2005

Journal of Actuarial Practice, Volume 12, 2005

Colin Ramsay , Editor
University of Nebraska - Lincoln, cramsay@unl.edu

Journal of Actuarial Practice Volume 12, 2005

ARTICLES

Journal of Actuarial Practice

EDITORIAL POLICY

The aim of this international journal is to publish articles pertaining to the "art" and/or "science" involved in contemporary actuarial practice.

The Journal welcomes articles providing new ideas, strategies, or techniques (or articles improving existing ones) that can be used by practicing actuaries. One of the goals of the Journal of Actuarial Practice is to improve communication between the practicing and academic actuarial communities. In addition, the Journal provides a forum for the presentation and discussion of ideas, issues (controversial or otherwise), and methods of interest to actuaries.

The Journal publishes articles in a wide variety of formats, including technical papers, commentaries/opinions, discussions, essays, book reviews, and letters. The technical papers published in the Journal are neither abstract nor esoteric; they are practical and readable. Topics suitable for this journal include the following:

AIDS	financial reporting	pricing issues
annuity products	group insurance	product development
asset-liability matching	health insurance	reinsurance
cash-flow testing	individual risk taking	reserving issues
casualty ratemaking	insurance regulations	risk-based capital
credibility theory	international issues	risk theory
credit insurance	investments	social insurance
disability insurance	liability insurance	solvency issues
expense analysis	loss reserves	taxation
experience studies	marketing	valuation issues
FASB issues	pensions	workers' compensation

REVIEW PROCESS

A paper submitted to the Journal first is screened for suitability. If it is deemed suitable, copies are sent to several independent referees. The name of the author(s) of the paper under consideration is usually anonymous to the referees, and the identities of referees are never revealed to the author(s).

The paper is reviewed for content and clarity of exposition. Papers do not have to contain original ideas to be acceptable. On the basis of the referee reports, the editor makes one of the following decisions: (1) accept subject to minor revisions, (2) accept subject to major revisions, or (3) reject.

The editor sends the author(s) of the decision and along with copies of the referees' reports. The referee process is expected to take three to four months (depending on the length of the paper).

See inside back cover for instructions to authors.

Absalom Press, Inc.
P.O. Box 22098, Lincoln, NE 68542-2098, USA.

Journal of Actuarial Practice ISSN 1064-6647

Risk-Based Regulatory Capital for Insurers: A Case Study

Christian Sutherland-Wong* and Michael Sherris†

Abstract‡

We study the issues in determining regulatory capital requirements using advanced modeling by assessing and comparing capital requirements under the two alternative approaches. A dynamic financial analysis (DFA) model is used for this case study. These issues are of current international interest as regulators, insurers, and actuaries face the significant issues involved with the introduction of risk-based capital for insurers.

Key words and phrases: *insurer solvency, standardized solvency assessment, advanced modeling, dynamic financial analysis*

*Christian Sutherland-Wong, BCom (Hons.), A.I.A.A., graduated from University of New South Wales. His research interests include capital regulation for general insurers. He has worked as an actuarial analyst with Ernst & Young, Tillinghast-Towers Perrin, and Macquarie Bank and is currently working as a management consultant with Bain & Company.

Mr. Sutherland-Wong's address is: Bain & Company; Level 35, Chifley Tower; 2 Chifley Square; Sydney; AUSTRALIA. E-mail: *Christian.Sutherland-Wong@Bain.com*

†Michael Sherris B.A., M.B.A., F.I.A.A., F.I.A., F.S.A., is a professor and head of actuarial studies at University of New South Wales (NSW). He has 20 years experience as an academic and, prior to becoming an academic, worked in the financial services industry in life insurance, superannuation, corporate finance, investment funds management and money market areas. His current research is in quantitative risk management and modeling in insurance, finance and commodity markets including risk-based capital.

Prof. Sherris' address is: Actuarial Studies, Faculty of Commerce and Economics, University of NSW, Sydney, AUSTRALIA. E-mail: *m.sherris@unsw.edu.au*

‡The authors acknowledge the invaluable assistance of Scott Collings and Abhijit Apte, of Trowbridge Deloitte, and of David Koob, Brett Ward, and David Finnis. Sutherland-Wong acknowledges the financial support of APRA and the award of the Brian Gray Scholarship to support this research. Prof. Sherris acknowledges the financial support of Australian Research Council Discovery Grant DP0345036 and the UNSW Actuarial Foundation of The Institute of Actuaries of Australia.

1 Introduction

1.1 Background

Capital is defined as the excess of the value of an insurer's assets over the value of its liabilities. In practice, the value of the assets and liabilities is reported using statutory and regulatory requirements. Regulatory requirements are used for solvency assessment. Methods of determining economic capital have become the focus of insurers in recent years. Regulatory capital requirements for banks and insurers increasingly are becoming risk-based to reflect the economic impact of balance sheet risks. Giese (2003) discusses the concept of economic capital along with the recent developments in economic capital models.

However determined, capital provides a buffer that allows insurers to pay claims even when losses exceed expectations or asset returns fall below expectations. As described by the IAA (International Actuarial Association) Insurer Solvency Assessment Working Party (2004) a level of capital provides, among other things, a "rainy day fund, so when bad things happen, there is money to cover it."

Cummins (1988) and Butsic (1994) discuss the need for regulation in insurance. Butsic (1994) argues that if markets were perfectly efficient, capital regulation would not be necessary. Insurers could determine their own level of capital, and market forces would price premiums depending upon the riskiness of an insurer becoming insolvent. Fully informed consumers would diversify their insurance policies across insurers taking into account the risk of insurer default. Taylor (1995) and Sherris (2003) use economy wide models to explore equilibrium insurance pricing and capitalization. Sherris (2003) shows that in a complete and frictionless market model the level of capital will be reflected in the market price of premiums for insurance and there is no unique optimal level of capital for an insurer.

In reality the complete and perfect markets assumptions do not hold. There is information asymmetry between consumers and insurers. As the costs of insurer insolvency can be significant, insurers do not report their level of default risk even though this is often assessed by rating agencies. For this form of market failure, as described by Frank and Bernanke (2001, pp. 297-312), an efficient way for insurers to demonstrate financial soundness is to meet regulated levels of capital prescribed. This regulatory capital serves as protection for consumers against the adverse effects of insurer insolvency.

Another factor that is important in practice is the existence of government or industry-based guarantee funds that compensate policyholders in the event of insurer insolvency. There is no formal arrangement of guarantee of insurers in Australia. These guarantee funds are taken into account in considering the risk-based capital that insurers hold where they exist. They also may generate moral hazard if the cost of such guarantee funds is not reflected in the premiums charged to insurers. This is an area that is not addressed in detail in this paper. If they were to be included in the model, then the capital requirements would be the requirements after allowing for the government or industry support.

The IAA Insurer Solvency Assessment Working Party has developed a global framework for risk-based capital for insurers. In their 2004 working paper entitled "A Global Framework for Insurer Solvency Assessment," the working party advocates two methodologies for regulatory capital determination: the standard approach and the advanced approach. The standard approach applies industry wide risk factor charges to the calculation of the insurer's capital requirement, while the advanced approach allows insurers to use a dynamic financial analysis (DFA) model to calculate their capital requirement, better reflecting the insurer's risks.

Banks have been increasingly moving to the use of internal models for capital requirements under Basel.[1] Insurers in a number of countries will be faced with similar requirements as regulators adopt a more risk-based capital approach to regulation. Against this background, the issues in implementing risk-based capital are of significant interest to insurers and actuaries at an international level.

1.2 Capital Regulation in Australia

The Australian Prudential Regulation Authority (APRA) is the primary capital regulator of non-life (property and casualty) insurers in Australia. APRA reviewed its approach to regulating non-life insurance companies and recently released a new set of prudential standards. These standards contain a new methodology for determining a non-life insurer's minimum capital requirement. The new capital requirements more closely match regulatory capital to an insurer's risk profile, otherwise known as risk-based capital.[2]

[1]See Basel Committee on Banking Supervision papers "Operational Risk" (January 2001), "Overview of the New Basel Capital Accord" (January 2001), and "Working Paper on the Regulatory Treatment of Operational Risk" (September 2001). Full details are available from the Bank for International Settlements web site <http://bis.org>.

[2]For further information on the background to the APRA general insurance reform, refer to Gray (1999 and 2001) and IAA Insurer Solvency Assessment Working Party (2004).

Non-life insurers are able to calculate the minimum capital required in one of two ways:

> An insurer may choose one of two methods for determining its *minimum capital requirement* (MCR). Insurers with sufficient resources are encouraged to develop an in-house capital measurement model to calculate the MCR (this is referred to as the *internal model based method* (IMB)). Use of this method, however, will be conditional on APRA's and the Treasurer's prior approval and will require insurers to satisfy a range of qualitative and quantitative criteria. Insurers that do not use the IMB method must use the **prescribed method.**[3]

APRA's prescribed method is in line with the standard approach of the IAA Insurer Solvency Assessment Working Party's, while the IMB method is in line with the advanced approach. The solvency benchmark for the new APRA standards is a maximum probability of insolvency in a one year time horizon of 0.5%.

The IAA Insurer Solvency Assessment Working Party considers that the prescribed method should produce a more conservative (higher) value for the minimum capital requirement, as it should determine a minimum level applicable to all insurers licensed to conduct business. The IMB method should produce a lower minimum capital requirement but would only be available as a capital calculation methodology to larger, more technically able insurers with effective risk management programs.

1.3 The Purpose of this Study

This paper presents the results of a case study of the assessment of regulatory capital for non-life insurers in Australia. The case study highlights the issues involved in determining the capital requirements advocated by the IAA Insurer Solvency Assessment Working Party and demonstrates the challenges of the internal model based approach for insurers. It also highlights shortcomings of the prescribed method. The comparative levels of capital required under the prescribed method and the IMB method are important for insurers considering the use of internal model-based methods. Regulators adopt an approach such that insurers using either method should meet minimum levels of capital

[3] APRA's Prudential Standard GPS 110.

that ensure a consistent probability of insolvency across different insurers. This may have shortcomings when the probability of ruin over a single year horizon is used as a risk measure. In practice most insurers that develop internal models will consider risk measures that take into account the ruin probability and the severity of ruin. It is also important to consider longer horizons than the one year adopted by the regulators and used in this study.

Our study aims to compare the MCRs under the two methodologies.[4] In order to do this we use techniques that insurers would use in practice. The approach used is as follows. We develop a model of a typical, large non-life insurer with five business lines: (i) domestic motor, (ii) household, (iii) fire and industry-specific risk (ISR), (iv) public liability, and (v) compulsory third party (CTP) insurance. A dynamic financial analysis (DFA) model is used for the IMB method capital requirement, and this is compared to capital levels calculated under the prescribed method. The DFA model is used to allocate capital to each of the risks considered using a method adopted by practitioners. The model insurer's business mix, asset mix and business size are changed to examine the effect on capital requirements.

The main results of the analysis are as follows: Based on the liability volatility assumptions developed by leading industry consultants, the IMB method was found to produce a higher MCR than the prescribed method. From the insurer's perspective, this indicates a possible incentive to use the prescribed method in practice. It was also found that the prescribed method capital requirements were inadequate to ensure a ruin probability in one year of less than 0.5% for the entire general insurance industry. This illustrates the difficulty in developing prescribed method requirements that reflect insurer differences.

Finally, the liability volatility assumptions have a significant impact on the results produced by the internal model. There was no consensus on insurance liability volatility assumptions suitable for capital requirements for the Australian business. Consulting firms had developed and published estimates using their own experience and knowledge. A rigorous study is required to quantify these assumptions more precisely. This is an important area for future research. The assumptions adopted in this study and examined for sensitivity, however, are the best estimates available.

From an international perspective this study identifies challenges for risk-based capital requirements in insurance. Prescribed methods, although easier to apply, are more difficult to develop, especially if con-

[4] Readers are referred to Collings (2001) and IAA Insurer Solvency Working Party (2004) for prior work in this area.

sistent treatment of different insurers is important. On the other hand, implementing internal model-based capital requirements requires that the issue of the calibration of models and consistency in assumptions used for different classes of business be properly addressed. An internal model can deal with the many interactions between the assets and liabilities and many of the most important risks, but this will only be the case if the models are based on a sound estimation of risks from actual data. This is an area that requires attention before regulators can use this approach with the confidence that is necessary for such an important aspect of insurer risk management.

2 The Prophet DFA Model

This study uses a DFA model to determine the capital requirement under the IMB method. The DFA model used was developed using Prophet, a DFA software package produced by Trowbridge Consulting. The Prophet DFA model is used by several large non-life insurers in Australia for internal management purposes. Other DFA software packages commonly used in the Australian non-life insurance industry include Igloo (developed by The Quantium Group), Moses (developed by Classic Solutions) and TASPC (developed by Tillinghast Towers-Perrin Consulting). Although these various software packages have different features, we do not expect significant differences in the results from using a different DFA software package based on the simplified assumptions used in the model.

The Prophet DFA model calibrated for this study uses typical assumptions for this purpose. It was not developed to meet the requirements for approval by APRA and the Treasurer for use in the IMB method. The Prophet model is broadly representative of current industry best practice in general insurance DFA modeling.

2.1 Description

The Prophet DFA model consists of an economic model and an insurance model. The key interaction between the two models is inflation, which affects both the asset returns in the economic model and the claims and expenses in the insurance model. We describe the main features of the DFA model for completeness. Other models will differ in details but are broadly similar to the model described here.

2.1.1 The Economic Model

Prophet uses The Smith Model® (TSM®) to model the economic environment. TSM® is a proprietary economic model that forecasts a range of economic variables including bond yields, equity returns, property returns, inflation and the exchange rate. The key features of the model are that TSM® ensures that all initial prices and projections are arbitrage-free and that markets are efficient. Historical data are used to calibrate TSM® to derive the necessary parameters for the projections, including the risk premium and covariance matrix parameters that ensure efficiency in markets.[5] We should emphasize that we are not advocating the use of any particular model or software. We use typical software and assumptions as would be used by an insurer in practice in order to assess the impact of capital requirements and to draw conclusions about the alternate approaches. The economic model used may impact the results through the assumptions made about inflation and how this is incorporated into the liability model.

2.1.2 The Insurance Model

The insurance model is disaggregated into separate models for each of the insurer's business lines' liabilities. Assets, liabilities not relating to a specific line, and interactions between business lines are modeled at the insurer entity level.

Opening Financial Position: The opening financial position for the insurer is an input and covers the details of the insurer's liabilities and assets. From this opening position projections are simulated for the insurer's asset returns, claims for each business line, expenses, and reinsurance recoveries.

Asset Returns: Asset returns are projected based on the assumed asset allocation and the simulations from the economic model.

Claims: There are four stochastic claims processes in the model: run-off claims (outstanding claims); new attritional losses; new large claims; and new catastrophe claims. Attritional and large claims are modeled separately for each business line, while catastrophe claims are modeled by the catastrophe event.

- **Run-off Claims:** The opening value for the outstanding claims reserve equals the expected discounted value of the inflated

[5] For further details on The Smith Model® visit <http://thesmithmodel.com>.

run-off claims. The expected run-off claims are input into the model in the form of a run-off triangle. For each accident year the run-off claims are assumed to follow a lognormal distribution with a variance parameter for each business line and each accident year as input into the model.

Run-off patterns used in DFA model case study also are available from the authors. These were developed from assumed industry run-off patterns. A summary of the key factors is given in Table 1.

Table 1
Cumulative Payment Development Patterns by Business Line (Uninflated)

	Business Line				
Period	Motor	Home	F&ISR	Pub. Liab.	CTP
1	81%	86%	20%	11%	5%
2	100%	98%	90%	23%	11%
3	100%	99%	95%	37%	24%
4	100%	99%	98%	54%	41%
5	100%	100%	99%	69%	59%
6	100%	100%	100%	81%	73%
7	100%	100%	100%	88%	82%
8	100%	100%	100%	92%	88%
9	100%	100%	100%	95%	93%
10	100%	100%	100%	97%	96%
11	100%	100%	100%	98%	98%
12	100%	100%	100%	99%	99%
13	100%	100%	100%	99%	99%
14	100%	100%	100%	99%	100%
15	100%	100%	100%	100%	100%

Notes: Period = Development Period, F&ISR = Fire & ISR, Pub. Liab. = Public Liability.

- **New Attritional Losses:** Ultimate attritional losses from new claims are assumed to follow a lognormal distribution, with a specified payment pattern. Inflation and superimposed inflation also are included. Correlations between business lines are modeled by a specified correlation matrix that is put into the model. Parameters for attritional losses are given in Table 2.

Table 2
Attritional Claims Parameters
Lognormal Distribution Claims as a Percentage of GEP

	Motor	Home	F&ISR	Pub. Liab.	CTP
μ	-25.4%	-81.7%	-70.8%	-74.8%	-36.3%
σ	22.4%	30.6%	26.0%	27.0%	20.9%

Notes: F&ISR = Fire & ISR, Pbl. Liab. = Public Liability.

- **New Large Claims:** A collective risk model is used to model large claims. The frequency of claims is modeled as a Poisson process, and a lognormal distribution is used to model large claims severity as, for example, in Klugman, Panjer, and Willmot (1998, Chapter 4, pp. 291–384). Let K_i^L be the number of large claims for business line i, X_{ik}^L the size of the k^{th} large claim in business line i, Z_i^L is the aggregate large claims for business line i with

$$K_i^L \sim \text{Poisson}(\lambda_i)$$
$$X_{ik}^L \sim \text{Lognormal}\left(\mu_i, \sigma_i^2\right)$$
$$Z_i^L = \sum_{k=1}^{K_i^L} X_{ik}^L$$

and parameter values given in Table 3. The assumptions for the large claims payment pattern, inflation and superimposed inflation are identical to those used in the modeling of attritional claims.

- **New Catastrophe Claims:** Catastrophe claims are modeled based upon similar principles to the collective risk model with some modifications. Four catastrophe types are modeled separately. For each catastrophe, a Poisson frequency process was used to model the number of catastrophe events per year, and an empirical distribution was used to model the claim severity from the event. For each event, there is a primary and a secondary severity process modeled, with the primary process being larger than the secondary process. The

Table 3
Poisson Frequency and Lognormal Severity

	Fire & ISR	Public Liability	CTP
λ	0.53	1.36	1.42
μ	7.72	8.22	8.67
σ	0.46	0.25	0.22

Notes: F&ISR = Fire & ISR, Pbl. Liab. = Public Liability.

key difference between the modeling of large and catastrophe claims is that catastrophes are considered as events and are not specific to any business line. We assume

$$K_j^C \sim \text{Poisson}\left(\lambda_j\right)$$
$$X_{jk}^C \sim \text{Empirical Distribution}$$
$$Y_{jk}^C \,\Big|\, X_{jk}^C \sim \text{Empirical Distribution.}$$

Each business line is assigned a fixed percentage of either the primary severity or the secondary severity for each catastrophe type.

$$\mathrm{PC}_{ij} = P_j^C \times A_i$$
$$\mathrm{SC}_{ij} = S_j^C \times B_i$$

where PC_{ij} is the aggregate primary claims for business line i from catastrophe type j, $100A_i\%$ of primary severity for business line i, P_j^C is the aggregate primary severity for catastrophe type j which equals $\sum_{k=1}^{K_j^C} X_{jk}^C$, SC_{ij} is the aggregate secondary claims for business line i from catastrophe type j, $100B_i\%$ of secondary severity for business line i,

$$S_j^C = \sum_{k=1}^{K_j^C} Y_{jk}^C \,\Big|\, X_{jk}^C$$

is the aggregate secondary severity for catastrophe type j, K_i^C is the number of catastrophe events for catastrophe type

j, X^C_{jk} is the primary severity of the k^{th} event for catastrophe type j, and $Y^C_{jk} \mid X^C_{jk}$ is the secondary severity of the k^{th} event for catastrophe type j. Estimated parameters for the catastrophe model used in the base case DFA model are as follows:

- for the small catastrophe claims parameters the Poisson frequency was $\lambda = 4.3$ and mean and standard deviation of the empirical severity of GEP were 0.88% and 0.43%, respectively, and
- the impact to each line of business as percentage of severity was motor (50%), home (100%), and Fire & ISR (190%).

The other parameters used are given in Table 4.

Table 4
Large Catastrophe Type Claims:
Parameter Impact on Each Business Line

Type		Motor	Home	Fire & ISR
1	% Primary Severity		6	8
1	% Secondary Severity	64.6		
2	% Primary Severity		9	8
2	% Secondary Severity	1.6		
3	% Primary Severity		85	84
3	% Secondary Severity	33.8		

Expenses: There are three categories of expenses in the model: acquisition expense; commission; and claims handling expense. Acquisition expense and commission are expressed as a fixed percentage of premiums. Claims handling expense is a fixed percentage of claims. Expenses vary across business lines.

Reinsurance: The model allows for individual excess of loss (XoL) reinsurance to cover large claims and catastrophe reinsurance to cover catastrophes. Proportional reinsurance is not explicitly modeled, so in effect attritional claims can be viewed to be net of proportional cover. For both reinsurance contracts there is a cost of cover, a deductible amount, an upper limit, and a specified number of reinstatements for the contract.

2.2 Key Interactions and Correlations

An important aspect of DFA modeling is accounting for the many interactions and correlations between variables in the model. It is particularly important when considering the tail-end of the distribution of insurance outcomes given that extreme losses are often driven by several variables behaving unfavorably. For example, a one in two hundred year loss for an insurer could occur when both a catastrophe event causes very high insurance claims and at the same time asset markets underperform. In the Prophet DFA model there are four key interactions that are modeled: between assets and liabilities; claims and expenses; attritional claims across business lines; and between catastrophe claims across business lines.

Relationship between Assets and Liabilities: Inflation is important in the relationship between assets and liabilities. Consumer price index (CPI) and average weekly earnings (AWE) inflation are projected by TSM®. Inflation impacts asset returns, as TSM® assumes markets are efficient and incorporates a risk premium and covariance matrix to relate inflation with other asset prices. The impact of TSM®'s projected inflation on liabilities is through claims inflation in the insurance model.

Relationship between Claims and Expenses: Claims handling expenses are modeled as a fixed percentage of claims. Thus, claims handling expenses are perfectly correlated with claims incurred.

Relationship between Attritional Claims across Business Lines: A correlation matrix is specified to model the relationship between the attritional claims of different business lines. The parameter values for the correlation matrix (based on the Tillinghast study) used for the DFA study base case are given in Table 5. This gives the ρ_{ij} for the correlation between line of business i and line of business j.

Relationship between Catastrophe Claims across Business Lines: As catastrophes are modeled as events that can impact multiple business lines, there exists a correlation between catastrophe claims across different business lines. For business lines that are impacted by either the primary or secondary severity distribution of a given catastrophe event, there will be perfect correlation between claims from that catastrophe event. In the case where one line is impacted by the primary severity distribution and another is impacted by the secondary severity distribution, there will be a

Table 5
Parameter Values for the Correlation Matrix

	Motor	Home	F&ISR	Pub. Liab.	CTP
Motor	1.00	0.75	0.40	0.00	0.55
Home	0.75	1.00	0.35	0.00	0.00
Fire & ISR	0.40	0.35	1.00	0.00	0.00
Public Liability	0.00	0.00	0.00	1.00	0.35
CTP	0.55	0.00	0.00	0.35	1.00

Notes: F&ISR = Fire & ISR, Pub. Liab. = Public Liability.

positive correlation (but less than one). Lines that are not affected by a given catastrophe event will have zero correlation with lines that are affected. Modeling of dependence in insurance is a topic of current research. We have not included more detailed models of dependence in this case study. We aimed to use current industry practice which is currently largely based on correlations. Even using correlations is problematic because there is no current agreement on the assumptions to use.

3 The Data and DFA Model Assumptions

The data sources used to create the model insurer came from:

- APRA's June 2002 Selected Statistics on the General Insurance Industry (APRA statistics);
- Tillinghast's report "Research and Data Analysis Relevant to the Development of Standards and Guidelines on Liability Valuation for General Insurance;"
- Trowbridge's report "APRA Risk Margin Analysis;"
- Allianz Australia Insurance Limited (Allianz);
- Promina Insurance Australia Limited (Promina); and
- Insurance Australia Group Limited (IAG).

The model insurer created is not representative of any of the insurers that provided data for the study. Full details of the model assumptions are provided in Sutherland-Wong (2003) and available from the authors on request. Brief details are provided below.

The following data items were used for the model insurer:

Number of Business Lines: Five business lines were included. This was considered large enough for an in-depth analysis without unduly complicating the analysis. To ensure a broad mix of business lines, two of the five were chosen to be short tail (Domestic Motor and Household), two long tail (Public Liability and CTP), and one of intermediate policy duration (fire and ISR). The largest business lines from the APRA statistics (by gross written premium) for each of these categories were chosen.

Size of Business Lines: The business size was set so that the model insurer had a 10% market share from the APRA statistics (by gross written premiums) in each business line.

Expected Claims: The expected claims for each line of business were set to a level to produce an expected after-tax return of 15% on capital based on an assumed capital level of 1.5 times the MCR calculated under the prescribed method. The payment pattern, premium assumptions, and inflation assumptions were used to solve for the expected claims for each business line to meet this target.

Claims Volatility: The volatility assumption used for each business line determines the insurance outcome at the 99.5th percentile and therefore directly impacts the MCR. Rather than using individual insurer data for these assumptions, we used statistics that were more representative of the broader Australian general insurance industry.

The Tillinghast and Trowbridge reports both include estimates of the coefficients of variation (CVs) of the insurance liabilities of the Australian general insurance industry. The reported CVs in these reports were vastly different, however, with the Tillinghast numbers being generally twice as large as the Trowbridge numbers. Table 6 provides details of CVs used in this DFA case study based on the Tillinghast report.

Thomson (2003) outlines the initial risk margins that insurers have adopted since the new standards came in force from July 2002. He reports that for short tail lines, insurers were generally aligned with the lower Trowbridge numbers. For long tail lines, the numbers were consistently lower than the Tillinghast report. There was a great deal of variation in the risk margins adopted within each business line, however, suggesting that there is no

Table 6
Liability CVs by Business Line

	Type of Liability CV	
	Outstanding Claims	Premium
Motor	12.4%	21.7%
Home	18.9%	33.1%
Fire & ISR	18.9%	28.4%
Public Liability	23.7%	29.6%
CTP	21.8%	27.2%

real consensus among the industry on the appropriate level for risk margins. It generally would be in the interest of insurers to adopt lower risk margins in order to report a lower liability value and also a lower capital requirement.

