in the presence of covariates. Section 2 contains a brief summary of the main features of GLMs which are of potential interest in modelling various aspects of the claims process. Particular attention is drawn to the rich variety of modelling distributions which are available and to the parameter estimation and model fitting techniques based on the concepts of quasi-likelihood and extended quasi-likelihood. Sections 3 and 4 focus respectively on the modelling of the claim frequency and claim severity components of the process in the presence of covariates. An overview of the potential of GLMs as a means of modelling these two aspects of the claims process is discussed. Relevant published applications are referenced, although an exhaustive search of the literature has not been conducted. A number of the suggested modelling techniques are illustrated in Section 5.

2. GLMs. QUASI-LIKELIHOOD. EXTENDED QUASI-LIKELIHOOD

Focus intially on independent response variables $\{Y_i\colon i = 1, 2, \ldots, n\}$ with either density or point mass function, as the case may be, of the type

$$f(y_i \mid \theta_i, \phi_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)\right\} \tag{2.1}$$

for specified functions $a(.)$, $b(.)$ and $c(.)$, where θ_i is the canonical parameter and ϕ_i the dispersion parameter. The cumulant function $b(.)$ plays a central role in characterising many of the properties of the distribution. It gives rise to the cumulant generating function, K, of the random variable Y_i, assuming it exits, according to the equation

$$K_{Y_i}(t) = \frac{b\{a(\phi_i)t + \theta_i\} - b\{\theta_i\}}{a(\phi_i)}. \tag{2.2}$$

Our immediate concern therefore is with distributions with at most two parameters.

Let $\mu_i = E(Y_i)$ throughout. Comparison of the density or point mass function of a standard distribution with expression (2.1) establishes membership or otherwise of this class of distributions. It also determines the specific nature of the canonical parameter θ_i and function $a(.)$ up to a constant, as well as the nature of the dispersion parameter ϕ_i and the other two functions $b(.)$ and $c(.)$. To uniquely determine θ_i and $a(.)$ it is also necessary to compare the variance of the standard distributions with the general expression (2.6) or, more specifically, expression (2.8) for the variance of Y_i.

For inference, the log-likelihood is

$$l = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)\right\}. \tag{2.3}$$

The identity

$$E\left\{\frac{\partial l_i}{\partial \theta_i}\right\} = 0 \Rightarrow E(Y_i) = \mu_i = b'(\theta_i) \tag{2.4}$$

where dash denotes differentiation. Thus, provided the function $b'(.)$ has an inverse, which is defined to be the case, the canonical parameter $\theta_i = b'^{-1}(\mu_i)$, a known function of μ_i.

The identity

$$E\left\{\frac{\partial^2 l_i}{\partial \theta_i^2}\right\} + E\left\{\left(\frac{\partial l_i}{\partial \theta_i}\right)^2\right\} = 0 \Rightarrow \mathrm{Var}(Y_i) = b''(\theta_i)\, a(\phi_i)$$

the product of two functions. Noting that $b''(.)$ is a function of the canonical parameter θ_i and hence of μ_i, the identity

$$b''(\theta_i) = V(\mu_i) \tag{2.5}$$

is established and hence the so-called variance function $V(.)$ defined. Hence the variance or second cumulant is

$$\text{(2.6)} \qquad \mathrm{Var}\,(Y_i) = \kappa_2^{(i)} = V(\mu_i)\, a(\phi_i).$$

The other function $a(.)$ is commonly of the type

$$\text{(2.7)} \qquad a(\phi_i) = \frac{\phi}{\omega_i}$$

with constant scale parameter ϕ and prior weights ω_i so that

$$\text{(2.8)} \qquad \mathrm{Var}\,(Y_i) = \frac{\phi V(\mu_i)}{\omega_i}.$$

This is assumed to be the case throughout. We remark that by setting $\phi = 1$, $1/\omega_i = \phi_i$, the reciprocals of the weights may also be re-interpreted as non-constant scale parameters ϕ_i.

We shall also have occasion to examine the degree of skewness in the Y_is. Here the identity

$$E\left\{\frac{\partial^3 l_i}{\partial \theta_i^3}\right\} + 3E\left\{\frac{\partial^2 l_i}{\partial \theta_i^2}\,\frac{\partial l_i}{\partial \theta_i}\right\} + E\left\{\left(\frac{\partial l_i}{\partial \theta_i}\right)^3\right\} = 0 \Rightarrow E\{(Y_i - \mu_i)^3\} = b'''(\theta_i)\, a^2(\phi_i)$$

so that, in terms of the variance function $V(.)$, on using equation (2.5), the third cumulant of Y_i is

$$\kappa_3^{(i)} = V\frac{dV}{d\mu_i}\,\{a(\phi_i)\}^2.$$

Hence the coefficient of skewness

$$\text{(2.9)} \qquad \frac{\kappa_3^{(i)}}{\{\kappa_2^{(i)}\}^{3/2}} = V^{-1/2}\frac{dV}{d\mu_i}\,\{a(\phi_i)\}^{1/2}.$$

The expressions for the second and third cumulants can also be derived from the cumulant generating function (2.2).

Covariates may be either explanatory variables, or explanatory factors, or a mixture of both. In all three cases, covariates enter through a linear predictor

$$\eta_i = \sum_j x_{ij}\beta_j$$

with known covariate structure (x_{ij}) and unknown regression parameters β_j and are linked to be mean, μ_i, of the modelling distribution through a monotonic, differentiable (link) function g with inverse g^{-1}, such that

$$g(\mu_i) = \eta_i \qquad \text{or} \qquad \mu_i = g^{-1}(\eta_i).$$

To fit such a model structure, maximum likelihood estimates for the $\beta_j s$ are normally sought. These are obtained through the numerical solution of the equations

$$\sum_{i=1}^{n} \omega_i \frac{y_i - \mu_i}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0 \qquad \forall j \tag{2.10}$$

derived by setting the partial derivatives

$$\frac{\partial l}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

of the log-likelihood with respect to the unknown parameters β_j to zero. Equations (2.3), (2.4), (2.5) and (2.7) are needed in the evaluation of the first two partial derivative terms on the right hand side. These estimates are sufficient in the case of the canonical link function, defined by $g = b'^{-1}$.

To broaden the genesis of equations (2.10) by relaxing the constraints imposed by the full log-likelhood assumption (2.3) and its associated distribution assumption (2.1), define

$$q = q(\underline{y}; \underline{\mu}) = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} \omega_i \int_{y_i}^{\mu_i} \frac{y_i - s}{\phi V(s)} \, ds \tag{2.11}$$

to be the quasi-likelihood (strictly quasi-log-likelihood) function. Then by setting the partial derivatives of q (rather than l) with respect to β_j to zero, equations (2.10) are again reproduced. Equations (2.10) are called the Wedderburn quasi-likelihood estimating equations. The resulting quasi-likelihood parameter estimates have similar asymptotic properties to maximum likelihood parameters estimates and are identical to maximum likelihood parameter estimates for the class of distributions defined by equation (2.1). This latter class of distributions includes the binomial, Poisson, gamma and inverse Gaussian distributions, all of which are of potential interest in a claims context. The individual details are summarised in Table 2.1. The overriding feature of both the quasi-likelihood expression (2.11) and the Wedderburn quasi-likelihood estimating equations (2.10) is that a knowledge of only the first and second moments is required of the modelling distribution of the $Y_i s$. Hence, by this means, it is possible to relax the full log-likelihood assumption (2.3) and extend the range of distributions which can be readily linked to covariates in practice with an attendant shift in emphasis from maximum likelihood estmation to maximum quasi-likelihood estimation. This has important implications for the claims process which are discussed in context later.

The goodness-of-fit of different hierarchical model predictor structures is monitored, in the first instance, by comparing the differences in model deviances. To do this, compare the current model structure, denoted by c, and whose fitted values are denoted by $\hat{\mu}_i$; with the full or saturated model structure, denoted by f, and which is characterised by the fitted values $\tilde{\mu}_i = y_i$, the perfect fit. Let $\hat{\theta}_i$ and $\tilde{\theta}_i$ denote the corresponding values of the canonical parameter, defined by $\theta_i = b'^{-1}(\mu_i)$, the inverse of b'. Since we are concerned here exclusively with changes to the structure

TABLE 2.1

DETAILS OF SPECIFIC GLM DISTRIBUTIONS

	Poisson	binomial	gamma	inverse Gaussian
d.f./p.m.f.	$\dfrac{e^{-\mu}\mu^y}{y!}$	$\binom{m}{y} p^y(1-p)^{m-y}$	$\dfrac{1}{\Gamma(\nu)}\left(\dfrac{\nu}{\mu}\right)^{\nu} y^{\nu-1} e^{-\nu y/\mu}$	$\dfrac{1}{\sqrt{2\pi\tau y^3}}\exp\left\{\dfrac{-1}{2\tau\mu^2 y}(y-\mu)^2\right\}$
parameters	$\mu > 0$	$p\varepsilon(0,1)$	$\mu, \nu > 0$	$\mu, \tau > 0$
range	$y = 0, 1, 2, \ldots$	$y = 0, 1, 2, \ldots, m$	$y > 0$	$y > 0$
canonic par.	$\theta = \log(\mu)$	$\theta = \log\left(\dfrac{p}{1-p}\right)$	$\theta = -\mu^{-1}$	$\theta = -\dfrac{1}{2}\mu^{-2}$
scale par.	$\phi = 1$	$\phi = 1$	$\phi = \nu^{-1}$	$\phi = \tau$
weights	$\omega = 1$	$\omega = 1$	$\omega = 1$	$\omega = 1$
$b(\theta)$	$\exp(\theta)$	$m\log(1+e^{\theta})$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$c(y, \theta)$	$-\log(y!)$	$\log\left\{\binom{m}{y}\right\}$	$\nu\log(\nu) + (\nu-1)\log(y)$ $-\log\{\Gamma(\nu)\}$	$-\dfrac{1}{2}\left\{\log(2\pi\tau y^3) + \dfrac{1}{\tau y}\right\}$
$\mu(\theta)$	$\exp(\theta)$	$\dfrac{me^{\theta}}{1+e^{\theta}}$	$-\dfrac{1}{\theta}$	$(-2\theta)^{-1/2}$
$V(\theta)$	μ	$\mu\left(1-\dfrac{\mu}{m}\right)$	μ^2	μ^3
κ_3	μ	$\mu\left(1-\dfrac{\mu}{m}\right)\left(1-\dfrac{2\mu}{m}\right)$	$2\mu^3\phi^2$	$3\mu^5\phi^2$

of the predictor, the scale parameter ϕ remains the same throughout. Then define

$$(2.12)\quad d^*(\underline{y};\underline{\hat{\mu}}) = -2(l^{(c)} - l^{(f)}) = -2\sum_{i=1}^{n}\frac{\omega_i}{\phi}\{(y_i\hat{\theta}_i - b(\hat{\theta}_i)) - (y_i\tilde{\theta}_i - b(\tilde{\theta}_i))\},$$

minus twice the log-likelihood ratio, of c relative to f, based on equations (2.3) and (2.5), to be the scaled deviance and

$$(2.13)\qquad d(\underline{y};\underline{\hat{\mu}}) = \phi d^*(\underline{y};\underline{\hat{\mu}})$$

to be the (unscaled) deviance of the current model c. Using the identity

$$\int_{y_i}^{\mu_i}\frac{y_i - s}{b''(b'^{-1}(s))}\,ds = \int_{\tilde{\theta}_i}^{\theta_i}(y_i - b'(t))\,dt$$

it follows from equations (2.12), (2.13) and (2.5) that the expression for the deviance can be written as

$$(2.14) \qquad d(\underline{y};\underline{\mu}) = \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} 2\omega_i \int_{\mu_i}^{y_i} \frac{y_i - s}{V(s)}\, ds = -2\phi q(\underline{y};\underline{\mu})$$

where $q(\underline{y};\underline{\mu})$ is the quasi-likelihood function. Hence in common with the construction of the quasi-likelihood and quasi-likelihood estimating equations, a knowledge of only the first and second moments is required of the modelling distribution of the $Y_i s$ to construct the model deviance.

A trivial re-arrangement of equation (2.14) implies the the quasi-likelihood, q, satisfies

$$-2q = \sum_{i=1}^{n} \frac{d_i}{\phi}.$$

To accommodate inference on any parameters, such as ϕ, which might be present in the variance of the response variables Y_i, define the extended quasi-likelihood (strictly the extended-quasi-log-likelihood) q^+ where

$$(2.15) \qquad -2q^+ = \sum_{i=1}^{n} \frac{d_i}{\phi} + \sum_{i=1}^{n} \log\{\phi V(y_i)\} \qquad \left(+ \sum_{i=1}^{n} \log(2\pi/\omega_i)\right).$$

Note that this expresssion is minus twice the log-likelihood for independent normally distributed responses $Y_i \sim N(\mu_i, \sigma^2)$, for which $\sigma^2 = \phi$, $V(\mu_i) = 1$; but is not an exact log-likelihood expression for any other case. The final term in the brackets is constant for a given data set, and may be omitted.

Diagnostic checks, based on a thorough graphical analysis of residuals, are conducted before the final adoption of a specific model structure. Deviance residuals

$$r_i = \operatorname{sign}(y_i - \hat{\mu}_i) \cdot \sqrt{d_i}$$

where d_i is ith component of the deviance defined in equation (2.14), are advocated. A suitable estimate for the constant scale parameter ϕ, if required, is provided by the moment estimator based on generalized Pearson residuals

$$(2.16) \qquad \hat{\phi} = \frac{1}{\nu} \sum_{i=1}^{n} \omega_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

where ν denotes the number of degrees of freedom associated with the fit.

Implementation is possible using the GLIM software package, BAKER & NELDER (1985) which is expressly designed to fit models of this type, while the reader is referred to the text by MCCULLAGH & NELDER (1989) for further detail.

3. CLAIM FREQUENCY

Claim frequency data are denoted throughout by (u, n_u, e_u), comprising the observed number of claims, n_u, accruing from exposures, e_u, defined for a set of units $\{u\}$. Typically the units are of the type $u \equiv (i_1, i_2, i_3, \ldots)$, a cross-classified grid of cells defined for preselected levels of appropriate covariates, often rating factors. A number of different possible modelling scenarios can be implemented.

Focus first on targetting the underlying or expected claim rates, denoted by λ_u, based on the Poisson modelling assumption $N_u \mid \lambda_u \sim \text{Poi}\,(e_u \lambda_u)$, with independence over all cells or units u, and where n_u denotes the realisation of the random variable N_u. Here it is assumed that the claim rates, λ_u, are constant within cells. In the notation of Section 2, the responses $Y_i \equiv N_u$ with

mean $\mu_u = E(N_u) = e_u \lambda_u$, variance function $V(\mu_u) = \mu_u$, scale parameter $\phi = 1$

and log-likelihood

$$l = \sum_{u=1}^{n} \{-\mu_u + n_u \log(\mu_u)\} + \text{constant}. \tag{3.1}$$

Two link functions are of particular interest in this context, namely the log-link and the parameterised power-link.

To implement the canonical log-link, for which

$$\eta_u = \log(\mu_u) = \log(e_u) + \log(\lambda_u) = \log(e_u) + \sum_j x_{uj}\beta_j$$

the vector of $\log(e_u)$ terms is declared as an offset. Such terms from part of the linear predictor and are characterised by a known regression coefficient with value one. Thus the target, λ_u, is linked to covariates through the relationship

$$\log(\lambda_u) = \sum_j x_{uj}\beta_j \Leftrightarrow \lambda_u = \exp\left(\sum_j x_{uj}\beta_j\right),$$

giving rise, possibly, to a multiplicative model structure for rating factors.

A number of applications appear in the literature. Thus McCullagh & Nelder (1983 & 1989), using data provided by the Lloyd's Register of Shipping concerning damage incidents caused to the forward section of cargo-carrying vessels, model the reported number of damage incidents classified by the three factors- ship type, year of construction, period of operation. To allow for possible inter-ship variability in accident proneness, over-dispersion is introduced into the model through the retention of the scale parameter which is then estimated as described in Section 2, rather than setting its value to one. This modelling refinement has an impact on the standard errors of the parameter estimates but not on the parameter estimates themselves (the solutions to equations (2.10)). Andrade e Silvo (1989), Brockman & Wright (1992) and Boskov (1992) have each applied this same model to motor claims data using a variety of potential rating factors in the predictor. Centeno & Andrade e Silvo (1991) discuss the case when there are certain fixed linear relationships between covariates in the predictor. Stroinski & Curri (1989) discuss the selection of rating factors in automobile claim frequency modelling. Renshaw & Haberman (1992) have modelled both sickness inception and sickness recovery rates as well as death rates from sickness with the predictor reflecting both age at sickness inception and, where applicable, sickness duration. A feature of some of this work involves the use of break-point predictor terms in which the positions of the knots or hinges are determined by deviance profiles, constructred by scanning the positional choices of the knots. RENSHAW (1991) has also demonstrated the potential for this model in the graduation of the force of mortality in the construction of life tables.

To implement the parameterised power-link function in this context, the alternative form of the log-likelihood expression:

$$l = \sum_{u=1}^{n} e_u \left\{ -\lambda_u + \frac{n_u}{e_u} \log(\lambda_u) \right\} + \text{constant};$$

obtained by substituting $\mu_u = e_u \lambda_u$ into expression (3.1), is exploited. This implies the declaration of $y_u = n_u / e_u$ as Poisson responses with prior weights e_u, while the predictor link is denoted

$$\lambda_u^{\gamma} = \sum_j x_{uj} \beta_j \Leftrightarrow \lambda_u = \left(\sum_j x_{uj} \beta_j \right)^{1/\gamma}$$

with link parameter γ. The case $\gamma = 1$ corresponds to the identity-link, while the case $\gamma = 0$ corresponds to the log-link. The optimum value of γ for a specific predictor structure is determined by constructing the deviance profile over the viable range of values of γ. Examples of this are to be found in RENSHAW & HABERMAN (1992) and in RENSHAW (1990).

The Poisson model (with $\phi = 1$) assumes that the claim rate, λ_u, is constant within cells. Heterogeneity across risks as opposed to time heterogeneity discussed by BERG & HABERMAN (1992) is historically introduced into the claim frequency process by modelling λ_u as a random variable. Focus on the weighted Poisson responses Y_u $(= N_u / e_u)$ with $Y_u \sim \text{Poi}(\lambda_u)$ so that

$$E(Y_u) = E\{E(Y_u \mid \lambda_u)\} = E(\lambda_u), \tag{3.2}$$

$$\text{Var}(Y_u) = E\{\text{Var}(Y_u \mid \lambda_u)\} + \text{Var}\{E(Y_u \mid \lambda_u)\}$$

and hence

$$\text{Var}(Y_u) = E(\lambda_u) + \text{Var}(\lambda_u). \tag{3.3}$$

Note that when λ_u is constant, $E(\lambda_u) = \lambda_u$, $\text{Var}(\lambda_u) = 0$ and the within cell homogeneous Poisson model is reproduced. For the heterogeneous case, $\text{Var}(Y_u) > E(Y_u)$, that is, the model is over dispersed. There are a number of feasible practical possibilities available:

1) Allow for heterogeneity through the introduction of a constant scale parameter ϕ as described in some of the applications identified above.
2) Allow for heterogeneity through the introduction of non-constant scale parameters ϕ_u and generate their values through the introduction of a second stage

GLM chosen to model identifiable patterns of heterogeneity across cells; a technique known as joint modelling. An example, applied to life insurance, is to be found in RENSHAW (1992).