The Tillinghast numbers were used in the analysis, as they represented a more conservative view of variability in the industry. The Trowbridge numbers were used as an alternative scenario in the analysis to determine the impact of these assumptions.

Payment Pattern: The payment pattern data were derived from typical insurer data.

Asset Mix: The asset mix for the model insurer was representative of the industry average investment mix. Details on the asset mix assumed are given in Table 7 for the base case.

Table 7
Asset Mix

	Proportion Invested
Cash	15%
Equities	20%
Fixed Interest	55%
Index Bonds	10%
Total	100%

Reinsurance: The reinsurance for each business line was based upon typical insurer data. For the long tail lines, individual XoL contracts were designed to cover most of the large claims. For the short tail lines (including fire and ISR), catastrophe XoL contracts were designed to set the maximum event retention (MER) of the insurer to equal $15 million.

Superimposed Inflation: The external factors that affect the run-off claims are inflation and superimposed inflation. The inflation level is derived from the economic model, while superimposed inflation is modeled as a stochastic two-state process. The superimposed inflation process consists of a normal superimposed inflation state and a high superimposed inflation state, with a transition probability matrix determining the movement between these two states. The process is described as follows: Let P denote the transition probability matrix where p_{ij} is the probability of moving from state i to state j, I_N is the superimposed inflation rate in the normal state, I_H is the superimposed inflation rate in the high state with

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

$I_N \sim U(a_N, b_N)$ and $I_H \sim U(a_H, b_H)$. Using typical insurer data, the estimated parameter values used in the model are:

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

$I_N \sim U(-0.02, 0.04)$ and $I_H \sim U(0.05, 0.15)$.

4 Assessment of the DFA model

The model was designed to broadly represent best practice in applying DFA models to capital analysis and to be consistent with the way that practitioners would model the business lines. The parameters of the model were set to capture the features of a typical insurer. The model also can be assessed against APRA's Guidance Note GGN 110.2, which sets out the qualitative and quantitative requirements for an internal model. The key quantitative risks that an internal model must capture, as specified by the Guidance Note, fall under the broad categories of investment risk, insurance risk, credit risk, and operational risk.

TSM® is used in Prophet to capture the dynamics of the economic market and the subsequent impact on an insurer's investment portfolio. While no stochastic asset model currently available is perfect, TSM® is representative of best practice in economic forecasting and assessment of investment risk.

The Guidance Note specifies a range of risks relating to the insurance business that need to be included in the model. These risks include outstanding claims risk, premium risk, loss projection risk, concentration risk, and expense risk. Prophet allows for these risks using the assumed variability in its three claims processes: attritional, large and catastrophe claims. Attritional claims are assumed to follow a lognormal distribution. This assumption is common industry practice for modeling claims. The lognormal assumption can be inadequate for capturing the true variability in claims processes, particularly when analyzing the tail-end of the distribution of claims. Modeling dependencies between business lines with a standard (linear) covariance assumption may not adequately capture the dependence in tail outcomes. Although not commonly used in industry practice, copulas are an increasingly useful method of measuring tail dependencies. Venter (2001) and Embrechts, McNeil, and Straumann (2000, Chapter 6, pp. 71–76.) provide a good coverage of the use of copulas in modeling tail dependencies in insurance.

The Prophet DFA model does attempt to capture the variability in claims at the tail-end of the distribution by including separate models for large claims and catastrophe claims. Dependencies between business lines in these tail outcomes are captured in part by the impact of catastrophe events on multiple business lines. The catastrophe model has a similarity to frailty models used to construct copulas. How well the model captures the tail risk in practice is an empirical issue that needs further research.

Concentration risk is a component of loss projection risk relating to the uncertainty of the impact of catastrophic events. This risk is accounted for by the catastrophe model. The excess of loss catastrophe reinsurance assumptions in the model limit the impact of concentration risk.

Expense risk is accounted for because claims handling expenses are expressed as a percentage of claims incurred. Although some unexpected expense increases may be independent of the amount of claims, there is normally a significant level of correlation between claims and claims handling expenses. Assuming expenses and claims are perfectly correlated results in a conservative allowance for the expense risk of the insurer—under circumstances in the tail when claims are higher than expected, so too will be claims handling expenses.

Like all businesses, insurers face the credit risk that parties who owe money to them may default. For an insurer, the key sources of credit risk arise from their investment assets, premium receivables, and reinsurance recoveries. Credit risk relating to investment assets is implicitly covered in The Smith Model. The Prophet DFA model calibrated in this study does not account for the risk of default in premiums or reinsurance owed. Thus, the MCR calculated by the IMB method using the Prophet DFA model will not include a charge for these risks. To compensate for this, in calculating the total MCR for the IMB method, the charge from the prescribed method for outstanding premiums and reinsurance recoveries is included.

Guidance Note GGN 110.2 highlights operational risk as a quantitative risk that should be included in an insurer's capital measurement. Operational risk, however, is a particularly difficult risk to quantify and is an area of ongoing research in both insurance and banking. APRA's prudential standards include a Guidance Note for operational risk, GGN 220.5, which outlines the qualitative measures an insurer should pursue to manage operational risk, but does not provide any guidance on how to quantify the risk for capital calculation.

The Prophet DFA model calibrated in this study does not account for operational risk. There is no well-accepted model nor sufficient data and analysis to properly assess insurer operational risk. The prescribed method does not have a charge for operational risk. The Basel Committee's Working Paper on the Regulatory Treatment of Operational Risk (2001) reports that operational risk should make up 12% of a bank's minimum required capital. Giese (2003) uses a survey of banks and non-life insurers to report that on average banks allocate approximately 30% of their capital to operational risk, while non-life insurers allocate approximately 16%. In the absence of an agreed approach to allocating capital to operational risk for non-life insurers, however, it was decided that no additional charge would be made. Given the comparative nature of this study, this assumption does not impact on the conclusions drawn or the significance of the results.

5 Methodology

A model insurer was created to be representative of a typical large non-life insurer operating in Australia. The Prophet DFA model was used to project future insurance outcomes under different assumptions. Six thousand (6000) simulations were performed for each set of assumptions and used to estimate the required capital to ensure a ruin probability over a one year horizon of 0.5%. The number of simulations was determined so that the standard error of the capital requirement estimate was small compared to the capital amount. The capital requirement calculated by the Prophet DFA model was then compared to the MCR under the prescribed method for the model insurer.[6] A summary of the prescribed method capital charges for Australia is provided in Tables 8 and 9.

Table 8
Outstanding Claims and Premium Liability Capital Charges for Direct Insurers

Class of Business	Risk Capital Factor: Outstanding Claims	Risk Capital Factor: Premium Liability
Home, Motor, and Travel	9%	13.5%
F&ISR and Others	11%	16.5%
CTP, Liability, and Professional Indemnity	15%	22.5%

Notes: Motor includes commercial and domestic; Liability includes public, employer, and product liabilities; F&ISR and Others include Fire & ISR, Marine, Aviation, Consumer Credit, Mortgage, Accident.

The following five sets of assumptions were examined to assess their impact on different types of insurers with different balance sheet structures. As the assumptions for the liability volatilities currently used differ significantly, it was important to examine the impact of these differences. In each case, only the assumption listed is changed from the base case.

[6] See APRA GPS 110 for the insurance and investment risk capital charges.

Table 9
Investment Capital Charges

Type of Charge	Charge
Cash and debt obligations of the commonwealth government, an Australian state or territory government, or the national government of a foreign country where the security has a Grade 1 counterparty rating or, if not rated, the long-term foreign currency counterparty rating of that country is Grade 1; GST receivables (input tax credits);	0.5%
Any debt obligation that matures or is redeemable in less than one year with a Grade 1 or 2 rating; cash management trusts with a Grade 1 or 2 rating;	1.0%
Any other debt obligation that matures or is redeemable in one year or more with a Grade 1 or 2 rating; reinsurance recoveries, deferred reinsurance expenses, and other reinsurance assets due from reinsurers with a Grade 1 or 2 counterparty rating;	2.0%
Unpaid premiums due less than six months previously, unclosed business, any other debt obligation with a rating of Grade 3; reinsurance recoveries, deferred reinsurance expenses, and other reinsurance assets due from reinsurers with a counterparty rating of Grade 3;	4%
Any other debt obligations with a counterparty rating of Grade 4; reinsurance recoveries, deferred reinsurance expenses, and other reinsurance assets due from reinsurers with a counterparty rating of Grade 4;	6%
Any other debt obligations with a counterparty rating of Grade 5; reinsurance recoveries, deferred reinsurance expenses, and other reinsurance assets due from reinsurers with a counterparty rating of Grade 5; listed equity instruments (including subordinated debt), units in listed trusts, unpaid premiums due more than six months previously:	8%
Direct holdings of real estate, unlisted equity instruments (including subordinated debt), units in unlisted trusts (excluding cash management trusts listed above), other assets not specified elsewhere in this table;	10%
Loans to directors of the insurer or directors of related entities (or a director's spouse), unsecured loans to employees exceeding $1,000; assets under a fixed or floating charge;	100%
Goodwill (including any intangible components of investments in subsidiaries), other intangible assets, future income tax benefits, assets in this category are zero weighted because they are deducted from Tier 1 capital when calculating an insurer's capital base; see GGN 110.1.	0%

1. **Alternative Liability Volatility Assumptions:** The model was run using the Trowbridge volatility assumptions. This was to indicate the sensitivity of the capital requirements to a change in volatility based on an alternative view on the variability of business lines. As both sets of volatility assumptions have been proposed it is of interest to examine the resulting difference.

2. **Riskier Asset Mix Assumption:** The model was run with the insurer having a significantly higher proportion of investment assets in equities. This was designed to indicate the MCR required for insurers in the industry holding significant levels of riskier assets. This also allows a comparison of the significance of the investment capital charge for the IMB and prescribed methods.

3. **Short Tail Insurer Assumption:** The insurer was assumed to only sell short tail business lines. Assets and liabilities were scaled back to reflect the smaller overall insurer size, while all other assumptions remained unchanged. Because some insurers have predominantly short tail business, this will identify the significance of the short tail capital charge for the comparison between the IMB and prescribed methods.

4. **Long Tail Insurer Assumption:** The insurer was assumed to only sell long tail business lines. Assets and liabilities were scaled back to reflect the smaller overall insurer size, while all other assumptions remained unchanged. This will identify the significance of the long tail capital charge. And,

5. **Smaller Insurer Assumption:** In this case the insurer was assumed to have premiums equal to 2.5% of the gross written premiums from the APRA statistics. The liability variability assumptions were adjusted according to the Tillinghast report to account for the smaller business size. Assets and liabilities were also scaled back and all other assumptions remained unchanged.

In order to compare the IMB and prescribed methods, it is necessary to allocate the MCR to lines of business. To do this we use a technique adopted by practitioners. Myers and Read (2001) have proposed an allocation of capital to lines of business based on marginal changes in business mix. Sherris (2004) shows that, under the assumptions of complete and frictionless markets, there is no unique capital allocation to line of business unless an assumption about rates of return or surplus ratios also is made. In this case study we have set the liability parameters to generate a constant rate of return across lines of business.

A numerical estimation procedure was used to allocate capital to line of business. The procedure was as follows:

Step 1: The size of business line 1 was reduced by 1%.

Step 2: The marginal change in the MCR was calculated, and this amount was allocated to business line 1.

Step 3: Steps 1 to 2 were repeated for business lines 2 to 5.

Step 4: Steps 1 to 3 were repeated 100 times until all the business line sizes were reduced to zero and the MCR was reduced to zero.

The capital allocated to each line was calculated as the sum of all of the marginal capital allocations for each line of business. Using a 1% reduction each time was sufficiently small so that the capital allocation was found to be independent of which line was reduced first. In addition, the capital allocated to each line of business is such that, as an additional small amount of each liability is added, the overall insurer one year ruin probability is maintained. This is equivalent to using the ruin probability for the total company as a risk measure when determining capital allocation. In other words, the capital allocated to each line of business is such that for the insurer the overall ruin probability is constant.

6 Capital Requirements and Model Results

6.1 Model Insurer—Base Case

The model was run for the base case assumptions. The Prophet DFA model produced a distribution of insurance outcomes. For each of these outcomes, the amount of assets in excess of the technical reserves required at the start of the year to ensure that the insurer's assets are equal to their liabilities at the end of the year was determined. This represents a distribution of capital requirements. The Prophet MCR was determined as the 99.5^{th} percentile of this distribution of capital requirements. By taking this capital requirement, the probability that total assets will exceed liabilities at the end of the year will be 99.5%, using the same simulations. This value was \$309.4M with a standard error of \$10.9M. The standard error was calculated using the Maritz-Jarrett method. Details of the method for computing the standard error are in Wilcox (1997, Chapter 3, p. 41). The distribution of capital requirements is shown in Figure 1.

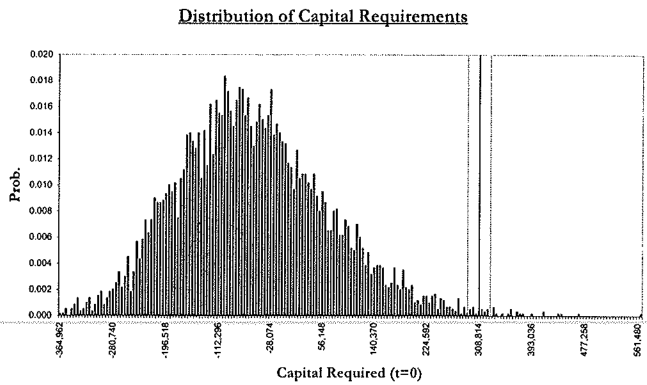

Figure 1: Distribution of Base Case Prophet Capital Requirements

The results of the determination of the MCR by both the internal model and the prescribed method are given in Table 10. As the Prophet DFA model does not make an allowance for credit risk, the overall MCR for the model insurer was determined as the sum of the internal model capital requirement plus the credit risk capital charge from the prescribed method. This capital requirement is the MCR calculated under the IMB method and is shown in Table 10.

Table 10
Base Case Minimum Capital Requirement (MCR)
Comparison Between IMB Method and Prescribed Method

	Base Case
IMB Method	
Prophet MCR	309,396,000
Adjustment for Credit Risk	28,705,000
Total MCR	338,101,000
IMB Standard Error	10,912,000
Prescribed Method	
Total MCR	233,323,000

The MCR calculated by the IMB method was found to be significantly larger than the MCR under the prescribed method. The MCR calculated by the IMB method represents the risk-based level of capital required to ensure a ruin probability in one year of 0.5%. The prescribed method is found to produce a capital requirement insufficient to ensure a probability of ruin over a one year time horizon of 0.5%.

To understand each method's treatment of the various risks, we break down each of the MCRs by line of business and by risk type. The capital charge components that make-up the prescribed method's MCR are presented in Table 11.

Table 11
Prescribed Method MCR Capital Charges (in $1,000s)

	Motor	Home	F&ISR	Pbl. Liab.	CTP	Totals
Risks						
Invest.						36,687
Credit						28,705
Concen.						15,000
Liabilities						
OCLMS	5,314	2,381	3,740	10,188	69,667	91,290
Premium	20,696	9,192	8,335	6,160	17,258	61,640
Total						233,323

Notes: F&ISR = Fire & ISR, Pbl. Liab. = Public Liability, Invest. = Investment, Concen. = Concentration, OCLMS = Outstanding Claims Liability.

Under the prescribed method, the total capital charges relating to liability risks (outstanding claims, premium, and concentration risk) equal $167.9M. The long tail business lines account for 61.5% of this charge, while the short tail lines (including fire and ISR) account for 38.5% of the charge.[7] The Prophet internal model capital requirement was allocated to individual business lines using the numerical approach set out earlier. The resulting allocation is shown in Table 12. For the MCR calculated by the IMB method, long tail lines account for 67.4% of capital while short tail lines (including fire and ISR) account for 32.6%. Although this allocation gives a slightly higher capital weighting to long tail lines than the prescribed method, the differences are small.

There are, however, significant differences in capital allocations for each business line. The prescribed method allocates the same percent-

[7]The concentration risk charge is allocated only to the short tail and fire & ISR lines.

Table 12
Base Case Allocation of
IMB Method MCR to Business Lines

Business Line	Capital Allocated In $1,000s	Percent Of Total
Motor	28,291	9.1
Home	60,965	19.7
Fire & ISR	11,550	3.7
Public Liability	7,266	2.3
CTP	201,323	65.1
Total	309,396	100.0

age charge to both household and motor insurance. As the model insurer has approximately half the level of household insurance as motor insurance, the prescribed method capital charge is approximately half. The allocation of the MCR calculated by the IMB method to the household line, however, is more than double the capital allocated to the motor line. This is due to the higher CV of 33% for household insurance vs. 22% for motor based on the Tillinghast report. The difference between the capital allocations under the two methods is illustrated in Figure 2.

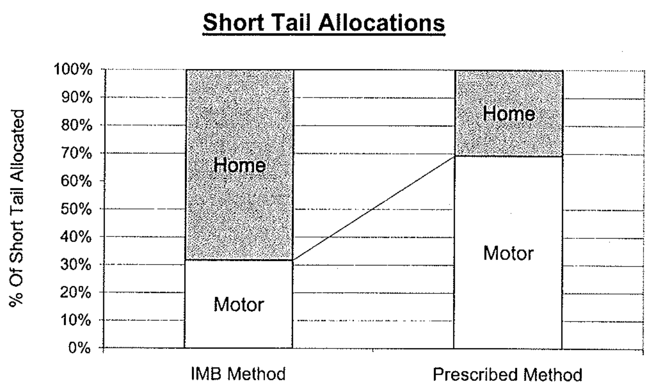

Figure 2: Relative Short Tail Capital Allocations

In considering these allocations of capital it is worth emphasizing that we are comparing a prescribed method with a method that was designed to ensure an equal expected rate of return to capital across lines of business. These differences will only be of real significance if company management were to use these results in their business strategy or decision making. In practice, these allocations are used for a variety of purposes including pricing as well as decisions about which lines of business to grow and to limit.

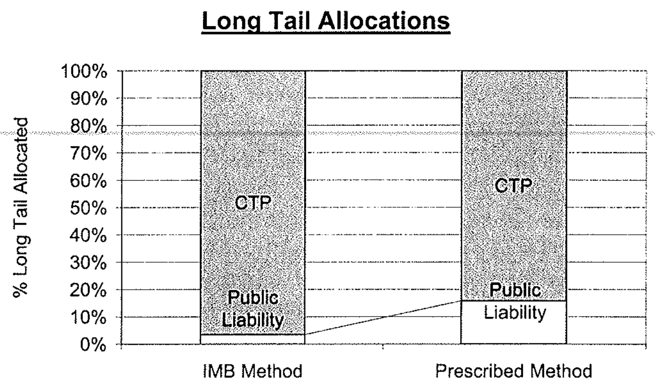

Figure 3: Relative Long Tail Capital Allocations

For public liability and CTP insurance the prescribed method gives the same allocation of capital charge percentages to each of these lines so that the difference in the prescribed method capital charged for the model insurer is due to the relative sizes of the business lines (51.8% of capital is allocated to CTP with 9.7% allocated to public liability). For the internal model allocation the capital allocated to CTP is much higher (65.1%). For public liability it is much smaller (2.3%).

Figure 3 illustrates the differences between the capital allocations under the two methods for the long tail lines. The difference in this case is driven largely by the diversification effects from each line. Public liability insurance has a moderate correlation (35%) with CTP and zero correlation with all other business lines. This results in the public liability line providing large diversification benefits to the model insurer. CTP on the other hand is assumed to have a 50% correlation with motor insurance so the diversification benefits to the model insurer are diminished and a higher level of capital, therefore, is allocated to this line.

The prescribed method has difficulty handling correlations between lines of business and differences in insurer business mixes. This is a strength of the internal model, although the assumptions underlying the correlations between business lines need to be considered carefully.

6.2 Alternative Assumptions

Table 13 summarizes the results from the alternative assumptions. The table shows the capital requirements from the IMB and prescribed methods for the base case and for each of the alternative assumptions.

6.2.1 Alternative Volatility Assumptions

Adopting the lower CVs from the Trowbridge report dramatically reduces the MCR calculated under the IMB method by $214.8M. The internal model results are extremely sensitive to the volatility assumptions for the insurance liabilities. An insurer who uses the Trowbridge CVs for the volatility of their business will require an MCR under the IMB method that is significantly lower than the MCR calculated under the prescribed method. Without an extensive study of liability volatility to validate these assumptions, it is open to insurers who can use the internal model approach to adopt volatility assumptions in line with these levels.

6.2.2 Riskier Asset Mix

As expected, the MCR under both the IMB and prescribed methods increase when the insurer's proportion of invested assets in equities is increased to 80%. There is a difference in increase for each method, however. Under the IMB method, the MCR increases by $61.0M, while under the prescribed method the increase was much less at only $49.0M. The capital charge for equities in the prescribed method may not be sufficient to allow for the impact of these securities on ruin probabilities.[8] Because asset risk, especially asset mismatch risk, is a major risk run by insurers, a prescribed method should not encourage insurers to adopt a riskier investment strategy. The above result suggests that the prescribed method in Australia may have an incentive for insurers to invest in equities.

---

[8] This is based on the assumption that TSM® is a realistic model of asset returns.

Table 13
Summary of MCR Comparisons for Alternative Assumptions

	Prescribed Method MCR Capital Charges in $1,000s					
	Base Case	Trowbridge CV	80% Equities	Tail Only Short	Tail Only Long	Small Insurer
IMB Method						
Prophet	309,396	94,586	370,414	209,196	228,828	139,951
Credit Risk	28,705	28,705	28,705	10,745	12,513	5,577
Total	338,101	123,291	399,119	219,941	241,341	145,528
Standard Error	10,912	3,469	13,517	5,221	9,056	5,413
Prescribed Method						
Investment Risk	36,391	36,391	85,366	9,517	22,656	9,098
Credit Risk	28,705	28,705	28,705	10,745	12,513	5,577
OSC Liability	91,290	88,237	91,290	7,890	80,999	28,808
Premium Liability	61,641	59,415	61,641	30,885	23,789	15,876
Concentration	61,641	59,415	61,641	30,885	23,789	15,876

It is interesting to note that had the Trowbridge CV assumptions been used, then changing the asset mix from 20% equities to 80% equities would have increased the MCR by a greater amount of \$83.7M. The reason for the larger increase under the Trowbridge assumptions is related to the relative size of the various risks and their impact on ruin probability.

The Trowbridge assumptions have lower insurance liability volatility, so that fewer of the outcomes at the 99.5^{th} percentile of the capital required distribution are due to high claims costs. Instead, the outcomes at the 99.5^{th} percentile are more often due to low asset returns. This leads to a higher proportion of the overall capital under the IMB method being attributed to asset risk when insurer liability volatility assumptions are lower. This in turn creates a greater disparity between the prescribed method's and IMB method's charges for asset risk.

6.2.3 Short Tail Insurer

Removing the long tail lines from the insurer reduces the overall insurer's size along with the MCR. Under the IMB method the MCR reduces by \$118.2M, while under the prescribed method the reduction is significantly larger at \$159.3M. The internal model allocates more capital to each of the insurer's liabilities than is charged by the prescribed method. Rather than comparing absolute changes in MCR, it is more interesting to compare the relative changes. The same will hold for the long tail and small insurer scenarios.

Figure 4 illustrates the MCRs under the different assumptions relative to the base case. For the short tail insurer, the IMB method has a reduction in its MCR of 65% of its original size, while under the prescribed method the MCR reduces to 32% of its original size. The internal model is allocating a greater amount of capital to a purely short tail insurer compared to the prescribed method. The reasons for this highlight some further shortcomings of the prescribed method. By reducing the number of lines of business, diversification benefits are lost. This is accounted for in the IMB method but not by the prescribed method, which has constant capital charges independent of the business mix. The result is that the capital calculated under the IMB method is higher than under the prescribed method.

The short tail capital charges (relative to other capital charges) under the prescribed method also may charge less for the risk of those lines than the internal model. This would be consistent with Collings's (2001) finding that as an insurer increases its business mix with short

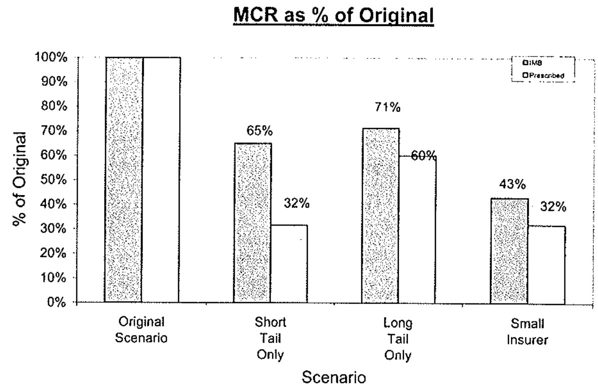

Figure 4: Alternative Assumptions MCRs as Percentage of Base Case

tail lines,[9] it will have a relatively larger capital increase under the IMB method than the prescribed method.

This means that insurers will have an incentive to write short tail lines if they are using the prescribed method. If there is a relative advantage in capital required for short tail lines, this also may lead to underpricing of these lines.

6.2.4 Long Tail Insurer

In Figure 4 we note that the MCR calculated by the IMB method reduces to 71% of its original size, while under the prescribed method the MCR reduces to 60% of its original size for the case of a long tail insurer. The internal model allocates a higher level of capital to a purely long tail insurer than the prescribed method.

The same two effects as for the short tail insurer appear to apply to the case of the long tail insurer. Once again there is a loss of some diversification benefits for the purely long tail insurer leading to the higher relative MCR under the IMB method than under the prescribed method. The long tail capital charges (relative to other capital charges) under the prescribed method charge less for the risk of those lines than the internal model. This is inconsistent with Collings (2001) findings

[9] Collings (2001) used motor insurance as an example of a short tail line.

that increasing the business mix with long tail lines[10] led to a greater relative MCR under the prescribed method than under the IMB method.

Regardless of the relative impact of long tail lines of business, however, it is clear that the prescribed method can not deal adequately with differences among business mix of insurers. Applying the prescribed method will lead to incentives for insurers to change their business mix to optimize their regulatory capital position. Lines of business with too low capital charges will be increased, leading to potential price cuts that can not be justified if proper risk allowance were to be made.