3) Allow for heterogeneity by nominating a specific distribution for the claim rate λ_u. Thus commonly in the claims context, λ_u is given a gamma distribution with mean $E(\lambda_u)$ and variance

$$\text{(3.4)} \qquad \operatorname{Var}(\lambda_u) = \frac{1}{\nu} \{E(\lambda_u)\}^2 = \frac{1}{\nu} \{E(Y_u)\}^2 \qquad \nu > 0,$$

on using equation (3.2). Then, it is well known that Y_u has the negative binomial distribution, for which the

$$\text{mean } \mu_u = E(Y_u), \text{ variance function } V(\mu_u) = \mu_u + \frac{1}{\nu}\mu_u^2, \text{ scale parameter } \phi = 1$$

on substituting expression (3.4) into equation (3.3). Note that as $\nu \to \infty$, then for finite μ_u, the distribution reverts to the Poisson distribution. Another possibility, BESSON & PARTRAT (1992), TREMBLAY (1992), is to assign an inverse Gaussian distribution with mean $E(\lambda_u)$ and scale parameter τ. It then follows from the relevant column of Table 2.1 that

$$\operatorname{Var}(\lambda_u) = \tau\{E(\lambda_u)\}^3 = \tau\{E(Y_u)\}^3 \qquad \tau > 0$$

so that Y_u now has the Poisson-inverse Gaussian distribution with

$$\text{mean } \mu_u = E(Y_u), \text{ variance function } V(\mu_u) = \mu_u + \tau\mu_u^3, \text{ scale parameter } \phi = 1.$$

This reverts to the Poisson distribution as $\tau \to 0$. Neither of these cases are members of the class of distributions defined by expression (2.1) so that their implementation lead to quasi-likelihood estimators for the $\beta_j s$ in the predictor. If these models are to be implemented, explicity expressions are needed for the deviance components defined in equation (2.14). These are

$$2\omega_u \left\{ y_u \log\left(\frac{y_u}{\mu_u}\right) + (y_u + \nu)\log\left(\frac{\mu_u + \nu}{y_u + \nu}\right) \right\}$$

for the negative binomial distribution, and

$$2\omega_u \left\{ y_u \log\left(\frac{y_u}{\mu_u}\right) + \frac{y_u}{2}\log\left(\frac{1 + \tau\mu_u^2}{1 + \tau y_u^2}\right) + \frac{1}{\sqrt{\tau}} \sin^{-1}\sqrt{\frac{\tau(\mu_u - y_u)^2}{1 + (\tau y_u \mu_u)^2 + \tau(y_u^2 + \mu_u^2)}} \right\}$$

for the Poisson-inverse Gaussian distribution. Implementation also requires a knowledge of the variance function parameters ν and τ. This is discussed in Section 5.

Focus secondly on targetting the probability of a claim (or at least one claim), denoted by p_u, based on the binomial modelling assumption $N_u | p_u \sim \text{bin}(e_u, p_u)$, with independence over all cells or units u, where again n_u denotes a realisation of the random variable N_u. In the notation of Section 2, the responses $Y_i \equiv N_u$ with mean $\mu_u = E(N_u) = e_u p_u$, variance function $V(\mu_u) = \mu_u\left(1 - \frac{\mu_u}{e_u}\right)$, scale parameter $\phi = 1$ and log-likelihood

$$l = \sum_{u=1}^{n} \left\{ n_u \log\left(\frac{\mu_u}{e_u}\right) + (e_u - n_u)\log\left(\frac{e_u - \mu_u}{e_u}\right)\right\} + \text{constant}.$$

The canonical log-odds or logit link

$$\eta_u = \log\left(\frac{\mu_u}{e_u - \mu_u}\right) = \log\left(\frac{p_u}{1 - p_u}\right) \Leftrightarrow p_u = \frac{e^{\eta_u}}{1 + e^{\eta_u}}$$

with linear predictor

$$\eta_u = \sum_j x_{uj}\beta_j$$

is likely to be of interest in a non-life claim frequency context, while its application in this context would appear to be somewhat limited. An application of its use in targetting the probability of at least one claim in the context of (Belgium) car insurance claims is given by BEIRLANT et al. (1991). A number of researchers, including COUTTS (1984), have used this predictor-link structure to target claim proportions, over a network of cells but with estimation by weighted least squares. The binomial modelling distribution assumption, used in conjunction with the logit and other link functions, has wide application in the construction of life tables, RENSHAW (1991).

4. CLAIM SEVERITY

Claim severity or loss distributions, defined on the positive real line, are invariably positively skewed. There is an extensive literature, see for example, HOGG and KLUGMAN (1984), documenting the modelling of homogeneous batches of empirical claim amounts by specific parameterised distributions. These include the gamma, Pareto, log-normal, log-gamma and Weibull distributions only the first of which is of the type defined by expression (2.1). HABERMAN & RENSHAW (1989) have indicated how certain loss distributions, not of the type defined by expression (2.1), may be fitted in the absence of covariates by the adaption of the associated likelihood function in a special way. Here we address the question: which loss distributions are capable of sustaining covariates?

Mean claim amounts are denoted throughout by x_u, categorised over a set of units $\{u\}$. Thus data summaries take the form (u, n_u, x_u) where x_u denotes the claim average in cell u based on n_u claims. The independence of the number n_u and the claim average x_u within each unit u is assumed. Again typically the units $u \equiv (i_1, i_2, i_3, \ldots)$, a cross-classified grid of cells defined for preselected levels of appropriate covariates, often rating factors. Denoting the underlying expected claim severity in cell u by μ_u and assuming the independence of individual claim amounts, the cells means X_u are modelled as the responses of a GLM with $E(X_u) = \mu_u$ and $\mathrm{Var}(X_u) = \phi V(\mu_u)/n_u$. Covariates defined on $\{u\}$ enter through a linear predictor, linked to the mean μ_u.

Focus first on the gamma distribution. Precedence for its use in this context is to be found in McCullagh and Nelder (1983 & 1989) in which a re-analysis of the celebrated car insurance data of Baxter, Coutts and Ross (1979) is presented. The data comprise (u, n_u, x_u) the number n_u and average cost of claims x_u, cross-classified by policy holder's age, car group and vehicle age. Modelling is based on independent gamma distributed individual claim amounts, for which the mean responses, X_u, satisfy

$$\text{mean } \mu_u = E(X_u), \text{ variance function } V(\mu_u) = \mu_u^2, \text{ scale parameter } \phi = \nu^{-1} > 0$$

with weights n_u so that $\mathrm{Var}(X_u) = \phi\mu_u^2/n_u$. In the analysis, a linear predictor η_u, composed of the additive main effects of the three factors only, is linked to μ_u through the canonical reciprocal link function. Factor interaction terms are found not to be significant. Use of the reciprocal-link function, a member of the parameterised family of power link function

$$\mu_u^\gamma = \eta_u \tag{4.1}$$

with $\gamma = -1$, is justified on the basis of the deviance profile constructed by allowing for incremental changes in γ. The model proposed by Mack (1991) for rating automobile insurance makes identical distributional assumptions to these (formulated in terms of cell sums rather than in terms of cell averages) but restricts the modelling to the log-link, the limiting form of the power link as γ tends to zero in order to focus on a multiplicative structure. The detail is presented in terms of two rating factors so that $u \equiv (i, j)$ with model structure

$$\log(\mu_{ij}) = \alpha_i + \beta_j;$$

while the maximum likelihood parameter estimating equations discussed by Mack (1991) are special cases of equations (2.10). Implementation is readily achieved using the GLIM software package. Mack (1991) also goes on to apply the same model structure in the claims reserving context. Brockman & Wright (1992) use the identical model structure to Mack (1991) in their analysis of the severity of motor claims data.

Focus next on the Pareto distribution with density

$$f(x \mid \beta, \lambda) = \frac{\beta\lambda^\beta}{(\lambda + x)^{\beta+1}}, \qquad x > 0$$

and parameters $\beta, \lambda > 0$. It follows that

$$E(X) = \frac{\lambda}{\beta - 1}, \qquad \mathrm{Var}(X) = \frac{\beta\lambda^2}{(\beta-1)^2(\beta-2)}$$

provided $\beta > 2$. Introducing the reparameterisation

$$\mu = \frac{\lambda}{\beta - 1}, \qquad \phi = \frac{\beta}{\beta - 2};$$

a 1 : 1 mapping $(\beta, \lambda) \mapsto (\mu, \phi)$ with domain $\boldsymbol{R}_{>2} \times \boldsymbol{R}_{>0}$ and image set $\boldsymbol{R}_{>0} \times \boldsymbol{R}_{>1}$, implies that we can construct a GLM based on independent Pareto distributed claim amounts for which the mean responses, X_u, satisfy

mean $\mu_u = E(X_u)$, variance function $V(\mu_u) = \mu_u^2$, scale parameter $\phi > 1$

and weights n_u so that again $\mathrm{Var}(X_u) = \phi \mu_u^2 / n_u$. Apart from the mild extra constraint on the scale parameter, these details are identical to those of the GLM based on independent gamma responses, and the two different modelling assumptions lead to essentially identical GLMs. They differ only in respect of the nature of the parameter estimation criterion, maximum likelihood in the case of the gamma response model and maximum quasi-likelihood in the case of the Pareto response model.

Focus next on the generalization of these distributional assumptions through the introduction of the parameterised power variance function

$$V(\mu) = V(\mu, \zeta) = \mu^{\zeta}. \tag{4.2}$$

The gamma and Pareto distributions are essentially identical special cases with $\zeta = 2$. The characteristics of this family of distributions are discussed in detail by, for example, JORGENSEN (1987). Simplifying the notation slightly by writing equation (2.4) as

$$\mu = \mu(\theta) = b'(\theta) \qquad \text{with inverse} \qquad \theta = \mu^{-1}(\mu),$$

it follows that the cumulant function $b(.)$, corresponding to a specific variance function $V(.)$, is determined by solving the equations

$$\frac{db}{d\theta} = \mu(\theta), \qquad \frac{d}{d\mu} \mu^{-1}(\mu) = \frac{1}{V(\mu)}.$$

First, the solution of the second of these equations determines $\mu^{-1}(.)$. This is then inverted to provide the expression for the right hand side of the first equation, which is then solved in turn for $b(.)$. For the power variance function defined by equation (4.2), the special solution of these equation obtained by setting the arbitrary constant of integration to zero is given by

$$b(\theta) = \begin{cases} \exp(\theta) & \zeta = 1 \\ \dfrac{\alpha - 1}{\alpha} \left(\dfrac{\theta}{\alpha - 1} \right)^{\alpha} & \zeta \neq 1, 2 \\ -\log(-\theta) & \zeta = 2 \end{cases} \tag{4.3}$$

where

$$\alpha = \frac{\zeta - 2}{\zeta - 1}. \tag{4.4}$$

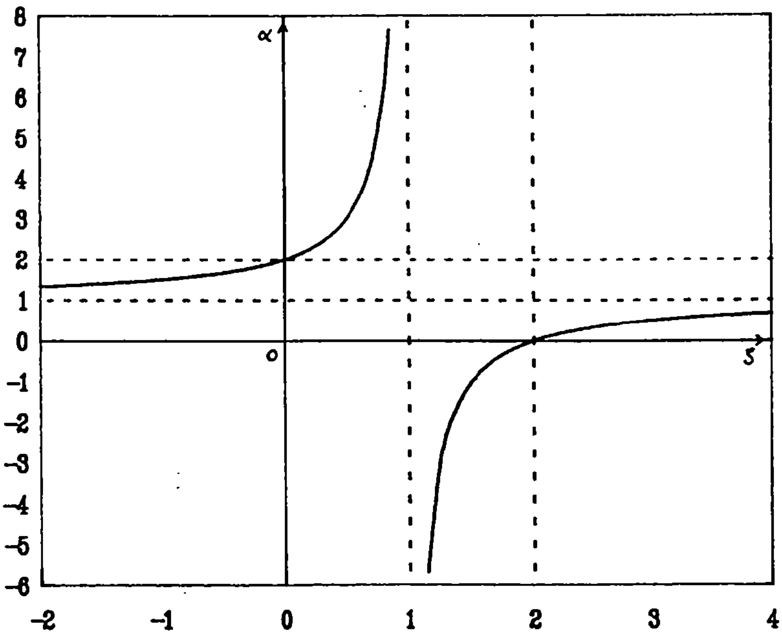


FIGURE 4.1. Graph of α against ζ.

Equation (4.3) characterises the properties of the distribution in question, while equation (4.4) is reproduced graphically in Figure 4.1.

For $\zeta > 2$ $(0 < \alpha < 1)$, $b(\theta)$ is the cumulant function of an extreme stable distribution with index α, see EATON, MORRIS and RUBIN (1971). The cumulant generating function and hence moment generating function is obtained by substituting expression (4.3) into equation (2.2). It generates GLMs with parameterised power variance function, equation (4.2) with $\zeta > 2$; has positive support, and is positive skewed. Equations (4.2), (2.7) and (2.9) determine the coefficient of skewness

$$\zeta \mu_u^{\frac{\zeta}{2}-1} \left(\frac{\phi}{\omega_u} \right)^{1/2} \qquad \zeta > 2 .$$

The family of distributions has therefore all the major characteristics of a loss distribution. It includes the inverse Gaussian distribution ($\zeta = 3$) and has the gamma distribution ($\zeta = 2$) as a limiting case. It represents a generalisation of these two cases. COUTTS (1984) suggests the potential of the two specific cases in the claim severity modelling context. For a given predictor-link structure, the optimum value for $\zeta \geq 2$ is determined by scanning (minus twice) the extended-quasi-likelihood profile, expression (2.15) namely

$$-2q^+ = \sum_{u=1}^{n} \frac{d_u}{\phi} + \sum_{u=1}^{n} \log \{\phi V(y_u)\}$$

for incremental changes in ζ. To compute these values, ϕ is estimated by expression (2.16) and (4.2), while evaluation of the integral in expression (2.14) for the power variance function given by equation (4.2), yields the deviance components

$$d_u = 2\,\omega_u \left\{ -\frac{x_u}{1-\zeta}\,(\mu_u^{1-\zeta} - x_u^{1-\zeta}) + \frac{1}{2-\zeta}\,(\mu_u^{2-\zeta} - x_u^{2-\zeta}) \right\} \qquad \zeta \neq 1, 2 .$$

Implementation is possible using the GLIM software package. McCullagh and Nelder (1989) illustrate the extended-quasi-likelihood profile for the Baxter et al. (1980) car insurance data set, which is optimal in the vicinity of $\zeta = 2.4$. They also demonstrate for these data how contour plots of the extended-quasi-likelihood determine the joint optimum position for the parameters (γ, ζ) when the parameterised power link function (4.1) is used in combination with the parameterised power variance function (4.2).

So far we have dealt with the cases $\zeta \geq 2$. The case $2 > \zeta > 1$ $(\alpha < 0)$ is also of considerable interest but in the slightly different context of aggregate claims. It is discuss by Renshaw (1993). Of the remaining cases, $\zeta = 1$ reproduces the Poisson modelling distribution, $1 > \zeta > 0$ $(\alpha > 2)$ does not generale GLMs with distributions of the type defined by equation (2.1), $\zeta = 0$ generates the normal model, while finally $0 > \zeta$ $(1 < \alpha < 2)$ generates extreme stable distributions with support on the whole of the real line which, for this and other reasons, are of no practical consequence here.

Other claim severity modelling distributions capable of supporting covariates are the log-normal and the log-gamma distributions.

5. AN APPLICATION

The motor insurance claims experience for a recent calendar year, made available by a leading U.K. insurance company, is available for analysis. By way of illustration, the data have been edited to read as follows:

$$(t, u, e_u^{(t)}, n_u^{(t)}, x_u^{(t)})$$

where

t- claim type (1- insured vehicle, 2- third party property, 3- third party injury, 4- others)

$u \equiv (i, j, k, l, m)$ – units or cells based on 5 cross-classified rating factors

$pa : i = 1, 2, 3, 4, 5$ – polyholders age (levels arranged in increasing age bands)

$vt : j = 1, 2, 3, 4, 5$ – vehicle type (levels arranged in perceived order of increasing risk)

$va : k = 1, 2, 3, 4,$ – vehicle age (levels arranged in increasing age bands)

$rd : l = 1, 2, 3, 4, 5$ – rating district (levels arranged in perceived order of increasing risk)

$cd : m = 1, 2, 3, 4,$ – no claims discount (4 levels, arranged in order of increasing discount)

$$e_u^{(t)} - \text{exposures}$$

$$n_u^{(t)} - \text{number of claims}$$

$$x_u^{(t)} - \text{mean claim severity}.$$

The independence of the number and the claim average within each cell for each claim type is assumed. The banding of the rating factors is deliberately left ill defined, and only selective features of the ensuing analysis presented by way of illustration. Other groupings of the rating factors are possible.

```
Residual histogram.
+------------+-----:---------:---------:---------:---------:---------:-+
|[-5.00,-4.50)| 1   |
|[-4.50,-4.00)| 1   |
|[-4.00,-3.50)| 0   |
|[-3.50,-3.00)| 3   |
|[-3.00,-2.50)| 20  |+++
|[-2.50,-2.00)| 34  |+++++
|[-2.00,-1.50)| 110 |++++++++++++++++
|[-1.50,-1.00)| 196 |++++++++++++++++++++++++++++
|[-1.00,-0.50)| 282 |++++++++++++++++++++++++++++++++++++++++
|[-0.50, 0.00)| 347 |++++++++++++++++++++++++++++++++++++++++++++++++++
|[ 0.00, 0.50)| 360 |+++++++++++++++++++++++++++++++++++++++++++++++++++
|[ 0.50, 1.00)| 254 |++++++++++++++++++++++++++++++++++++
|[ 1.00, 1.50)| 179 |++++++++++++++++++++++++++
|[ 1.50, 2.00)| 110 |++++++++++++++++
|[ 2.00, 2.50)| 51  |+++++++
|[ 2.50, 3.00)| 19  |+++
|[ 3.00, 3.50)| 9   |+
|[ 3.50, 4.00)| 3   |
|[ 4.00, 4.50)| 0   |
+------------+-----:---------:---------:---------:---------:---------:-+
                   0        70       140       210       280       350
Residual histogram.
+------------+-----:---------:---------:---------:---------:---------+
|[-5.40,-4.80)| 2   |
|[-4.80,-4.20)| 7   |+
|[-4.20,-3.60)| 16  |++
|[-3.60,-3.00)| 49  |+++++++
|[-3.00,-2.40)| 64  |+++++++++
|[-2.40,-1.80)| 121 |+++++++++++++++++
|[-1.80,-1.20)| 211 |++++++++++++++++++++++++++++++
|[-1.20,-0.60)| 285 |+++++++++++++++++++++++++++++++++++++++++
|[-0.60, 0.00)| 302 |+++++++++++++++++++++++++++++++++++++++++++
|[ 0.00, 0.60)| 334 |++++++++++++++++++++++++++++++++++++++++++++++++
|[ 0.60, 1.20)| 212 |++++++++++++++++++++++++++++++
|[ 1.20, 1.80)| 173 |+++++++++++++++++++++++++
|[ 1.80, 2.40)| 104 |+++++++++++++++
|[ 2.40, 3.00)| 60  |+++++++++
|[ 3.00, 3.60)| 23  |+++
|[ 3.60, 4.20)| 4   |+
|[ 4.20, 4.80)| 7   |+
|[ 4.80, 5.40)| 1   |
|[ 5.40, 6.00)| 0   |
+------------+-----:---------:---------:---------:---------:---------+
                   0        70       140       210       280
```

FIGURE 5.1. Deviance residual histograms. Poisson claim frequency (top), gamma claim severity models.

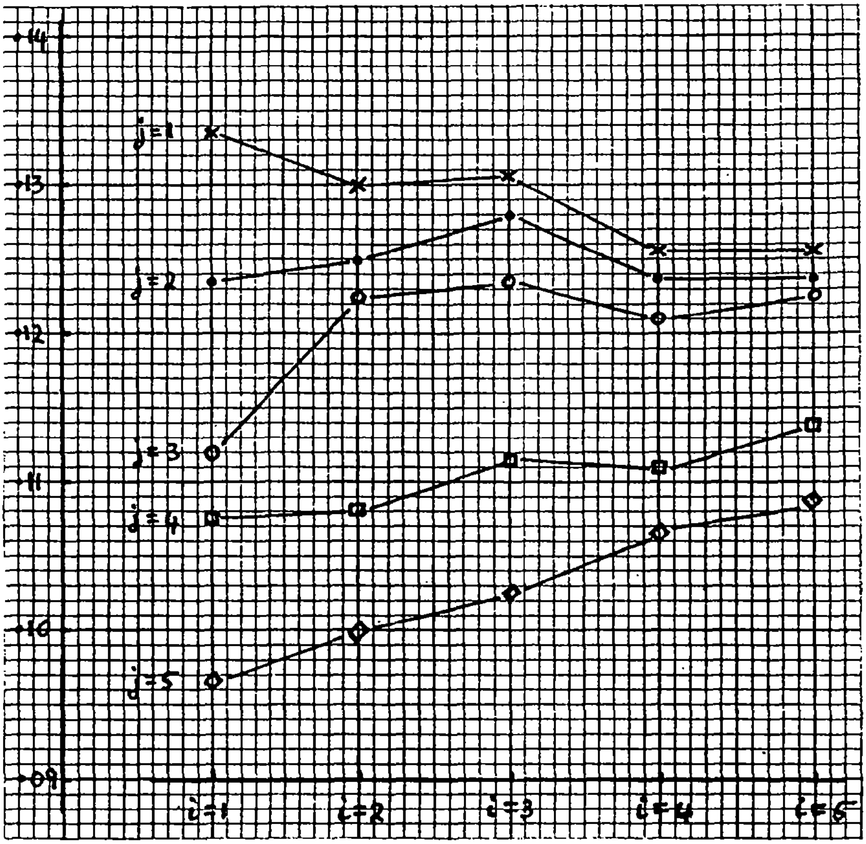


FIGURE 5.2. Contribution of the interactive factors $pa * vt$ to the linear predictor. Claim severity model.

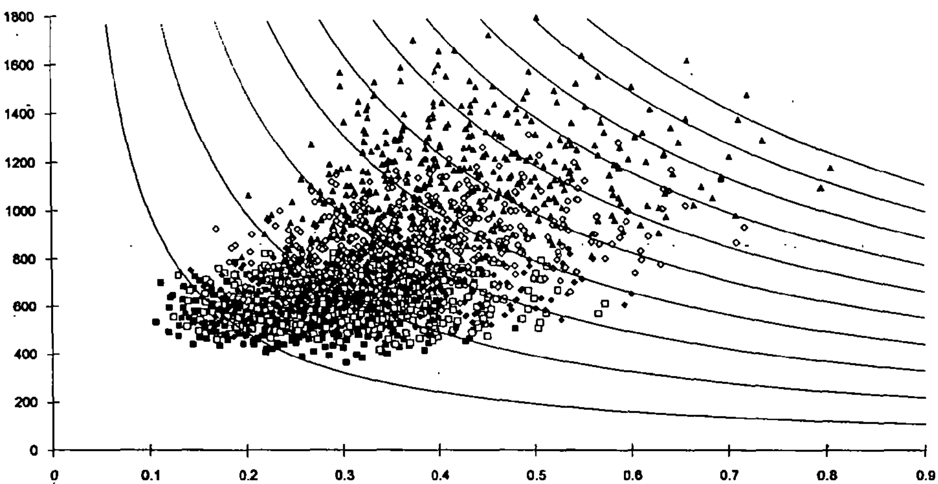


FIGURE 5.3. Expected claim severity $\hat{\mu}_u^{(1)}$ plotted against expected claim rate $\hat{\lambda}_u^{(1)}$. Classification by specific rating factor.

The modelling of the claim frequency and claim severity patterns across units u, for different claim types, provides estimates $\hat{\lambda}_u^{(t)}$ and $\hat{\mu}_u^{(t)}$ of the expected claim rates and expected claim severities respectively. The contribution, $rp_u^{(t)}$, to the risk premium for claim type t is then the product $\hat{\lambda}_u^{(t)}\hat{\mu}_u^{(t)}$ and the risk premium, rp_u, determined by summation over all risk types t. Thus we have the sequences of mappings:

$$(t, u, e_u^{(t)}, n_u^{(t)}, x_u^{(t)}) \mapsto (t, u, \hat{\lambda}_u^{(t)}, \hat{\mu}_u^{(t)}) \mapsto (t, u, rp_u^{(t)}) \mapsto (u, rp_u)$$

where

$$rp_u^{(t)} = \hat{\lambda}_u^{(t)}\hat{\mu}_u^{(t)} \qquad \text{and} \qquad rp_u = \sum_t rp_u^{(t)}.$$

We focus attention on the first mapping which represents the modelling stage and illustrate the application of various of the suggested modelling assumptions for damage to insured vehicle claim types ($t = 1$).

Consider first the Poisson claim frequency and gamma claim severity models, each with log-link functions and predictor structures composed of main effects and paired interaction terms. The improvements in the quality of the fits, monitored by the changes in deviance, as first main effects and then interaction effects are added to the predictor structures are examined. One such sequence of fits is reproduced in Table 5.1. An examination of such tables coupled with an examination of the resulting parameter estimates and their standard errors for each fitted model prompted the adoption of the predictor structure (expressed in GLIM notation)

$$pa * (vt + rd + cd) + va$$

comprising all five main effects and three second order interactions, all involving policyholders age pa, for the claim frequency model structure; and

$$pa * vt + va + rd + cd$$

comprising all five main effects and just one second order interaction term for the claim severity model structure. The various deviance residual plots are also highly supportive of the two fits. By way of illustration, only the deviance residual histograms are reproduced in Figure 5.1. The impact of the single interaction term on the claim severity model structure with parametric representation

$$pa * vt + va + rd + cd: \; \eta_{ijklm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \delta_l + \varepsilon_m$$

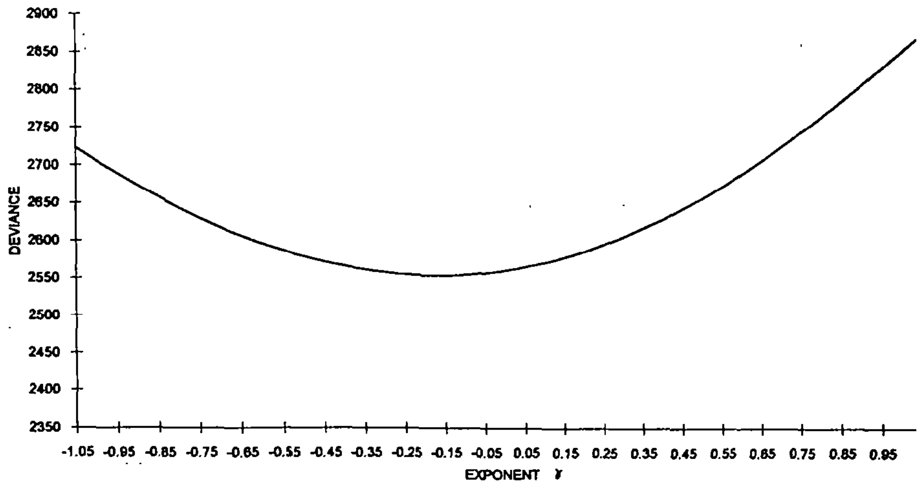
is illustrated in Figure 5.2 in which the estimated values of $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ are plotted against each level i of pa for each level j of vt. Without the interaction terms $(\alpha\beta)_{ij}$ a series of parallel 'lines' would result as the structure is then additive in these factors on the log scale. For this model the expected claim severities are determined by $\mu_{ijklm} = \exp(\eta_{ijklm})$. It can be informative to plot the resulting estimated claim frequencies $\hat{\lambda}_u^{(1)}$ against their corresponding estimated claim severities $\hat{\mu}_u^{(1)}$. One such scatter plot is illustrated in Figure 5.3. In addition the contours displayed represent those of constant risk premium levels, $\hat{\lambda}_u^{(1)}\hat{\mu}_u^{(1)} = \text{constant}$. Here also the impact of the different levels of a rating factor are highlighted by the use of a different symbol to represent each level of this factor. The clustering of the different symbols is informative.

TABLE 5.1

EXAMINATION OF MAIN EFFECTS AND TWO FACTOR INTERACTIONS ON THE DEVIANCES

	claim frequency		claim severity		
	dev.	diff.	dev.	diff.	d.f.
1	15371		12111		
		3295		5268	4
+ *vt*	12076		6842.5		
		4359		253.2	4
+ *pa*	7716.8		6589.3		
		327.3		773.2	3
+ *va*	7389.5		5816.1		
		2039		192.8	4
+ *rd*	5350.3		5623.3		
		2172		572.3	3
+ *cd*	3178.1		5050.9		
		95.57		355.7	16
+ $vt \cdot pa$	3082.5		4695.2		
		26.67		19.3	12
+ $vt \cdot va$	3055.8		4675.9		
		17.85		35.8	16
+ $vt \cdot rd$	3038.0		4640.1		
		70.88		41.3	12
+ $vt \cdot cd$	2967.1		4598.8		
		91.83		31.5	12
+ $pa \cdot va$	2875.3		4567.3		
		69.45		70.15	16
+ $pa \cdot rd$	2805.8		4497.1		
		457.4		35.1	12
+ $pa \cdot cd$	2348.4		4462.0		
		24.79		18.6	12
+ $va \cdot rd$	2323.6		4443.4		
		28.16		22.6	9
+ $va \cdot cd$	2295.4		4420.8		
		42.63		29.9	12
+ $rd \cdot cd$	2252.8		4390.9		

Reverting next to the power-link in combination with the same predictor structures as above, the resulting deviance profiles, constructed over a range of values of the power index, are reproduced in Figure 5.4. For the Poisson claim frequency model, the optimum power is at $\gamma = -0.17$, which is sufficiently close to zero to lend support to the log-link. Indeed if the one remaining paired interaction term involving the rating factor *pa* (and *va*) is added to the predictor structure, then the optimum value of the power is essentially at zero. For the gamma claim severity model the optimum value of the power in the link is at $\gamma = -0.31$, somewhat intermediate between the canonical reciprocal-link and the log-link. If the one interaction term is omitted from the model structure, the optimum power value shifts much closer to be reciprocal-link, a similar conclusion to that reported in McCullagh and Nelder (1989) in their reanalysis of the Baxter et al. (1979) data set involving a main effects structure.



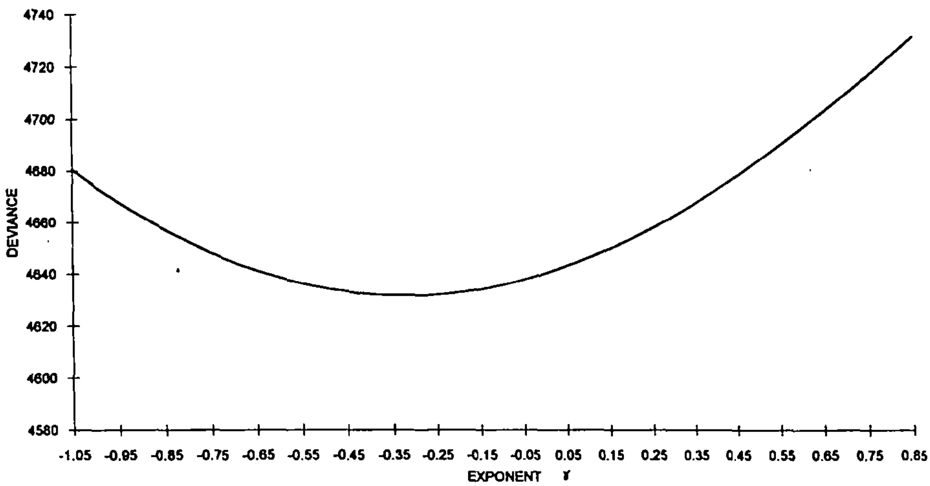


FIGURE 5.4. Deviance profiles, power links. Poisson claim frequency (top), gamma claim severity models.

For the claim severity model with power variance function and exponent $\zeta \geq 2$, in combination with the log-link and the above predictor structure, the deviance profile over $\zeta \geq 2$ has a similar U-shape to Figure 5.4 with a minimum at $\zeta = 2.63$. Thus the optimum positions of both the exponents in the power link and in the power variance function have so-far been chosen separately by allowing for variation in just one of the exponents while keeping the other exponent fixed. The joint optimum positions of the two exponents (γ, ζ) when the power link function (4.1) is used in combination with the power variance function (4.2) is determined by

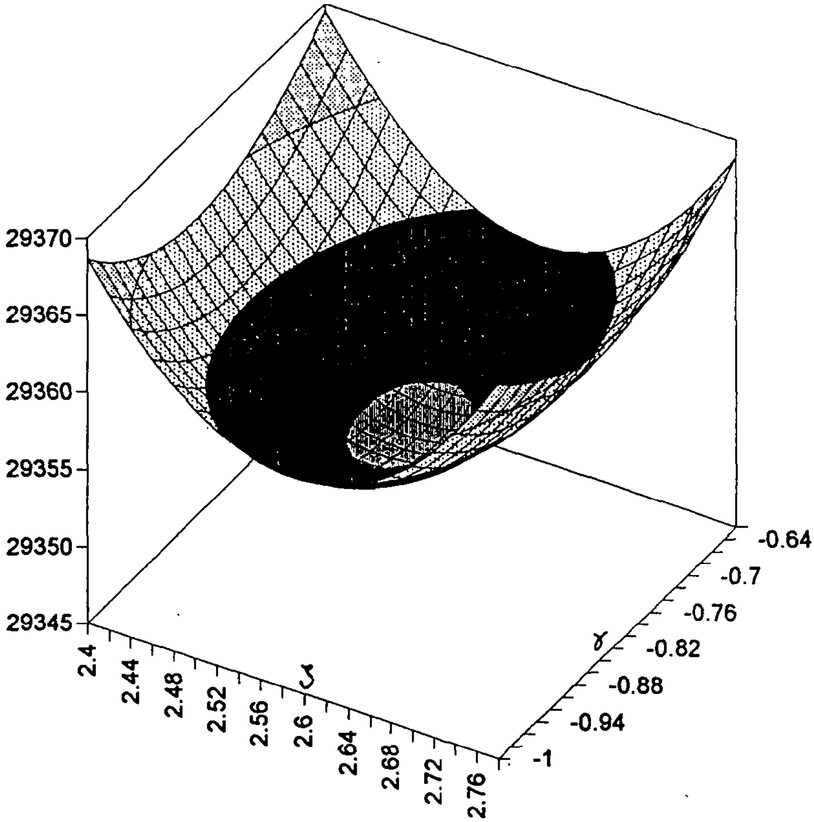


FIGURE 5.5. Extended quasi deviance profile, power link and variance function. Link exponent γ, variance function exponent ζ.

scanning the extended quasi-deviance profile defined by equation (2.15), part of which is reproduced in Figure 5.5 showing an optimum at $(-0.75, 2.54)$.

We revert finally to the introduction of heterogeneity into the claim frequency model through the use of either the Poisson-gamma or Poisson-inverse Gaussian distributions as described in Section 3. Each of the choices involves an unknown parameter, denoted respectively by ν or τ, in the corresponding variance function. One possible strategy for setting the value of the unknown parameter might be to optimise the extended quasi-likelihood but further work is needed in this respect.

6. SUMMARY

Discussion has focused on providing an overview of the variety of response variables available for modelling both the claim frequency and claim severity components of the claims process in general insurance in the presence of rating factors. Working within the rich class of GLMs it is necessary to specify only the first two moments of the associated response variables rather than the full likelihood in order to effect implementation. By this means, it is indicated how suitably selected parameterised variance functions can be used to model heterogeneity in the claim frequency process and to provide a parameterised family of claim response variables which include the gamma response variable as a limiting case. Additional modelling flexibility is achieved through the introduction of the parameterised power link function which includes the log-link as a special case. The salient characteristics in the implementation of these features are illustrated.

REFERENCES

ANDRADE E SILVA, J.M. (1989) An application of Generalized Linear Models to Portuguese Motor Insurance. *Proceedings XXI ASTIN Colloquium, New York*, 633.

BAKER, R.J. and NELDER, J.A. (1985) The GLIM System, Release 3.77, Generalized Linear Interactive Modelling Manual. *Oxford: Numerical Algorithms Group.*

BAXTER, L.A., COUTTS, S.M. and ROSS, G.A.F. (1979) Applications of Linear Models in Motor Insurance. *21st International Congres of Actuaries* **2**, 11.

BEIRLANT, J., DERVEAUX, V., DE MEYER, A.M., GOOVAERTS, M.J., LABIE, E. and MAENHOUDT, B. (1991) Statistical Risk Evaluation Applied to (Belgian) Car Insurance. *Insurance: Mathematics and Economics* **10**, 289.

BERG, M.P. and HABERMAN, S. (1992) Trend Analysis and Prediction Procedure for Time Nonhomogeneous Claim Process. *Submitted.*

BESSON, J.-L. and PARTRAT, Ch. (1992) Trend et Systèmes de Bonus-Malus. *ASTIN Bulletin* **22**, 11.

BOSKOV, M. (1992) Private communications.

BROCKMAN, M.J. and WRIGHT, T.S. (1992) Statistical Motor Rating: Making Effective use of Your Data. *J.I.A.* **119**, 457.

CENTENO, L. and ANDRADE E SILVA, J.M. (1991) Generalized Linear Models under Constraints. *Proceedings XXIII ASTIN Colloquium, Stockholm.*

COUTTS, S.M. (1984) *Motor Insurance Rating, An Actuarial Approach. J.I.A.* **111**, 87.

EATION, M.L., MORRIS, C. and RUBIN, H. (1971) On Extreme Stable Laws and Some Applications. *J. Appl. Prob.* **8**, 794.

HABERMAN, S. and RENSHAW, A.E. (1989) Fitting Loss Distributions using Generalized Linear Models. *Proceedings XXI ASTIN Colloquium, New York*, 149.

HOGG, R.V. and KLUGMAN, S.A. (1984) Loss Distributions. *John Wiley & Sons.*

JORGENSEN, B. (1987) Exponential Dispersion Models (with discussion). *J. R. Statist. Soc. B* **49**, 127.

MACK, T. (1991) A Simple Parameteric model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. *ASTIN Bulletin* **22**, 93.

MCCULLAGH, P. and NELDER, J.A. (1983, 1989) Generalized Linear Models. *Chapman & Hall.*

RENSHAW, A.E. (1990) Graduation by Generalized Linear Modelling Techniques. *Proceedings XXII ASTIN Colloquium, Montreux.*

RENSHAW, A.E. (1991) Actuarial Graduation Practice and Generalized Linear & Non-Linear Models. *J.I.A.* **118**, 295.