6.2.5 Small Insurer

Figure 4 also shows the effect of changing the size of the insurer. In this case the insurer is assumed to reduce to 25% of its original size. Under the prescribed method, the MCR reduced by a similar amount to 32% of its original size.[11] The percentage capital charges under the prescribed method are independent of insurer size. For the IMB method, while the size of the insurer decreased, the overall volatility of each of the business lines is assumed to increase. This is based on the assumption that smaller business portfolios have greater independent variance and that pooling of insurer risks reduces relative volatility within a class of business. The volatility assumptions in an internal model should depend on the size of the business line, with higher volatility assumed for smaller lines. The overall MCR under the IMB method reduced to 43% of its original size.

7 Risk-Based Capital Regulation of Insurers

7.1 Impact of Volatility Assumptions

Our results show a strong dependence of an internal model's output on the insurance liability's volatility assumptions. Of all the sensitivities performed, the greatest change in MCR resulted from changing from the original Tillinghast insurance liability CVs to the Trowbridge CVs.

Insurers would be expected to prefer to have a lower regulatory capital requirement. Insurers in the industry that have liability volatility

[10]Collings (2001) used public liability insurance as an example of a long tail line.

[11]The MCR under the prescribed method did not reduce to 25% of its original size because the risk margins for a smaller insurer are higher and the concentration charge was assumed to remain constant at $15M.

similar to the Tillinghast CVs are unlikely to adopt an internal model to calculate their MCR. Insurers that have liability volatility similar to the Trowbridge CVs have an incentive to adopt an internal model to lower their MCR. As yet no insurer in Australia has elected to use an internal model-based approach. This may be for a number of reasons. One of these could be that the prescribed method produces lower capital requirements than would be required if they were to adopt an internal model. If this were the case, then those insurers who use these levels of capital to price their insurance contracts could be undercharging compared to the premium rate required to generate the level of ruin probability considered appropriate by APRA using the internal model approach.

The importance of the assumed CVs in determining an insurer's MCR indicates a clear need for an assessment of the level of volatility across business lines and an understanding as to how this varies across companies. Thomson (2003) commented on APRA's disappointment with the general lack of justification by actuaries in the risk margins they adopted for their first reporting under the new APRA requirements. It appears that Australian insurers have yet to understand fully the true level of volatility in their businesses and have yet to reach agreement on best practice in calculating volatility. This is expected to be an important issue for any regulator to address, regardless of country, in the introduction of risk-based regulatory capital requirements.

7.2 Issues with the Prescribed Method

Considering the results in Figure 4, it is evident that the prescribed method does not prescribe a level of capital that is adequate to ensure a ruin probability of 0.5% for all insurers, regardless of size or mix of business. This is based on the presumption that the internal model used in this study represents an insurer's realistic business situation. The model used has been developed to be close to the realistic situation and reflects industry best practice. The MCR calculated by the internal model should be representative of the actual level of capital required to ensure a ruin probability in one year of 0.5%. Although the IAA Insurer Solvency Assessment Working Party (2004) state that the prescribed method should be conservative to make sure that it is representative of all insurers that conduct business, in the Australian general insurance industry this does not appear to be the case.

An important part of implementing the risk-based capital requirements is the calibration of the prescribed method capital charges. APRA calibrated the current capital charges at an industry-wide level so that the total MCR of the industry increased by a factor of 1.4 to 1.5 times the previous level. This was a substantial increase in capital requirements across the industry. The impact of the changes differs between insurers depending on their experience relative to the industry. Insurers with lower volatility experience are effectively treated the same as those with higher volatility experience. Even if the charges are adequate for the average insurer, they may be inadequate for insurers with greater than average volatility. Assuming APRA wants to secure an industry-wide solvency requirement of a 0.5% ruin probability in one year, then it will need to increase the prescribed capital charges for insurance liabilities. Given that the last change in regulatory requirements increased the capital requirement in the industry by around 50%, a further increase is a politically contentious issue.

This is the situation that is likely to face many regulators at an international level when they consider the introduction of these risk-based capital requirements. There are likely to be poorly capitalized insurers that will no longer be able to operate under these more stringent requirements. At the same time capital strong insurers will be expected to meet the requirements. Given the difficult capital situation that has been faced by the insurance industry at an international level, the adoption of risk-based capital for insurers may take longer and require more attention to capital-weak insurers than otherwise.

An important issue for the Australian regulator will be to consider the liability capital charges that should be increased and to what extent. Collings (2001) found that short tail lines had a relatively higher capital requirement under the IMB method compared to the prescribed method and vice versa for long tail lines. While the results from this study are broadly consistent with this, the differences between long tail and short lines are less distinct.

At an individual line level, our internal capital allocation showed that the household line was allocated a significantly larger amount of capital than the motor line. This was driven by the higher CV assumption for household from the Tillinghast report. Differences in household and motor volatility suggest that it is inappropriate for household and motor to have identical capital charges.

Differences in the capital allocations between CTP and public liability were assumed to be driven largely by diversification effects. Smaller insurers and insurers with less diversified business mixes are undercharged under the prescribed method to a greater extent than larger and well-diversified insurers. This strongly indicates the need to include diversification benefits in the capital requirements, concentration charges for less diversified insurers, or varying capital charges based upon business size.

Further sophistication to the prescribed method must be weighed against the benefits of simplicity in the method. It is clear, however, that using a prescribed method that is out of line with the actual risk-based capital requirements will produce incentives for insurers to behave out of line with the economics of the business. This is a critical issue if this approach leads to an incentive for insurers to underprice or grow riskier business lines.

7.3 Investment Risk Capital Charges

Our results demonstrated that the capital required for higher levels of equity investment was greater under the IMB method than under the prescribed method. The prescribed capital charge for equities under the model assumptions is insufficient to cover the risk. This is consistent with the findings of Collings (2001) that the prescribed method is less responsive to increases in equity investment than is the IMB method. An adequate charge to cover equity risk at the 99.5^{th} percentile would need to be larger than the current prescribed charge of 8%,[12] particularly for insurers with investment portfolios that are not well diversified.

On the asset side, the prescribed method provides little incentive for insurers with a well-diversified investment portfolio. As an example to illustrate this point consider the property investment capital charge. The capital charge for property investment is 10%, while the capital charge for listed equity is 8%. For insurers that perceive there to be relatively higher risk-adjusted returns to be gained from equity than from property, there is an incentive to overweight their investment in equity. This is despite the fact that there can often be considerable diversification benefits of holding equity and property together. Collings (2001) provides another example by considering the diversification and immunization benefits of holding appropriate amounts of government bonds and cash. While there is an optimal amount of each of these securities to hold that minimizes overall volatility for the insurer, the capital charges under the prescribed method do not distinguish between the two asset classes and charge a constant amount of 0.5%.

In order to ensure the MCR under the prescribed method provides an industry-wide solvency requirement of a 0.5% ruin probability in one year, APRA will need to change the investment capital charges.

[12] 8% is the prescribed capital charge for listed equity securities.

For risky assets such as equity, the current capital charges should be increased. APRA also should provide incentives for insurers to hold well-diversified asset portfolios. This could be achieved by offering diversification discounts or alternatively a more stringent investment concentration charge.[13]

7.4 Incentive to Use an Internal Model

The opening section of APRA's capital standards states that APRA encourages insurers with sufficient resources to adopt an internal model for calculation of their MCR. APRA has a desire for insurers to begin to adopt the IMB method in line with its aim for insurers to more closely match their capital requirements with their individual risk characteristics. The results of the analysis of the capital requirements that we have undertaken indicate that there is no incentive to adopt the IMB method, especially if an insurer has insurance liability CVs in line with the Tillinghast report. The prescribed method's capital charges would need to increase to the extent that the internal model would produce a lower MCR. Alternatively there needs to be a much closer examination of the volatility of insurer liabilities and a more careful calibration of the prescribed method capital charges.

There are other reasons why insurers would not adopt an internal model for the MCR calculation. Even though risk management and measurement techniques in non-life (property and casualty) insurance have vastly improved over the last decade and DFA modeling has become an important part of internal management for many large insurers, developments in these areas are still occurring. The IMB method requires an internal model with a very high degree of sophistication to adequately address all the material risks of an insurer and their complex interrelationships. There also needs to be the actuarial and risk management human resource skills to ensure proper implementation and interpretation of results. The internal model used in this study was based on simplifying assumptions, and the internal model for a real-world insurer would be far more complex. Even with an adequate internal model, the assumptions required in the model need far more careful attention. A greater understanding and consensus of the underlying volatility of insurance liabilities is a major requirement for non-life insurers in order to adopt an internal model for MCR calculation.

Even as actuaries develop the necessary skills and capabilities to adequately implement an internal model for the IMB method, there will

[13] The current investment concentration charge only applies to Grades 4 and 5 debt and does not apply to concentrated holdings in other securities.

no doubt exist further obstacles from other stakeholders in the general insurance industry. The black-box stigma attached to internal models is likely to be an area that actuaries will need to overcome in order to convince general insurance senior management and the regulators to trust the internal model's output for management purposes and MCR calculation.

Industry experts have identified another obstacle to the IMB method. Financial analysts involved in the trading of general insurance company shares may not have the confidence in the insurer's management to rely on them determining their own regulatory capital requirements. Financial analysts may not be willing to rely upon the MCR calculated under the IMB method.

Differences in approaches to internal modeling also may make it difficult to compare the MCR output from one insurer's internal model with another insurer. Comparison across different insurers is important for regulatory reasons and to avoid opportunistic insurers taking advantage of differences in models. Financial analysts and regulators may prefer to make MCR comparisons based upon the prescribed method where the formula is fixed and insurer judgment does not impact the results. This leaves open the need to develop prescribed method charges that are more risk-based.

8 Closing Comments

8.1 Summary and Conclusions

The IAA Insurer Solvency Assessment Working Party (2004) has advocated two methods for non-life insurers to calculate their capital requirement: the standard approach and the advanced approach. In Australia, these dual capital requirements are known as the prescribed method and the IMB method. This study explores the implications of these new capital requirements.

From APRA's perspective, the aim is to meet a regulatory objective of requiring that insurers hold a level of capital to ensure a minimum ruin probability across the industry. It is important that the prescribed method adequately charges risks to meet this objective for all non-life insurers licensed to do business in Australia.

This study compared the MCRs calculated under the two methods and analyzed the prescribed method's capital charges using a model representative of industry best practice. Despite this, simplifying assumptions were made in the model's calibration. A lack of consensus remains as to the insurance liability volatility assumptions.

The results of this study, however, have highlighted some significant issues for both regulators and insurers. For the model insurer studied, the MCR calculated under the IMB method was significantly larger than the MCR under the prescribed method. The implication of this result is that despite APRA's desire for insurers to adopt an internal model for MCR calculation, there is an incentive for insurers to use the prescribed method to produce a lower MCR. This also highlights the need to develop prescribed methods that are consistent with the underlying risk of the insurer. To do this, the need for a diversification allowance is very important.

The results were shown to be highly sensitive to the insurance liability volatility assumptions. It is arguable, however, that the current capital charge levels in Australia are too low in order for the prescribed method to ensure a ruin probability in one year of less than 0.5% across the entire general insurance industry. This is likely to be very difficult to achieve. Differences between insurers of different sizes and with different business mixes should at least be considered more carefully in any revision of the prescribed method capital charges.

There is a strong case for including either diversification benefits or more stringent concentration charges in the prescribed method to address the risk reduction associated with a well diversified business mix and asset portfolio and to give a more consistent treatment of insurers with different characteristics.

The internal model's results rely heavily on its volatility assumptions. There is a major need for a study to be carried out using insurer level data to develop a consensus in the industry as to the level of insurance liability volatility that should be allowed in internal models for capital determination.

We can only conclude that there is much to be done by regulators and insurers if they are to adopt risk-based capital requirements. Some countries have taken a step along this path already. Australia has been one of the first countries to introduce a risk-based regulatory regime for non-life insurers, and its experience is no doubt of great interest to insurers, actuaries and regulators internationally. We have analyzed the capital requirements with a view to identifying lessons for others. There is still a long way to go before insurers will be in a position to confidently adopt the IMB method for the MCR calculation even in Australia.

8.2 Areas for Further Research

Our results depend to some extent on the degree to which the model and the assumptions used are representative of actual insurers. Our aim has been to use an internal model that broadly represents the insurer's business situation and parameters and assumptions based on industry best practice. We would expect any insurer that used a model similar to the one that we have used for the IMB approach would come to similar conclusions.

In this study, many simplifying assumptions were made in the model's calibration. We are not aware of any comprehensive study that has been completed that examines and assesses the appropriate insurance liability volatility assumptions taking into account actual insurer data and allowing for insurer-specific characteristics. This is a critical area of research required for risk-based regulatory capital if internal models are to be used with any confidence.

The modeling of claims correlation is another important area for further research. Dependency models need to be further considered. Brehm (2002) outlines a formal quantitative approach for estimating correlation from data. The Tillinghast and Trowbridge reports use a much more qualitative approach. Copulas also have great potential for modeling insurance liability dependencies, especially for tail events.

Despite these issues, the case study presented here identifies the issues and gives guidance for any insurer considering internal modeling for risk-based capital. There are important lessons at an international level because the approach adopted in Australia is similar to that proposed at the international level.

References

APRA. *General Insurance Prudential Standards and Guidance Notes.* Sydney, NSW: Australian Prudential Regulation Authority, July 2002. Available on line at <http://apra.gov.au>.

Artzner, P., Delbaen, F., Eber, J-M., and Heath, D. "Coherent Measures of Risk." *Mathematical Finance* 9 (1999): 203-228.

Bateup, R. and Reed, I. "Research and Data Analysis Relevant to the Development of Standards and Guidelines on Liability Valuation for General Insurance." Sydney, NSW: Tillinghast Towers-Perrin, November 2001. Prepared for the Institute of Actuaries of Australia and APRA and available on line at <http://actuaries.asn.au>.

Brehm, P.J. "Correlation and the Aggregation of Unpaid Loss Distributions." *Casualty Actuarial Society Fall Forum* (2002): 1-24.

Butsic, R.P. "Solvency Measurement for Property-Liability Risk-Based Capital Applications." *The Journal of Risk and Insurance* 61, no. 4 (1994): 656-690.

B&W Deloitte "How TSM Works—Calibration and Projection." Prepared for DFA Clients, July 2003.

B&W Deloitte "Prophet DFA User Manual." Prepared for DFA Clients, January 2003.

Casualty Actuaries Society. "Report of the Operational Risks Working Party to GIRO 2002." Available on line at <http://www.casact.org>.

Collings, S. "APRA's Capital Adequacy Measures: The Prescribed Method v the Internal Model Method." Presented at the Institute of Actuaries of Australia *XIII*th General Insurance Seminar, November 2001. Available at <http://www.actuaries.asn.au>.

Collings, S. and White, G. "APRA Risk Margin Analysis." Presented at the Institute of Actuaries of Australia *XIII*th General Insurance Seminar, November 2001. <http://www.actuaries.asn.au>.

Cummins, J.D. "Risk-Based Premiums for Insurance Guaranty Funds." *The Journal of Finance* 43, no. 4 (1988): 823-839.

Duffie, D. and Pan, J. "An Overview of Value-at-Risk." *Journal of Derivatives* (Spring 1997): 7-49.

Dynamic Financial Analysis Committee. *DFA Handbook.* Arlington, VA: Casualty Actuarial Society, September 1995.

Embrechts, P., McNeil, A.J. and Straumann, D. "Correlation: Pitfalls and Alternatives." In *Extremes and Integrated Risk Management* (Paul Embrechts, Editor). London, England: Risk Books, 2000.

Financial Services Authority. "Enhanced Capital Requirements and Individual Capital Assessments for Non-Life Insurers." London, England: Financial Services Authority, 2003.

Frank, R.H. and Bernanke B.S. *Principles of Microeconomics.* Whitby, Ontario: McGraw Hill-Irwin, 2001.

Giese, G. "Economic Capital versus Regulatory Capital–A Market Benchmark." *Risk*, Basel II, 16, no. 5, May 2003. Available online at <http://www.risk.net>.

Gray, B. "Prudential Supervision of General Insurance–Policy Discussion Paper." Australian Prudential Regulation Authority, March 2001. Available online at <http://www.apra.gov.au>.

Gray, B. "Study of the Prudential Supervisory Requirements for General Insurers in Australia." Australian Prudential Regulation Authority, September 1999. Available online at <http://www.apra.gov.au>.

Holton, G. *Value-at-Risk: Theory and Practice.* San Diego, CA: Academic Press, 2003.

IAA Insurance Regulation Committee. "Report of Solvency Working Party." International Actuarial Association, February 2002. Available online at <http://www.actuaries.org>

IAA Insurer Solvency Assessment Working Party. "A Global Framework for Insurer Solvency Assessment." International Actuarial Association, May 2003. Available online at <http://www.actuaries.org>

IAA Insurer Solvency Assessment Working Party. "A Global Framework for Insurer Solvency Assessment: Appendix B—Non-life Case Study." International Actuarial Association, 2004. Available online at <http://www.actuaries.org>

Klugman, S.A., Panjer, H.H., and Willmot, G.E. *Loss Models: From Data to Decisions.* New York, NY: John Wiley and Sons, 1998.

Mack, T. "Distribution Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *ASTIN Bulletin* 23, no. 2 (1993): 213–225.

Myers, S.C. and Read Jr., J.A. "Capital Allocation for Insurance Companies." *Journal of Risk and Insurance* 68, no. 4 (2001): 545–580

Rockafellar, R.T. and Uryasev, S. "Conditional Value-at-Risk for General Loss Distributions." *Journal of Banking and Finance* 26 (2002): 1443–1471.

Sherris, M. "Equilibrium Insurance Pricing, Market Value of Liabilities and Optimal Capitalization." Presented at the 13th International AFIR Colloquium, Maastricht, Netherlands, September 2003. Available online at <http://www.actuarial.unsw.edu.au>.

Sherris, M. "Solvency, Capital Allocation and Fair Rate of Return in Insurance." Actuarial Studies Working Paper, University of New South Wales, Australia, 2004. Available online at <http://www.actuary.unsw.edu.au/files/sherriscapall.pdf>

Slater, D. and Gillott, N. "Calibration of the General Insurance Risk Based Capital Model." Prepared for the Financial Services Authority by Watson Wyatt UK, July 2003. Available at <http://www.fsa.gov.uk>.

Stevens, A. and Tadros, R. "Getting the Solvency Balance Right." *Risk* 16, no. 5 (May 2003). Available online at <http://www.risk.net>.

Sutherland-Wong, C. "A Comparison of Methods for APRA's General Insurance Capital Requirements." BCom Honours Thesis, Actuarial Studies, University of News South Wales, 2003.

Taylor, G. "An Equilibrium Model of Insurance Pricing and Capitalization." *Journal of Risk and Insurance* 62, no. 3 (1995): 408–446.

Thomson, R. "The Approved Actuary Regime in General Insurance - An APRA Reflection." *Actuary Australia* (July 2003): 12-13. Available online at <http://www.actuaries.asn.au>

Venter, G. "Tails of Copulas." Presented at the ASTIN Colloquium, Washington, USA, July 2001. <http://www.actuaries.org>

Wilcox, R.R. *Introduction to Robust Estimation and Hypothesis Testing.* San Diego, CA: Academic Press, 1997.

A New Hybrid Defined Benefit Plan Design

Wayne E. Dydo*

Abstract

Traditional defined benefit plans can be difficult to understand and complex to administer. Hybrid plans (cash balance and pension equity) arose in part to address the former issue, but at a price of greater administrative and litigation risk. I introduce a design for defined benefit pension plans that is easy to communicate to participants, allows for accrual patterns that closely replicate those of the two most common forms of hybrid pension plans, and avoids the controversial nondiscrimination issues that currently trouble sponsors of hybrid plans. The design defines the benefit as a fixed percentage of pay payable over a period of time, which period is built up over a participant's employment. When translated into a lifetime pension commencing at normal retirement age, an interesting pattern of accrual rates develops. Numerical examples and illustrations are provided, along with suggested uses for this type of plan.

Key words and phrases: *hybrid pension plan, actuarial equivalent, Section 417(e), Internal Revenue Code*

*Wayne E. Dydo, Ph.D., F.S.A., E.A., M.A.A.A., is an actuary at Watson Wyatt Worldwide. He received his Ph.D. in mathematics from the University of Notre Dame in 1973. Shortly thereafter he joined Aetna Life and Casualty, attained his F.S.A. in 1976, and joined Watson Wyatt in 1977. His expertise includes design, implementation and funding of qualified defined benefit plans, development of special studies with respect to analyses of the adequacy of pension plan benefits, funding, and compliance with government-imposed requirements, cost analyses such as for union negotiations, and research and strategic planning for pension and other deferred compensation matters, forecast studies, and experience studies. Current research interest involves matching cash flows from a bond portfolio to expected benefit plan cash flows using short selling and linear programming techniques.

Dr. Dydo's address is: Watson Wyatt Worldwide, 1001 Lakeside Avenue, Suite 1900, Cleveland, OH 44114-1172, U.S.A. E-mail: *drwedwed@earthlink.net*

1 Introduction

A nontraditional or hybrid defined benefit plan exhibits some of the characteristics of a defined contribution plan from the perspective of the participant, but is funded and administered like a traditional defined benefit plan. The benefit formulation under a hybrid plan is something other than a lifetime annuity commencing at normal retirement age. The cash balance hybrid establishes a notional account for each participant that each year is credited with a percentage of pay and an interest credit, both at plan-specified rates. The pension equity hybrid defines the benefit as a lump sum equal to an accumulated percentage of final average earnings. The accumulated percentage results from annual percentage credits that generally vary by age or a combination of age and service.

The popularity of hybrid plans can be traced to two features: (i) the presentation of the benefit as a lump sum that participants understand, and (ii) the accrual pattern, i.e., the way the benefits build up during a participant's employment with the plan's sponsor. A hybrid plan's accrual pattern typically provides more benefit during the earlier years of employment than is the case with traditional defined benefit designs.

Despite the popularity and appeal of hybrid designs, the courts, regulators, and politicians are attempting to significantly alter the utility of these types of plans. For example, in July 2003 the federal district court for the southern district of Illinois ruled in the Cooper v. IBM case[1] that both the pension equity plan and the cash balance formulas in IBM's plan violate federal age discrimination law because the rate of accrual for the age 65 annuity benefit declines with age. This decline occurs because of the time value of money: \$1,000 invested at age 25 will produce a higher lifetime annuity at age 65 than \$1,000 invested at age 45, assuming a positive investment return each year.

In this paper I introduce a new defined benefit plan design, called an *annuity certain plan*, which employers may find appealing given the current legal and political complications surrounding hybrid defined benefit plans. The annuity certain plan design has several attractive attributes:

- it reduces mortality risk for the sponsor,
- it simplifies the plan valuation,
- it encompasses a simple benefit formulation that participants can understand, and

[1] Cooper v. IBM Personal Pension Plan, 274 F.Supp.2nd 1010 (S.D. Ill. 2003).

- its benefit formulation depends solely on service and annuity factors at normal retirement age.

This last feature implies that benefit accrual rates do not decrease on account of age, thus avoiding a principal objection currently being raised against hybrid plans.

The paper is organized as follows: Section 2 gives an overview of the annuity certain plan design. More details of the design and uses of annuity certain plan are described in Section 3. Closing comments are given in Section 4.

2 Annuity Certain Plans

2.1 An Overview

By law,[2] a tax-qualified defined benefit plan is required to provide a lifetime benefit commencing at retirement. As this commitment extends over decades, it subjects both the plan and the participant to mortality risk, albeit in opposite ways. The plan could end up paying benefits over longer than expected time periods if the participant lives longer than expected, whereas a participant who dies prematurely forfeits substantial benefit value. Providing a lump sum option, however, allows the plan to remove the mortality risk if the participant chooses the lump sum option, but at a high cost to the plan, especially in a low interest rate environment.

The annuity certain plan attempts to reduce the employer's mortality risk by defining the benefit as a temporary, but guaranteed, benefit. The plan may be thought of as an extended severance plan. The participant receives a defined percentage of pay (career average or final average) payable for a fixed period of time commencing at normal retirement age. As the payments continue whether the participant is alive or dead, these payments are called an "annuity certain." The period of time over which the benefit is paid will vary by the amount of time the employee has worked for the employer (service): the longer the service, the longer the benefit will be paid. If benefits are allowed to start before normal retirement age, the size of the retirement benefit is reduced using a fixed interest rate.

A defined benefit plan must express the plan's benefit as a lifetime annuity commencing at normal retirement age, which for the remainder of this paper is assumed to be age 65. This benefit is called the

[2]Section 411(a)(7) of the Internal Revenue Code of 1986, as amended; Treas. Reg. Section 1.411(a)-7(a)(1).

accrued normal retirement benefit. The accrued normal retirement benefit would be determined from the pension benefit payable for the calculated period of time (I refer to this as the annuity certain benefit) using an actuarial equivalence assumption set of unisex mortality (as prescribed in section 417(e) of the Internal Revenue Code) and a fixed low interest rate (possibly different from the rate used to reduce the benefit for early commencement).

Determining the participant's monthly retirement benefits paid from the termination of employment is not particularly difficult. At the moment of the termination of employment, let

$n =$ Participant's years of service;

$\text{Ben} =$ Participant's monthly retirement benefits;

$f_{65} =$ Fraction of participant's final average monthly salary;

$m =$ Number of months credited for each year of service; and

$\text{N} =$ Total number of months credited.

The participant's monthly retirement benefits paid from the termination of employment is

$$\text{Ben} = f_{65} \times \text{Final Average Monthly Salary}, \tag{1}$$

which is paid for $\text{N} = mn$ months.

For example, suppose the plan pays $f_{65} = 45\%$ of the participant's final average monthly salary for $m = 4$ months for each year of service, commencing at age 65. A participant with $n = 20$ years of service would lead to $\text{N} = 80$ months of pension payment, with each payment equal to 45% of the participant's monthly final average salary. These payments would start when the participant reaches age 65 and continue thereafter for the 80-month period irrespective of the survival of the participant.

A short period of payment for the short-service employee may seem inadequate, but keep in mind that the payments are substantial, equaling 45% of average monthly salary. Below I show that the equivalent lifetime annuity benefit for an employee with 5 years of service is 5.65% of average salary, which is a typical lifetime pension for short service.