RENSHAW, A.E. (1992) Joint Modelling for Actuarial Graduation and Duplicate Policies. *J.I.A.* **119**, 69.

RENSHAW, A.E. (1993) An Application of Exponential Dispersion Models to Premium Rating. *ASTIN Bulletin* **23**, 145.

RENSHAW, A.E. and HABERMAN, S. (1992) Graduations Associated with a Multiple State Model for Permanent Health Insurance. *Under review.*

STROINSKI, K.J. and CURRIE, I.D. (1989) Selection of Variables for Automobile Insurance Rating. *Insurance: Mathematics and Economics* **8**, 35.

TREMBLAY, L. (1992) Using the Poisson Inverse Gauassian in Bonus-Malus Systems. *ASTIN Bulletin* **22**, 97.

ARTHUR E. RENSHAW
Department of Actuarial Science & Statistics,
The City University, London.

WORKSHOP

A COMPARATIVE ANALYSIS OF 30 BONUS-MALUS SYSTEMS

BY JEAN LEMAIRE AND HONGMIN ZI

Wharton School
University of Pennsylvania

ABSTRACT

The automobile third party insurance merit-rating systems of 22 countries are simulated and compared, using as main tools the stationary average premium level, the variability of the policyholders' payments, their elasticity with respect to the claim frequency, and the magnitude of the hunger for bonus. Principal components analysis is used to define an "Index of Toughness" for all systems [1].

1. INTRODUCTION

Most countries have now introduced merit-rating or bonus-malus systems (BMS) in third party liability automobile insurance rating. Such systems penalise policyholders at fault in accidents by surcharges, and reward claim-free years by discounts. This study uses simulation to compare the BMS in force in six East Asian countries (Hong Kong, Japan, South Korea, Malaysia-Singapore, Taiwan, Thailand), fourteen European countries (Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, The Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom), as well as Kenya and Brazil. Several of these countries have recently modified their system. In these cases, both the old and the new BMS are studied, to investigate the impact of the recent modifications.

The regulatory environments in the selected countries are extremely diversified, from total freedom (like in the U.K., where each insurer is free to design its own BMS) to government-imposed systems (like in Switzerland, where all companies have to use the same BMS), with many intermediate situations (Denmark, for instance, where insurers apply BMS rules quite loosely). Obviously the approach to bonus-malus design depends on regulation. If a tariff is imposed by the government and every insurer has to use it, there is no commercial pressure to match the premiums to the risks by making use of every available relevant information. Supervising authorities may choose, for socio-political reasons, to exclude from the tariff structure certain risk factors, even though they may be significantly correlated to losses. The government may then seek to correct for the inadequacies of the *a*

[1] The authors wish to express their most sincere thanks to Azleen Ahmad,Bjorn Ajne, Alain Chevreau, Ted Chung, Freddy Corlier, Thatsanee Dharmpipit, Hans Gerber, Walther Neuhaus, Meha Patel, Danny Quant, Joakim Hertig, Peter Johnson, Gi-Taig Jung, Edward Levay, Harri Lonka, Thomas Mack, Riccardo Ottaviani, Ermanno Pitacco, Shuji Naito, Roberto Westenberger, and Chen Yeh-Lai, who have kindly provided detailed information about the systems in force in their respective countries. Special thanks to Ted Chung, for providing us with extensive loss data, and to Peter Johnson, for thoughtful personal comments.

priori system by using a "tough" BMS. In a free market, carriers need to use a rating structure that matches the premiums to the risks as closely as possible, or at least as closely as the rating structures used by competitors. This entails using virtually every available classification variable correlated to the risks, since failing to do so would mean sacrificing the chance to select against competitors, and incurring the risk of suffering adverse selection by them. Therefore, the use of more *a priori* classification variables is expected in free market countries, which decreases the need for a sophisticated BMS.

Despite these major differences in perspective, the comparison of BMS across countries may prove to be interesting, if only to allow countries to evaluate how "severe" their BMS is, compared to neighbours. This article extends and updates the results of a preceding study (LEMAIRE, 1988a), where 13 BMS were analysed. Two main reasons motivate this update:

(i) Several countries have modified their BMS since 1988, enforcing stiffer penalties in case of claims.

(ii) While the earlier study focussed on insurance companies, the emphasis of the present research is the policyholder. For instance, this research evaluates the evolution of the average premium and its variability, as a function of the policyholder's claim frequency. The earlier study evaluated the insurer's premium income, by introducing a density function for the claim frequencies in the portfolio (the structure function) as well as a model for the number of new insureds and policy terminations. The two approaches lead to very different results. In most countries the constant flow of new drivers subsidies existing policies; the average premium level in an open portfolio is higher than the expected premium paid by an average policyholder.

All BMS are summarised in the Appendix. Section 2 presents the tools used in the analysis: the relative stationary average premium level, the coefficient of variation of premiums, as a function of time and claim frequency, the elasticity of premiums with respect to claim frequency, and the average claim retention to avoid future surcharges. There is a significant positive correlation between these measures. In Section 3, factor analysis is used to summarise the data, and define an "Index of Toughness" for all systems, as the score along the first principal component. Comments for some BMS are found in Section 4.

2. Tools for the Comparison of the Systems

All BMS were simulated, assuming that the number of at-fault claims for a given policyholder conforms to a Poisson distribution, with parameter λ. All values of λ between 0 and 1 were considered. In many countries, the average claim frequency in a typical portfolio is at or below 10%. This average value was selected as benchmark for summary presentations.

In a few countries, the starting class in the BMS depends on exogenous variables like the age of the driver, or the annual mileage of the car. All simulations were performed assuming a new policyholder, driving annually less than 15,000 kilometres in a passenger car, without business use. Assumptions specific to single countries are described in the Appendix.

Tool #1: The Relative Stationary Average Premium Level

An apparently inescapable consequence of the implementation of a BMS is a progressive decrease of the observed average premium level, due to a concentration of policyholders in the high-discount classes. With claim frequencies averaging 10% or less, it would be necessary to penalise each claim by nine classes to maintain a balanced distribution of policyholders among the classes. Because such severe penalties seem commercially impossible to enforce, most policies tend to cluster in the lowest BMS classes.

For all systems, the average premium level of a policyholder with claim frequency 10% was simulated for 30 years, the maximum period most BMS seem to take a reach stationarity. Figure 1 presents the evolution of the mean premium

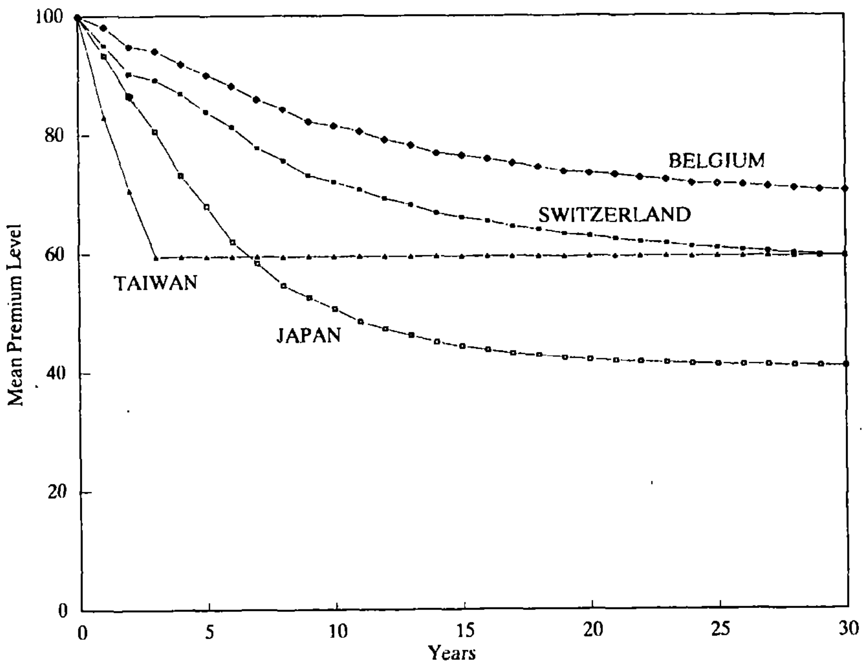


FIGURE 1. Evolution of mean level.

level for the selected countries. For a simple system like the Taiwanese, the premium decreases abruptly in the first few years, the time it takes for the best policyholders to reach the highest discount. The system then stabilises rapidly. For the more "sophisticated" systems the premium decreases in a much smoother way, and the steady state is not reached until at least 30 years have elapsed.

Given the wide variety of systems in force, stationary average levels are difficult to compare. Therefore, a "Relative Stationary Average Level" (RSAL) was defined as

$$\mathrm{RSAL} = \frac{\text{stationary average level} - \text{minimum level}}{\text{maximum level} - \text{minimum level}}$$

Expressed as a percentage, this is an index that determines the relative position of the average policyholder, when the lowest premium is set equal to zero and the highest to 100. A low value of RSAL indicates a high clustering of policies in the lowest BMS classes. A high RSAL suggests a better spread of policies among classes. Table 1 ranks all systems according to the RSAL. The top three countries on this list have very simple, bonus-only, systems: in case of a claim, a policyholder loses the entire discount accumulated over several years.

TABLE 1

RELATIVE STATIONARY AVERAGE LEVEL FOR ALL SYSTEMS

Rank	Country	RSAL
1	Kenya	28.79%
2	Spain	25.67%
3	Malaysia	21.17%
4	Finland (new)	16.04%
5	Sweden	14.20%
6	Netherlands	11.78%
7	U.K. (protected)	11.37%
8	Taiwan	9.55%
9	Finland (old)	8.46%
10	Hong Kong	8.35%
11	Thailand	8.03%
12	U.K. (unprotected)	7.07%
13	Portugal	6.75%
14	Norway (old)	6.61%
15	Switzerland (new)	6.47%
16	Germany (new)	5.85%
17	Japan (new)	4.63%
18	Belgium (new)	4.05%
19	Denmark	3.78%
20	Switzerland (old)	2.90%
21	France	2.12%
22	Norway	2.11%
23	Brazil	1.85%
24	Korea	1.37%
25	Luxembourg (new)	1.36%
26	Italy (new)	1.30%
27	Luxembourg (old)	1.01%
28	Japan (old)	0.88%
29	Belgium (old)	0.74%
30	Italy (old)	0.01%

Note: In theory, the value of the RSAL cannot be computed for Norway, as there is no maximum premium level. In practice, however, very few policyholders have more than three claims in a given year: the probability that a driver with claim frequency 0.10 has 4 or more claims in a year is 3.8×10^{-6}. Therefore, high-malus classes in Norway are very sparsely populated, all the more so as malus evasion seems to be tolerated by insurers. It was therefore assumed that no driver can have a malus exceeding three claims above starting level.

All BMS carry an implicit penalty for new drivers, since the premium level of the access class is substantially higher than the average stationary premium level. Table 2 ranks all systems according to this first-year surcharge.

TABLE 2
IMPLICIT SURCHARGE FOR NEWCOMERS

Rank	Country	Surcharge
1	Germany (new)	+212.97%
2	Norway (new)	+195.80%
3	Denmark	+189.50%
4	Norway (old)	+159.13%
5	Sweden	+158.89%
6	Netherlands	+146.29%
7	Japan (old)	+144.12%
8	Finland (old)	+143.39%
9	Finland (new)	+142.57%
10	Korea	+135.51%
11	Hong Kong	+122.04%
12	Japan (new)	+121.76%
13	Italy (new)	+121.38%
14	Luxembourg (new)	+100.89%
15	U.K. (unprotected)	+ 98.75%
16	Switzerland (old)	+ 94.10%
17	Luxembourg (old)	+ 92.25%
18	U.K. (protected)	+ 84.65%
19	France	+ 77.55%
20	Malaysia	+ 76.65%
21	Kenya	+ 74.60%
22	Taiwan	+ 68.20%
23	Switzerland (new)	+ 67.88%
24	Italy (old)	+ 64.26%
25	Brazil	+ 52.33%
26	Thailand	+ 50.55%
27	Belgium (new)	+ 41.87%
28	Belgium (old)	+ 39.26%
29	Spain	+ 28.70%
30	Portugal	+ 26.95%

Several countries at the bottom of the list have, in addition to these implicit increases, explicit penalties for inexperienced drivers: surcharges in France, a deductible after a claim in Belgium and Switzerland. Of course the implicit surcharge for new drivers is not related to the overall toughness of the system; it is a measure of the degree of cross-subsidization between young and experienced drivers.

Tool #2: The coefficient of variation of the insured's premiums

Insurance consists in a transfer of risk from the policyholder to the carrier. Without experience rating, the transfer is total (perfect solidarity): the variability of insureds' payments is zero. With experience rating, personalised premiums from the policyholder will vary from year to year according to claims history; cooperation between drivers is weakened. Solidarity between policyholders can be evaluated by a measure of the variability of annual premiums. The coefficient of variation (standard deviation divided by mean) was selected, as it is a dimension-less parameter. There is thus no need for currency conversions.

The Actuarial Institute of the Republic of China kindly provided us with market-wide observed loss distributions, property damage and bodily injury, for accident years 1987 to 1989. These distributions are very well represented by a Log-normal model (LEMAIRE, 1993). Assuming that the aggregate claims process is Compound Poisson with Log-normal severities (BOWERS et al., 1986, chapter 11), its coefficient of variation is found to average 6.40. While loss distributions in other countries of course differ from the Taiwanese experience, the coefficient of variation is not likely to be affected much.

Table 3 ranks all countries according to the stationary coefficient of variation of payments, for a policyholder with claim frequency 0.10. These figures are divided by 6.40 in the last column, to indicate the percentage of the original coefficient of variation retained by the policyholder. They show that, even for the most severe systems, insureds are only asked to carry a small part of the variability of the process, 7.18% for the new Swiss system, on top of the list.

TABLE 3

COEFFICIENT OF VARIATION OF PREMIUMS

Rank	Country	Coef. of variation	Percentage retained
1	Switzerland (new)	0.4595	7.18%
2	Norway (old)	0.3900	6.09%
3	Kenya	0.3835	5.99%
4	Finland (new)	0.3834	5.99%
5	Sweden	0.3769	5.89%
6	Netherlands	0.3523	5.50%
7	Japan (new)	0.3283	5.13%
8	Taiwan	0.3162	4.94%
9	Malaysia	0.3075	4.80%
10	Denmark	0.3017	4.71%
11	Switzerland (old)	0.2700	4.21%
12	Finland (old)	0.2570	4.02%
13	Germany (new)	0.2536	3.96%
14	Hong Kong	0.2518	3.93%
15	U.K. (unprot)	0.2419	3.78%
16	Luxembourg (new)	0.2147	3.35%
17	Belgium (new)	0.2128	3.32%
18	France	0.2049	3.20%
19	Norway (new)	0.2049	3.20%
20	Portugal	0.1956	3.06%
21	Thailand	0.1925	3.01%
22	Spain	0.1533	2.40%
23	Korea	0.1271	1.99%
24	Japan (old)	0.1261	1.97%
25	U.K. (prot)	0.1260	1.97%
26	Luxembourg (old)	0.1075	1.68%
27	Italy (new)	0.0934	1.46%
28	Belgium (old)	0.0586	0.92%
29	Brazil	0.0304	0.48%
30	Italy (old)	0.0046	0.07%

Figure 2 shows the evolution of the coefficient of variation with time, for a benchmark policyholder, for the selected systems. Typically, the coefficient of variation starts at zero for the first policy year, increases until the best policyholders reach the maximum discount, then decreases until stationarity is reached. Figure 3 shows the coefficient of variation as a function of the claim frequency.

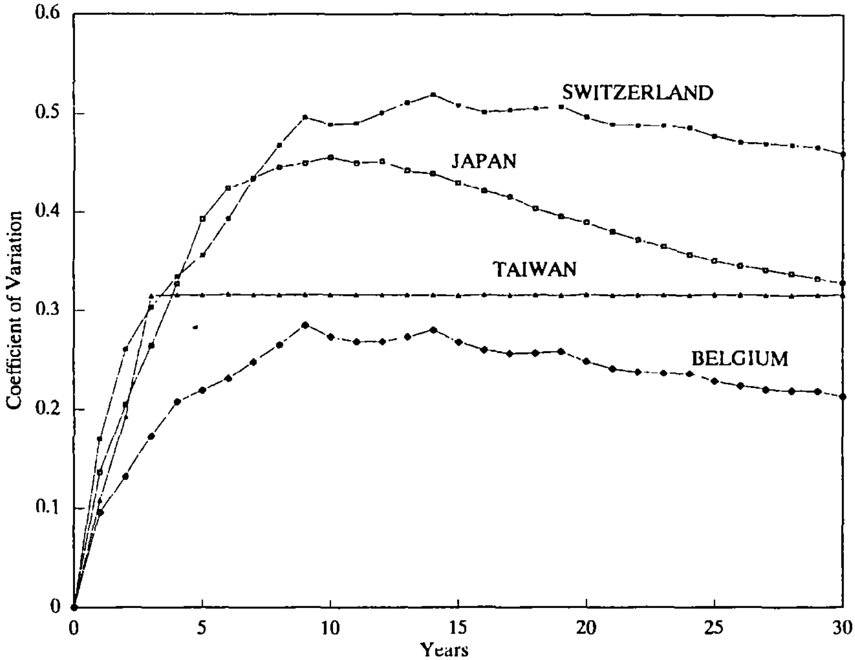


FIGURE 2. Evolution of coefficient of variation.

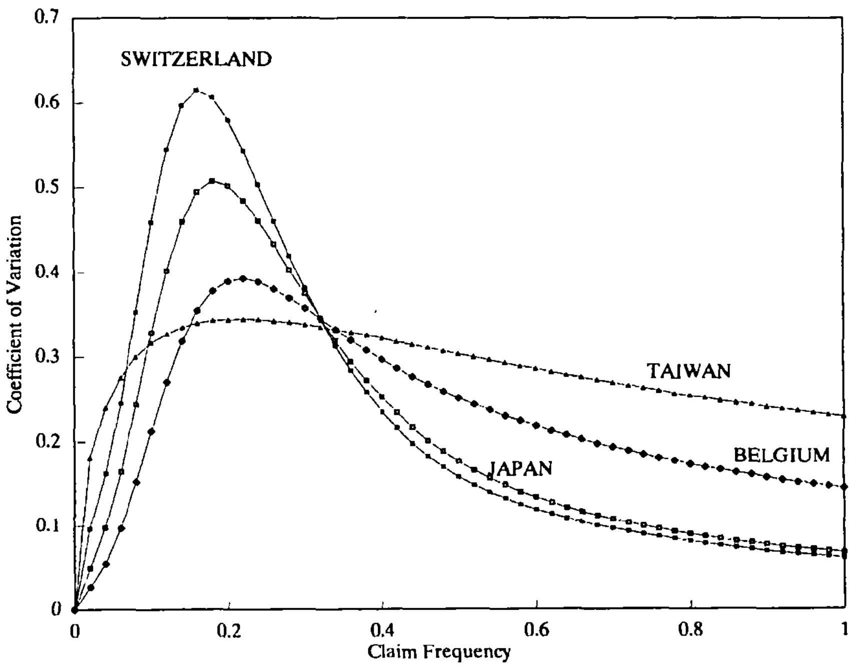


FIGURE 3. Coefficient of variation.