Commencement of retirement benefits at 55 or later could be allowed, but the annuity certain benefit must be reduced for each year retirement age precedes age 65. If i is the discount interest rate, let f_x denote the fraction of participant's final average monthly salary at termination of employment at age x that is used to determine the benefit paid. It follows that

$$f_x = f_{65} \times (1 + i)^{-(65-x)}.$$

For example, if the discount rate were $i = 6\%$, then instead of the 45% of average salary payable at 65, the monthly benefit would be reduced to 25.13% of average salary if payments were to commence at age 55. The length of the certain payment period does not change.

As noted above, a tax-qualified defined benefit plan must express an accrued normal retirement benefit in the form of a lifetime annuity. This accrued normal retirement benefit is defined as the actuarial equivalent of the accrued annuity certain benefit payable at normal retirement age (65) using 4% interest and section 417(e) mortality. The relatively low interest rate minimizes the likelihood of section 417(e)-required benefit improvements.

Specifically, section 417(e) of the Internal Revenue Code requires the present value of any optional form of benefit that is not paid over the lifetime of the participant be at least as great as the present value of the normal retirement benefit computed using the prescribed mortality table and interest rate (currently the 30-year Treasury security rate). For the annuity certain plan this means the annuity certain benefit under the plan must be at least as great as the actuarially equivalent annuity certain benefit computed from the accrued normal retirement benefit using the section 417(e) required interest and mortality.

If the annuity certain plan provides a benefit only at normal retirement age, section 417(e) will be satisfied so long as the plan's actuarial equivalent interest rate is less than or equal to the section 417(e) rate. A 4% plan rate should satisfy this requirement. If a reduced annuity certain benefit is provided before normal retirement age, then section 417(e) will be satisfied as long as the early commencement discount rate is not too large. A careful examination of the actuarial formulas involved in determining whether or not section 417(e) is satisfied will show that the rate needs to be tested only at the earliest retirement age and for the longest expected annuity certain period under the plan.

Table 1 shows the age-55 ratio of the minimum annuity certain benefit under 417(e) to the annuity certain benefit under the plan formula. Various section 417(e) interest rates and plan early commencement discount rates are illustrated. The certain period is 15 years and the plan's actuarial equivalence rate is 4%. A ratio of 100% or lower means that section 417(e) is satisfied. For example, if the plan uses a 6% early commencement discount rate and allows benefit commencement as early as age 55, then for all plan benefits less than or equal to 15 years in duration, section 417(e) is satisfied so long as the section 417(e) rate is 5.1% or higher. Notice that if the plan's formula allows for four months of payment for each year of service, then 45 years of service would be needed to accrue the 15-year period benefit.

Table 1
Ratio of 417(e) Minimum Benefit to Plan Formula Benefit For 15-Year Annuity Certain Benefit at Age 55

417(e)	Early Commencement Discount Rate					
Rate	4.50%	5.00%	5.50%	6.00%	6.50%	7.00%
4.50%	92.9%	97.5%	102.2%	107.2%	112.3%	117.7%
4.75%	90.2%	94.6%	99.2%	104.0%	109.0%	114.3%
5.10%	86.6%	90.8%	95.2%	99.8%	104.6%	109.6%
5.25%	85.1%	89.2%	93.5%	98.1%	102.8%	107.7%
⋮						

The accrued normal retirement benefit, expressed as a percentage of final average annual pay, can be determined as follows: let

n = Number of years participant has worked for employer;

m = Number of months of payment earned for each year of employment; and

x = Participant's retirement age, $x = 55, 56, \ldots, 65$;

r = Actuarial equivalence rate, which is the interest rate used in converting a benefit from one form to another;

f_x = Percentage of average monthly earnings payable at age x. Generally, if i denotes the early commencement discount rate and 65 is the normal retirement age, then

$f_x = f_{65}(1+i)^{-(65-x)}$;

$P(x,n)$ = Participant's accrued retirement benefit percentage at age x; and

$k = (1+r)^{-m/12}$.

Note that the actuarial equivalence rate, r, is an effective annual rate specified in the plan document that is used to determine the value of a benefit form under the plan in order to convert the form into an equivalent form. The quantity $\ddot{a}^{(r)}_{\overline{m/12}|r}$ is the value of 1 payable in equal monthly installments at the beginning of each month for m months at an effective annual rate r. Thus $12 f_x \ddot{a}^{(r)}_{\overline{m/12}|r}$ represents the value of the benefit earned for the first year of service expressed as a percentage of final average monthly earnings and payable monthly beginning at age x. The value of the benefit earned for the second year of service is k times this amount because k represents the discounted value (at rate r) of 1 deferred for $m/12$'s of a year. Similarly, the value of the benefit earned for each successive year of service is k times the value of the benefit earned for the prior year of service. Dividing by $12\ddot{a}_x^{(12)}$ converts the benefit into a lifetime annuity at age x. It follows that

$$\begin{aligned} P(x,n) &= f_x \times \left(\frac{\ddot{a}^{(r)}_{\overline{m/12}|r}}{\ddot{a}_x^{(12)}} \right) \left(1 + k + \cdots + k^{n-1}\right) \\ &= \left(\frac{\ddot{a}^{(r)}_{\overline{m/12}|r}}{\ddot{a}_x^{(12)}} \right) \left(\frac{1-k^n}{1-k} \right) f_x. \end{aligned} \tag{2}$$

Formula (2) presents the participant's accrued lifetime annuity benefit payable at age x as a percentage of final monthly average earnings given the participant has n years of service. Formula (2) shows that each year's accrual is k times the previous year's accrual and is independent of the age of the participant when the benefit is accrued. The factor k is the ratio of the annual accrual in one year to the annual accrual in the prior year for a given commencement age. So the series of accrual rates for a given commencement age forms a decreasing geometric series: the longer the payment period per year of service, the greater the rate of decrease. This decreasing series of accrual rates ensures that the accrual rules of section 411(b) are satisfied and also distinguishes this type of plan from other hybrids. Also notice that these accrual rates (percentages of final average pay paid as a lifetime annuity) depend only on service and age at commencement, not current age, so the plan satisfies the requirement of the law that the rate of accrual is independent of age (e.g., the accrued normal retirement benefit for a 40-year-old with ten years of service is the same as for a 50-year-old with ten years of service).

Table 2 shows the accrued normal retirement benefit expressed as a percentage of final average annual pay using the above example of the plan paying $f_{65} = 45\%$ of monthly average pay, $m = 4$ months of payment per year of service, $r = 4\%$ actuarial equivalence rate, $i = 6\%$ early commencement discount rate.

Table 2
$P(x, n)$ for Various Years of Service (n) And Commencement Ages (x)

Years of	Commencement Age (x)			
Service (n)	65	62	60	55
1	1.16%	0.90%	0.76%	0.51%
5	5.65%	4.38%	3.71%	2.49%
10	10.95%	8.49%	7.19%	4.83%
20	20.55%	15.94%	13.51%	9.07%
30	28.98%	22.47%	19.05%	12.79%

3 Design and Uses of Annuity Certain Plans

3.1 Other Design Features

The key variables of the plan are the pay replacement percentage, f; the period of payment per year of service, m; and the normal retirement age, 65. The replacement percentage could be integrated, (e.g., 30% for average pay up to covered compensation and 45% for excess pay). The periods of payment could vary by years of service, e.g., six months of payment for the first five years of service and four months thereafter, depending upon the desired accrual pattern.

Optional annuity forms probably should be restricted to the straight life and 50% joint and surviving spouse forms. With respect to pre-annuity-commencement death benefits, there are a number of choices. As a minimum, the plan would need to provide the pre-retirement surviving spouse annuity required under the law: 50% of the accrued normal retirement benefit, reduced to an early commencement date selected by the spouse using plan actuarial equivalence assumptions, and converted into a 50% joint and survivor annuity as if the participant had survived to that date and selected the 50% joint and survivor annuity as the optional form. Or the plan could provide the full accrued annuity certain benefit to the spouse, if the spouse survives to the earliest retirement date for the participant. Or the plan could provide the full annuity certain benefit to any designated beneficiary.

3.2 Comparison with Cash Balance and Pension Equity Plans

A typical age-graded cash balance plan might have the benefit structure shown in the first row of Table 3. With respect to a pension equity plan, the second row of Table 3 shows a reasonable age-graded design, where the entries in the table are the points (percentages) of final five-year average pay that serve to define the lump sum value under the plan.

Table 3
Pay Credits for a Typical Age-Graded Plan Design

Plan	Age Groups						
	< 30	30–34	35–39	40–44	45–49	50–54	≥ 55
CBP	5%	6%	7%	8%	9%	11%	13%
PEP	5%	6%	8%	9%	11%	14%	18%

Notes: CBP = Cash balance plan, and PEP = Pension equity plan.

Table 4
Months of Payment Accrual
For a Service-Graded Annuity Certain Plan

Service (in Years)	1–5	6–10	11–15	16–20	21–25	≥ 26
Accrual (in Months)	6	6	5	5	4	3

Now consider a service-graded annuity certain plan as defined in Table 4. Assume that the benefit is 40% of final five-year average pay and a 4% interest rate is used to define the reduction for pre-65 commencement and to determine actuarial equivalence. If the value of the annuity certain at ages under 65 were to be paid as a lump sum using the 4% rate and if the cash balance plan credits 5% interest each year at year end and pay credits are also made at year end, then the lump sum value of the annuity certain plan, the accumulated value of the cash balance account, and the value of the account under the pension equity plan for an employee hired at age 35 and achieving 4% annual pay increases are almost identical as shown in Figure 1. Similar results (almost identical value curves) are obtained for other hire ages.

These examples show that the annuity certain plan can be designed to produce benefits that have lump sum values similar to those produced by the two most popular types of hybrid plans.

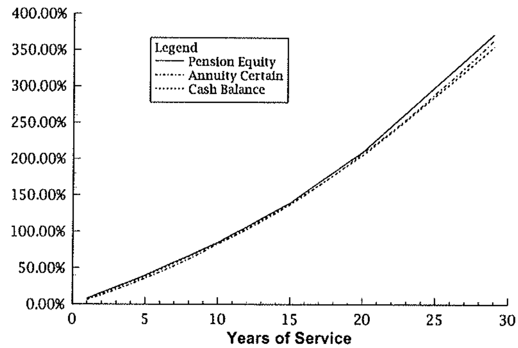

Figure 1: End of Service Value as a Percentage of Final Average Pay

With respect to valuation, the determination of funding and expense amounts could be handled easily in a spreadsheet. If no pre-retirement decrements were assumed and retirement were assumed to occur only at normal retirement age, the normal cost would simply be the discounted value (at valuation interest rate) of the number of projected payments expected to be earned during the year. Actuarial accrued liability would be the discounted value of the number of payments earned to date.

3.3 Uses of Annuity Certain Plans

Annuity certain plans might be appealing to professional service corporations whose defined contribution plans' contribution limits have been reached. An annuity certain plan would be a straightforward way for the corporation to defer additional compensation, especially because of the guaranteed series of payments (assuming he or she does not convert to a lifetime annuity) and the relative ease of valuation. Investment risk could be reduced significantly if the plan were to be funded with conservative fixed-income investments of appropriate maturities.

This design might also be used for a supplemental executive retirement plan for a larger company. Suppose, for example, that the company wished to encourage executive retirement at age 60, and that its qualified pensions plans (either defined benefit or defined contribution) provided no subsidies for pre-age 65 retirement. A 60% plan with six months of payment for each year of service, with a maximum of 15 years counted, would allow for such early retirement by allowing the executive to defer commencement of his qualified benefits for up to 7.5 years. Used in this manner, the annuity certain plan might be viewed as a way to provide early retirement subsidies not available in the qualified plan.

In general, then, the annuity certain plan may be beneficial for executives retiring at earlier than typical ages. Executives tend to retire at earlier ages than other employees and the annuity certain form of payment provides a bridge that would allow the executives to manage their investments more aggressively over a longer period or, perhaps, delay commencement of benefits provided by other plans until normal retirement age

Also, the simplicity of the design should make it easy for executives and their financial planners to place a value on the benefit. It has been our experience that financial planners are often confused by traditional defined benefit plans. Measuring the value of the benefit from an annuity certain arrangement does not require any background in life contingencies.

The current legal and regulatory objections to hybrid plans do not apply to the annuity certain plan because benefits are based solely on service and are converted into a single life annuity using only an age-65 annuity factor. Larger corporations looking for a defined benefit plan with a hybrid-like accrual pattern coupled with an easy-to-understand formulation should consider a plan of this type.

Finally, a plan of this type could be designed to complement a company's 401(k) plan by allowing deferral of the date of benefit commencement for the 401(k) plan. This deferral period could provide employees a safeguard against market decline in the several years immediately before and after retirement.

4 Closing Comments

Defining a retirement benefit as a limited number of monthly payments, paid regardless of survivorship, linked to both the time period of employment and the earnings either throughout or near the end of the employment period, and commencing on a fixed date are the defining features of an annuity certain plan. This type of defined benefit plan can provide value accrual patterns similar to those of today's typical hybrid plans. In addition, it should not run afoul of actual and proposed legal restrictions.

Journal of Actuarial Practice Vol. 12, 2005

A Primer on Duration, Convexity, and Immunization

Leslaw Gajek,* Krzysztof Ostaszewski,† and Hans-Joachim Zwiesler‡

Abstract§

The concepts of duration, convexity, and immunization are fundamental tools of asset-liability management. This paper provides a theoretical and practical overview of the concepts, largely missing in the existing literature on the subject, and fills some holes in the body of research on the subject. We not present new research, but rather we provide a new presentation of the underlying theory, which we believe to be of value in the new North American actuarial education system.

Key words and phrases: *duration, convexity, M-squared, immunization, yield curve, term structure of interest rates.*

*Leslaw Gajek, Ph.D., is a professor of mathematics. He holds a Ph.D. in statistics from the Technical University of Lodz and Habilitation in statistics from the Polish Academy of Sciences. He was President of the Social Security Administration of Poland from 1999 to 2001.

Dr. Gajek's address is: Instytut Matematyki, Politechnika Łódzka, Wólczańska 215, 93-005 Łódź, POLAND. E-mail: *gal@p.lodz.pl*

†Krzysztof Ostaszewski, Ph.D., F.S.A., C.F.A., is a professor of mathematics and director of the actuarial program at Illinois State University. He holds a Ph.D. in mathematics from the University of Washington. He was a Fulbright Research Fellow in 1995 and a Fulbright Senior Specialist in 2003 and 2004.

Dr. Ostaszewski's address is: Actuarial Program, Illinois State University, Normal, IL 61790-4520, U.S.A. E-mail: *krzysio@ilstu.edu*

‡Hans-Joachim Zwiesler, Ph.D., is a professor of actuarial science. He is also the Managing Director of the Center for Applications of Finance and Insurance at the University of Ulm and the President of the Institute of Finance and Actuarial Sciences in Ulm. He received his Ph.D. and Habilitation in Actuarial Science from the University of Ulm.

Dr. Zwiesler's address is: Institut Fuer Finanz-Und Aktuarwissenschaften, Universitaet Ulm, Helmholzstrasse 22, Ulm D-89081, GERMANY. E-mail: *zwiesler@mathematik.uni-ulm.de*

§The work of Drs. Gajek and Ostaszewski was partly supported by the KBN Grant 2H02B 011 24.

1 Introduction

The concepts of duration and convexity are commonly used in the field of asset-liability management. They are important because they provide key measures of sensitivity of the price of a financial instrument to changes in interest rates and they help develop methodologies in interest rate risk management. Traditional approaches used by financial intermediaries often allowed for borrowing at short-term interest rates, relatively lower, and investing at longer-term interest rates, relatively higher, hoping to earn substantial profits from the difference in the level of the two interest rates. Interest rate risk management utilizing the concepts of duration and convexity helps point out the dangers of such a simplistic approach and develops alternatives to it. Thus, the thorough understanding of these two concepts must be an important part of the education of today's actuaries. In North America, the introduction of the concepts of duration and convexity now occur fairly early in the actuarial examination process. The new Society of Actuaries examination system starting in May 2005 will introduce these concepts in the new Financial Mathematics (FM) examination at the level of old Course 2 Society of Actuaries examination. They also are presented in the Society of Actuaries Course 6 examination, as well as Casualty Actuarial Society Examination 8, based on the more theoretical approach of Panjer (1998) and the more practical ones of Fabozzi (2000) or Bodie, Kane, and Marcus (2002).

There is a split in the way duration and convexity are generally presented in the finance and actuarial literature: from a theoretical perspective as rates of change or from a practical perspective as weighted average time to maturity (for duration) or weighted average square of time to maturity (for convexity). These two perspectives are naturally connected, but the nature of connection are not explicitly discussed in the educational actuarial literature.

The objective of this paper is to fill the existing void and give a general overview of the two fundamental concepts. This paper is presented at the level where it is accessible to students who have completed three semesters of calculus and one or two semesters of probability, i.e., at the level of the current Course P Society of Actuaries examination on probability, and have a working knowledge of the theory of interest as presented in the text by Kellison (1991). We hope that this paper will allow future actuaries to combine the theoretical and the practical approaches in their education and training.

2 Duration

2.1 Duration as Derivatives

Duration is a measure of the sensitivity of a financial asset to changes in interest rates. It is based on the assumption of using only one interest rate, which commonly is interpreted as a flat yield curve assumption. As a change in an interest rate amounts to a parallel shift in a flat yield curve, use of duration also commonly is said to assume a parallel shift in the yield curve.

For a given interest rate i, let δ denote the corresponding force of interest, which satisfies $\delta = \ln(1+i)$. Thus if P is the price of a financial asset, we often write P as a function of the interest rate i as $P(i)$ or as a function of the force of interest δ, $P(\delta)$. This notation is necessitated by the simultaneous use of the interest rate and the force of interest in our presentation.

Definition 1. *The duration of a security with price P is*

$$D(P) = -\frac{1}{P}\frac{\mathrm{d}P}{\mathrm{d}i} = -\frac{d}{\mathrm{d}i}\ln(P). \tag{1}$$

We should emphasize the following features of this definition: (i) it makes no assumptions about the type or structure of the security; (ii) it applies whether or not the cash flows of the security are dependent on interest rates; (iii) it applies whether or not the security is risk-free; and (iv) it applies whether or not the security contains interest rate options. This definition applies to *all* securities, including bonds, mortgages, options, stocks, swaps, interest-only strips, etc. Later in this paper we will analyze this definition under some specific assumptions about the security.

The term $-\mathrm{d}P/\mathrm{d}i$ usually is termed the *dollar duration* of the security. We propose to abandon this term for a less restrictive one: *monetary duration*, which we believe to be better because of lack of reference to a specific national currency.

Because of the standard approximation of the derivative with a difference quotient, we see that for sufficiently small Δi:

$$D(P) \approx \frac{P(i-\Delta i) - P(i)}{P(i)\,\Delta i} = \frac{\dfrac{P(i-\Delta i) - P(i)}{P(i)}}{\Delta i}. \tag{2}$$

Equation (2) means that duration gives us the approximate ratio of the percentage loss in the value of the security per unit of interest rates, a commonly used approximation. Note also that because the loss in the value of the security $[P(i - \Delta i) - P(i)]/P(i)$ is expressed as percentage and Δi is in percent per year (if the interest rate used is annual, a common standard), the unit for duration is a year (or, in general, the time unit over which the interest rate is given).

Instead of defining duration in terms of the derivative with respect to the interest rate, one could define duration with respect to the force of interest as follows:

Definition 2. *The Macaulay duration of a security with price P is*

$$D_M(P) = -\frac{1}{P}\frac{\mathrm{d}P}{\mathrm{d}\delta} = -\frac{d}{\mathrm{d}\delta}\ln P. \tag{3}$$

Clearly these two definitions of duration are connected because

$$\frac{\mathrm{d}P}{\mathrm{d}i} = \frac{\mathrm{d}P}{\mathrm{d}\delta}\frac{\mathrm{d}\delta}{\mathrm{d}i}.$$

Hence it follows that

$$D(P) = \frac{1}{1+i}D_M(P). \tag{4}$$

Suppose we have n securities, and let $\mathrm{Dur}(P_k)$ be either the duration or Macaulay duration of the k^{th} security whose price is $P_k > 0$, for $k = 1, 2, \ldots, n$. If a security has price $P > 0$ that is a linear combination of the prices of these n securities, i.e.,

$$P = b_1 P_1 + b_2 P_2 + \cdots + b_n P_n \tag{5}$$

where the b_ks are constants, then it follows directly from the definition of duration or Macaulay duration that:

$$\mathrm{Dur}(P) = \sum_{k=1}^{n} b_k \frac{P_k}{P}\mathrm{Dur}(P_k). \tag{6}$$

2.2 Duration as Weighted Averages

Let A_t denote the known non-zero cash flow at time t produced by a security under consideration, and let $\mathcal{T}$ denote the set of future time points at which the security's cash flow occurs. For simplicity we further assume that A_t does not depend on i. Throughout this paper we say a security has deterministic cash flows when its cash flows do not depend on the interest rate. At first, we will assume that the cash flows are discrete and that there is only one interest rate regardless of maturity (i.e., the yield curve is flat).

Then the present value of the security, i.e., its price, is:

$$P = \sum_{t \in \mathcal{T}} \frac{A_t}{(1+i)^t}. \tag{7}$$

In this case, monetary duration is given by:

$$-\frac{\mathrm{d}P}{\mathrm{d}i} = \sum_{t \in \mathcal{T}} \frac{tA_t}{(1+i)^{t+1}} = \frac{1}{(1+i)} \sum_{t \in \mathcal{T}} t\,\mathrm{PV}(A_t), \tag{8}$$

where

$$\mathrm{PV}(A_t) = \frac{A_t}{(1+i)^t}$$

is the present value of the cash flow A_t. The duration of this security is therefore:

$$D(P) = \frac{1}{P} \sum_{t \in \mathcal{T}} \frac{tA_t}{(1+i)^{t+1}} = \frac{1}{(1+i)} \sum_{t \in \mathcal{T}} t\,w_t \tag{9}$$

where w_t is the weight function

$$w_t = \mathrm{PV}(A_t)/P. \tag{10}$$

Thus duration turns out to be a weighted average time to maturity, modified by the factor $1/(1+i)$. For this reason, the concept of duration as introduced in equation (1) is commonly called *modified duration* for securities with deterministic cash flows. For securities with cash flows that are dependent on interest rates, which causes the cash flows to be random in nature if interest rates are random, duration is most often termed *effective duration*. For such securities, however, duration still is defined as in equation (1).

The weighted average time to maturity concept is actually the original idea of duration. For a security with deterministic cash flows, Macaulay (1938) defined duration as

$$D_M(P) = \frac{1}{P} \sum_{t \in \mathcal{T}} \frac{tA_t}{(1+i)^t} = \sum_{t \in \mathcal{T}} t\,w_t. \tag{11}$$

If the weights w_t are positive, we can introduce a discrete random variable T with probability distribution with $\mathbb{P}(T=t) = w_t$. It then becomes clear from equation (11) that the Macaulay duration is the expected value of T for this probability distribution, i.e., $D_M(P) = \mathbb{E}(T)$.

We quickly can see from equation (11) that the duration of a single payment at a future time t is $t/(1+i)$ and its Macaulay duration is t.

2.3 Duration Using Nominal Rates

Duration can be calculated with respect to nominal interest rates, such as semi-annual rates ($i^{(2)}$), quarterly rates ($i^{(4)}$), or ($i^{(12)}$). Recall the definition of a nominal rate, $i^{(m)}$, as given in Kellison (1991):

$$\left(1+\frac{i^{(m)}}{m}\right)^m = (1+j)^m = 1+i = e^{\delta}$$

where $j = i^{(m)}/m$. Therefore

$$\frac{\mathrm{d}i}{\mathrm{d}i^{(m)}} = \left(1+\frac{i^{(m)}}{m}\right)^{m-1} = \frac{1+i}{1+\frac{i^{(m)}}{m}} \tag{12}$$

and

$$\frac{\mathrm{d}\delta}{\mathrm{d}i^{(m)}} = \frac{\mathrm{d}\delta}{\mathrm{d}i}\cdot\frac{\mathrm{d}i}{\mathrm{d}i^{(m)}} = \frac{1}{1+\frac{i^{(m)}}{m}}. \tag{13}$$

The definition of duration with respect to $i^{(m)}$ is

$$D^{(m)}(P) = -\frac{1}{P}\frac{\mathrm{d}P}{\mathrm{d}i^{(m)}}. \tag{14}$$

It is easy to prove that

$$D^{(m)} = \frac{1+i}{1+\frac{i^{(m)}}{m}}D = \frac{1}{1+\frac{i^{(m)}}{m}}D_M. \tag{15}$$

It is not common to consider the case of continuous stream of payments for calculation of duration, because such securities do not exist in reality. We briefly consider such hypothetical securities for purely theoretical purposes. Suppose a security has continuous cash flows of $A_t\mathrm{d}t$ in $(t, t+\mathrm{d}t)$, with a constant force of interest δ. The security's price, P, is given by

$$P(\delta) = \int_0^{\infty} e^{-t\delta}A_t\mathrm{d}t \tag{16}$$

assuming the integral exists. Its Macaulay duration is:

$$D_M\left(P\left(\delta\right)\right) = -\frac{P'\left(\delta\right)}{P\left(\delta\right)} = \frac{\int\limits_0^\infty te^{-t\delta}A_t\mathrm{d}t}{\int\limits_0^\infty e^{-t\delta}A_t\mathrm{d}t}. \tag{17}$$

2.4 Some Examples

Thus far we have assumed that the security's cash flows do not depend on the interest rate. What if there is such dependence? We will now consider a few examples of such securities.

Example 1. Consider a discrete security paying a cash flow $A_t = e^{t\delta}$ at a single time t. Its price (present value) is $P = 1$. As $-\mathrm{d}P/\mathrm{d}\delta = 0$, and the duration of this security is zero. From equation (6), any linear combination of instruments like this, paying the accumulated value of a monetary unit at a given interest rate, also will have duration of zero.

Example 2. Similarly, if a discrete security with a single cash flow of $A_t = e^{(t-1)\delta}$ at time t, its price is $P = e^{-\delta}$ and duration of 1.

These two examples illustrate the well-known fact that floating rate securities[1] indexed to a short term rate (i.e., rate that resets somewhere between times 0 and 1 year) have durations between 0 and 1. By using the same argument, one can show that the duration of a floating rate security that resets every n-years and with no restrictions on the level of the new rate after reset (so that the new rate can fully adjust to the market level of the interest rates) is the same as the duration on an otherwise identical n-year bond.