Tool #3 :The efficiency of the bonus-malus system

Consider two policyholders, one with a claim frequency of 0.10, the other with a λ of 0.11. Over a long period of time, the second driver should pay 10% more premiums than the first. A BMS with this property is called perfectly efficient. In practice, however, the mean premium increase will in most cases be much lower than 10%. If the increase is, say, 2% instead of 10%, the system's efficiency is said to be 20%. Denoting $P(\lambda)$ the mean stationary premium for a claim frequency λ, the efficiency $\mu(\lambda)$ of the BMS is defined as

$$\mu(\lambda) = \frac{dP(\lambda)/P(\lambda)}{d\lambda/\lambda}$$

It is the elasticity of the mean stationary premium with respect to the claim frequency: the relative increase of the premium, divided by the relative increase of the claim frequency. It measures the response of the system to a change in the claim frequency. This concept was first introduced in actuarial science by LOIMARANTA (1972).

Ideally, the efficiency should be close to 1 for the most common values of λ. Table 4 indicates the efficiency of all systems for a policyholder with claim

TABLE 4

EFFICIENCY

Rank	Country	Efficiency
1	Switzerland (new)	0.449
2	Finland (new)	0.403
3	Sweden	0.298
4	Netherlands	0.275
5	Norway (old)	0.263
6	Germany (new)	0.257
7	Kenya	0.237
8	Japan (new)	0.232
9	Switzerland (old)	0.208
10	France	0.200
11	Belgium (new)	0.195
12	Finland (old)	0.194
13	Luxembourg (new)	0.183
14	Malaysia	0.165
15	Denmark	0.165
16	Taiwan	0.136
17	Hong Kong	0.133
18	U.K. (unprotected)	0.129
19	Norway (new)	0.127
20	Portugal	0.111
21	Thailand	0.081
22	Spain	0.079
23	Korea	0.078
24	Italy (new)	0.063
25	Luxembourg (old)	0.058
26	Japan (old)	0.052
27	U.K. (protected)	0.051
28	Belgium (old)	0.024
29	Brazil	0.011
30	Italy (old)	0.001

frequency 0.10. On top of the list are countries (Switzerland, Finland, The Netherlands, and Belgium) that have recently modified their BMS, by adopting tougher transition rules. Figure 4 shows the efficiency of the selected systems as a function of λ.

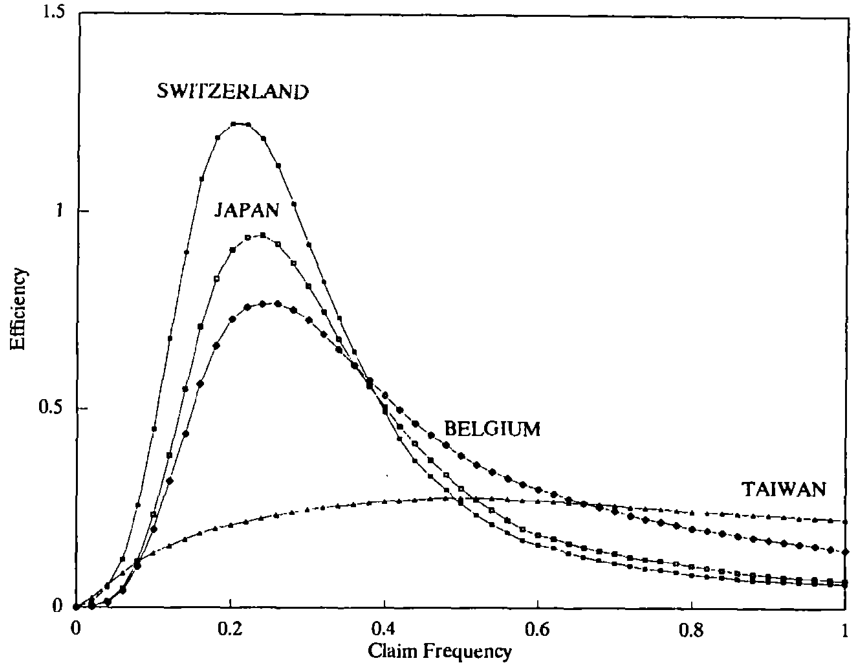


FIGURE 4. Efficiency.

Tool #4: The average optimal retention

A well-known side-effect of BMS is the "hunger for bonus", the tendency of policyholders to pay small claims themselves, and not to report them to their carrier, in order to avoid future premium increases. A severe BMS will of course lead to a large bonus hunger inducement.

The optimal hunger for bonus associated with each BMS can be calculated using an algorithm based on dynamic programming (LEMAIRE, 1985, chapter 18). For each class of the system, the algorithm computes the optimal retention level, the level under which it is the policyholder's interest to not report a claim. Calculations require the following input:

(i) A discount factor, to compare present payments (the claim indemnified) with future savings (surcharges avoided). This factor includes not only inflation, but also policyholders' personal characteristics such as income increase anticipation and impatience rate. The selected factor was 0.90;

(ii) A loss distribution. Since bodily injury claims have to be reported to the police and the insurer, a property damage only distribution should be used here. The 1989 Taiwanese property damage loss distribution can be accurately fitted by a Log-normal distribution, with parameters $\mu = 8.7876$ and $\sigma^2 = 1.3569$. Since five years have elapsed since 1989, and since Taiwanese loss amounts are probably below worldwide averages, a 60% inflation factor was applied. It increases μ to $8.7876 + \ln(1.60) = 9.2576$, while leaving σ^2 unchanged;

(iii) A claim frequency, set at 10%; and

(iv) A conversion factor, that enables the comparison of widely different BMS, and premiums expressed in many different currencies. Since the class at level 100 is situated at quite different positions, premium levels were rescaled by a multiplicative factor, in such a way that the average premium collected, if all claims are reported, is the same for each country. The basic units of Table 5 are such that the average collected premium, using an expense ratio of 40% of the

TABLE 5

AVERAGE OPTIMAL RETENTIONS

Rank	Country	Average optimal retention (Basic units)	Average optimal retention (Percentage of average premium)
1	Taiwan	10,879	315.92%
2	Kenya	6,959	202.08%
3	Finland (new)	6,882	199.84%
4	Norway (old)	6,641	192.85%
5	Switzerland (new)	6,406	186.03%
6	Sweden	5,873	170.26%
7	Netherlands	5,799	168.40%
8	Germany (new)	5,451	158.29%
9	Malaysia	5,032	146.12%
10	Finland (old)	4,915	142.74%
11	Portugal	4,815	139.83%
12	Denmark	4,431	128.58%
13	Hong Kong	3,823	111.01%
14	U.K. (unprotected)	3,818	110.88%
15	Switzerland (old)	3,749	108.87%
16	Norway (new)	3,300	95.83%
17	Belgium (new)	3,001	87.14%
18	Luxembourg (new)	2,886	83.81%
19	Japan (new)	2,791	81.04%
20	Thailand	2,624	76.20%
21	France	2,524	73.28%
22	Spain	2,384	69.21%
23	Korea	2,145	62.28%
24	Luxembourg (old)	1,442	41.87%
25	U.K. (protected)	1,393	40.45%
26	Belgium (old)	1,286	37.34%
27	Italy (new)	1,181	34.28%
28	Japan (old)	712	20.68%
29	Brazil	370	10.74%
30	Italy (old)	19	0.55%

gross premium, is 3,443.6. The knowledge of the average premium effectively collected in each country would then enable the calculation of optimal retentions in that country's currency. Table 5 ranks all systems according to the average optimal retention: the optimal retention for each class is weighted by its stationary class probability. Figures are provided both in basic units and in percentages of the average premium.

3. An Index of Toughness

All four measures defined in Section 2 can be used to evaluate the "mildness" or "toughness" of a BMS. A system that penalises claims heavily will exhibit high RSAL, coefficient of variation of premiums, efficiency, and optimal retentions. These four measures, presented in Tables 1, 3, 4 and 5 for a benchmark policyholder, are however highly positively correlated, as shown in Table 6.

TABLE 6

Correlations between the four measures of toughness

	RSAL	Coef. of variation	Efficiency	Average retention
RSAL	1	.4748	.3167	.4813
CV		1	.9009	.8378
Efficiency			1	.6853
Retention				1

Principal components analysis was used to summarise these data. The first principal component, or factor, explains 72.60% of the total variance, the second 18.71%. Correlations between the first two factors and the four variables are indicated in Table 7.

TABLE 7

Factor Pattern — Correlations between variables and factors

	Factor 1	Factor 2
RSAL	.6155	−.7777
Coef. of variation	.9673	−.1591
Efficiency	.8837	−.3428
Average retention	.8993	−.0243

The first principal component is heavily correlated with efficiency, average retention, and the coefficient of variation. It is less correlated with RSAL. It can clearly be used as a measure of the toughness of a BMS, with the coefficient of variation as the best substitute variable for this index. Standardized factor scores for all 30 systems are provided in Table 8. They rank all systems according to "toughness". Obviously, this ranking does not imply any judgment about the

TABLE 8

FIRST FACTOR SCORES FOR ALL SYSTEMS

A measure of toughness

Rank	Country	Factor score
1	Switzerland (new)	1.7917
2	Finland (new)	1.7794
3	Kenya	1.6942
4	Sweden	1.2791
5	Taiwan	1.1585
6	Norway (old)	1.0974
7	Netherlands	1.0610
8	Malaysia	0.7948
9	Germany (new)	0.5044
10	Finland (old)	0.3427
11	Japan (new)	0.2710
12	Denmark	0.1912
13	Switzerland (old)	0.1060
14	Hong Kong	0.0100
15	U.K. (unprotected)	−0.0683
16	Spain	−0.1116
17	Portugal	−0.1339
18	Belgium (new)	−0.1604
19	Luxembourg (new)	−0.2831
20	France	−0.2886
21	Norway (new)	−0.3934
22	Thailand	−0.4754
23	U.K. (protected)	−0.8170
24	Korea	−0.9310
25	Luxembourg (old)	−1.1475
26	Italy (new)	−1.2003
27	Japan (old)	−1.2102
28	Belgium (old)	−1.4146
29	Brazil	−1.6210
30	Italy (old)	−1.8248

quality of the systems. "Tough" is not to be considered as a synonym of "good" (or "bad"). Also, the rankings could have been somewhat different, had another benchmark claim frequency been selected.

Assuming that factor scores are normally distributed, a percentile on the standard normal distribution can be assigned to each system. For instance, the new Swiss system has a factor score situated 1.7917 standard deviations to the right of the mean. That corresponds to percentile 96.36 in the "universe" of BMS.

Factor scores are computed by the formula

$$\text{SCORE} = 0.26255 \times \left(\frac{\text{EFF} - 0.16193}{0.10769}\right) + 0.26719 \times \left(\frac{\text{RET} - 3784.37}{2382.47}\right) + $$

$$+ 0.28739 \times \left(\frac{\text{CV} - 0.23087}{0.11398}\right) + 0.18086 \times \left(\frac{\text{RSAL} - 7.4757}{7.2557}\right)$$

That formula can be applied to rank any BMS not considered among the 30 analyzed here. For instance, the BMS in force in Germany in the early 1980s was not used in the construction of the above formula. It can nevertheless be positioned on Table 8. It has a RSAL of 1.74%, an efficiency of 0.163, an average retention of 2900, and a coefficient of variation of 0.1865. Its factor score is evaluated at -0.3530, which ranks this system 21st on our Index of Toughness.

From Table 7, and the above formula, it is apparent that the RSAL is a mediocre tool to evaluate toughness. This is probably due to the fact that it is strongly influenced by the premium for the upper class, a class which is sparsely populated for the sophisticated systems. Alternative definitions of the RSAL could eliminate the influence of the classes with low occupation, at the expense of some arbitrariness. This would most probably result in a higher ranking of systems with many classes like the Belgian and the Swiss BMS.

An important remark is that the coefficient of variation is very close to the first principal component as a measure of the toughness of a BMS, since the correlation between the two is 0.9673. Calculating the value of the Index of Toughness necessitates the computation of the values taken by the four tools, and their weighted average. Using the coefficient of variation as an alternate measure is much simpler and loses little in accuracy. The rank correlation between the two measures (Tables 3 and 8) is 0.9653.

The most striking conclusion of the study of Table 8 is the position of the second-generation BMS. With the exception of Norway, all the countries that recently changed their system made it much tougher. Switzerland jumps from the 13th to the 1st rank, Finland from the 10th to the 2nd, Japan from the 27th to the 11th, etc.

4. Comments

4.1. Belgium

The old Belgian system, in force since 1971, exemplified the problems faced by insurers using a mild BMS: a strong clustering of policies in the high-discount classes. With only a two-class penalty for the first claim, the system was in fact designed for an average claim frequency of 1/3. The much lower claims frequencies observed since the 1974 first oil shock created an increasing lack of financial balance, with over 75% of the policyholders in one of the three lowest classes in 1983, and less than 1% of insureds in the malus zone. For instance, one company allowed BEF 713 millions in maluses in 1983, while recovering only 3 millions in maluses, thus producing an average discount of 32.84%. This led the Professional Union of Insurance Companies to set up a study group and suggest a new system to the regulatory authorities (see Lemaire, 1988b). The new system was implemented in 1992. It penalises the first claim by 4 classes.

The new system has a special rule, that no policyholder can be in the malus zone after 4 consecutive claim-free years. This makes the BMS non-markovian, as it requires insurers to memorise the past behaviour of the policyholders for three years. The study of the BMS necessitates the subdivision of several classes into four sub-classes, adding a digit specifying the number of consecutive claim-free years (see LEMAIRE, 1985, chapter 17, for a description of the procedure). The impact of the special transition rule is evidenced in Table 9; a driver in class 18.0 (who had an accident last year) has an optimal retention of 288.16% of the average premium. This retention increases to 457.52% for an insured in class 18 with three claim-free years.

TABLE 9

OPTIMAL RETENTIONS — BELGIAN BMS

Class	Optimal retention	Class	Optimal retention
0	38.41%	16.6	254.05%
1	56.50%	16.3	305.99%
2	76.59%	17.7	252.17%
3	98.26%	17.2	296.85%
4	117.80%	17.3	360.03%
5	137.34%	18.0	288.16%
6	156.05%	18.1	326.98%
7	174.03%	18.2	382.01%
		18.3	457.52%
8	190.40%	19.0	257.56%
9	208.83%	19.1	304.64%
10	224.98%	19.2	369.69%
11	239.38%	19.3	457.52%
12	254.56%	20.0	228.29%
13	273.65%	20.1	283.74%
14	285.46%	20.2	359.28%
15	269.02%	21.0	196.03%
		21.1	260.11%
		22	147.31%

The impact of the stronger transition rules is evident in our overall ranking. Belgium moves from the 28th to the 18th place. Still, the new system still has a slightly negative score on the first factor. The new BMS has to be classified as "average".

4.2. Japan

Up to April 1993, Japanese insurers used a BMS that was unique in the world in the sense that any claim involving bodily injury was penalised as two property damage claims. (Korea is the only other country where penalties depend on claim severity). That system was extremely mild, ranking 27th in the "toughness" scale. Once a policyholder had reached the highest discount class, his first claim was not penalised, as the premium level remained at 40. Even two claims in a single year only raised the premium level from 40 to 45. As the penalty for a property damage claim was two classes only, the system was "designed" for claim frequencies around 1/3. The efficiency was extremely high for claims frequencies around 0.33, culminating at 1.165 for $\lambda = 0.29$. The old Japanese BMS was a rare example of an "over-efficient" system, for specific values of λ.

The transition rules are now tougher, and the BMS ranks in 11th place. Table 10 shows that optimal retentions have considerably increased in all but the top upper classes. The simulation predicts a somewhat better spread of policies among the classes, with 61% of all drivers (instead of 74%) eventually occupying class 1.

TABLE 10

OPTIMAL RETENTIONS — JAPANESE BMS

Class	Optimal retention (old)	Stationary class probability (old)	Optimal retention (new)	Stationary class probability (new)
1	5.80%	.7409	20.13%	.6095
2	16.13%	.0794	39.65%	.0608
3	34.95%	.0879	68.36%	.0714
4	64.33%	.0333	113.84%	.0865
5	107.37%	.0283	169.00%	.0382
6	159.73%	.0116	230.80%	.0306
7	216.07%	.0084	294.55%	.0317
8	265.54%	.0040	350.40%	.0205
9	309.07%	.0028	399.33%	.0141
10	347.46%	.0014	437.20%	.0103
11	380.76%	.0009	474.27%	.0085
12	409.51%	.0005	496.78%	.0067
13	423.46%	.0003	508.49%	.0042
14	427.26%	.0002	383.44%	.0030
15	281.77%	.0001	252.43%	.0024
16	137.39%	.0001	123.02%	.0016
Average	20.68%		81.04%	

4.3. Switzerland

In January of 1990, Swiss insurers modified their BMS, keeping all of its former characteristics while adding a penalty class for each claim. This made the Swiss system the toughest system in the world. The impact of the change in the transition rules on optimal retentions and on the stationary distribution of policyholders is shown in Table 11. The decision to enforce a strong BMS was probably influenced by the fact that Swiss insurers are only allowed to use one *a priori* classification variable (the engine displacement, with over 70% of all vehicles in one class), as well as a deductible for young drivers.

4.4. Taiwan

Taiwan has adopted a simple system. Its unique characteristic (shared with Thailand) is that all surcharges are erased after a single claim-free year, and that all discounts are eliminated following a single claim. As a result, optimal retentions are very high in all classes, and Taiwan ranks first in average optimal retention. (For most other countries, retentions can be extremely high, but in sparsely-populated high-malus classes. Low retentions in the best classes results in a lower weighted average retention).

TABLE 11

OPTIMAL RETENTIONS — SWISS BMS

Class	Optimal retention (old)	Stationary class probability (old)	Optimal retention (new)	Stationary class probability (new)
0	68.87%	.6512	98.12%	.5396
1	104.07%	.0648	136.14%	.0489
2	135.67%	.0781	170.16%	.0535
3	164.42%	.0972	200.87%	.0700
4	190.84%	.0250	235.40%	.1084
5	223.00%	.0220	273.13%	.0255
6	259.94%	.0224	314.14%	.0230
7	300.96%	.0156	358.08%	.0207
8	346.17%	.0054	404.68%	.0264
9	386.78%	.0045	446.43%	.0314
10	423.44%	.0047	490.69%	.0079
11	464.83%	.0039	537.17%	.0064
12	510.23%	.0009	585.85%	.0060
13	558.85%	.0010	636.41%	.0090
14	610.66%	.0013	681.26%	.0100
15	656.67%	.0008	716.85%	.0023
16	688.55%	.0002	750.10%	.0020
17	719.47%	.0003	778.53%	.0022
18	746.42%	.0003	629.71%	.0028
19	565.56%	.0002	476.50%	.0023
20	381.43%	.0001	321.01%	.0009
21	189.38%	.0001	159.38%	.0009
Average	108.87%		186.03%	

TABLE 12

OPTIMAL RETENTIONS — TAIWANESE BMS

Class	Optimal retention	Stationary class probability
1	339.72%	.7403
2	339.72%	.0782
3	223.63%	.0862
4	195.00%	.0000
5	195.00%	.0906
6	195.00%	.0046
7	195.00%	.0001
8	195.00%	.0000
9	195.00%	.0000
Average	315.92%	

Another consequence of the strong transition rules is the high variability of the premium for the policyholders with a low claim frequency ($\lambda < 0.10$), who constitute a majority (see Fig. 3).