Example 3. Consider a security that is an n-year certain annuity-immediate with level payments of $1/m$ made m times per year for n years, i.e., payments are made at times $1/m$, $2/m$, ...,$(nm-1)/m$. Assuming a constant interest rate to maturity of i, the price of this security is $P = a^{(m)}_{\overline{n}|i}$. It follows that the Macaulay duration of this security is:

$$\begin{aligned} D_M\left(a^{(m)}_{\overline{n}|i}\right) &= \frac{1}{a^{(m)}_{\overline{n}|i}} \sum_{k=1}^{nm} \frac{k}{m} \cdot \frac{1}{m} \cdot (1+i)^{-\frac{k}{m}} \\ &= \frac{1}{d^{(m)}} - \frac{n}{((1+i)^n - 1)}, \end{aligned} \tag{18}$$

[1]Floating rate securities are securities whose coupons *reset*, i.e., change in a manner consistent with the market level of interest rates.

where $d^{(m)} = 1 - (1+i)^{-\frac{1}{m}}$.

Example 4. On the other hand, if the security under consideration is a continuous annuity paid for n years, then its price is $P = \bar{a}_{\overline{n}|\,i}$ and its Macaulay duration is

$$D_M\left(\bar{a}_{\overline{n}|\,i}\right) = \frac{1}{\delta} - \frac{n}{e^{n\delta} - 1}. \tag{19}$$

The price and duration follow directly from those in example 3 above by letting $m \to \infty$ in $a^{(m)}_{\overline{n}|\,i}$ and in equation (18).

Note that the second terms in (18) and in (19) are identical. When $n \to \infty$, the limit is $1/d^{(m)}$, which is the price of a discrete perpetuity-immediate, and $1/\delta$, which is the price of a continuous perpetuity. Note that the duration of a continuous perpetuity is its price.

Example 5. What would be the Macaulay duration of a perpetuity-due? As every payment of such a perpetuity arrives exactly an m^{th} of a year before the corresponding payment of a perpetuity-immediate, its Macaulay duration is

$$\frac{1}{d^{(m)}} - \frac{1}{m} = \frac{1}{i^{(m)}}.$$

Thus the Macaulay duration of a perpetuity-due is the price of the corresponding perpetuity-immediate, while the Macaulay duration of a perpetuity-immediate is the price of the corresponding perpetuity-due.

Example 6. Finally, consider a security that is a risk-free bond with principal value of one dollar, maturing n years from now, paying an equal coupon of $r^{(m)}/m$ per unit of principal value m times a year at the end of each m^{th} of a year, with $i^{(m)}$ being the nominal annual yield interest rate compounded m times a year at the time of bond issue and i being the annual effective interest rate. The price of this bond is

$$P = r^{(m)} a^{(m)}_{\overline{n}|\,i} + (1+i)^{-n}$$

and its Macaulay duration, calculated here as a weighted-average time to maturity as in equation (11), is:

$$
\begin{aligned}
D_M(P) &= \frac{1}{P}\left[\sum_{k=1}^{nm}\left(\frac{k}{m}\cdot\frac{r^{(m)}}{m}\cdot(1+i)^{-\frac{k}{m}}\right)+n\cdot(1+i)^{-n}\right] \\
&= \frac{1}{P}\left[r^{(m)}\,(Ia)^{(m)}_{\overline{n}|i}+n(1+i)^{-n}\right] \\
&= \frac{1}{P}\left[\frac{r^{(m)}}{i^{(m)}}\ddot{a}^{(m)}_{\overline{n}|i}+\left(\frac{r^{(m)}}{i^{(m)}}-1\right)n(1+i)^{-n}\right].
\end{aligned}
\tag{20}
$$

If the bond is currently trading at par then $r^{(m)} = i^{(m)}$ so that the price of the bond is $P = 1$ and its Macaulay duration reduces to

$$D_M(P) = \ddot{a}^{(m)}_{\overline{n}|i}. \tag{21}$$

2.5 Effective Duration

In the above examples there was a direct functional relationship between the cash flows and interest rate. In practice, however, securities have complex relationships between cash flows and interest rates, and one cannot generally write a direct functional relationship between the cash flows and interest rate. In such cases duration is usually estimated rather than directly calculated.

The standard approximation approach is to use the Taylor series expansion of the price as a function of interest rate:

$$P(i+\Delta i) = P(i) + \frac{dP}{di}\Delta i + \frac{1}{2}\frac{d^2P}{di^2}(\Delta i)^2 + \cdots \tag{22}$$

Ignoring terms involving $(\Delta i)^2$ and higher yields

$$-\frac{dP}{di}\frac{1}{P} \approx \frac{P(i) - P(i+\Delta i)}{(\Delta i)\,P(i)} \tag{23}$$

and

$$-\frac{dP}{di}\frac{1}{P} \approx \frac{P(i-\Delta i) - P(i)}{(\Delta i)\,P(i)}. \tag{24}$$

We obtain a commonly used approximation of duration by averaging the right side of equations (23) and (24) yields

$$D_E(P) \approx \frac{P(i-\Delta i) - P(i+\Delta i)}{2P(i)\,(\Delta i)}. \tag{25}$$

Because this approximation can deal with any interest rate and/or any default options embedded in the security, $D_E(P)$ is often called an *option-adjusted duration* or *effective duration*.

3 Convexity

For any security with price P, the quantity:

$$C(P) = \frac{1}{P}\frac{d^2P}{\mathrm{d}i^2} \tag{26}$$

is called the *convexity* of the security, and

$$C_M(P) = \frac{1}{P}\frac{d^2P}{\mathrm{d}\delta^2} \tag{27}$$

is called the *Macaulay convexity* of the security. As $PD_M(P) = -\mathrm{d}P/\mathrm{d}\delta$, the monetary duration of the security, we also have:

$$C_M(P) = -\frac{1}{P}\frac{d}{\mathrm{d}\delta}\left(P \cdot D_M(P)\right) = D_M^2(P) - \frac{dD_M(P)}{\mathrm{d}\delta}. \tag{28}$$

The quantity

$$M^2(P) = \frac{d^2(\ln(P))}{\mathrm{d}i^2} = -\frac{dD(P)}{\mathrm{d}i} \tag{29}$$

is called the *M-squared* of the security, while

$$M_M^2(P) = \frac{d^2(\ln P)}{\mathrm{d}\delta^2} = -\frac{dD_M(P)}{\mathrm{d}\delta} = C_M(P) - D_M^2(P) \tag{30}$$

will be termed the *Macaulay M-squared.*

For a security with discrete deterministic cash flows so that $P = \sum_{t\in\mathcal{T}} A_t e^{-\delta t}$, we have

$$\begin{aligned} C_M(P) &= \frac{1}{P}\sum_{t\in\mathcal{T}} t^2 e^{-\delta t} A_t \\ &= \sum_{t\in\mathcal{T}} t^2 w_t \end{aligned} \tag{31}$$

where w_t is defined in equation (10),

$$\begin{aligned} M_M^2(P) &= \frac{1}{P}\sum_{t\in\mathcal{T}}(t - D_M(P))^2 e^{-\delta t} A_t \\ &= \sum_{t\in\mathcal{T}} w_t (t - D_M(P))^2, \end{aligned} \tag{32}$$

and

$$\begin{aligned} \frac{dM_M^2(P)}{\mathrm{d}\delta} &= -\frac{1}{P}\sum_{t\in\mathcal{T}}(t - D_M(P))^3 e^{-\delta t} A_t \\ &= -\sum_{t\in\mathcal{T}} w_t (t - D_M(P))^3. \end{aligned} \tag{33}$$

Similar expressions can be developed for $C(P)$, $M^2(P)$, and $dM^2(P)/\mathrm{d}i$.

Equation (32) allows for a relatively simple and intuitive interpretation of Macaulay duration, Macaulay convexity, and Macaulay M-squared of a deterministic security. As we stated before, assuming cash flows are positive, Macaulay duration is the expected time to cash flow with respect to the probability distribution whose probability function (or probability density function, in the case of continuous payments) is $f_T(t) = w_t$. Macaulay convexity is the second moment of this random variable, and Macaulay M-squared is the variance of it. This means that Macaulay duration can be interpreted intuitively as the expected time until maturity of cash flows of a security, Macaulay M-squared is the measure of dispersion of the cash flows of the said security, and Macaulay convexity is a sum of Macaulay M-squared and the square of Macaulay duration.

By the chain rule of calculus,

$$\frac{\mathrm{d}P}{\mathrm{d}i} = \frac{1}{(1+i)}\frac{\mathrm{d}P}{\mathrm{d}\delta} \tag{34}$$

and

$$\begin{aligned} \frac{d^2P}{\mathrm{d}i^2} &= \frac{d}{\mathrm{d}i}\left(\frac{1}{1+i}\frac{\mathrm{d}P}{\mathrm{d}\delta}\right) \\ &= -\frac{1}{(1+i)^2}\frac{\mathrm{d}P}{\mathrm{d}\delta} + \frac{1}{(1+i)^2}\frac{d^2P}{\mathrm{d}\delta^2}, \end{aligned} \tag{35}$$

which means that

$$C = \frac{1}{(1+i)^2} D_M + \frac{1}{(1+i)^2} C_M. \tag{36}$$

For $M^2 = C - D^2$, we easily can prove that

$$\begin{aligned} M^2 &= \frac{1}{(1+i)^2} C_M + \frac{1}{(1+i)^2} D_M - \frac{1}{(1+i)^2} D_M^2 \\ &= \frac{1}{(1+i)^2} M_M^2 + \frac{1}{(1+i)^2} D_M. \end{aligned} \tag{37}$$

For a security with discrete deterministic cash flows A_t (at time t) and price P given in equation (7), then

$$C = \frac{1}{(1+i)^2} \sum_{t \geq 0} t\,(t+1)\, w_t = \frac{1}{(1+i)^2} C_M + \frac{1}{(1+i)^2} D_M. \tag{38}$$

If this security consists of a single payment at time t, then its Macaulay convexity is t^2 and its convexity is

$$C = \frac{t^2}{(1+i)^2} + \frac{t}{(1+i)^2} = \frac{t\,(t+1)}{(1+i)^2}, \tag{39}$$

its M_M^2 is 0, and its M^2 is $t/\,(1+i)^2$.

Again, we suppose there are n securities. This time, however, we let $\mathrm{Conv}\,(P_k)$ be either the convexity or Macaulay convexity of the k^{th} security whose price is $P_k > 0$, for $k = 1, 2, \ldots, n$. If a security has price $P > 0$ given by equation (5), where the b_ks are constants, then it follows directly from the definition of convexity or Macaulay convexity that:

$$\mathrm{Conv}\,(P) = \sum_{k=1}^{n} b_k \frac{P_k}{P} \mathrm{Conv}\,(P_k)\,. \tag{40}$$

If a security has embedded options (such as direct interest rate options, prepayment option, or the option to default), then the only practical calculation of convexity is as an approximation. Using the Taylor series expansion of equation (22) and ignoring terms in powers of $(\Delta i)^3$ and higher yields

$$\begin{aligned} P\,(i + \Delta i) - P\,(i) &\approx \frac{\mathrm{d}P}{\mathrm{d}i} \Delta i + \frac{1}{2} \frac{d^2 P}{\mathrm{d}i^2} (\Delta i)^2 \\ P\,(i - \Delta i) - P\,(i) &\approx \frac{\mathrm{d}P}{\mathrm{d}i} (-\Delta i) + \frac{1}{2} \frac{d^2 P}{\mathrm{d}i^2} (-\Delta i)^2 \end{aligned}$$

which are summed to give the following approximation to

$$\frac{d^2P}{di^2}(\Delta i)^2 \approx P(i-\Delta i) - 2P(i) + P(i+\Delta i).$$

It follows that

$$C = \frac{d^2P}{di^2}\frac{1}{P} \approx \frac{P(i-\Delta i) - 2P(i) + P(i+\Delta i)}{P(i)(\Delta i)^2}, \tag{41}$$

which is a popular approximation to C that is used for securities with interest sensitive cash flows.

For nominal interest rates, the convexity measure with respect to $i^{(m)}$ is based on the following result:

$$\begin{aligned}\frac{d^2P}{d\left(i^{(m)}\right)^2} &= \frac{d}{di^{(m)}}\left(\frac{dP}{di^{(m)}}\right) = \frac{d}{di^{(m)}}\left(\frac{1}{1+\frac{i^{(m)}}{m}}\frac{dP}{d\delta}\right)\\ &= \frac{1}{\left(1+\frac{i^{(m)}}{m}\right)^2}\frac{d^2P}{d\delta^2} + \frac{1}{\left(1+\frac{i^{(m)}}{m}\right)^2}\frac{1}{m}\left(-\frac{dP}{d\delta}\right).\end{aligned}$$

Therefore, convexity with respect to $i^{(m)}$, $C^{(m)}$, is

$$C^{(m)} = \frac{1}{\left(1+\frac{i^{(m)}}{m}\right)^2}C_M + \frac{1}{\left(1+\frac{i^{(m)}}{m}\right)^2}\frac{1}{m}D_M. \tag{42}$$

It is worthwhile to note that for $m \to \infty$, equation (42) becomes equation (36). For $m = 1$, the right side of equation (42) reduces to C_M, indicating consistency in both boundary cases.

Let us illustrate the concepts of duration and convexity with a simple example.

Example 7. Consider a bond whose current price is 105 and whose derivative with respect to the yield to maturity is -525. The yield to maturity is an annual effective interest rate of 6%. Then the duration of the bond is:

$$-\frac{1}{P}\cdot\frac{dP}{di} = -\frac{1}{105}\times(-525) = 5.$$

Because the effective measure of duration is equal to the Macaulay duration divided by $1+i$, we also can calculate the Macaulay duration of this bond as $5 \times 1.06 = 5.30$. Now suppose that for the same bond, the second derivative of the price with respect to the interest rate is 6720. Then its convexity is:

$$\frac{1}{P}\cdot\frac{d^2P}{di^2} = \frac{1}{105}\times 6720 = 64.$$

4 Classical Immunization

Assume that a financial intermediary has assets, $A(i)$, and liabilities, $L(i)$, that depend on the interest rates. Then the surplus, or capital, of the intermediary, $S(i)$, is defined as

$$S(i) = A(i) - L(i).$$

Though in practice the surplus value may be established not by the market, but by the regulatory or accounting principles, it is important that managers of a financial intermediary understand the relationship of surplus value (market value) to interest rate changes.

Redington (1952) proposed an integrated treatment of assets and liabilities through the study of the surplus function $S(i)$. Suppose the objective of the financial intermediary is to prevent the surplus level from changing solely due to interest rate changes. One possible approach to achieving this objective is to structure the assets and liabilities so that the change in the value of S to be close to zero for infinitesimal changes in interest rates, i.e., to have $\Delta S \approx 0$ for $\Delta i \approx 0$. This implies that the financial intermediary must set

$$\frac{dS}{di} = \frac{dA}{di} - \frac{dL}{di} = 0, \tag{43}$$

i.e., the monetary duration of assets must be equal to the monetary duration of liabilities. If, additionally, the financial intermediary wants to ensure that slight interest rate changes yield an increase in the level of its surplus, the following condition must hold:

$$\frac{d^2S}{di^2} > 0, \tag{44}$$

i.e., the surplus is a convex function of the interest rate. This convexity can be achieved by having assets of greater monetary convexity than that of liabilities.

Suppose, instead, the intermediary was more concerned with protecting the ratio of its assets to liabilities, rather than protecting the actual surplus level.[2] In such a case, the intermediary would be interested in setting the derivative with respect to the interest rate of the ratio of assets and liabilities to zero, while keeping its second derivative positive. As the natural logarithm is a strictly increasing function, however, we can transform this ratio as follows:

[2]This may be a result of the common regulatory concern with capital ratio (i.e., ratio of surplus to assets) or management's desire to control risk by monitoring the capital ratio.

$$R(i) = \ln\left(\frac{A(i)}{L(i)}\right).$$

To protect the surplus ratio level, we set $dR/di = 0$, i.e.,

$$\frac{d\ln(A(i))}{di} = \frac{d\ln(L(i))}{di}, \tag{45}$$

or, equivalently, set the duration of assets equal to the duration of liabilities and simultaneously set $d^2R/di^2 > 0$, i.e.,

$$\frac{d^2\ln(A(i))}{di^2} > \frac{d^2\ln(L(i))}{di^2}, \tag{46}$$

i.e., ensure that the M^2 for the assets is greater than the M^2 for the liabilities. When durations of assets and liabilities are equal, greater M^2 is equivalent to greater convexity, so this condition can be restated as convexity of assets exceeding convexity of liabilities. The approach of equations (45) and (46) is the most common form of classical immunization and is considered to be the standard for applications of immunization.

We should note that classical immunization has many critics, including the present authors, because it violates the no-arbitrage principle of pricing capital assets (Gajek and Ostaszewski, 2002, 2004; Ostaszewski, 2002; and Ostaszewski and Zwiesler, 2002, as well as Panjer, 1998, Chapter 3). The more commonly quoted criticisms of classical immunization include the following:

- Immunization assumes one interest rate, i.e., flat yield curve, which only moves in parallel shifts;
- Immunization assumes only instantaneous infinitely small change in the yield curve, and, of course, such changes are not usually experienced in practice; and
- Immunization requires continuous costly rebalancing due to the continuous changes in the underlying values of the assets and liabilities that result in changes in durations and convexities.

Interestingly, many problems with immunization can be avoided with relatively small modification of the idea. Instead of trying to unrealistically assure that $\Delta S = S(i + \Delta i) - S(i)$ is always nonnegative, one can instead try to bound ΔS from below by a (possibly negative) quantity that can be made as large as possible via a proper choice of the asset portfolio. We will briefly outline this approach. Note that

$$\Delta S = S(i + \Delta i) - S(i) = \sum_{t>0} \frac{S_t}{(1+i)^t} \left(\frac{(1+i)^t}{(1+i+\Delta i)^t} - 1 \right), \tag{47}$$

where S_t is the net surplus cash flow at time t. Hence, by the Schwartz Inequality, we have:

$$\Delta S \geq -\left(\sum_{t>0} \frac{S_t^2}{(1+i)^{2t}} \right)^{\frac{1}{2}} \left(\sum_{t>0} \left(\frac{(1+i)^t}{(1+i+\Delta i)^t} - 1 \right)^2 \right)^{\frac{1}{2}}. \tag{48}$$

Therefore, the change in surplus value is bounded from below by a product of two quantities: the first one depending on the portfolio structure, and the second one depending only on the change in the interest rate. It is clear from (48) that ΔS might be negative, but if we find a way to decrease the quantity:

$$\sum_{t>0} \frac{S_t^2}{(1+i)^{2t}},$$

which can be termed the immunization risk measure, then we can reduce the risk of decline in surplus value, at least in the worst case scenario. This approach is analyzed in detail by Gajek and Ostaszewski (2004).

Suppose that your company is planning to fund a liability of \$1 million to be paid in five years. Assume that the current yield on bonds of all maturities is 4%. Your company can invest in a one-year zero-coupon bond or a ten-year zero-coupon bond to fund this liability. Find the amounts of the two bonds that should be purchased in order to match the duration of the liability. Will such duration-matched portfolio immunize the liability?

The present value of the liability is:

$$\frac{1000000}{1.04^5} \approx 821927.11.$$

The Macaulay duration of the liability is five. Its duration is

$$\frac{5}{1.04} \approx 4.76190476.$$

Let us write w for the portion of the asset portfolio invested in the one-year zero-coupon bond. Then $1 - w$ is the portion invested in the ten-year zero-coupon bond. The duration of the asset portfolio is the weighted average of durations of those two zero-coupon bonds, i.e.,

$$\frac{1}{1.04} \times w + \frac{10}{1.04} \times (1 - w) = \frac{10}{1.04} - \frac{9w}{1.04}.$$

In order to match the duration of the liability, we must have

$$\frac{10}{1.04} - \frac{9w}{1.04} = \frac{5}{1.04}.$$

Therefore, $9w = 5$, and

$$w = \frac{5}{9} \approx 55.56\%.$$

In order to match durations, we must invest 55.56% of the portfolio in the one-year zero-coupon bond and 45.44% in the ten-year zero-coupon bond.

Immunization requires that the asset portfolio has convexity in excess of that of the liability. The convexity of the liability is:

$$\frac{5 \times 6}{1.04^2} \approx 27.7366864.$$

The convexity of the asset portfolio is:

$$\frac{5}{9} \times \frac{1 \times 2}{1.04^2} + \frac{4}{9} \times \frac{10 \times 11}{1.04^2} \approx 46.2278107.$$

Therefore, the asset portfolio has convexity in excess of that of the liability, and the portfolio is immunized.

5 Yield Curve and Multivariate Immunization

5.1 The Yield Curve

So far we have assumed the same interest rate for discounting cash flows for all maturities. In practice, however, the rates used for discounting cash flows for various maturities differ. This can be seen by comparing the actual interest rates for *pure discount bonds*, also known as *zero coupon bonds*, i.e., bonds that make only one payment at maturity, and no intermediate coupon payments. These bonds are discounted at different rates that depend on their remaining term to maturity.

The *yield curve* or *term structure of interest rates* is the pattern of interest rates for discounting cash flows of different maturities. The specific functional relationship between the time of maturity and the corresponding interest rate is usually called the *yield curve*, especially when represented graphically, while *term structure of interest rates* is the general description of the phenomenon of rates varying for different maturities. When longer term bonds offer higher yield to maturity rates than shorter term bonds (as is usually the case in practice) the pattern of yield rates is termed an *upward sloping yield curve.* If yield to maturity rates are the same for all maturities, we call this pattern a *flat yield curve.* Finally, a rare, but sometimes occurring, situation when longer term yield to maturity rates are lower than shorter-term ones, is termed an *inverted yield curve.*

When practitioners estimate the yield curve, they begin with the yield rates of bonds that are perceived to be risk-free. In the United States, the most common bonds utilized as risk-free bonds are those issued by the federal government, i.e., United States Treasury Bills (those with maturities up to a year), Treasury Notes (those with maturities between one and ten years), and Treasury Bonds (those with maturities of ten years or more). But this explanation does not make it clear what interest rate is used in the yield curve for each maturity. There are three ways to define the yield curve (and term structure of interest rates):

1. Assign to each term to maturity the yield rate of a risk-free bond with that term to maturity and trading at par, i.e., trading at its redemption value. The resulting yield curve is termed the *bond yield curve*;

2. Assign to each maturity the yield rate on a risk-free zero-coupon bond of that maturity. This yield curve is called the *spot curve*, and the interest rates given by it are called *spot rates*; and

3. Use the short-term interest rates in future time periods implied by current bond spot rates.

Let us explain the concepts of short-term interest rates and forward rates. A *short-term interest rate* (or *short rate*) refers to an interest rate applicable for a short period of time, up to one year, including the possibility of an instantaneous rate over the next infinitesimal period of time. A *spot interest rate* (or *spot rate*) for maturity n periods, s_n, is an interest rate payable on a loan of maturity n periods that starts immediately and accumulates interest to maturity, $n = 1, 2, \ldots$. A single period *forward interest rate* (or *forward rate*), f_t, is an interest rate payable on a future loan that commences at time t until time $t + 1$, $t = 0, 1, 2, \ldots$.

If we use the one-year rate as the short rate for the purpose of deriving forward rates, we have the following relationship:

$$(1+s_n)^n = (1+f_0)(1+f_1)\dots(1+f_{n-1}). \tag{49}$$

We also have

$$1+f_{n-1} = \frac{(1+s_n)^n}{(1+s_{n-1})^{n-1}}. \tag{50}$$

The yield curve also can be studied for the continuously compounded interest rate, i.e., for the force of interest, δ_t, which is expressed as a function of time.

The distinction between the spot rate and the forward rate is best explained by presenting their mathematical relationship. If δ_t is the spot force of interest for time t and φ_t is the forward force of interest at time t, then the accumulated value at time t of a monetary unit invested at time 0 is:

$$\left(e^{\delta_t}\right)^t = e^{\int_0^t \varphi_s ds}. \tag{51}$$

Therefore we have

$$\delta_t = \frac{1}{t}\int_0^t \varphi_s ds,$$

i.e., the spot rate for time t is the mean value of the forward rates between times 0 and t. By the fundamental theorem of calculus,

$$\varphi_t = t\frac{\mathrm{d}\delta_t}{\mathrm{d}t} + \delta_t. \tag{52}$$

This shows us that $\varphi_t > \delta_t$ if and only if $\mathrm{d}\delta_t/\mathrm{d}t > 0$.

We will illustrate the use of spot and forward rates with a simple example. Suppose a 4%, 1000 par, annual coupon bond with a four-year maturity exists in a market in which the spot rates are:

- 1 year spot rate is $s_1 = 3.0\%$,
- 2 year spot rate is $s_2 = 3.5\%$,
- 3 year spot rate is $s_3 = 4.0\%$,
- 4 year spot rate is $s_4 = 4.5\%$.

Then the value of this bond is the present value of its cash flows discounted using the spot rates:

$$\frac{40}{1.03} + \frac{40}{1.035^2} + \frac{40}{1.04^3} + \frac{1040}{1.045^4} \approx 983.84.$$

For the same date, we also can calculate the corresponding one-year forward rates at times 0, 1, 2, 3 (i.e., from time 0 to time 1, from time 1 to time 2, from time 2 to time 3, and from time 3 to time 4) as follows:

- The forward rate from time 0 to time 1 is $f_1 = 3.0\%$, same as the one year spot rate.
- The forward rate from time 1 to time 2, denoted by f_2, is derived from the condition

  $$(1 + 0.03)\,(1 + f_2) = 1.035^2,$$

  so that

  $$1 + f_2 = \frac{1.035^2}{1.03} \approx 1.04002427,$$

  and

  $$f_2 \approx 4.002427\%.$$

- The forward rate from time 2 to time 3, denoted by f_3, is derived from the condition

  $$(1 + s_2)^2\,(1 + f_3) = 1.035^2 \times (1 + f_3) = 1.04^3,$$

  so that

  $$1 + f_3 = \frac{1.04^3}{1.035^2},$$

  and

  $$f_3 \approx 5.007258\%.$$

- The forward rate from time 3 to time 4, denoted by f_4, is derived from the condition

  $$(1 + s_3)^3\,(1 + f_4) = 1.04^3 \times (1 + f_4) = 1.045^4,$$

  so that

  $$1 + f_4 = \frac{1.045^4}{1.04^3},$$

  and

  $$f_4 \approx 6.014469\%.$$

5.2 Multivariate Immunization

To address some of the weaknesses of classical immunization, Ho (1990) and Reitano (1991a, 1991b) developed a multivariate generalization of duration and convexity. They replaced the single interest rate parameter i by a yield curve vector $\vec{i} = (i_1, ..., i_n)$, where the coordinates of the yield curve vector correspond to certain set of key rates. Reitano (1991a) wrote: "For example, one might base a yield curve on observed market yields at maturities of 0.25, 0.5, 1, 2, 3, 4, 5, 7, 10, 20 and 30 years." The price function is then $P(i_1, ..., i_n)$. Instead of analyzing derivatives with respect to one interest rate variable, one could use multivariate calculus tools to study the price function.