On all other measures, Taiwan ranks about average. The overall ranking of the system is 5th. The maximum efficiency of 0.278 is low, and is only obtained for a high value of the claim frequency ($\lambda = 0.49$).

REFERENCES

BOWERS, N., HICKMAN, J., GERBER, H., JONES, D. and NESBITT, C. (1986) "Actuarial Mathematics". Society of Actuaries, Schaumburg, Illinois.

LEMAIRE, J. (1985) "Automobile Insurance: Actuarial Models". Kluwer, Boston.

LEMAIRE, J. (1988a) "A comparative Analysis of most European and Japanese Bonus-malus systems". J. of Risk and Insurance LV, 660–681.

LEMAIRE, J. (1988b) "Construction of the New Belgian Motor Third Party Tariff Structure". *ASTIN Bulletin* **18**, 99–112.

LEMAIRE, J. (1993) "Selecting a Fitting Distribution for Taiwanese Automobile Losses". Unpublished manuscript.

LOIMARANTA, K. (1972) "Some Asymptotic Properties of Bonus Systems". *ASTIN Bulletin* **6**, 233–245.

NEUHAUS, W. (1988) "A bonus-malus system in automobile insurance". *Insurance: Mathematics & Economics* 7, 103–112.

APPENDIX

DESCRIPTION OF ALL BONUS-MALUS SYSTEMS

This appendix provides a summary description of all BMS analysed in this paper. For each BMS, we provide the number of classes, all premium levels, the starting levels, and a short description of the transition rules: the number of classes decreased following a claim-free year, and the number of classes increased following claims. Special rules and assumptions are mentioned. A perfectly accurate description of all BMS would necessitate a full presentation of the transition table, and require many more pages. The obvious regulatory trend in most countries is towards more freedom. So it is probable that, by the time this article is published, the BMS described here will co-exist with many other systems.

1-2. BELGIUM —Old system (1971)

* Number of classes: 18
* Levels: 60, 65, 70, 75, 80, 85, 90, 95, 100, 100, 105, 110, 115, 120, 130, 140, 160, 200
* Starting level: 85 for pleasure use and commuting, 100 for business use
* Claim-free: −1. Cannot be above level 100 after 4 consecutive claim-free years.
  First claim: +2. Subsequent claims: +3

New system (1992)

* Number of classes: 23
* Levels: 54, 54, 54, 57, 60, 63, 66, 69, 73, 77, 81, 85, 90, 95, 100, 105, 111, 117, 123, 130, 140, 160, 200
* Starting level: 85 for pleasure use and commuting,100 for business use
* Claim-free: −1. Cannot be above level 100 after 4 consecutive claim-free years.
  First claim: +4. Subsequent claims: +5

3. BRAZIL

* Number of classes: 7
* Levels: 65, 70, 75, 80, 85, 90, 100
* Starting level: 100
* Claim-free: − 1
  Each claim: + 1

4. DENMARK

* Number of classes: 10
* Levels: 30, 40, 50, 60, 70, 80, 90, 100, 120, 150
* Starting level: 100
* Claim-free: − 1
  Each claim: + 2

5-6. FINLAND — Old System

* Number of classes: 14
* Levels: 40, 50, 50, 50, 50, 60, 60, 70, 80, 100, 110, 120, 130, 150
* Starting level: 120
* Claim-free: − 1
  First claim: from + 6 (lowest classes) to + 1 (highest classes)
  Subsequent claims: + 3

New system

* Number of classes: 17
* Levels: 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 100, 100
* Starting level: lowest 100
* Claim-free: − 1
  First claim: + 3 or + 4. Subsequent claim: + 4 or + 5

7. FRANCE

* Number of classes: 351
* Levels: all integers from 50 to 350
* Starting level: 100.
* Claim-free: 5% reduction. Cannot be above level 100 after 2 consecutive claim-free years.
  Each claim: 25% increase, 12.5% if shared responsibility.
* A recent modification is that the first claim of a policyholder who was at the lowest level for at least 3 years is not penalised.

8. GERMANY — Old System

* Number of classes: 18
* Levels: 40, 40, 40, 40, 40, 45, 50, 55, 60, 65, 70, 85, 100, 125, 175, 175, 200, 200

* Starting level: 175, or 125 if licensed for at least three years
* Claim-free: – 1 or to level 100, if more favourable
  Each claim: from + 1 or + 2 (highest levels) to + 4 or + 5 (lowest levels)

9. New System

* Number of Classes: 22
* Levels: 30, 35, 35, 35, 40, 40, 40, 40, 40, 45, 45, 50, 55, 60, 65, 70, 85, 100, 125, 155, 175, 200
* Starting level: 175 or 125, depending on experience and other cars in the same household.
* Claim-free: – 1, except in the upper classes.
  Each claim: from + 1 (upper classes) to + 9 (lowest class)

10. HONG KONG

* Number of classes: 6
* Levels: 40, 50, 60, 70, 80, 100
* Starting level: 100
* Claim-free: – 1
  First claim: + 2 or + 3. Subsequent claims: all discounts lost

11-12. ITALY — Old system

* Number of classes: 13
* Levels: 70, 70, 70, 75, 80, 85, 92, 100, 115, 132, 152, 175, 200
* Starting level: 115
* Claim-free: – 1
  Each claim: + 1

New System (1991)

* Number of classes: 18
* Levels: 50, 53, 56, 59, 62, 66, 70, 74, 78, 82, 88, 94, 100, 115, 130, 150, 175, 200
* Starting level: 115
* Claim-free: – 1
  First claim: + 2. Subsequent claim: + 3

13-14. JAPAN

* Number of classes: 16
* Levels: 40, 40, 40, 42, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150
* Starting level: 100
* Claim-free: – 1

Old System (1984)

Each claim: + 2 Property Damage, + 4 Bodily Injury

New System (1993).

Each claim: + 3

* 12.5% of all claims have bodily injury implications.

15. KENYA

* Number of classes: 7
* Levels: 40, 50, 60, 70, 80, 90, 100
* Starting level: 100
* Claim-free: – 1
  Each claim: all discounts lost

16. KOREA

* Number of classes: 37
* Levels: 40, 45, 50, 55, 60, ..., 210, 215, 220
* Starting level: 100
* Claim-free: the premium level generally decreases by 10. Moving down is however only allowed after 3 claim-free years. The policy cannot be above level 100 after 3 claim-free years.
  Each claim: Property damage claims are penalised by 0.5 or 1 penalty point, depending on the cost. Bodily injury claims are penalised by 1 to 4 points, depending on the type of injury. Serious offenses are assessed supplementary points, up to 3. The premium increase is 10 levels per penalty point, with a few exceptions.
* As data concerning the distribution of injuries were not available, it was assumed that all claims were penalised by one point, by far the most probable value.

17-18. LUXEMBOURG — Old system

* Number of classes: 22
* Levels: 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 100, 105, 110, 115, 120, 130, 140, 160, 180, 200, 225, 250
* Starting level: 100
* Claim-free: – 1. Cannot be above level 100 after 4 consecutive claim-free years
  Each claim: + 2

New system

* Two new classes, at levels 47.5 and 45, have been added.
* Each claim: + 3

19. MALAYSIA — SINGAPORE

* Number of classes: 6
* Levels: 45, 55, 61.67, 70, 75, 100
* Starting level: 100
* Claim-free: – 1
  Each claim: all discounts lost

20. THE NETHERLANDS (1981)

* Number of classes: 14
* Levels: 30, 32.5, 35, 37.5, 40, 45, 50, 55, 60, 70, 80, 90, 100, 120
* Starting level: 70 to 100, depending on age and annual mileage
* Claim-free: – 1
  Each claim: + 3 to + 5

21-22. NORWAY — Old system

* Number of classes: infinite
* Levels: 30, 40, 50, 60, 70, ...
* Starting level: 100
* Claim-free: – 1 or level 120, if more favourable
  First claim: + 2 (highest levels) or + 3 (3 lowest levels).
  Subsequent claims: + 2

New system

Several BMS currently coexist. The following system was launched in 1987 by a leading company (see NEUHAUS, 1988)

* Number of classes: infinite
* Levels: all integers from 25 up
* Starting level: 80, for drivers aged at least 25 insuring their privately owned vehicle. 100 for all others.
* Claim-free: 13% discount.
  Each claim: fixed amount premium increase (NOK 2,500 in 1988). The penalty cannot however exceed 50% of the basic premium. The penalty is reduced by half for the drivers who have had between five and nine consecutive claim-free years at level 25, for their first claim. It is waived for drivers who have had at least ten consecutive years at the 25 level, for their first claim. An extra deductible is enforced if the claimant is at a higher level than 80, prior to the claim.

23. PORTUGAL

* Number of classes: 6
* Levels: 70, 100, 115, 130, 145, 200
* Starting level: 100

* Claim-free: − 1 after two-consecutive claim-free years
  Each claim: + 1

24. SPAIN

* Number of classes: 5
* Levels: 70, 80, 90, 100, 100
* Starting level: highest 100
* Claim-free: − 1
  Each claim: all discounts lost
* The use of this BMS has now been discontinued by most insurers, as complete rating freedom now exists.

25. SWEDEN

* Number of classes: 7
* Levels: 25, 40, 50, 60, 70, 80, 100
* Starting level: 100
* Claim-free: − 1. Level 25 is only awarded after 6 consecutive claim-free years.
  Each claim: + 2
* A fixed premium of SEK 100 (about 10% of the average premium) is not affected by the BMS.

26-27. SWITZERLAND

* Number of classes: 22
* Levels: 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 110, 120, 130, 140, 155, 170, 185, 200, 215, 230, 250, 270
* Starting level: 100
* Claim-free: − 1

Old system

Each claim: + 3

New system (1990)

Each claim: + 4

28. TAIWAN

* Number of classes: 9
* Levels: 50, 65, 80, 100, 110, 120, 130, 140, 150
* Starting level: 100
* Claim-free: − 1 or to level 80, if more favourable
  Claims: if k claims, to level $100 + 10k$

29. THAILAND

* Number of classes: 7
* Levels: 60, 70, 80, 100, 120, 130, 140
* Starting level: 100
* Claim-free: − 1 or to level 80, if more favourable
  First claim: to level 100. Two or more claims: to level 120 or + 1 (least favourable)

30-31. UNITED KINGDOM (Typical BMS)

* Number of classes: 7
* Levels: 33, 40, 45, 55, 65, 75, 100
* Starting level: 75
* Claim-free: − 1
  First claim: + 3 (level 33), + 2 (levels 40 and 45), + 1.
  Subsequent claims: + 2
* As British insurers enjoy complete tariff structure freedom, many BMS coexist. Many insurers have recently introduced "protected discount schemes": policyholders who have reached the maximum discount may elect to pay a surcharge, usually in the [10%-20%] range, to have their entitlement to discount preserved in case of a claim. More than two claims in five years result in disqualification from the protected discount scheme. Both the protected and unprotected forms are analysed.

JEAN LEMAIRE AND HONGMIN ZI
Department of Insurance and Risk Management,
Wharton School, University of Pennsylvania,
3641 Locust Walk, Philadelphia, PA 19104-6218, U.S.A.

ADDITIVITY OF CHAIN-LADDER PROJECTIONS

BY BJÖRN AJNE

Skandia, Stockholm

ABSTRACT

In this paper some results are given on the addivity of chain-ladder projections. Given two claims development triangles, when do their chain-ladder projections add up to the projections of the combined triangle, that is the triangle being the element-wise sum of the two given triangles?

Necessary and sufficient conditions for equality are given. These are of a fairly simply form and are directly connected to the ordinary chain-ladder calculations. In addition, sufficient conditions of the same form are given for inequality between the combined projection vector and the sum of the two original projections vectors.

KEYWORDS

Chain-ladder projections.

1. INTRODUCTION

Consider two claims development triangles C and D. C consists of positive elements $C(i,j)$, where i denotes the accident years and runs from 0 to n. The index j denotes the development year. For each i it runs from 0 to $n-i$. Thus n denotes the calendar year at the end of which the triangle C is observed, the oldest accident year observed being year number zero. For D the same things hold true with $C(i,j)$ exchanged for $D(i,j)$.

The triangles C and D are thought of as corresponding to two different subportfolios. The elements $C(i,j)$ and $D(i,j)$ are thought of as **accumulated** claims data for accident year i at the end of development year j, be it claims numbers or claims payments or payments plus known reserves. Below they are referred to as amounts.

If now we fill out the triangles into full squares using the ordinary chain-ladder method, $C(i,n)$ and $D(i,n)$ will for each accident year i be the projected final accumulated amounts for that year. $C(0,n)$ and $D(0,n)$ are already there, being the final amounts for the base year. Adding $C(i,n)$ and $D(i,n)$ for all i, we will get the projected final accumulated amounts for the combined portfolio.

This, however, we can also get in another way. We can add the two triangles C and D to get a third triangle E with elements $E(i,j)$, being sums of the corresponding $C(i,j)$ and $D(i,j)$. Then we do the chain-ladder on E to obtain projected final accumulated amounts $E(i,n)$ for the combined portfolio.

The purpose of this paper is to study under what circumstances the two methods will give the same result. This is done in Section 3. When these circumstances are not present sufficient conditions will be given for one method to be more prudent than the other one. This is done in Section 4.

The paper is an improved version of a paper presented to the 23rd ASTIN Colloquium (AJNE, 1991) with simpler proofs and somewhat more far-reaching results.

The practical application is rather the opposite way round to that described above. We are given the total portfolio. When should we contemplate dividing it up into subportfolios in order to get more prudent estimates of its final amounts?

In the appendix an illustration is given in the form of four pairs of simple development triangles (C, D).

Among other things, the question of additivity of claims reserving methods is treated in an lecture given by Hans Bühlmann at the 24th ASTIN Colloquium in Cambridge (BÜHLMANN, 1993).

2. SOME CHAIN-LADDER FORMULAS

Let us recall some chain-ladder calculations. We do it for the triangle C, the corresponding being valid for D and E.

Chain-ladder is performed using quotients between accumulated amounts as link-ratios. That is, putting

(1) $$f(j) = \sum_{k=0}^{n-j} C(k,j) \Big/ \sum_{k=0}^{n-j} C(k,j-1) \qquad \text{for} \qquad j = 1 \ldots n$$

we have

(2) $$C(i,n) = C(i,n-i)\, f(n-i+1)\, f(n-i+2) \ldots f(n)$$

The factors $f(j)$ describe the estimated distribution of the claims amounts over the development years, assumed to be one and the same for all accident years in the underlying model. The distribution of the accumulated amounts is given by $U(0), U(1), \ldots, U(n)$ where

(3) $$U(j) = 1/f(j+1)\, f(j+2) \ldots f(n) \qquad \text{for} \qquad j = 0 \ldots (n-1)$$
$$U(n) = 1$$

From (2) and (3) it follows that

(4) $$C(i,n-i) = C(i,n)\, U(n-i)$$

Denote by $C(i)$ the sum of the first $(i+1)$ projected amounts. Also, denote by $C(\cdot,j)$ the jth column sum (in the original triangle) and by $C'(\cdot,j)$ the same sum with the term $C(n-j,j)$ omitted. That is

(5) $$C(\cdot,j) = \sum_{k=0}^{n-j} C(k,j) \qquad C'(\cdot,j) = \sum_{k=0}^{n-j-1} C(k,j)$$

$$C(i) = \sum_{k=0}^{i} C(k, n) \tag{6}$$

By induction it is proved that

$$C(i, n) = C(i-1)\, C(i, n-i)/C'(\cdot, n-i) \tag{7}$$

Formula (7) yields a rapid recursive calculation of the projections $C(i, n)$ for $i = 1 \dots n$. On the author's part it goes back to an observation made by Kjell Andersson (ANDERSSON, 1992).

From (4) and (7) we find

$$C'(\cdot, n-i) = C(i-1)\, U(n-i) \tag{8}$$

$$C(\cdot, n-i) = C(i)\, U(n-i) \tag{9}$$

Formulas (4) and (9) are contained in a theorem by Thomas Mack (MACK, 1991).

3. NECESSARY AND SUFFICIENT CONDITIONS FOR EQUALITY

We now bring all three triangles C, D and E into play. For D and E we use a notation corresponding to (5) and (6) above. The estimated cumulative distribution of claims amounts over development years, corresponding to U for the triangle C, is denoted by V for the triangle D.

Theorem 1: The necessary and sufficient conditions for the chain-ladder projections to be additive,

$$E(i, n) = C(i, n) + D(i, n) \qquad \text{for all} \qquad i,$$

is that for each positive i at least one of the following two equalities (a) and (b) holds true

(a) $U(n-i) = V(n-i)$

(b) $C(i, n)/(C(0, n) + \dots + C(i-1, n)) = D(i, n)/(D(0, n) + \dots + D(i-1, n))$

Proof:

We want to compare $E(i, n)$ with $C(i, n) + D(i, n)$.

For $i = 0$, equality trivially holds as all three entities are then elements of the base triangles.

Now consider the case when i is positive. Applying (7) to $E(i, n)$ and observing that the E-triangle is the sum of the C- and D-triangles, we get

$$E(i, n) = E(i-1)\, (C(i, n-i) + D(i, n-i))/(C'(\cdot, n-i) + D'(\cdot, n-i))$$

We then apply (4) to the numerator and (8) to the denominator to get

$E(i, n) =$
$E(i-1)(C(i,n)\,U(n-i) + D(i,n)\,V(n-i))/(C(i-1)\,U(n-i) + D(i-1)\,V(n-i))$

Dividing through by $C(i, n) + D(i, n)$, and in the right hand member also both multiplying and dividing by $C(i-1) + D(i-1)$, we finally get

(10) $\quad E(i,n)/(C(i,n) + D(i,n)) = Q(i) \times E(i-1)/(C(i-1) + D(i-1))$

where $Q(i)$ is the quotient between

(11) $\quad (C(i,n)\,U(n-i) + D(i,n)\,V(n-i))/(C(i,n) + D(i,n))$

and

(12) $\quad (C(i-1)\,U(n-i) + D(i-1)\,V(n-i))/(C(i-1) + D(i-1))$

The last two expressions are the averages of $U(n-i)$ and $V(n-i)$ using as weights, in the first case $C(i, n)$ and $D(i, n)$, and in the second case $C(i-1)$ and $D(i-1)$. Also remember that

(13) $\quad C(i-1) = C(0,n) + \dots + C(i-1,n)$

(14) $\quad D(i-1) = D(0,n) + \dots + D(i-1,n)$

(15) $\quad E(i-1) = E(0,n) + \dots + E(i-1,n)$

Now the argument begins. First assume that the projections are additive so that

(16) $\quad E(i,n) = C(i,n) + D(i,n) \qquad \text{for all} \qquad i$

Then, from (10), $Q(i) = 1$ for each positive i. According to (11) and (12) this means that either

(17) $\quad U(n-i) = V(n-i)$

or else, according to the interpretation of (11) and (12) as averages,

(18) $\quad C(i,n)/D(i,n) = C(i-1)/D(i-1)$

Conversely, if for each positive i at least one of (17) and (18) is true, then $Q(i) = 1$ and (16) follows by induction from (10) and the fact that (16) is true for $i = 0$.

Condition (18) may be written

(19) $C(i,n)/(C(0,n) + \dots + C(i-1,n)) = D(i,n)/(D(0,n) + \dots + D(i-1,n))$

This finishes the proof.