There is one objection that could be raised with respect to this approach. For example, when analyzing a deterministic function of several variables $f(x_1, x_2, \ldots, x_n)$, it is implicitly assumed that the variables x_j and x_k are mutually independent, i.e., $\partial x_j / \partial x_k = 0$. This is definitely *not* the case when various maturity interest rates are considered. Nevertheless, one can study such multivariate models for the purpose of better understanding their properties.

The quantities $\partial \ln P / \partial i_k$ are termed *partial durations* (Reitano, 1991a, 1991b) or *key-rate durations* (Ho, 1990). The *total duration vector* is:

$$-\frac{P'(i_1, ..., i_n)}{P(i_1, ..., i_n)} = -\frac{1}{P(i_1, ..., i_n)} \left(\frac{\partial P}{\partial i_1}, ..., \frac{\partial P}{\partial i_n} \right). \tag{53}$$

One also can introduce the standard notion of directional derivative of $P(i_1, ..., i_n)$ in the direction of a vector $\vec{v} = (v_1, ..., v_n)$:

$$P'_{\vec{v}}(i_1, ..., i_n) = \vec{v} \bullet \left(\frac{\partial P}{\partial i_1}, \ldots, \frac{\partial P}{\partial i_n} \right) \tag{54}$$

where the "$\bullet$" refers to the dot product of the vectors. The second derivative matrix also can be used to define the *total convexity*:

$$\frac{P''(i_1, ..., i_n)}{P(i_1, ..., i_n)} = \frac{1}{P(i_1, ..., i_n)} \left[\frac{\partial^2 P}{\partial i_k \partial i_l} \right]_{1 \le k, l \le n}. \tag{55}$$

One now can view the surplus of an insurance firm as a function of the set of key interest rates chosen. Applying multivariate calculus, we can obtain the two immunization algorithms that are directly analogous to the one-dimensional case:

- To protect the absolute surplus level, set the first derivative (gradient) of the surplus function to zero, i.e.,

$$S'(i_1,...,i_n) = \vec{0}, \quad \text{or, equivalently} \quad A'(i_1,...,i_n) = L'(i_1,...,i_n) \tag{56}$$

where $\vec{0}$ is the zero vector, with all its components being zero and with the symbols A, L referring to assets and liabilities, respectively. In addition we must make the second derivative matrix, $S''(i_1,...,i_n)$, positive definite.

- To protect the relative surplus level (i.e., surplus ratio), set:

$$-\frac{A'(i_1,...,i_n)}{A(i_1,...,i_n)} = -\frac{L'(i_1,...,i_n)}{L(i_1,...,i_n)} \tag{57}$$

and make the total convexity matrix positive definite.

It should be noted (Panjer, 1998, Chapter 3) that key-rate immunization with respect to a large number of key-rates, large enough to be effectively exhaustive of all possible rates determining the yield curve, forces the immunized portfolio toward an exact cash flow match for the corresponding liabilities. While such cash flow matching does provide complete protection against interest rate risk, it is generally more expensive than an immunizing portfolio; if cash flow matching were our objective, this entire analysis would have been unnecessary.

6 Closing Comments

Duration, convexity, and immunization too often are taught in a simplified or even simplistic way and from a perspective somewhat conflicting with that of actuarial practice. We hope that this primer will be a useful tool for practicing actuaries, and others interested in measures of sensitivity with respect to interest rates.

This paper covers some of the material currently included in the Financial Mathematics examination in the new actuarial education system in North America effective in 2005, and we hope that our work can be of value to candidates studying for that examination.

References

Bodie, Z., Kane, A. and Marcus, A. *Investments,* Fifth Edition. Columbus, OH: Irwin/McGraw-Hill, 2002.

Fabozzi, F.J. (Editor) *Handbook of Fixed Income Securities,* Sixth Edition. Columbus, OH: Irwin/McGraw-Hill, 2000.

Gajek, L. and Ostaszewski, K. *Plany Emerytalne: Zarzadzanie Aktywami i Zobowiazaniami* (In Polish, translates as: *Pension Plans: Asset-Liability Management*). Warsaw, Poland: Wydawnictwo Naukowo-Techniczne, May 2002.

Gajek, L. and Ostaszewski, K. *Financial Risk Management of Pension Plans.* Amsterdam, The Netherlands: Elsevier, 2004.

Ho, T.S.Y. *Strategic Fixed Income Investment.* Homewood, IL: Dow Jones-Irwin, 1990.

Macaulay, F. *Some Theoretical Problems Suggested by the Movement of Interest Rates, Bond Yields, and Stock Prices in the U.S. since 1856.* New York, NY: National Bureau of Economic Research, 1938.

Ostaszewski, K.M. *Asset-Liability Integration.* Society of Actuaries Research Monograph. Schaumburg, IL: Actuarial Education and Research Fund, 2002.

Ostaszewski, K. and Zwiesler, H.J. "Gefahren von Duration-Matching-Strategien." (In German, translates as: "The Perils of Asset-Liability Matching Strategies.") *Blaetter der Deutschen Gesellschaft fuer Versicherungsmathematik* 25, no. 3 (April 2002): 617–627.

Panjer, H.H. (Editor) *Financial Economics.* Schaumburg, IL: The Actuarial Foundation, 1998.

Redington, F. M., "Review of the Principles of Life Office Valuations." *Journal of the Institute of Actuaries* 78, 350, part 3, (1952): 286–340.

Reitano, R.R. "Multivariate Duration Analysis." *Transactions of the Society of Actuaries* 43 (1991a): 335–376.

Reitano, R.R. "Multivariate Immunization Theory." *Transactions of the Society of Actuaries* 43 (1991b): 393–428.

Journal of Actuarial Practice Vol. 12, 2005

Modeling Clusters of Extreme Losses

Beatriz Vaz de Melo Mendes* and Juliana Sá Freire de Lima†

Abstract‡

We model extreme losses from an excess of loss reinsurance contract under the assumption of the existence of a subordinated process generating sequences of large claims. We characterize clusters of extreme losses and aggregate the excess losses within clusters. The number of clusters is modeled using the usual discrete probability models, and the severity of the sum of excesses within clusters is modeled using a flexible extension of the generalized Pareto distribution. We illustrate the methodology using a Danish fire insurance claims data set. Maximum likelihood point estimates and bootstrap confidence intervals are obtained for the parameters and statistical premium. The results suggest that this cluster approach may provide a better fit for the extreme tail of the annual excess losses amount when compared to classical models of risk theory.

Key words and phrases: *reinsurance, excess of loss, cluster of extremes, extreme value theory*

*Beatriz Vaz de Melo Mendes, Ph.D., is an associate professor of statistics. She obtained her Ph.D. in statistics in 1995 from Rutgers University, USA. Her research interests include robustness, extreme value theory, long memory models, and their applications to finance, insurance, and reinsurance.

Dr. Mendes' address is: Statistics Department, Federal University at Rio de Janeiro, RJ, BRAZIL. E-mail: *beatriz@im.ufrj.br*

†Juliana Sá Freire de Lima is a Ph.D. student in statistics at the Federal University at Rio de Janeiro, Brazil. She currently works at IRB-Brasil (Brazilian Re-insurance Institute). Her research interests include extreme value theory, Bayesian methods, and their applications in insurance and reinsurance.

Ms. De Lima's address is: Statistics Department, Federal University at Rio de Janeiro, RJ, BRAZIL. E-mail: *jsflima@im.ufrj.br*

‡The authors gratefully acknowledge financial support from CNPq of Brazil. We wish to thank the anonymous referees and the editor for their helpful comments and suggestions that have improved this paper.

1 Introduction

Of great concern to insurers is the risk arising from catastrophic claims. Often such claims represent a relatively large proportion of the aggregate claim amount (see Embrechts, Klüppelberg, and Mikosch, 1997, page 4). Thus, insurers may seek protection through various types of reinsurance arrangements such as excess of loss reinsurance.

In this paper we address the problem of modeling the reinsurer's total losses arising from excess of loss reinsurance contracts. The classic excess of loss (XL) with a given retention level u can be described as follows: let X_i denote the size of the i^{th} claim, $Z_i = \min(u, X_i)$ denote amount covered by the cedent (the insurer), and $Y_i = \max(0, X_i - u)$ denote the amount covered by the reinsurer, then $X_i = Z_i + Y_i$. If there are N claims in the contract period, then the aggregate claim amount paid by the reinsurer is the compound sum S,

$$S = \sum_{i=1}^{N} Y_i. \tag{1}$$

Typically the number of claims N is modeled by a negative binomial ($\text{NB}(k, p)$) or a Poisson ($\text{Poisson}(\lambda)$) distribution, and Y follows a gamma or a Pareto distribution. S has been widely studied in actuarial risk theory; see, for example, Sundt (1982), Embrechts, Maejima, and Teugels (1985), McNeil (1997), Berglund (1998), and Klugman, Panjer, and Willmot (2004, Chapter 6).

Consider the two-dimensional random process $\{T_i, X_i\}$, $i = 1, 2, \ldots$ where T_i and X_i are the time and size of the i^{th} claim, respectively. Whenever it is realistic to assume that the X_is are independent and identically distributed (iid) and independent of the T_is, the problem of modeling the insurer's aggregate excess losses S may be split in two parts: modeling the number of excess losses N occurring during the period and modeling the severity of the individual claim excess Y_i. In practice, unfortunately, the iid assumption may not hold because the two-dimensional random process may possess another subordinated process that may induce the occurrence of a sequence of large claims that occur in groups or clusters. Examples of such subordinated processes are floods, earthquakes, and hurricanes.

To overcome the problem of local dependence (i.e., short range occasional temporal dependence), we propose to identify clusters of extreme losses and define a new variable A_k to denote the sum of excess losses within the k^{th} cluster of extreme losses. It is now reasonable to assume that the iid assumption holds for the A_ks. By modeling separately the number of clusters of excesses C and the severity of the aggregated excess losses A_k, we have an annual excess losses amount of S where

$$S = \sum_{j=1}^{C} A_j, \tag{2}$$

where the A_js are iid and independent of C, the random number of clusters.

There exist alternative approaches to dealing with the problem of dependent risks. For example, Heilmann (1986) studied stop-loss cover under relaxation of the independence assumption. Kremer (1998) provided formulae and examples for calculating the premium of generalized largest claims reinsurance covers in the case of dependent claim sizes. Schumi (1989) developed a method for calculating the distribution of the total excess losses amount when losses come from different sources. The key point Schumi analyzed is that the two distributions involved, i.e., the excess over retention limits and the excess over the retained annual aggregate, are not independent. Goovaerts and Dhaene (1996) also relaxed the independence assumption and showed that the same compound Poisson approximation for the aggregate claims distribution still performs well when the dependency between two risks i and j is caused by the dependency between the Bernoulli random variables I_i and I_j, where I_i indicates the occurrence of at least one claim for risk i.

To model the aggregated excess A_i, we use distributions from extreme value theory. More specifically, we use the modified generalized Pareto distribution, a powerful and flexible extension of the generalized Pareto distribution. This modified generalized Pareto distribution was obtained in Anderson and Dancy (1992) as a limit result based on a point process representation. In this representation, the (one-dimensional) marginals are be a Pareto type distribution.

Three models of the size of the i^{th} excess loss are compared:

Model 1 assumes Y_i follows a generalized Pareto distribution and the number of claims N is a negative binomial or a Poisson distribution.

Model 2 assumes the severity of the aggregated excess losses A_k follows a modified generalized Pareto distribution and the number of clusters C is a negative binomial or a Poisson distribution.

Model 3 assumes Y_i follows a gamma distribution and the number of claims N is a negative binomial or a Poisson distribution.

The distribution of the (annual) excess losses amount S is obtained by convolutions. Results indicate that the proposed Model 2 may yield more conservative estimates for premiums.

Our models may be used by insurers to search for alternative choices for the retention limit. In a related work, McNeil (1997) fitted the generalized Pareto distribution to insurance losses that exceed high thresholds using Model 1. He considers the sensitivity of inference to the choice of the threshold value and also discusses dependence in the data and other issues such as seasonality and trends.

The remainder of this paper is organized as follows. In Section 2 we formally introduce our proposed models of the annual excess loss amount by considering sums of excess losses within clusters. We provide some background from extreme value theory that justifies the dependence in the data, the (de)clustering technique, and the use of the modified generalized Pareto distribution as an alternative to distributions often used in classical actuarial risk modeling. Estimation methods and statistical tests are also discussed. In Section 3 we illustrate the methodology using the Danish fire insurance claims data. Two empirical rules are used to define clusters of excess losses. Distributions are fitted to the excess and aggregated excess data to obtain the distribution of S. The three models are then compared. Confidence intervals for parameter estimates and for the statistical premium are obtained using bootstrap techniques. In Section 4 we consider a higher retention level and model the upper extreme tail of the fire insurance claims. Finally, in Section 5 we give our conclusions.

2 Modeling Clusters of Excesses Using Extreme Value Theory

Extreme value theory is concerned with the behavior of extremes from a stochastic process $\{X_1, X_2, \ldots\}$. The modeling structure proposed is motivated by the asymptotic results of Mori (1977) and Hsing (1987) with respect to a two-dimensional point process of excesses over a high threshold u, which governs both the loss size and their arrivals. Mori and Hsing have shown that under weak long-range mixing conditions, large values of the strictly stationary sequence $\{X_1, X_2, \ldots\}$ occur in clusters, and the two-dimensional point process converges to a non-Poisson process. They showed that, for the class of possible limiting distributions for the two-dimensional point process, the peak excess within a cluster converged weakly to a generalized Pareto distribution.

As discussed in Anderson and Dancy (1992) and Anderson (1994), under an extreme event and for u sufficiently high, the tail behavior of the sum of excesses beyond u should also be of Pareto type. Anderson and Dancy (1992) proposed the modified generalized Pareto distribution and applied the methods to the analysis of atmospheric ozone levels.

We propose to characterize clusters of extreme claims and to model the sum of excess losses within a cluster using the modified generalized Pareto distribution, $G_\xi(y)$, given by

$$G_\xi(y) = \begin{cases} 1 - \left(1 + \xi \left(\frac{y}{\psi}\right)^\theta\right)^{-1/\xi}, & \text{for } \xi \neq 0 \text{ and } y > 0; \\ 1 - e^{-\left(\frac{y}{\psi}\right)^\theta}, & \text{for } \xi = 0 \text{ and } y > 0; \end{cases} \tag{3}$$

where $\theta > 0$, and $\psi > 0$ is a scale parameter. The generalized Pareto distribution may be obtained from equation (3) by putting $\theta = 1$ and $\xi > 0$, and the Weibull distribution corresponds to $\xi = 0$. Fitting the modified generalized Pareto distribution to the data is equivalent to taking a Box-Cox transformation (that is, to consider a new variable Y^θ, see Hoaglin, Mosteller, and Tukey (1983)) and modeling the transformed data using a generalized Pareto distribution. We chose to fit the modified generalized Pareto distribution, which allows for simultaneous estimation of all parameters and for standard statistical tests of nested models (sub-models obtained by making restrictions on the parameters of the full model, see Bickel and Doksum, 1977).

Figure 1 illustrates the flexibility of the modified generalized Pareto density, with its varying shapes and heavy/long tails. In both plots $\xi = 0.3$, $\psi = 1$, and θ varies from $\theta = 0.2$ up to $\theta = 2.5$. When $\theta < 1$ the densities are strictly decreasing with heavier tails; $\theta = 1$ corresponds to the generalized Pareto distribution; and when $\theta > 1$ the densities possess a positive mode.

We have seen that short range dependence of excess losses results in clusters of extreme claims. The frequency and size of these clusters depend on the retention level and on the definition of a cluster. In practice, the choice of the retention level u is made directly between insurer and reinsurer, thus making the definition of a cluster the only unresolved issue.

How should clusters be defined? The answer depends on the type of data being used. For example, financial data and environment data certainly allow for different definitions. We have not found a formal rule in the literature. Coles (2001), however, suggests using an empirical rule that, for a given u, defines consecutive excesses over u as belonging to the same cluster. Under Coles's method a new cluster starts

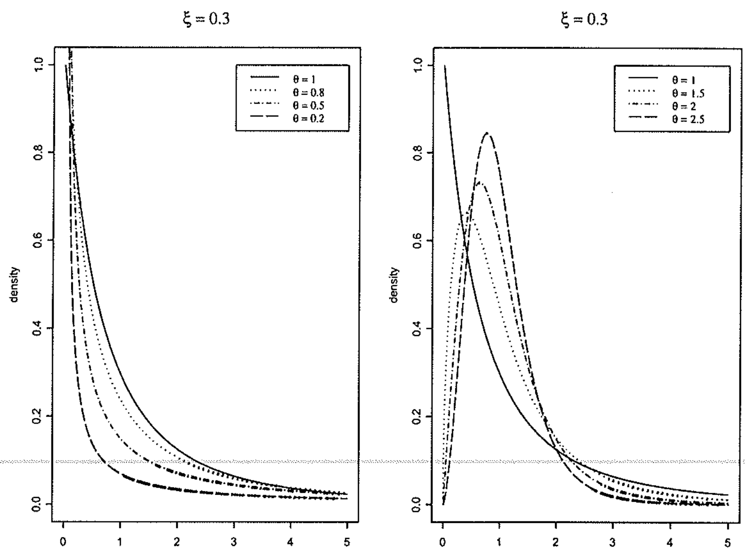

Figure 1: The Modified Generalized Pareto Density for $\xi = 0.3$, Scale $\psi = 1$, and Varying Values of θ

after r consecutive values have fallen below u, for some pre-specified value of r. Coles's method of cluster identification is also known as the runs method. For more details on cluster identification see Reiss and Thomas (1997) and Embrechts, Klüppelberg, and Mikosch (1997).

There is a trade off between choosing a small r (which hurts the independence assumption between clusters) and choosing a large r (which include data not generated by the same subordinated process). For any given data set it is advisable to experiment with different choices for r (and u) for cluster determination then check the results for robustness.

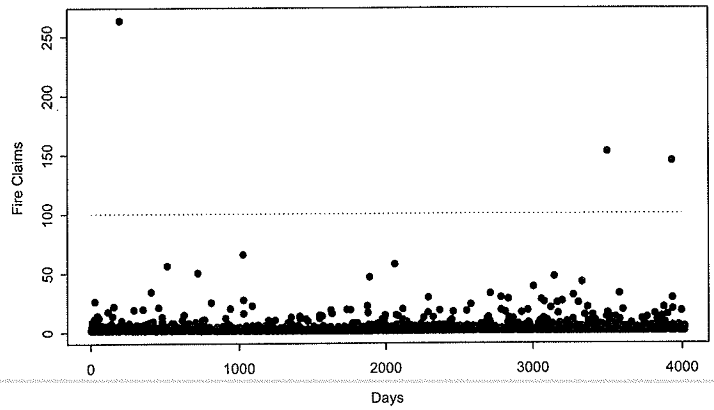

Figure 2: Time Series of Danish Fire Insurance Claims

3 Illustration of Our Methodology

3.1 The Data Set

Our methodology is illustrated using Danish fire insurance claims data,[1] which consist of 2167 observations of fire insurance claims in millions of Danish Kroner (1985 prices) from 1980 to 1990. Figure 2 shows a time series plot of the data: size of claim (the y-axis) versus the total number of days measured from the baseline of 01/01/1980 up to the time of occurrence (the x-axis). There are only three very extreme observations, and, according to McNeil (1997), the data show no clustering. In spite of that, this data set is used to illustrate the usefulness of the proposed modeling structure and to experiment with two declustering strategies and two retention levels.

Let us define the k^{th} empirical mean excess as the mean of the k largest excess observations. Figure 3 shows the empirical mean excess function of the data set, which is a plot of the k^{th} empirical mean excess

[1]This data set was kindly made available to us by Paul Embrechts of ETH Zurich. It has been used by several authors, including Embrechts, Klüppelberg, and Mikosch (1997) and McNeil (1997).

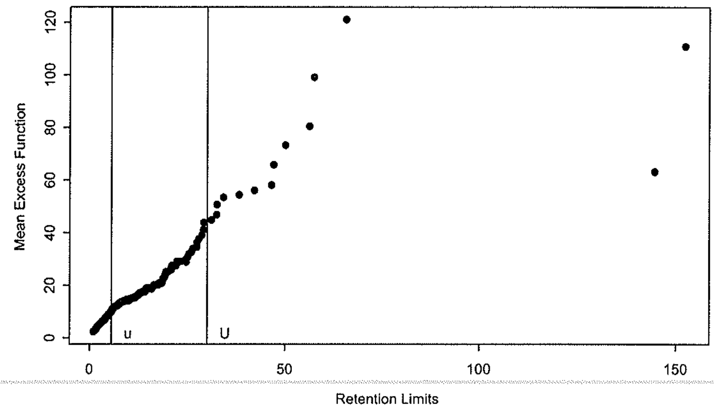

Figure 3: The Empirical Mean Excess Function of the Danish Fire Insurance Data

versus the $k + 1^{\text{th}}$ largest observation. This plot may also be used as an exploratory technique for choosing a threshold. The increasing linear aspect of the graph indicates that a generalized Pareto distribution with $\xi > 0$ might be a valid approximation to the entire data set.

To help in choosing a retention limit we order the claim sizes from smallest to largest. We observe that the largest ten percent of claims sizes (i.e., the 217 largest claims) add up to almost half (46%) of the total claim amount, which is 7,335.486 million Danish kroners. This suggests taking the 90 percentile of the empirical distribution as a first choice for the retention limit u, i.e., $u = 5.561735$. A second value of the retention level, $u = 30$, is determined by examining the empirical mean excess function. Both thresholds are shown in Figure 3. As mentioned earlier, the choice of retention limit must also take into account other insurance company factors such as operational costs and the amount of capital in reserve.

Throughout the rest of Section 3, we assume $u = 5.561735$ and there are 217 excess losses. This excess of loss data show a long tail with three extreme observations.

3.2 Estimation and Tests

The full modified generalized Pareto distribution (MGPD) model, i.e., MGPD(ψ, ξ, θ), is fitted via maximum likelihood to data from the excess losses random variable Y_i and from the aggregated excesses random variable A_i. We use the three constrained models: (i) the Weibull distribution (i.e., MGPD$(\psi, 0, \theta)$); (ii) the generalized Pareto distribution (GPD) (i.e., MGPD$(\psi, \xi, 1)$); and (iii) the unit exponential distribution (i.e., MGPD$(\psi, 0, 1)$). For the sake of comparisons, we also fit a gamma distribution with mean ξ/ψ and variance ξ/ψ^2.

Although there are other commonly used estimation methods such as the method of moments (e.g., Embrechts, Klüpelberg, and Mikosch, 1997) and Bayesian methods (e.g., Reiss and Thomas, 1999), we use maximum likelihood estimation due to its desirable asymptotic properties. The likelihood ratio test is used to discriminate between the nested models. The best model is then compared to the gamma fit using the AIC and BIC criteria, which are criteria based on a penalized log-likelihood (Bickel and Doksum, 1977).

The Poisson distribution with mean λ (Poi(λ)), and the negative binomial distribution with mean $kp/(1-p)$ and variance $kp/(1-p)^2$ (i.e., NB(k, p)) are fitted by maximum likelihood to both N and C. The Pearson chi-square test for discrete data, which is a measure of departure between the observed and expected frequencies of claims (or clusters) under the model (Bickel and Doksum, 1977), is used to assess the quality of each fit and to choose the best model. The distribution of S is obtained by convolutions and the normal approximation. Graphical tools, such as the qq-plot, are also employed to check the adequacy of all fits.

Overall emphasis is placed on accurately fitting the tail of the claim distribution, as this is crucial for obtaining good estimates of the net premium and the statistical premium.

3.3 Fitting Y and N

Table 1 shows the maximum likelihood estimates of the parameters of the distributions fitted to the data. It also shows the log-likelihood value (LL), the mean, and the variance of each fitted model. The likelihood ratio tests indicate the full modified generalized Pareto distribution model yields the best fit to the excess losses. The AIC and BIC tests reject the gamma fit in favor of the modified generalized Pareto distribution. Graphical analysis of the modified generalized Pareto distribution fit (not shown here) indicates a good adherence of all observations but the three extreme ones.

The Poisson and the negative binomial are fitted by maximum likelihood to the 11 observations of the number of excess losses N. The Pearson's chi-square test indicates the negative binomial distribution assumption for N is reasonable. The estimates are $\widehat{\mathbb{E}[N]} = 19.7747$ and $\widehat{\mathbb{V}\mathrm{ar}[N]} = 34.8145$, giving the distribution of N as NB(26, 0.568).

Table 1
Maximum Likelihood Fit for Various Models of Y_i
Using the 217 Excess Losses Data and Retention Limit $u = 5.5617$

Model	LL	$\hat{\psi}$	$\hat{\xi}$	$\hat{\theta}$	$\widehat{\mathbb{E}[N]}$	$\widehat{\mathbb{V}\mathrm{ar}[N]}$
MGPD	-662.5155	3.6270	0.1966	0.7450	9.6178	373.6700
Weibull	-665.2370	3.4697	0.0000	0.6430	9.5738	185.1200
GPD	-669.4158	4.4600	0.5900	1.0000	10.8780	∞
EXPON	-716.7387	10.000	0.0000	1.0000	10.0000	100.0000
Gamma	-673.3982	0.0510	0.5100	—	10.0000	196.0800

Notes: MGPD = modified generalized Pareto distribution, GPD = generalized Pareto distribution, EXPON = exponential distribution

Summarizing, the best fit for the severity and the number of excess losses over the retention limit $u = 5.561735$ are, respectively, the MGPD($\hat{\psi} = 3.6270, \hat{\xi} = 0.1966, \hat{\theta} = 0.7450$) and NB(26, 0.568), which we will call Model 1. Under Model 3 the severity has the classical gamma distribution with parameters $\hat{\psi} = 0.0510$ and $\hat{\xi} = 0.5100$, and N is NB(26, 0.568), also shown in Table 1. The 95% non-parametric bootstrap confidence intervals for the parameter estimates of the two models, based on 5000 replications of the data, are given in the first and third rows of Table 1.