$C(i, n)$and $D(i, n)$ are our estimated total claims amounts for accident year i for the two subportfolios. We will use either member of (19) as a measure of **the rate of increase** (in claims volume) of the corresponding portfolio at accident year i.

If (17) holds for all i, or if (19) holds for all i, then the sufficient condition of Theorem 1 is fulfilled. We thus have the following two corollaries.

Corollary 1: If the two subportfolios are equally long-tailed, then the chain-ladder projections are additive.

Corollary 2: If the two subportfolios have the same rate of increase for each accident year, then the chain-ladder projections are additive.

4. SUFFICIENT CONDITIONS FOR INEQUALITY

If, instead of (17), we have

$$U(n-i) \leq V(n-i) \qquad \text{for all positive} \qquad i \tag{20}$$

then the subportfolio C will have an estimated accumulated distribution of claims amounts over development years which increases to one at a slower rate than that of D. We will then say that subportfolio C is **at least as long-tailed** as subportfolio D.

If, instead of (19), we have

$$C(i,n)/C(i-1) \leq D(i,n)/D(i-1) \qquad \text{for all positive} \qquad i \tag{21}$$

we will say that D **increases at least as fast** as C.

If this is the case, we will also have

$$C(i,n)/D(i,n) \leq C(i-1)/D(i-1)$$

Theorem 2: If one of two subportfolios is at least as long-tailed as, and increases (in claims volume) at least as fast as, the other one, then the chain-ladder projections of the combined portfolio are less than or equal to the sums of the corresponding projections of the two subportfolios. If one of the subportfolios is at least as long-tailed as the other one, while the latter increases at least as fast as the first one, then the chain-ladder projections of the combined portfolio are greater than or equal to the sums of the corresponding projections of the two subportfolios.

Proof:

If both (20) and (21) are fulfilled, then for the averages (11) and (12), which define the quotient $Q(i)$, we find

1) $U(n-i)$ is less than or equal to $V(n-i)$
2) The weight given to $U(n-i)$ in the numerator is less than or equal to the weight given to it in the denominator.

Thus $Q(i)$ is greater than or equal to one for all positive i. From (10) it then follows by induction that

$$E(i,n)/(C(i,n)+D(i,n)) \geq 1 \qquad \text{for all} \qquad i$$

Arguing in the same way, we see that if C is at the same time more (or equally) long-tailed and faster (or equally) increasing as compared to D, that is (21) with reversed inequality sign and (20) hold true, then

$$E(i,n)/(C(i,n)+D(i,n)) \leq 1 \qquad \text{for all} \qquad i$$

This finishes the proof.

It may be noted that we have introduced only partial orderings between development triangles, in that the inequality signs in (20) and (21) in general may go in opposite directions for different i.

5. CONCLUSIONS

We have given a partial answer to the question of Section 1 on which method to use. The answer is almost self-evident, at least a posteriori. Assume, for instance, that we add together a long-tailed business which decreases in volume and a short-tailed, increasing one. The long-tailed character of the early accident years will give high lag-factors for the later development years. These lag-fators will then grossly overestimate the final amounts for the dominating short-tailed business of the later accident years. That is, the combined method will give the highest projections.

An example in the opposite direction may be a motor comprehensive account where no division is made between third party claims and hull damage claims. If the third party claims take an increasing share of the total business, a separation of the two types of claims into different development triangles would certainly have been desirable from a prudent point of view.

Even if a certain degree of prudence is to be recommended, the goal is not to have as large reserves as possible, but to have as correct reserves as possible. So, in conclusion, the lesson to be learnt from this exercise in the following one.

If one part of a portfolio can be assumed to differ significantly from the rest of the portfolio with respect to both long-tailedness *and* rate of change of the claims volume, that part should be treated separately in making chain-ladder projections. Returning to prudence, this is especially important if it is at the same time more long-tailed and faster increasing than the rest of the portfolio.

REFERENCES

AJNE, B. (1991) *A Note on the Additivity of Chain-ladder Projections*. Paper presented to the 23rd ASTIN Colloquium (Speaker's corner).

ANDERSSON, K. (1992) *A Direttissima in Chain-ladder and the Danger of Direttissimas*. Personal communication.

BUHLMANN, H. (1993) *Claims Reserves: Theory and Practice*. Mimeographed paper.

MACK, T. (1991) A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. *ASTIN Bulletin* **21**, 93–109.

APPENDIX

Below four pairs of simple development triangles (C, D) are exhibited. For the first three pairs, chain-ladder projectins do add. This means that the projected amounts corresponding to the combined triangle E are the sums of the corresponding projections for C and D, in accordance with the results of Section 3.

For the fourth pair, treating C and D separately will give more prudent projections for the combined portfolio. This means that the projected amounts of E are less than or equal to the sums of the corresponding projections for C and D (with inequality sign in at least one place). This is in accordance with one of the two sufficient conditions of Section 4.

In all the cases there are three accident years 0, 1 and 2. These are observed through development years 0 to 2, 0 to 1 and 0 only, respectively. Thus, in the notation of the main paper, $n = 2$. Projected amounts are shown within parentheses. The amounts in the third column are the chain-ladder projections.

Case 1

C			*D*			*E*		
100	200	300	100	250	375	200	450	675
100	300	(450)	100	250	(375)	200	550	(825)
160	(400)	(600)	100	(250)	(375)	260	(650)	(975)

$\text{Proj}\,(E) = \text{Proj}\,(C) + \text{Proj}\,(D)$

In this case C and D are equally long-tailed, the link-ratios (lag-factors) of formula (1) being $f(1) = 2.5$ and $f(2) = 1.5$. So projections add because of Corollary 1 of Section 3.

Case 2

C			*D*			*E*		
100	200	300	10	100	150	110	300	450
100	300	(450)	40	150	(225)	140	450	(675)
260	(650)	(975)	65	(325)	(487.5)	325	(975)	(1462.5)

$\text{Proj}\,(E) = \text{Proj}\,(C) + \text{Proj}\,(D)$

In this case C and D have the same rate of increase (but not the same link ratios), as shown by the fact that the third columns are proportional to each other. So projections add because of Corollary 2 of Section 3.

Case 3

C			D			E		
100	200	300	200	300	450	300	500	750
200	400	(600)	200	300	(450)	400	700	(1050)
300	(600)	(900)	400	(600)	(900)	700	(1200)	(1800)

Proj (E) = Proj (C) + Proj (D)

In this case none of the above-mentioned circumstances are present but additivity follows from Theorem 1 in Section 3. For $i = 1$, the equality (17) is fulfilled, as the link ratio $f(2) = 1.5$ for both C and D, making $U(1) = V(1)$. For $i = 2$, the equality (19) is fulfilled, as the quotient between the third element in column three and the sum of the first two ones is 1 for both C and D.

C is more long-tailed than D, as (20) is fulfilled with strict inequality for $i = 2$. It is also faster increasing than D as (21) is fulfilled with reversed inequality signs and strict inequality for $i = 1$. This illustrates why strict inequalities cannot be introduced in Theorem 2 in Section 4, without adding the rather pointless requirement that the necessary and sufficient condition of Theorem 1 must be fulfilled.

Case 4

C			D			E		
100	250	375	10	100	150	110	350	525
100	250	(375)	40	150	(225)	140	400	(600)
100	(250)	(375)	65	(325)	(487.5)	165	(495)	(742.5)

Proj (E) less than Proj (C) + Proj (D)

This case illustrates a normal use of Theorem 2 in Section 4. D is more long-tailed and faster increasing than C, and there is no equality sign in (21).

BJÖRN AJNE
Skandia, S-103 50 Stockhom.

SHORT CONTRIBUTIONS

DEDUCTIBLES AND THE INVERSE GAUSSIAN DISTRIBUTION

By Peter ter Berg

Interpolis, Tilburg, Netherlands

Keywords

Inverse Gaussian; Censoring; Truncation; Deductibles; Limits; Moments.

1. INTRODUCTION

The calculation of mean claim sizes, in the presence of a deductible, is usually achieved through numerical integration. In case of a Lognormal or Gamma distribution, the quantities of interest can easily be expressed as functions of the *cumulative* distribution function, with modified parameters. This also applies to the F-distribution, where the incomplete Beta function enters the scene; see for instance the appendix in Hogg and Klugman (1984).

The purpose of this paper is to derive an explicit formula for the first two moments of the Inverse Gaussian distribution, in the presence of censoring. For reasons of completeness we also consider truncation of the Inverse Gaussian distribution by an upper limit.

The tractability of the derivation depends in a crucial way on two properties of the Inverse Gaussian distribution. Firstly, the cumulative distribution function of the Inverse Gaussian can be written as a simple function using the Normal probability integral. Secondly, the moment generating function of a *censorized* Inverse Gaussian distribution boils down to an expression containing the cumulative Inverse Gaussian distribution. This manifests itself most clearly in case of life insurance where the quantity of interest is the expectation of a present value. In case of non-life insurance, where the dimension of the Inverse Gaussian random variable is money instead of time, a further step is required: differentiation of the moment generating function.

So, a natural order of this paper is to address ourselves first to the derivation for the life case and afterwards tackling the more laborious derivation for the non-life case.

2. MATHEMATICAL PRELIMINARIES

We denote the Inverse Gaussian density with mean μ and variance μ^2/ϕ by:

$$h(x \mid \mu, \phi) = [\mu\phi/2\pi x^3]^{1/2} \exp\{-\phi(x-\mu)^2/2\mu x\} \tag{2.1}$$

and its cumulative distribution function as:

(2.2) $$H(x \mid \mu, \phi) = N[(x-\mu)\sqrt{\phi/\mu x}] + e^{2\phi} N[-(x+\mu)\sqrt{\phi/\mu x}]$$

where N denotes the Normal probability integral:

$$N(z) = (2\pi)^{-1/2} \int_{-\infty}^{z} \exp(-1/2 t^2)\, dt$$

which can be evaluated by means of expansions such as given in ABRAMOWITZ and STEGUN (1970). Whenever the parameters do not enter explicitely in h or H we will assume these are μ and ϕ.

Observe that $e^{tx} h(x \mid \mu, \phi)$ is proportional with an Inverse Gaussian density:

(2.3) $$\exp(tx)\, h(x \mid \mu, \phi) = \exp(\phi - f)\, h(x \mid m, f)$$

where the auxiliary parameters m and f depend on t:

(2.4) $$m = \mu\phi/f$$
$$f = (\phi^2 - 2t\mu\phi)^{1/2}$$

Alternatively, we may say that the Esscher transform of (2.1) is $h(x \mid m, f)$. Integration of (2.3) over part of the positive axis is tractable using (2.2). Integrating (2.3) over the whole positive axis gives the moment generating function of (2.1) as:

$$E[e^{tX}] = \exp(\phi - f)$$

from which we easily see that the n-fold convolution of (2.1) is again an Inverse Gaussian density:

$$h^{n*}(x \mid \mu, \phi) = h(x \mid n\mu, n\phi)$$

a property which it has in common with the Gamma density and which formed the reason for HADWIGER (1942) to put (2.1) forward as a modelling tool in insurance and demography.

In case of deductibles or limits, this property is lost, however.

3. PRESENT VALUES IN LIFE INSURANCE

Consider a, not necessarily human, life duration X, with density (2.1). A lump sum B will be paid at moment X. With a discount factor $\exp(-\delta)$ the present value V of B at moment $D < X$ is:

(3.1) $$V = B \exp[\delta(D - X)]$$

In case there is an upper limit L for the moment of payment, (3.1) is valid as long as $D < X \leq L$ and for $X > L$, (3.1) is replaced by:

(3.2) $$V = B \exp[\delta(D - L)]$$

The expected value of V^τ, where $\tau = 1$ or 2 is of special economic interest, is then easy to derive. We have:

$$V^\tau = B^\tau \exp[\delta\tau(D - X)] \qquad D < X \leq L \tag{3.3}$$
$$= B^\tau \exp[\delta\tau(D - L)] \qquad L < X$$

Using (2.1-2-3-4) with $t = -\delta\tau$ results in:

$$E[V^\tau] = B^\tau[1 - H(D)]^{-1}\{Q[H(L \mid m, f) - H(D \mid m, f)] + R[1 - H(L)]\} \tag{3.4}$$

where the auxiliary variables Q and R are given by:

$$Q = \exp[\phi - f - tD]$$
$$R = \exp[t(L - D)]$$

Whenever $L \to \infty$, (3.4) simplifies to:

$$E[V^\tau] = B^\tau Q \frac{[1 - H(D \mid m, f)]}{[1 - H(D \mid \mu, \phi)]} \tag{3.5}$$

4. EXPECTED VALUES IN NON-LIFE INSURANCE

Now X represents the size of a monetary loss, which is modified to a claim size Y by a deductible D and a limit L:

$$Y = 0 \qquad X \leq D$$
$$= X - D \qquad D < X \leq L$$
$$= L - D \qquad L < X$$

So, the probability of a nilclaim is given by $H(D)$.

The moment generating function of Y can be written as:

$$M(t) = H(D) + R[1 - H(L)] + Q\{H(L \mid m, f) - H(D \mid m, f)\}$$

In order to derive $E[Y]$ and $E[Y^2]$, we have to differentiate $M(t)$ with respect to t, substituting $t = 0$ afterwards.

The following auxiliary results are helpful in this task:

$$dm/dt = m^2/f$$
$$df/dt = -m$$
$$dQ/dt = Q(m - D)$$
$$dH(z \mid m, f)/dt = 2m\{N[(z - m)\sqrt{f/mz}] - H(z \mid m, f)\}$$
$$dN[(m - z)\sqrt{f/mz}]/dt = z^2 f^{-1} h(z \mid m, f)$$

where z is a dummy variable, which does not depend on t.

After some rewriting, we arrive then at:

$$
\begin{aligned}
M'(t) = {} & R(L-D)\,[1-H(L)] + \\
& + Q(m+D)\,\{H(D\,|\,m,f) - H(L\,|\,m,f)\} + \\
& + 2Qm\,\{N[(m-D)\sqrt{f/mD}\,] - N[(m-L)\sqrt{f/mL}\,]\}
\end{aligned}
$$

$$
\begin{aligned}
M''(t) = {} & R(L-D)^2\,[1-H(L)] + \\
& + Q[m^2 f^{-1} - (m+D)^2]\,\{H(D\,|\,m,f) - H(L\,|\,m,f)\} + \\
& + 2Qmf^{-1}(m-2fD)\,\{N[(m-D)\sqrt{f/mD}\,] - N[(m-L)\sqrt{f/mL}\,]\} + \\
& + 2Qmf^{-1}[D^2 h(D\,|\,m,f) - L^2 h(L\,|\,m,f)]
\end{aligned}
$$

Now the main goal of this paper follows easily by substituting $t = 0$:

$$
\begin{aligned}
\text{(4.1)}\quad E[Y] &= M'(0) \\
&= (L-D)\,[1-H(L)] + \\
&\quad + (\mu+D)\,[H(D) - H(L)] + \\
&\quad + 2\mu\,\{N[(\mu-D)\sqrt{\phi/\mu D}\,] - N[(\mu-L)\sqrt{\phi/\mu L}\,]\}
\end{aligned}
$$

$$
\begin{aligned}
\text{(4.2)}\quad E[Y^2] &= M''(0) \\
&= (L-D)^2\,[1-H(L)] + \\
&\quad + [\mu^2\phi^{-1} - (\mu+D)^2]\,[H(D) - H(L)] + \\
&\quad + 2\mu\phi^{-1}(\mu - 2\phi D)\,\{N[(\mu-D)\sqrt{\phi/\mu D}\,] - N[(\mu-L)\sqrt{\phi/\mu L}\,]\} + \\
&\quad + 2\mu\phi^{-1}[D^2 h(D) - L^2 h(L)]
\end{aligned}
$$

If we let $L \to \infty$, (4.1) simplifies to:

$$
\begin{aligned}
\text{(4.3)}\quad E[Y] &= 2\mu N[(\mu-D)\sqrt{\phi/\mu D}\,] - (\mu+D)\,[1-H(D)] \\
&= (\mu-D)\,N[(\mu-D)\sqrt{\phi/\mu D}\,] + (\mu+D)\,e^{2\phi} N[-(\mu+D)\sqrt{\phi/\mu D}\,]
\end{aligned}
$$

which agrees with formula (15) in CHHIKARA and FOLKS (1977)[1].

The second moment (4.2) simplifies to:

$$
\begin{aligned}
\text{(4.4)}\quad E[Y^2] &= [(\mu+D)^2 - \mu^2\phi^{-1}]\,[1-H(D)] + \\
&\quad + 2\mu\phi^{-1}\{D^2 h(D) + (\mu - 2\phi D)\,N[(\mu-D)\sqrt{\phi/\mu D}\,]\}
\end{aligned}
$$

Whenever interest focusses on moments, conditionally on $X > D$, the formulae (4.1-2-3-4) must be divided by the probability $[1 - H(D)]$.

[1] I came across this reference after completion of this paper. It does not contain an explicit derivation of this result, however.

It is wellknown that deductibles have a loss eliminating effect. At the same time however, the coefficient of variation of the aggregate claim size distribution increases. A clear exposition of these matters can be found in chapter 5 of STERK (1979).

The availability of (4.3) and (4.4) enables a routine illustration of these findings with the Inverse Gaussian distribution.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (Eds.) (1970) *Handbook of Mathematical Functions*. New York: Dover Publications, Inc.

CHHIKARA, R. S. and FOLKS, J. L. (1977) The Inverse Gaussian Distribution as a Lifetime Model. *Technometrics* **19**, 461–468.

HADWIGER, H. (1942) Wahl einer Näherungsfunktion für Verteilungen auf Grund einer Funktionalgleichung. *Blätter für Versicherungsmathematik* **5**, 345–352.

HOGG, R. V. and KLUGMAN, S. A. (1984) *Loss Distributions*. New York: John Wiley & Sons, Inc.

STERK, H. P. (1979) *Selbstbeteiligung unter risikotheoretischen Aspekten*. Karlsruhe: Verlag Versicherungswirtschaft e.V.

PETER TER BERG
Burg. Rauppstraat 24, 5037 MH Tilburg, Netherlands.

A METHOD FOR MODELLING VARYING RUN-OFF EVOLUTIONS IN CLAIMS RESERVING

BY R.J. VERRALL

Department of Actuarial Science and Statistics
City University
London

ABSTRACT

This paper considers the application of state space modelling to the chain ladder linear model in order to allow the run-off parameters to vary with accident year. In the usual application of the chain ladder technique, the development factors are assumed to be the same for each accident year. This implies that the run-off shape does not alter with accident year. This paper shows how this assumption can be relaxed in order to allow a recursive smooth model to be applied, or for large changes in the shape of the run-off curve. It is possible for these changes to be modelled using external inputs, or for a multiprocess model to be used to detect changes in the run-off shape.

KEYWORDS

Chain ladder technique; Kalman filter; linear models; state space models.

1. INTRODUCTION

The claims reserving process is made up of two parts. The first part is the analysis of the data, and the fitting of suitable models. The second part consists of using the results of the modelling process to project future claims experience. This paper is concerned primarily with the model fitting procedure: the use of the models for forecasting will be discussed only briefly. In addition, attention is restricted to the development factors and no attempt is made to provide a comprehensive claims analysis procedure. Thus, this paper shows how to modify the chain ladder linear model when there are indications that the run-off pattern is changing. This change might be gradual or sudden: the value of the method presented here is that it allows the change to be incorporated into the reserving process.