3.4 Fitting A and C

First we must use a rule to define a cluster. The runs method is applied to the data, and two empirical rules are postulated:

- Rule 1 requires at least three consecutive days ($r = 3$) with no occurrence of claims exceeding u to separate clusters; and
- Rule 2 requires at least four consecutive days ($r = 4$) with no occurrence of claims exceeding u to separate clusters.

Rule 1 results in a data set of $C = 169$ clusters, while Rule 2 also results in a long right tail data set with $C = 158$ clusters. Both rules show a long tail. Table 2 gives the maximum likelihood estimates of the distributions fitted to the sum of excess losses within the 169 clusters under Rule 1.

Table 2
Maximum Likelihood Fit for Various Models of A_k
Under Rule 1 with 169 Clusters and Retention Limit $u = 5.5617$

Model	LL	$\hat{\psi}$	$\hat{\xi}$	$\hat{\theta}$	$\widehat{\mathbb{E}[A]}$	$\widehat{\mathbb{V}\mathrm{ar}[A]}$
MGPD	-563.5884	4.8634	0.2380	0.7960	12.325	618.93
Weibull	-566.4257	4.3804	0.0000	0.6640	12.349	299.02
GPD	-566.7906	6.2600	0.5200	1.0000	13.042	∞
EXPON	-600.4472	12.840	0.0000	1.0000	12.840	164.87
Gamma	-572.3500	0.0420	0.5400	—	12.857	306.12

Notes: MGPD = modified generalized Pareto distribution, GPD = generalized Pareto distribution, EXPON = exponential distribution

Under Rule 1, all tests indicate the modified generalized Pareto distribution is the best distribution for the aggregated excess losses. The best model for the independent sums of excess losses A over the retention limit $u = 5.561735$ and the number of clusters of excess losses C are the MGPD($\hat{\psi} = 4.8634, \hat{\xi} = 0.2380, \hat{\theta} = 0.7960$) and the negative binomial with parameters $\hat{k} = 34$ and $\hat{p} = 0.688$. This is called Model 2.

Under Rule 2 the statistical tests indicate the modified generalized Pareto distribution gives the best fit with parameter estimates of $\theta = 0.856$, $\xi = 0.306$, and $\psi = 5.693$. The moments of A_i are $\widehat{\mathbb{E}[A]} = 13.9184$, and $\widehat{\mathbb{V}\mathrm{ar}[A]} = 630.01$, which are different from those under Rule 1.

As expected, results change with the choices of cluster definition. Our objective in this section, however, is neither to find the best rule for this data set nor to find the best value for u. Again, our point here is that the differences in estimates of the pair A and C and the pair Y and N affect the estimation of the distribution of S (given in Section 4). We stress that whenever one suspects about dependence in the data, clustering should be investigated and modeled. Thus, we continue our analysis using just the aggregated data from the first rule.

4 Approximating the Distribution of S

Let $F(s) = \Pr[S \leq x]$. The exact expression for $F(s)$ is known only in a few special cases. If the severity distribution is arithmetic,[2] then an exact recursive formula may be available. In general, determining $F(s)$ is a challenging problem, so approximations are needed. Pentikäinen (1987) and Klugman, Panjer, and Willmot (2004, Chapter 6) provide an excellent discussion of several approximations used by actuaries.

Pentikäinen (1987) describes the normal power approximation, which is an improvement on the basic normal approximation. If μ_S, σ_S and γ_S are the mean, standard deviation, and coefficient of skewness of S, then the normal power approximation is

$$F(s) \approx \Phi\left[-\frac{3}{\gamma_S} + \sqrt{\frac{9}{\gamma_S^2} + 1 + \frac{6}{\gamma_S}\left(\frac{s-\mu_S}{\sigma_S}\right)}\right]$$

while the basic normal approximation is

$$F(s) \approx \Phi\left(\frac{s-\mu_S}{\sigma_S}\right)$$

where $\Phi(x)$ is the cdf of the standard normal distribution. The moments of S are determined using equations

$$\begin{aligned}
\mu_S &= \mathbb{E}[Y]\,\mathbb{E}[N] \\
\sigma_S^2 &= \mathbb{V}\mathrm{ar}[Y]\,\mathbb{E}[N] + (\mathbb{E}[Y])^2\,\mathbb{V}\mathrm{ar}[N] \\
\mathbb{E}\left[(S-\mu_S)^3\right] &= \mathbb{E}[N]\,\mathbb{E}\left[(Y-\mathbb{E}[Y])^3\right] + 3\mathbb{V}\mathrm{ar}[N]\,\mathbb{E}[Y]\,\mathbb{V}\mathrm{ar}[Y] \\
&\quad + \mathbb{E}\left[(N-\mathbb{E}[N])^3\right]\mathbb{E}\left[Y^3\right].
\end{aligned}$$

For clusters we replace Y and N by A and C, respectively.

Another approach is via simulation. This is done by simulating from the fitted distributions of Y and N (or A and C) and computing the convolutions for $s \geq 0$:

$$\mathbb{P}[S \leq s] = \mathbb{P}[N = 0] + \sum_{n=1}^{\infty} \mathbb{P}[Y_1 + \cdots + Y_n \leq s]\,\mathbb{P}[N = n]. \tag{4}$$

[2]A discrete distribution is said to be arithmetic with span $h > 0$ if it has a probability mass point at some point x_0 and its other probability mass points, if any, occur only at a subset of the points $x_j = x_0 + hj$ for $j = \ldots, -2, -1, 0, 1, 2, \ldots$.

To numerically approximate the distribution of S, we truncate the infinite sum at a very large value of N (or C). In the case of Model 3 the convolutions were obtained analytically.

Table 3 gives estimates of the mean, variance, and coefficient of skewness of S each for the three models. Table 4 provides estimates of the percentile premiums using simulations and the normal and normal power approximations. As expected, the light tail of the normal distribution underestimates the premiums attached to smaller probabilities. On the other hand, the normal power approximations provided results very close to those obtained by convolutions for Model 3, but overestimated the premiums for Models 1 and 2.

Table 3

Mean, Variance, and Skewness of S

Model	$\hat{\mu}_S$	$\hat{\sigma}_S^2$	$\hat{\gamma}_S$
1	190.2	10609.6	1.1363
2	190.0	12947.5	1.2945
3	197.8	7358.9	0.6879

Table 4

Percentile Premium Estimates Using Simulations, Normal Power, and Normal Approximations

	Convolutions		Normal Power		Normal	
Model	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$
1	317	374	334.7	392.9	322.2	359.6
2	327	394	351.6	419.0	335.9	377.2
3	313	354	314.1	355.7	307.7	338.8

Figure 4 shows, at the left side and for the three models, the plot of the percentile premium $\hat{P}_\alpha$ as a function of their corresponding cumulative probabilities $1 - \alpha$. For any fixed small exceedance probability, smaller premiums are predicted under Models 2 and 3 than under the proposed model given in equation (2). For example, for $\alpha = 0.02$, the premium values are 400, 460, and 500, respectively under Models 3, 1, and 2. At the right side we can see the corresponding densities, where we observe the heavier tail provided by Model 2. The estimates of the percentile premiums $\hat{P}_{0.10}$ and $\hat{P}_{0.05}$ are given in Table 4.

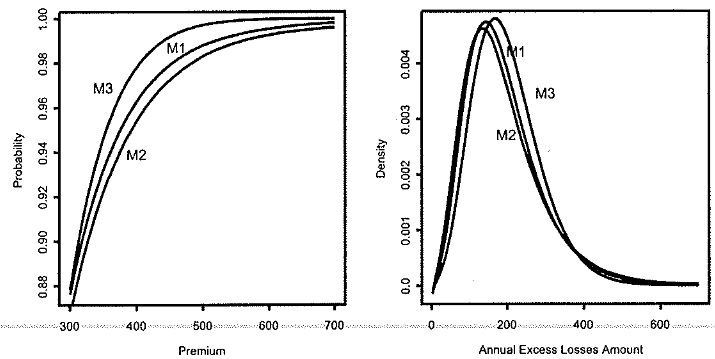

Figure 4: Percentile Premiums and Densities of S for the Three Models

It is always desirable to obtain lower and upper confidence limits for the statistical premiums. Using 5,000 replications of the data we obtained their 95% non-parametric bootstrap confidence intervals, shown in Table 5.

For this data set, the graphical analysis based on the fitted and empirical distributions did not provide a clear indication of the best fit for S, probably due to the small sample size of just 11 observations. We could observe a nice fitting of the extreme tail of S for the three models. The Kolmogorov goodness of fit test yielded the test statistic values of 0.1696, 0.1611, and 0.1776, respectively for Models 1, 2, and 3. Because the critical value at the 5% level is 0.398 for a sample of size 11, we keep the null hypothesis that S is well modeled by the three models. The slightly smaller value of the test statistic from Model 2, however, is an indication it provides the best fit.

The results for the Danish insurance data indicate that the modeling strategy proposed in this paper may provide a more accurate fit for the extreme tails of S. From the practical point of view, this may be seen as an advantage, as more conservative estimates of the statistical premium were obtained under Model 2.

Table 5
95% Bootstrap Confidence Intervals for Model Parameters and Percentile Premiums

Model	$\hat{\psi}$	$\hat{\xi}$	$\hat{\theta}$	$\hat{P}_{0.10}$	$\hat{P}_{0.05}$
1	[2.84,4.53]	[0.03,0.46]	[0.65,0.86]	[250,378]	[294,467]
2	[3.61,6.19]	[0.04,0.52]	[0.64,0.92]	[313,519]	[355,610]
3	[0.03,0.07]	[0.44,0.63]	—	[249,358]	[279,436]

5 Summary

In this paper we focused on the problem of modeling the annual excess loss amount S arising from the classical excess of loss contract. By assuming that a subordinated process may exist and would be responsible for a sequence of large claims, we proposed to characterize clusters of extreme losses and to aggregate the excesses within clusters. Following the classical approach taken in risk theory, we proposed to model S by modeling separately the sum of excess losses A within clusters and the number of clusters C. We discussed the influence of the declustering rules adopted and the effects of the retention level values chosen.

To model the aggregated excess claims A we proposed the flexible modified generalized Pareto distribution, an extension of the generalized Pareto distribution, a well known distribution from the extreme value theory. The modified generalized Pareto distribution allows for heavy/long tails and for different density shapes according to the value of its (modifying) parameter θ. We provided background from the extreme value theory to justify the presence of dependence in the data and the use of the modified generalized Pareto distribution as an alternative to distributions often found in classical homogeneous risk modeling in actuarial science.

The new modeling structure was applied to the Danish fire insurance claims data and compared to two classical approaches based on the excess losses and on the gamma and the generalized Pareto distributions. All models were fitted by the maximum likelihood methodology. The number of excess claims N and the number of independent clusters C were modeled by a negative binomial or a Poisson. Standard statistical tests were carried out to discriminate among nested models and to test goodness of fits.

All tests indicated the modified generalized Pareto distribution as the best fit for the excess and for the aggregated excess losses. We obtained the distribution of S by convolutions, normal power approximation and normal approximation. We found that the proposed procedure provided a better fit for the extreme tail of S, being more conservative in the estimation of the statistical premium. Confidence intervals for parameter estimates and for the statistical premium were obtained using bootstrap techniques.

Summarizing, results indicated that more accurate estimation of the distribution of the annual sum of excess losses may be obtained by modeling the local dependence and by using a more flexible distribution, able to accommodate different density shapes and longer tails.

Even though the modeling structure proposed in this paper may be used by the insurer to search for a suitable value for the retention limit, we did not focus on this issue. For any given data set, the analyst should carry out some type of sensitivity analysis, for example by experimenting with different choices of the threshold value and different rules for cluster definition. In practice, and for data showing stronger local dependence, this sensitivity analysis is highly recommended.

Future areas for further research include simulations of data possessing some known type of dependence structure to assess relationships between different types of dependence and strength of aggregation.

References

Anderson, C.W. and Dancy, G.P. "The Severity of Extreme Events." Research Report 92/593, Department of Probability and Statistics, University of Sheffield, England, 1992.

Anderson, C.W. "The Aggregate Excess Measure of Severity of Extreme Events." *Journal of Research of the National Institute of Standards and Technology* 99 (1994): 555-561.

Berglund, R.M. "A Note on the Net Premium for a Generalized Largest Claims Reinsurance Cover." *ASTIN Bulletin* 28, no. 1 (1998): 153-162.

Bickel, P. and Doksum, K. *Mathematical Statistics: Basic Ideas and Topics.* Berkeley, CA: University of California, 2001.

Coles, S. *An Introduction to Statistical Modeling of Extreme Values.* London, UK: Springer-Verlag, 2001.

Embrechts, P., Klüppelberg, C. and Mikosch, T. *Modelling Extremal Events for Insurance and Finance.* Berlin: Springer-Verlag, 1997.

Embrechts, P., Maejima, M. and Teugels, J.L. "Asymptotic Behaviour of Compound Distributions." *ASTIN Bulletin*, 15, no. 1 (1985): 45–48.

Goovaerts, M.J., and Dhaene, J. "The Compound Poisson Approximation for a Portfolio of Dependent Risks." *Insurance: Mathematics & Economics* 18 (1996): 181-185.

Heilman, W.R. "On the Impact of Independence of Risks on Stop Loss Premiums." *Insurance: Mathematics & Economics*, (1986): 197–199.

Hoagin, D.C., Mosteller, F. and Tukey, J. *Understanding Robust and Exploratory Data Analysis.* New York, NY: Wiley, 1983.

Hsing, T. "On the Characterization of Certain Point Processes." *Stochastic Processes and their Applications* 26 (1987): 297–316.

Klugman, S.A., Panjer, H.H., and Willmot, G.E. *Loss Models: From Data to Decisions.* New York: John Wiley & Sons, 2004.

Kremer, E. "Largest Claims Reinsurance Premiums for the Weibull Model." *Blätter der deutschen Gesettschaft für Versicherungsmathematik* (1998): 279–284.

McNeil, A.J. "Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory." *ASTIN Bulletin* 27 1997: 117–137.

Mori, T. "Limit Distribution of Two-Dimensional Point Processes Generated by Strong-Mixing Sequences." *Yokohama Mathematics Journal* 25 (1977): 155–168.

Pentikäinen, T. "Approximate Evaluation of the Distribution of Aggregate Claims." *ASTIN Bulletin* (1987): 15–39.

Reiss, R.D. and Thomas, M. *Statistical Analysis of Extreme Values.* Berlin: Birkhäuser-Verlag, 1997.

Reiss, R.D. and Thomas, M. "A New Class of Bayesian Estimators in Paretian Excess-of-Loss Reinsurance." *ASTIN Bulletin*, 29, no. 2 (1999): 339–349.

Schumi, J.R. "Excess Loss Distributions over an Underlying Annual Aggregate." *Casualty Actuarial Society Forum* (Fall 1989): 159–172.

Sundt, B. "Asymptotic Behaviour of Compound Distributions and Stop-Loss Premiums." *ASTIN Bulletin* 13 (1982): 89–98.

Journal of Actuarial Practice Vol. 12, 2005

Modeling Insurance Loss Data: The Log-EIG Distribution

Uditha Balasooriya,* Chan Kee Low,† and Adrian Y.W. Wong‡

Abstract§

The log-EIG distribution was recently introduced to the probability literature. It has positive support and a moderately long tail, and is closer to the lognormal than to the gamma or Weibull distributions. Our simulations show that data generated from a log-EIG distribution cannot be adequately described by lognormal, gamma, or Weibull distributions. The log-EIG distribution is a worthwhile candidate for modeling insurance claims (loss) data or lifetime data. Examples of fitting the log-EIG to published insurance claims data are given.

Key words and phrases: *claims distribution, optimal invariant selection procedure, Akaike information criterion, simulation, fitting distributions*

*Uditha Balasooriya, Ph.D., is an associate professor at Nanyang Technological University. His research interests include life testing, reliability, statistical, and actuarial modeling.

Dr. Balasooriya's address is: S3-B2B-57, Nanyang Business School, Nanyang Technological University, Nanyang Avenue, SINGAPORE 639798. E-mail: *auditha@ntu.edu.sg*

†Chan Kee Low, Ph.D., is an associate professor at Nanyang Technological University. His research interests include applications of econometrics and statistical techniques to problems in economics, insurance, and finance.

Dr. Low's address is: S3-B1C-100, School of Humanities and Social Sciences, Nanyang Technological University, Nanyang Avenue, SINGAPORE 639798. E-mail: *acklow@ntu.edu.sg*

‡Adrian Y.W. Wong, F.S.A., F.C.A.S., F.I.A., F.L.M.I., is an associate professor at Nanyang Technological University. He is a Fellow of the Society of Actuaries and his research interests include applications of actuarial techniques to problems in insurance and finance.

Professor Wong's address is: S3-B1C-87, Nanyang Business School, Nanyang Technological University, Nanyang Avenue, SINGAPORE 639798. E-mail: *awongyw@ntu.edu.sg*

§This research was partially supported by the Nanyang Technological University AcRF Grant, Singapore. The authors would like to thank the editor and three anonymous referees for their helpful comments and suggestions that have substantially improved the quality of the paper.

1 Introduction

In fitting distributions to insurance loss data, several families of distributions have been proposed. The common characteristics of these distributions are their skewness to the right and their long tails to capture occasional large values that are commonly present in insurance loss data. One fundamental question confronting actuaries, reliability analysts, and other researchers, however, is the approach used to select the best model for a given data set.

Various approaches have been proposed for discriminating between families of distributions. For example:

- Lehmann (1959) has provided the so-called most powerful invariant test, which is uniformly most powerful in the class of tests that are invariant under certain transformations of the data.

- There is the separate families test based on the Neyman-Pearson maximum likelihood ratio; see, for example, Cox (1962). The concept of separate families of distributions is important, as it is natural to consider competing families in model selection.

- Geisser and Eddy (1979) have proposed a synthesis of Bayesian and sample-reuse approach for model selection. The emphasis here is to obtain a model that yields the best prediction for future observations.

- The maximum likelihood ratio test was proposed by Dumonceaux, Antle, and Haas (1973) for selecting between two models with unknown location and scale parameters. This test has the advantage that the distribution of the ratio of the two likelihood functions does not depend on the location and scale parameters. Gupta and Kundu (2003) used this test to discriminate between Weibull and generalized exponential distributions.

- Marshall, Meza, and Olkin (2001) used maximum likelihood and Kolmogorov distance methods to compare selected lifetime distributions, including the gamma, Weibull, and lognormal.

- Quesenberry and Kent (2001) proposed a method for selecting between distributions based on statistics that are invariant under scale transformation of the data. As pointed out by Quesenberry and Kent, however, for selecting among distributions that involve both shape and scale parameters, an optimal invariant procedure does not always exist.

- Selection based on the goodness-of-fit test, such as Pearson chi-square and the Kolmogorov-Smirnov tests, often results in more than one family of distributions deemed to be fitting the data well. This approach therefore does not always lead to selecting the best distribution for a given set of data.

In a recent paper, Guiahi (2001) discussed the issues and methodologies for fitting alternative parametric probability distributions to samples of insurance loss data. When exact sizes of loss are available, Scollnik (2001 and 2002) discussed how the Bayesian inference software package WinBUGS can be used to model loss distributions. Cairns (2000) provides detail discussion on parameter and model uncertainty.

The degree of difficulty in discriminating between two distributions has been explained by Littell, McClave, and Often (1979) and Bain and Engelhardt (1980). The problem is that often more than one family of distributions may exhibit a good fit to a given set of data. Bain and Engelhardt have pointed out that even though two models may offer similar degree of fit to a data set (even for moderate sample sizes), it is still desirable to select the correct (or more nearly correct) model, if possible, because inferences based on the model will often involve tail probabilities where the effect of the model assumption will be more critical.

The concept of long-tailed (sometimes called "heavy-tailed") distribution conveys the idea of relatively large probability mass at extreme values of the random variable. In the literature, it seems that what constitutes a long-tailed distribution depends on the context of the problem at hand and the distributions that are compared. For example, in analyzing time-varying volatility of financial data, long-tailed distributions are described as having kurtosis measure larger than the normal distribution (see Campbell, Lo, and MacKinlay 1997, pp. 480–481).

In ruin theory, heavy-tailed distributions are sometimes defined as those that satisfy the Cramér-Lundberg theorem for the probability of ultimate ruin (see Embrechts, Klüppelberg and Mikosch 1997, p. 43). One approach to compare the tail behavior of two arbitrary density functions, $f(x)$, $g(x)$, is to examine the ratio $f(x)/g(x)$ as x tends to infinity. If $g(x)$ has a heavier (lighter) tail than $f(x)$, then the ratio approaches zero (infinity) as x tends to infinity; see, for example, Klugman, Panjer, and Willmot (2004, Chapter 4.3).

In loss modeling, the concern is usually with the tail of the distribution. Small losses do not cause as much concern as large ones, so it is important that the fitted distribution has sufficient probability mass in the tail to adequately capture the probability of large losses. This is particularly relevant in reinsurance where one is required to price a high-excess layer. For this reason, in practice the lognormal and Weibull distributions are more often used than the gamma distribution.

The objective of this paper is to investigate the performance of a new model, called the log-EIG distribution, proposed by Saw, Balasooriya, and Tan (2002) and to compare it with other commonly used distributions for fitting insurance losses and other applications. It appears that the log-EIG has some features that are somewhat different from the other commonly used distributions such as the gamma, lognormal, and Weibull. In this regard, the log-EIG distribution, which generally has a thicker tail than both the Weibull and gamma distributions, is a good candidate for modeling loss data. In selecting among competing distributions, we employ the Quesenberry and Kent (2001) selection criterion. Using a Monte Carlo simulation study, we investigate the usefulness of the log-EIG distribution and its features. We also illustrate the practical usefulness of this distribution through applications to three published insurance data sets. For two of these data sets, we show that the log-EIG fits the data best, when compared with the lognormal, gamma, and Weibull distributions.

2 Properties of the Log-EIG

Saw, Balasooriya, and Tan (2002) introduced the log-EIG as an alternative loss distribution with non-zero coefficient of skewness. Its probability density function (pdf) is given by

$$\begin{aligned} \mathrm{LEIG}(x,\theta_1,\theta_2) = {} & \frac{1}{\sqrt{2\pi}\theta_2 x}\left(\frac{\theta_1}{x}\right)^{1/(2\theta_2)} \\ & \times \exp\left[-2\left(\sinh\left(\frac{1}{2\theta_2}\ln\frac{x}{\theta_1}\right)\right)^2\right] \end{aligned} \tag{1}$$

for $x > 0$, where $\theta_i > 0$ for $i = 1, 2$; θ_1 is a scale parameter and θ_2 is a shape parameter. The cumulative distribution function (cdf) of the log-EIG takes the form

$$\begin{aligned} F_X(x) = {} & \Phi\left(\left(\frac{x}{\theta_1}\right)^{1/2\theta_2} - \left(\frac{\theta_1}{x}\right)^{1/2\theta_2}\right) \\ & + e^2\Phi\left(-\left(\frac{x}{\theta_1}\right)^{1/2\theta_2} - \left(\frac{\theta_1}{x}\right)^{1/2\theta_2}\right) \end{aligned} \tag{2}$$

where, as usual, $\Phi(\cdot)$ denotes the standard normal cdf. The mean and variance of the log-EIG distribution are

$$\text{Mean} = c\theta_1 K_{\theta_2-\frac{1}{2}}(1) \tag{3}$$

$$\text{Variance} = c\theta_1^2 \left[K_{2\theta_2-\frac{1}{2}}(1) - cK^2_{\theta_2-\frac{1}{2}}(1)\right] \tag{4}$$

where $c = e\sqrt{\frac{2}{\pi}}$, and

$$K_{\theta_2 k-\frac{1}{2}}(1) = \int_0^\infty \frac{1}{2}\, w^{\theta_2 k-\frac{3}{2}} \exp\left\{-\frac{(w+w^{-1})}{2}\right\} dw$$

is a modified Bessel function; see, for example, Zhang and Jin (1996). For convenience, the probability density functions of the gamma, lognormal, and Weibull together with their means and variances are given below: the gamma distribution with parameters α and γ has pdf

$$G(x,\alpha,\gamma) = \frac{x^{\alpha-1}}{\gamma^\alpha \Gamma(\alpha)} \exp(-\frac{x}{\gamma}),$$

with mean $\alpha\gamma$ and variance $\alpha\gamma^2$; the Weibull distribution with parameters λ and β has pdf

$$W(x,\lambda,\beta) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\lambda}\right)^{\beta}\right],$$

with mean $\lambda\Gamma\left(1+\frac{1}{\beta}\right)$ and variance $\lambda^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right]$, and the lognormal distribution with parameters μ and σ has pdf

$$\text{LN}(x,\mu,\sigma) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left\{-\frac{[\ln(x/\mu)]^2}{2\sigma^2}\right\}$$

with mean $\mu\exp\left(\frac{\sigma^2}{2}\right)$ and variance $\mu^2[\exp(2\sigma^2) - \exp(\sigma^2)]$.