In order to project future claims it is necessary to have as full an understanding of the pattern of claims experience as possible. The chain ladder technique, which is widely used, does not allow the run-off pattern to change from accident year to accident year. It is unlikely that evidence that the run-off pattern has changed will come to light when the chain ladder technique is used in its usual form. The purpose of this paper is to adapt the chain ladder technique to allow the development factors to evolve with accident year.

2. THE CHAIN LADDER LINEAR MODEL

In order to apply the recursive smoothing methods, we first write the chain ladder technique as a linear model. The original reference for this is KREMER (1982) and a useful later reference is RENSHAW (1989). We will not give full details here as these can be found in the above papers and in VERRALL (1989) and VERRALL (1991a).

The chain ladder technique is based on the following model:

$$E[C_{ij} \mid C_{i,j-1}, \ldots, C_{i1}] = \lambda_j C_{i,j-1} \tag{1}$$

where

C_{ij} = cumulative claims in accident year i, development year j.

λ_j = development factor for year j.

Define the incremental claims by Z_{ij} where

$$\begin{aligned} Z_{ij} &= C_{ij} - C_{i,j-1} \qquad j \geq 2 \\ Z_{i1} &= C_{i1} \end{aligned}$$

The logged incremental claims are denoted by Y_{ij} where $Y_{ij} = \log(Z_{ij})$. The chain ladder linear model is

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j \tag{2}$$

with the constraints that $\alpha_1 = \beta_1 = 0$.

Because the model has a row and column effect, the parameters are called the row parameters (α_i) and column parameters (β_j). The following relationship between the column parameters was derived in VERRALL (1991b):

$$\lambda_j = 1 + \frac{e^{\beta_j}}{\sum_{k=1}^{j-1} e^{\beta_k}} \tag{3}$$

The chain ladder technique estimates the development factors by $\tilde{\lambda}_j$, where

$$\tilde{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{i,j-1}} \tag{4}$$

(assuming we have a $n \times n$ run-off triangle).

This can be seen to be a weighted average of the estimate of the development factor for each row, the weights being the cumulative claims in development year $j-1$. The estimates from each row are

$$\frac{C_{1j}}{C_{1,j-1}}, \frac{C_{2j}}{C_{2,j-1}}, \ldots, \frac{C_{n-j+1,j}}{C_{n-j+1,j-1}}$$

and the weights are $C_{1,j-1}, C_{2,j-1}, \ldots, C_{n-j+1,j-1}$.

This gives the estimate $\tilde{\lambda}_j$ in equation (4). The chain ladder technique is often criticised because it does not allow the run-off shape to evolve, as it imposes the same development factors on each row. An alternative model would be

$$E[C_{ij} \mid C_{i,j-1}, \ldots, C_{i1}] = \lambda_{ij} C_{i,j-1} \tag{5}$$

This is obviously unreasonable since there would be far too many parameters. The model which will be studied in this paper lies between these two extremes. It will be assumed that the development parameters are similar from row to row, but not identical. The extension uses a state space model in a similar way to VERRALL (1989). However, that paper did not address the development factors in any detail, and the recursive relationship defined here has not been considered before. The estimates of the development factors in the chain ladder technique will be found from equation (3).

The next section describes the state space model which allows the development factors to vary from row to row.

3. THE STATE SPACE MODEL

This section contains a summary of the state space model which was derived in VERRALL (1989) and shows how to extend it to allow the run-off pattern to evolve stochastically. The data which make up the claims run-off triangle can be regarded as a time series, and in year t the data which are received are

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t-1} \\ \vdots \\ Z_{t,1} \end{bmatrix}$$

The chain ladder linear model, given by equation (2), can be written in matrix form for the data at time t as

$$\underline{Y}_t = F_t \underline{\theta}_t + \underline{e}_t \tag{6}$$

where F_t is the matrix which specifies the model

and $\underline{\theta}_t$ is the parameter vector at time t.

VERRALL (1989) gives the model for the basic chain ladder technique, but it is necessary to extend it to separate the development parameters in each accident year. Thus, it is necessary to differentiate between $\beta_{1,2}$ and $\beta_{2,2}$, where $\beta_{1,2}$ and $\beta_{2,2}$ are the original column parameter β_2, but in rows 1 and 2 respectively. Unfortunately, it is hard (and not helpful) to define the general form of the model at time t, but we can see the way it can be done by considering times $t = 2$ and 3 (say):

$$\begin{bmatrix} Y_{1,2} \\ Y_{2,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_{1,2} \end{bmatrix} + \begin{bmatrix} e_{1,2} \\ e_{2,1} \end{bmatrix} \tag{7}$$

(8)

$$\begin{bmatrix} Y_{1,3} \\ Y_{2,2} \\ Y_{3,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_{1,2} \\ \beta_{2,2} \\ \beta_{1,3} \end{bmatrix} + \begin{bmatrix} e_{1,3} \\ e_{2,1} \\ e_{3,1} \end{bmatrix}$$

These equations, the observation equations, form one part of the state space model. The difference between this model and the standard chain ladder linear model is that the column parameters are not the same in each row. This can be seen by considering two observations in the same column, for example $Y_{2,3}$ and $Y_{3,3}$. The standard model is:

$$Y_{2,3} = \mu + \alpha_2 + \beta_3 + e_{2,3}$$
$$Y_{3,3} = \mu + \alpha_3 + \beta_3 + e_{3,3}$$

and the new model is

$$Y_{2,3} = \mu + \alpha_2 + \beta_{2,3} + e_{2,3}$$
$$Y_{3,3} = \mu + \alpha_3 + \beta_{3,3} + e_{3,3}$$

The state space model connects $\beta_{2,3}$ to $\beta_{3,3}$, but does not insist that $\beta_{3,3} = \beta_{2,3}$. The connection is made by the system equations which, in their most general form, are as follows

(9) $$\underline{\theta}_{t+1} = G_t \underline{\theta}_t + H_t \underline{u}_t + \underline{w}_t .$$

It can be seen that the system equation relates the parameter vector at time $t+1$ to the parameter vector at time t. The matrix G_t governs the exact form of this relationship. The vector $\underline{u}_t$ contains any new parameters at time t which are not related to those at time $t-1$. In this case these will be the column parameters for the new column which is added to the triangle at time t. It is usual to use vague prior distributions for each input $\underline{u}_t$, reflecting the fact that there is unlikely to be information about the parameters before any relevant data are received. $\underline{w}_t$ is a zero-mean stochastic disturbance term.

The model is mostly defined by the form of the matrix G_t. In this paper, it is chosen so that the column parameters evolve in the following way:

Row	Column 2	3	...
2	$\beta_{2,2} = \beta_{1,2} + w_{2,2}$	$\beta_{2,3} = \beta_{1,3} + w_{2,3}$	...
3	$\beta_{3,2} = \beta_{2,2} + w_{3,2}$	$\beta_{3,3} = \beta_{2,3} + w_{3,3}$	
4	$\beta_{4,2} = \beta_{3,2} + w_{4,2}$		
⋮	⋮		

If the stochastic disturbance terms have zero variance, then the column parameters are identical in each row and the model reverts to the basic chain ladder linear model. The larger the variance, the more variation is allowed between the rows. These variances can be chosen by the user. They can be the same for each row and column, or can differ according to prior opinion on changes in the run-off pattern. For example, if there is evidence that the initial rows form a homogeneous group, but that there is then a change in the run-off pattern, larger variances terms can be included to allow this change to be reflected in the column parameters.

It has often been remarked in previous papers that the chain ladder technique is over-parameterised, due to using a separate parameter, α_t, for each row. This criticism is of the opposite form to that made of the development factors. The use of a separate parameter for each row effect makes too little connection between the rows, while the use of identical development factors for each row makes too great a connection between the rows. The usual way to overcome the problem with the row parameters is to use a recursive model for these parameters. This is defined by

$$\alpha_{t+1} = \alpha_t + v_t \tag{10}$$

where v_t is a zero-mean stochastic disturbance term.

Continuing with the illustration for the model at times $t = 2$ and 3, the system equation which relates the parameter vector at time $t = 3$ to that at time $t = 2$ is

$$\begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_{1,2} \\ \beta_{2,2} \\ \beta_{1,3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_{1,2} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} u_3 + \begin{bmatrix} 0 \\ 0 \\ v_2 \\ 0 \\ w_{2,2} \\ 0 \end{bmatrix} \tag{11}$$

We have now defined a state space model which allows the run-off pattern to change from row to row. This model can be fitted using the Kalman Filter, as was described in VERRALL (1989). The next section contains a numerical illustration of this model.

4. NUMERICAL ILLUSTRATION

As a numerical illustration, the data which has been used previously by the author is again used in this paper. It should be emphasised that this is for illustration purposes: the whole range of claims data is varied enough that comprehensive examples covering all possibilities are not feasible. The data is taken from TAYLOR and ASHE (1983).

357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
352118	884021	933894	1183289	445745	320996	527804	266172	425046	
290507	1001799	926219	1016654	750816	146923	495992	280405		
310608	1108250	776189	1562400	272482	352053	206286			
443160	693190	991983	769488	504851	470639				
396132	937085	847498	805037	705960					
440832	847631	1131398	1063269						
359480	1061648	1443370							
376686	986608								
344014									

with exposure factors

610	721	697	621	600	552	543	503	525	420

The exposures for each year of business are divided into the claims data before the analysis is carried out.

We now apply the model given by equation (11). This relates the row parameters recursively, and also allows the development parameters to evolve. Attention will be focussed on the development parameters, as it is the evolution of the run-off shape which is the subject of this paper.

In order to illustrate the effect of the model, the state variances have been chosen as follows.

$$\operatorname{var}(e_{ij}) = 0.116,\ \operatorname{var}(\alpha_t \mid \alpha_{t-1}) = 0.0289,\ \operatorname{var}(\beta_{ij} \mid \beta_{i-1,j}) = 0.01$$

These values have been chosen in line with VERRALL (1989). It should be emphasised that it is possible to estimate these from the data if that is appropriate. Also, they can be varied in order to gauge their effect. They do not have to be constant: sharp changes in the run-off shape can be modelled by putting a larger variance for the development factors at the appropriate point. Table 1 shows the estimates of μ and of the row parameters, α_i. Also shown for comparison purposes are the estimates from the standard model, given by equation (2). Table 2 shows the estimates of the column parameters, β_{ij}, for columns 2 to 10 in rows 1 to 9. The final row in this table shows the estimates from the standard model.

TABLE 1

	Standard Model	State Space Model
Overal Mean	6.106	6.126
Row Parameters	0.194	0.184
	0.149	0.168
	0.153	0.194
	0.299	0.291
	0.412	0.387
	0.508	0.469
	0.673	0.534
	0.495	0.524
	0.602	0.536

TABLE 2

	2	3	4	5	6	7	8	9	10
					State Space Model				
1	0.925	0.886	0.914	0.383	0.025	−0.175	−0.479	−0.074	−1.413
2	0.918	0.895	0.945	0.361	−0.0349	−0.135	−0.461	−0.0628	
3	0.920	0.907	0.964	0.361	−0.0797	−0.130	−0.447		
4	0.918	0.920	0.980	0.332	−0.050	−0.161			
5	0.895	0.942	0.951	0.352	−0.0264				
6	0.894	0.960	0.940	0.375					
7	0.890	0.990	0.944						
8	0.898	1.014							
9	0.897								
					Standard Model				
	0.911	0.939	0.965	0.383	−0.005	−0.118	−0.439	−0.054	−1.393

The column parameters can be converted into the more familiar development factors, using equation (3). For the state space model, this is applied to each row separately. The results are shown in Table 3.

TABLE 3

	2	3	4	5	6	7	8	9	10
				State Space Model					
1	3.522	1.688	1.419	1.174	1.104	1.077	1.053	1.075	1.018
2	3.504	1.698	1.432	1.168	1.097	1.080	1.053	1.076	
3	3.511	1.705	1.438	1.167	1.092	1.080	1.054		
4	3.505	1.716	1.443	1.161	1.094	1.077			
5	3.447	1.744	1.431	1.165	1.097				
6	3.444	1.758	1.423	1.169					
7	3.435	1.783	1.419						
8	3.454	1.799							
9	3.452								
				Standard Model					
	3.488	1.733	1.434	1.169	1.098	1.080	1.054	1.075	1.018

We will concentrate on the estimates of the parameters, and not show their standard errors (although these are available). It can be clearly seen that the development parameters have been allowed to evolve. The first parameter seems to be decreasing, while the second one is increasing. Patterns such as this can give useful insights into the changes in the run-off shape.

5. CONCLUSIONS

A state space model has been suggested which allows the development factors to evolve recursively. The model is not bound by the strong assumption made by the chain ladder technique that the run-off shape is the same for each accident year.

It may not be clear what parameter estimates should be used for forecasting the future development of the triangle. The most sensible estimates would be the latest ones. These are

3.452 1.799 1.419 1.169 1.097 1.077 1.054 1.076 1.018

compared with those of the ordinary chain ladder model:

3.488 1.733 1.434 1.169 1.098 1.080 1.054 1.075 1.018

The advantage of using the estimates from the dynamic model are that they are more likely to reflect the most recent run-off shape, which the best indication of future development. In particular, if large changes have occured in the development parameters, the straightforward estimates may be unreliable. The usual chain ladder technique does not weight the data according to the time since it was received. The first rows have the same effect on the estimates of the development parameters from this point of view as the more recent rows. The dynamic model gives more weight to recent data, by allowing the parameter to evolve.

It would also be straightforward to allow a sudden change in the run-off evolution by allowing the development factors to change suddenly. This can be done by using a suitably large variance for the stochastic disturbances.

ACKNOWLEDGEMENT

This research was performed under E.C. contract SPES-CT91-0063.

REFERENCES

KREMER, E. (1982) IBNR-Claims and the Two-Way Model of ANOVA. *Scand. Act. J.*, **Vol. 1**, 47–55.

RENSHWA, A. E. (1989) Chain Ladder and Interactive Modelling (Claims Reserving and Glim). *Journal of the Institute of Actuaries 1989*, **Vol. 116**, Part. III, 559–587.

TAYLOR, G.C. and ASHE, F.R. (1983) Second Moments of Estimates of Outstanding Claims. *J of Econometrics*, **Vol. 23**, 37–61.

VERRALL R.J. (1989) A State Space Representation of the Chain Ladder Linear Model. *Journal of the Institute of Actuaries 1989*, **Vol. 116**, Part. III, 589–610.

VERRALL, R.J. (1990) Bayes and Empirical Bayes Estimation for the Chain Ladder Model. *ASTIN Bulletin 1990*, **Vol. 20**, No. 2, 217–243.

VERRALL, R.J. (1991a) On the Unbiased Estimation of Reserves from Loglinear Models. *Insurance: Mathematics and Economics 1991*, **Vol. 10**, No. 1, 75–80.

VERRALL, R.J. (1991b) Chain Ladder and Maximum Likelihood. *Journal of the Institute of Actuaries 1991*, **Vol. 118**, Part. III, No. 471, 489–499.

R.J. VERRALL
Department of Actuarial Science and Statistics,
City University, London.

BOOK REVIEW

BJÖRN SUNDT (1993): *An Introduction to Non-Life Insurance Mathematics*. Third edition. Veröffentlichungen des Instituts für Versicherungswissenschaft der Universität Mannheim, Vol. 28, Verlag Versicherungswirtschaft, Karlsruhe, 215 pages, DM 32.—.

In ASTIN Bulletin 1987 we could read the book review on Sundt's first edition. Since then, two further editions have appeared, the third being published in late 1993. Like the preceding this third edition of the book covers the most important topics of non-life insurance mathematics: credibility theory, bonus systems, tariffication, the risk process, accumulated claims distribution, reserves, and utility theory. It is a very properly written introductory textbook into modern risk theory for students as well as for practitioners. Whenever new editions appeared, it resulted in a considerable improvement of the presentation.

In order to document the changes from one edition to the next, we cite from the prefaces (and this reflects indeed what is needed to be said about this edition):

> "The most extensive changes from the first edition of the book are the following: A new Section 6.8 on hierachical credibility and a new Chapter 8 on multiplicative rating models are included. In Chapter 7 on bonus systems, the asymptotic optimality criterion has been replaced with a non asymptotic one. In Chapter 9 a different proof of Lundberg's Inequality is given. To give a better flow in the presentation, the material on moment-generating functions, Laplace transforms, and convex functions has been transferred to appendices. A simple proof of Ohlin's Lemma is given in Appendix A. The discussion on the optimal choice of a compensation function has been extended and presented under more general conditions as it seemed that greater generality could be achieved without complicating the mathematics." (Second edition)
>
> "The most extensive change from the second edition to the third edition is the inclusion of exercises... In the exercises, I have to a large extent included questions where one should comment on assumptions or results. In a practically oriented subject like insurance mathematics, it is important that the material becomes not only mathematics, but that one also continuously considers questions like, what does this imply, are these assumptions realistic, does this result seem reasonable, etc. At an exercise course, such questions could be discussed between the teacher and the students.
>
> The text of the book has been much less changed than was the case with the second edition. Most of the changes have been aimed at simplifying and clarifying, and correcting errors in the second edition. The most important changes are the following: New material includes subsections 5F, 6.4D-E, 7B, 8.3D, 10.2B, 10.6C-D, and 12D, and Appendix A... The material in the old section on ruin theory in Chapter 9 has been reorganized and divided into three sections (Sections 9.4-6)." (Third edition)

There is only one point which could possibly be criticized: In each step, from edition one to two and from two to three, the size of the text and the formulas has been diminished. This together with a somewhat nonstandard style for sub- and superscripts (e.g. on p. 80 below) makes the reading considerably harder.

CHRISTIAN HIPP

ACTUARIAL VACANCY

CONCORDIA UNIVERSITY
DEPARTMENT OF MATHEMATICS & STATISTICS

The Department of Mathematics and Statistics at Concordia University is seeking to fill a tenure track position in Actuarial Mathematics at the Assistant Professor level starting August 15, 1995. Duties will include undergraduate and graduate level teaching, research and participation in the administration and expansion of the Actuarial programs.

Interested candidates should send a curriculum vitae and a brief description of field of interest and arrange for three letters of recommendation to be sent as early as possible to:

Dr. M. Belinsky, Chair
Department of Mathematics & Statistics
Concordia University
7141 Sherbrooke St. W.
Montreal, Quebec
Canada, H4B 1R6

Concordia University is committed to Employment Equity and encourages applications from women, aborginal peoples, visible minorities and disables persons. In accordance with Canadian immigration requirements, priority will be given to Canadian citizens and permanent residents of Canada.

GUIDELINES TO AUTHORS

1. Papers for publication should be sent in quadruplicate to one of the Editors:

Hans Bühlmann,
Mathematik, ETH-Zentrum,
CH-8092 Zürich, Switzerland.

D. Harry Reid,
Eagle Star Insurance Company Ltd,
The Grange, Bishop's Cleeve
Cheltenham Glos GL52 4XX, United Kingdom.

or to one of the Co-Editors:

Alois Gisler,
"Winterthur" Swiss Insurance Company,
P.O. Box 357, CH-8401 Winterthur, Switzerland.

David Wilkie
Messrs. R. Watson & Sons
Watson House, London Rd., Reigate, Surrey RH2 9PQ, 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 author(s), year, title, journal (in italics, cf. point 9), volume (in boldface, cf. point 9), and pages. For book references give author(s), year, title (in italics), publisher, and city.

   Examples:

   Barlow, R.E. and Proschan, F. (1975) *Mathematical Theory of Reliability and Life 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.