One can use the ratio of the density functions to show that the lognormal has a heavier tail than the gamma distribution, and that the log-EIG has a heavier tail than the gamma. For the case of Weibull, the ratio of the log-EIG pdf to the Weibull pdf is

$$\exp\left[\left(\frac{x}{\lambda}\right)^{\beta} - \frac{1}{2}\left(\frac{x}{\theta_1}\right)^{\frac{1}{\theta_2}} - \left(\frac{1}{2\theta_2}+\beta\right)\ln x\right].$$

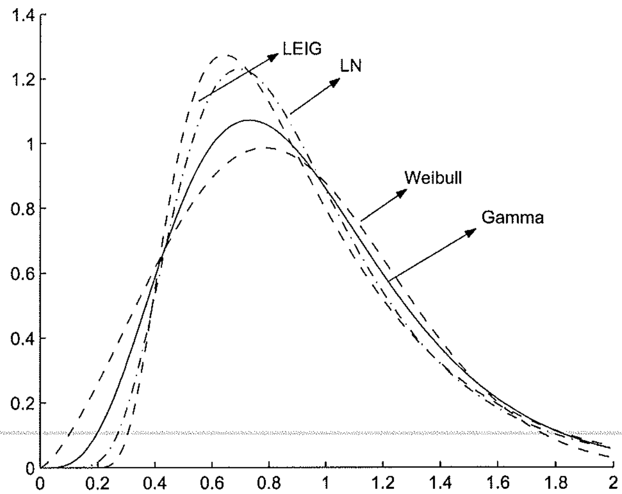

Figure 1: PDFs with Mean = 0.913149 and Variance = 0.166158

When $\beta > 1/\theta_2$ the above ratio approaches infinity when $x \to \infty$. Therefore, the log-EIG has a heavier tail than Weibull when $\beta > 1/\theta_2$.

The pdf of the gamma, log-EIG, lognormal and Weibull corresponding to a common mean and variance equal to 0.91315 and 0.16616, respectively, are shown in Figure 1. Notice that the log-EIG has the highest peak and they are all skewed to the right. Closeness of the log-EIG curve to the lognormal curve is clearly evident from Figure 1.

The functional form of the hazard function for log-EIG is analytically intractable. Saw, Balasooriya, and Tan (2002) have plotted the hazard function for several parameter values and show that it is generally non-monotone. Nevertheless, depending on the parameter values, the log-EIG distribution can accommodate a variety of situations corresponding to monotonic as well as non-monotonic failure rates.

Two important attributes of claim distributions are (i) the limited expected value (LIMEV) and (ii) the layered expected value (LAYEV). The limited expected value of a claim amount random variable X is

$$\text{LIMEV}_X(u) = \mathbb{E}[min(X, u)],$$

where u is the policy limit. In Table 1 we compare the LIMEV of the log-EIG, lognormal, gamma, and Weibull corresponding to u equal to the

Table 1
Limited Expected Values of Distributions with Fixed Mean and Variance at Selected Percentiles of the Log-EIG Distribution

	LEIG	LN	G	W
	Mean = 0.91315 and Variance = 0.16616			
	$\theta_1 = 1.0$	$\mu = 0.8338$	$\alpha = 5.0184$	$\lambda = 1.0302$
u (%tile)	$\theta_2 = 0.5$	$\sigma = 0.4263$	$\gamma = 0.1820$	$\beta = 2.3846$
1.1154 (P_{75})	0.7372	0.8249	0.8244	0.8259
1.7094 (P_{95})	0.8228	0.8967	0.9014	0.9065
2.2325 (P_{99})	0.8597	0.9092	0.9117	0.9129
	Mean = 1.0 and Variance = 1.0			
	$\theta_1 = 1.0$	$\mu = 0.7071$	$\alpha = 1.0$	$\lambda = 1.0$
u (%tile)	$\theta_2 = 1.0$	$\sigma = 0.8326$	$\gamma = 1.0$	$\beta = 1.0$
1.2441 (P_{75})	0.7342	0.7482	0.7118	0.7118
2.9221 (P_{95})	0.9353	0.9374	0.9462	0.9462
4.9841 (P_{99})	0.9855	0.9822	0.9932	0.9932
	Mean = 2.0 and Variance = 33.0			
	$\theta_1 = 1.0$	$\mu = 0.6576$	$\alpha = 0.1212$	$\lambda = 0.6955$
u (%tile)	$\theta_2 = 2.0$	$\sigma = 1.4915$	$\gamma = 16.500$	$\beta = 0.4226$
1.5477 (P_{75})	1.1818	0.7968	0.4548	0.6073
8.5385 (P_{95})	1.9833	1.5455	1.2852	1.4046
24.8412 (P_{99})	2.0000	1.8395	1.8248	1.8157

Notes: %tile = Percentile and $P_\epsilon = \epsilon^{\text{th}}$ percentile.

$75^{\text{th}}, 95^{\text{th}}$, and 99^{th} percentiles of the log-EIG when $\theta_1 = 1.0$, and $\theta_2 = 0.5, 1.0$, and 2.0. The parameter values of the competing distributions are chosen to give the same mean and variance of the log-EIG. When $\theta_1 = 1.0$ and $\theta_2 = 0.5$, the log-EIG has the smallest LIMEV among the competing distributions, whereas, when $\theta_1 = 1.0$ and $\theta_2 = 2.0$, it has the largest LIMEV. This seems to indicate that the tail thickness of the log-EIG is sensitive to changes in θ_2 values.

The layered expected claim, on the other hand, is the expected claims corresponding to different layers of insurance. Knowledge of the layered expectation is useful to insurers and reinsurers when pricing policies with deductibles and retention limits. If X is the incurred loss on a policy with a deductible L_d and a retention limit L_u, the claim amount Y paid by the insurer is given by

$$Y = \begin{cases} 0 & \text{if } X \le L_d \\ X - L_d & \text{if } L_d < X \le L_u \\ L_u - L_d & \text{if } X > L_u. \end{cases}$$

The layered expected claim is $\text{LAYEV}(L_d, L_u) = \mathbb{E}(Y)$, i.e.,

$$\text{LAYEV}(L_d, L_u) = \int_{L_d}^{L_u} (x - L_d)\mathrm{d}F_X(x) + (L_u - L_d)\int_{L_d}^{\infty} \mathrm{d}F_X(x),$$

where $F_X(x)$ is the cdf of X. The above equation can be expressed as

$$\text{LAYEV}(L_d, L_u) = \text{LIMEV}(L_u) - \text{LIMEV}(L_d).$$

In addition, the average amount per payment, AAPP, is given by:

$$\text{AAPP} = \frac{\text{LIMEV}(L_u) - \text{LIMEV}(L_d)}{P(X > L_d)}.$$

As the AAPP and $\text{LAYEV}(L_d, L_u)$ for the log-EIG are analytically complex, in Table 3 we present the AAPP and $\text{LAYEV}(L_d, L_u)$ for the competing distributions for selected L_d and L_u values corresponding to the $5^{\text{th}}, 75^{\text{th}}, 95^{\text{th}}$, and 99^{th} percentiles of the log-EIG distribution. We note from the tabulated values that the log-EIG is distinctly different from the other distributions for all the cases considered. This further indicates that the log-EIG represents a family of distributions which exhibit significant differences to the more commonly used lognormal, gamma, and Weibull distributions.

Saw, Balasooriya, and Tan (2002) have discussed the maximum likelihood estimation of the log-EIG parameters, which involves the solution of two nonlinear equations. As there are no closed-form solutions, numerical methods such as the Newton-Raphson[1] have to be used to obtain the maximum likelihood estimates.

In the case of grouped data, as is common for insurance loss data, maximum likelihood estimation may proceed along the same line as discussed in Hogg (1984, p. 122). Again, iterative methods are required to obtain maximum likelihood estimates. Alternatively, one could use other methods such as the minimum distance or minimum chi-square, as discussed in Hogg (1984, pp. 143–151).

[1] For more on the numerical solution of nonlinear equations see, for example, Burden and Faires (2001, Chapter 2).

Table 2
Average Amount per Payment and $\mathbb{E}(Y)$ for Selected Layers of the Loss Distributions with Fixed Mean and Variance

		LEIG	LN	G	W
		Mean = 0.91315 and Variance = 0.16616			
L_d †	L_u †	$\theta_1 = 1.0$, $\theta_2 = 0.5$	$\mu = 0.8338$, $\sigma = 0.4263$	$\alpha = 5.0184$, $\gamma = 0.1820$	$\lambda = 1.0302$, $\beta = 2.3846$
0.4291 (P_{05})	1.1154 (P_{75})	0.8348(0.2746)‡	0.5264(0.4951)	0.4433(0.4038)	0.4663(0.4119)
1.154(P_{75})	1.7094(P_{95})	0.8733(0.0856)	0.3646(0.0090)	0.2842(0.0077)	0.2701(0.0081)
1.7094 (P_{95})	2.2325 (P_{99})	0.3347(0.0154)	0.9522(0.2368)	0.2368(0.0104)	0.1807(0.0064)
		Mean = 1.0 and Variance = 1.0			
L_d	L_u	$\theta_1 = 1.0$, $\theta_2 = 1.0$	$\mu = 0.7071$, $\sigma = 0.8326$	$\alpha = 1.0$, $\gamma = 1.0$	$\lambda = 1.0$, $\beta = 1.0$
0.1841 (P_{05})	1.2441 (P_{75})	1.2604(0.5521)	0.7310(0.6923)	0.6535(0.5437)	0.6535(0.5437)
1.2441 (P_{75})	2.9221 (P_{95})	1.7442(0.2010)	0.9522(0.2368)	0.8133(0.2344)	0.8133(0.2344)
2.9221 (P_{95})	4.9841 (P_{99})	2.1796(0.0502)	1.2519(0.0553)	0.8728(0.0470)	0.8728(0.0470)
		Mean = 2.0 and Variance = 33.0			
L_d	L_u	$\theta_1 = 1.0$, $\theta_2 = 2.0$	$\mu = 0.6576$, $\sigma = 1.4915$	$\alpha = 0.1212$, $\gamma = 16.500$	$\lambda = 0.6955$, $\beta = 0.4226$
0.0339 (P_{05})	1.5477 (P_{75})	2.4897(1.1479)	0.9235(0.9019)	0.8735(0.4361)	0.7659(0.5794)
1.5477 (P_{75})	8.5385 (P_{95})	3.6231(0.8015)	3.3025(0.9347)	3.9169(0.8303)	3.2402(0.7973)
8.5385 (P_{95})	24.8412 (P_{99})	4.0259(0.0167)	8.5380(0.3656)	7.8038(0.5396)	7.3645(0.4111)

Notes: † Values in parentheses are percentiles of the LEIG distribution; ‡ Values in parentheses are $\mathbb{E}(Y)$.

Table 3

Percentage of Selections Among Different Groups of Candidate Models Using the QK Criterion when n = 50 and 100†

Model	Number of Candidate Models						
	4				2	2	2
LEIG	LEIG	LN	G	W	LN	G	W
$\theta_2 = 0.5$	74.95	14.95	9.05	1.05	22.80	14.10	3.85
	75.78	*20.42*	*3.80*	*0.00*	*23.92*	*8.31*	*0.50*
$\theta_2 = 1.0$	86.55	9.15	3.90	0.40	12.65	5.35	3.10
	85.40	*13.70*	*0.90*	*0.00*	*14.60*	*1.40*	*0.60*
$\theta_2 = 2.0$	75.55	21.85	0.00	2.60	23.80	1.85	4.50
	78.40	*21.30*	*0.00*	*0.30*	*21.60*	*0.30*	*0.60*
LN					LEIG	G	W
$\sigma = 0.5$	35.95	41.25	19.75	3.05	36.85	23.50	7.85
	27.60	*55.30*	*16.80*	*3.00*	*27.80*	*17.30*	*2.60*
$\sigma = 1.0$	31.90	56.35	10.70	1.05	48.75	11.95	9.95
	35.70	*60.50*	*3.80*	*0.00*	*35.70*	*3.80*	*2.00*
$\sigma = 2.0$	39.40	51.60	0.15	8.85	39.85	2.70	9.20
	35.20	*62.70*	*0.00*	*2.10*	*35.20*	*0.30*	*2.10*
G					LEIG	LN	W
$\gamma = 0.5$	1.45	0.75	68.30	29.50	3.05	4.35	31.60
	0.00	*0.00*	*75.00*	*25.00*	*0.00*	*0.30*	*25.00*
$\gamma = 1.0$	3.15	6.35	45.00	45.50	5.85	9.75	49.35
	0.00	*1.40*	*52.60*	*46.00*	*0.00*	*1.80*	*46.70*
$\gamma = 2.0$	11.85	8.80	47.95	31.40	15.80	18.90	31.40
	0.30	*8.40*	*64.50*	*26.80*	*1.50*	*8.70*	*26.80*
W					LEIG	LN	G
$\beta = 0.5$	6.25	5.20	23.20	65.35	8.85	10.65	23.20
	0.00	*3.00*	*15.70*	*81.30*	*1.10*	*3.00*	*15.70*
$\beta = 1.0$	2.75	6.30	47.20	43.75	5.85	9.95	51.95
	0.30	*2.70*	*49.70*	*47.30*	*1.60*	*3.10*	*51.60*
$\beta = 2.0$	1.75	2.50	25.40	70.35	6.55	10.30	29.65
	0.10	*0.00*	*18.50*	*81.40*	*0.70*	*2.80*	*18.60*

Notes: Italicized values refer to n = 100.

Table 3 (Contd.)
Percentage of Selections Among Different Groups of Candidate Models Using the QK Criterion when n = 50 and n = 100

Model	Number of Candidate Models								
	3			3			3		
LEIG	LEIG	LN	G	LEIG	LN	W	LEIG	G	W
θ_2 = 0.5	74.95	14.95	10.10	76.85	20.85	2.30	85.90	13.05	1.05
	75.78	*20.42*	*3.80*	*76.08*	*23.82*	*0.10*	*91.69*	*8.31*	*0.00*
θ_2 = 1.0	86.55	9.15	4.30	87.15	10.70	2.15	94.65	4.95	0.40
	85.40	*13.70*	*0.90*	*85.40*	*14.30*	*0.30*	*98.60*	*1.40*	*0.00*
θ_2 = 2.0	76.15	22.80	1.05	75.55	21.85	2.60	95.50	0.00	4.50
	78.40	*21.50*	*0.10*	*78.40*	*21.30*	*0.30*	*99.40*	*0.00*	*0.60*
LN	LEIG	LN	G	LEIG	LN	W	LN	G	W
σ = 0.5	35.95	41.25	22.80	36.75	55.50	7.75	76.50	20.45	3.05
	27.60	*55.30*	*17.10*	*27.80*	*69.60*	*2.60*	*65.70*	*34.00*	*0.30*
σ = 1.0	31.90	56.35	11.75	32.00	58.05	9.95	88.05	10.09	1.05
	35.70	*60.50*	*3.80*	*35.70*	*62.30*	*2.00*	*96.20*	*3.80*	*0.00*
σ = 2.0	39.70	57.65	2.65	39.40	51.60	9.00	90.80	0.15	9.05
	35.20	*64.50*	*0.30*	*35.20*	*62.70*	*2.10*	*97.90*	*0.00*	*2.10*

Notes: Italicized values refer to n = 100.

Table 3 (Contd.)
Percentage of Selections Among Different Groups of
Candidate Models Using the QK Criterion when n = 50 and n = 100

Model	Number of Candidate Models								
	3			3			3		
G	LEIG	LN	G	LEIG	G	W	LN	G	W
$\gamma = 0.5$	1.95	2.65	95.40	1.95	68.35	29.70	1.85	68.35	29.80
	0.00	*0.30*	*99.70*	*0.00*	*75.00*	*25.00*	*0.00*	*75.00*	*25.00*
$\gamma = 1.0$	3.15	6.80	90.05	5.60	46.90	47.50	9.30	45.20	45.50
	0.00	*1.80*	*98.20*	*0.00*	*53.30*	*46.70*	*1.40*	*52.60*	*46.00*
$\gamma = 2.0$	11.85	8.80	79.35	15.80	52.80	31.40	18.90	49.70	31.40
	0.30	*8.40*	*91.30*	*1.50*	*71.70*	*26.80*	*8.70*	*64.50*	*26.80*
W	LEIG	LN	W	LEIG	G	W	LN	G	W
$\beta = 0.5$	6.25	5.20	88.55	8.85	23.20	67.95	10.65	23.20	66.15
	0.00	*3.00*	*97.00*	*1.10*	*15.70*	*83.20*	*3.00*	*15.70*	*81.30*
$\beta = 1.0$	2.90	7.35	89.75	5.20	49.05	45.75	8.85	47.30	43.85
	0.30	*2.80*	*96.90*	*1.50*	*50.50*	*48.00*	*3.00*	*49.70*	*47.30*
$\beta = 2.0$	2.05	8.30	89.65	2.75	26.90	70.35	4.00	25.65	70.35
	0.10	*2.70*	*97.20*	*0.10*	*18.50*	*81.40*	*0.10*	*18.50*	*81.40*

Notes: Italicized values refer to n = 100.

3 Selection Procedure

For a given set of n observations $x_1, x_2, \cdots, x_n$, suppose it is required to choose one member from among a set of competing families of distributions $F_1, F_2, \cdots, F_k$ with scale and shape parameters, ϑ_i and ν_i, that best fits the data. Let f_i be the probability density function corresponding to F_i, $i = 1, 2, \ldots k$. The optimum invariant selection criterion of Quesenberry and Kent (2001) selects F_i which maximizes the selection statistic

$$S_i = \int_0^\infty f_i(tx_1, tx_2, \cdots, tx_n)\, t^{n-1} dt,$$

where $\vartheta_i = 1, i = 1, 2, \ldots, k$. Note, for a random sample $x_1, x_2, \ldots, x_n$, the above function can be expressed as a product of the f_i's, i.e.,

$$f_i(tx_1, tx_2, \cdots, tx_n) = \prod_{j=1}^n f_i(tx_j).$$

For the case of log-EIG where $\vartheta_i = \theta_1 = 1$ and $\nu_i = \theta_2$, it can be shown that the statistic, S_i, is given by

$$\left(\frac{e}{\sqrt{2\pi}\theta_2}\right)^n \prod_{j=1}^n \left(\frac{1}{x_j}\right)^{1+1/2\theta_2} \int_0^\infty \frac{1}{t^{1+n/2\theta_2}} \exp\left\{-\frac{1}{2}\left(t^{1/\theta_2}\phi + \frac{\psi}{t^{1/\theta_2}}\right)\right\} dt,$$

where $\phi = \sum_{j=1}^n x_j^{1/\theta_2}$ and $\psi = \sum_{j=1}^n x_j^{-1/\theta_2}$. The selection statistics for the other distributions can be similarly derived and are given in Quesenberry and Kent (2001).

When $\nu_1, \nu_2, \ldots, \nu_k$ are unknown, Quesenberry and Kent (2001) proposed that a suitable scale invariant estimate be substituted for ν_i. The selection criterion is then said to be suboptimal invariant. From extensive Monte Carlo studies involving the gamma, lognormal, and Weibull distributions, Quesenberry and Kent (2001) established that the proposed selection procedure performs well when selecting among families of distributions with shape and scale parameters.

For the log-EIG, lognormal, and Weibull distributions, when applying the suboptimal procedure, we substitute the shape parameter by its maximum likelihood estimates in the computation of S_i. Following Quesenberry and Kent (2001), for the gamma distribution we employ the approximate maximum likelihood estimate of the shape parameter proposed by Greenwood and Durand (1960); that is

$$\hat{v} = \begin{cases} \dfrac{0.5000876 + 0.1648852R - 0.0544274R^2}{R} & \text{for } 0 < R \le 0.5772, \\ \dfrac{8.898919 + 9.059950R + 0.9775373R^2}{R(17.79728 + 11.968477R + R^2)} & \text{for } 0.5772 < R \le 17, \end{cases}$$

where

$$R = \ln\left(\frac{\text{arithmetic mean of the observations}}{\text{geometric mean of the observations}}\right).$$

In selecting among probability models one also can use information theoretic criteria such as the Akaike information criterion (AIC) or some of its modifications such as the AIC with finite corrections (AICC) [Sugiura, 1978], or the Bayesian information criterion (BIC) [Schwarz, 1978]. For the four distributions considered in this paper, the AIC, AICC, and BIC give identical results because these distributions have the same dimension.[2] Thus, for comparing with the Quesenberry and Kent criterion (QK), we only report the selection results using the AIC criterion.

4 Simulation Results

In our simulation study, we generated 2,000 random samples of size $n = 50$ and 1,000 samples of size $n = 100$ from each of the four distributions gamma, log-EIG, lognormal, and Weibull. Random observations from the lognormal, gamma, and Weibull distributions were generated using MATLAB® standard routines for selected values of the parameters. For the log-EIG distribution, random observations were obtained by first generating inverse Gaussian variates using Dataplot and then transforming them to log-EIG variates using the relationships between the inverse Gaussian, exponential inverse Gaussian, and the log-EIG distributions; see Kanefuji and Iwase (1996) and Saw et al. (2002). It follows from these relationships that if Z is distributed as Inverse Gaussian with shape and location parameters both equal to 1, then $X = \theta_1 Z^{\theta_2}$ has a LEIG(θ_1, θ_2) distribution.

Table 3 presents percentages of selections among different groupings of candidate models consisting of 4, 3, and 2 competing distributions when the data are generated by the model indicated in the first

[2]When the competing models have the same number of parameters, they are said to have the same dimension; see Judge, Griffiths, Hill, Lütkepohl, and Lee (1985, pp. 870–873).

column of the table. The values in parentheses are percentages of selections when $n = 100$. For example, the entries 74.95, 14.95, 9.05, 1.05 at the beginning of the table mean that when the data are generated from a log-EIG distribution with parameters $\theta_1 = 1$ and $\theta_2 = 0.5$, the suboptimal selection procedure selected the log-EIG, lognormal, gamma, and Weibull as the population distribution 74.95%, 14.95%, 9.05%, and 1.05% of the time, respectively. The tabulated values under the heading '3' give the percentages of selections for groups of three competing distributions where the true population distribution is one of the competing members. The tabulated values under the heading '2' give the percentages of selections for the specified distribution under each heading, when compared with the population distribution indicated in the first column of the table. The entries therefore represent percentages of incorrect selections. For comparison, in Table 4 we present percentages of correct selection using the AIC selection criterion.

In distinguishing the log-EIG when it is the true population with all the alternative groupings of families considered, the lowest percentage of the correct selection is 74.95 (73.35) for the case when $\theta_2 = 0.5$ ($\theta_2 = 2.0$). To save space, note that throughout this section the figures in parentheses refer to the corresponding values for AIC criterion reported in Table 4. When data are generated from the lognormal, gamma, and Weibull distributions, the lowest percentage of correct selections are 41.25% (28.80%) when $\sigma = 0.5(\sigma = 0.5)$, 45.00% (42.15%) when $\gamma = 1.0(\gamma = 1.0)$ and 43.75% (46.50%) when $\beta = 1.0(\beta = 1.0)$, respectively. This seems to indicate that the log-EIG, the new addition to the location and scale family of distributions, has some features that are somewhat different from the other commonly used loss distributions.

From the tabulated values in Tables 3 and 4, we note that when the true distribution is log-EIG, among the other competing three distributions, the lognormal is selected more often than the gamma or Weibull. On the other hand, when the true distribution is lognormal, the log-EIG is selected more often than the gamma or Weibull in all the groupings considered. For example, when two distributions compete, and samples of size $n = 50$ are generated from lognormal with $\sigma = 0.5, 1.0, 2.0$, log-EIG is selected 36.85% (50.0%), 48.75% (48.75%), 39.85% (44.40%) versus 23.50% (24.60%), 11.95% (12.95%), 2.70% (3.75%) for G, and 7.85% (8.60%), 9.95% (10.75%), 9.20% (10.35%) for Weibull, respectively. The corresponding figures for lognormal when the samples are generated from log-EIG with $\theta_2 = 0.5, 1.0, 2.0$ are 22.80% (23.15%), 12.65% (23.50%), 23.80% (26.45%), versus 14.10% (14.25%), 5.35% (7.70%), 1.85% (2.70%) for G, and 3.85% (4.65%), 3.10% (4.30%), 4.50% (5.00%) for Weibull, respectively. The same pattern is observed for the case of $n = 100$ although the corresponding percentages of incorrect for log-EIG and lognormal are somewhat lower than when $n = 50$. These findings seem to indicate that the log-EIG is closer to the lognormal than to the gamma or Weibull distributions.

While both QK and AIC criteria yield high percentages of correction selections, the QK performs marginally better in most of the cases considered in this simulation study. The QK criterion, however, is computationally more involved than the AIC.

Next we consider the situation when data arise from a log-EIG distribution but the investigator considers choosing one of the gamma, lognormal or Weibull to fit the data. Table 5 gives the percentages of selections for gamma, lognormal, and Weibull by the suboptimal selection procedure for the competing groupings $\{G, \text{LN}, \text{Weibull}\}$, $\{G, \text{LN}\}$, $\{\text{LN}, \text{Weibull}\}$, and $\{G, \text{Weibull}\}$ when the data are generated from the log-EIG with various values of the shape parameter θ_2. Again as we observed earlier, the tabulated values clearly indicate that the lognormal distribution is the closest distribution to the log-EIG for all the θ_2 values considered. When only gamma and Weibull are considered, gamma appears to be closer to log-EIG for $\theta_2 = 0.5$ or 1.0, while Weibull is closer to log-EIG when $\theta_2 = 2.0$. This is consistent with the higher selection proportions for gamma when $\theta_2 = 0.5$ or 1.0 and higher selection proportion for Weibull when $\theta_2 = 2.0$ in the simulation results reported in Tables 3 and 4. Therefore, it seems that when gamma and Weibull compete to represent log-EIG, the selection depends on the shape parameter of the log-EIG from which the data arise.

The similarities/differences among the four distributions are further illustrated by Table 6 which compares selected percentile values of the distributions with the same mean and variance, i.e., given the first two moments of the distributions. The selected common means and variances correspond to the log-EIG when $(\theta_1, \theta_2) = (1.0, 0.5)$, (1.0, 1.0), (1.0, 2.0). The parameter values for the lognormal, gamma, and Weibull distributions for the given means and variances are reported in the table. From the table, it can be seen that the percentiles for lognormal are closer to that of the log-EIG than to the gamma or Weibull. Further, the percentiles for gamma are closer to the log-EIG than the Weibull for $(\theta_1, \theta_2) = (1.0, 0.5)$, (1.0, 1.0), while the converse is true when $(\theta_1, \theta_2) = (1.0, 2.0)$. These observations are consistent with the simulation results reported in Tables 3, 4, and 5 and provide some theoretical justification for the simulation results.

Table 4

Percentage of Selections Among Different Groups of Candidate Models Using the AIC Criterion when $n = 50$ and $100^{\dagger}$

Model	Number of Candidate Models: 4				2	2	2
LEIG	LEIG	LN	G	W	LN	G	W
$\theta_2 = 0.5$	75.20	14.05	9.55	1.20	23.15	14.25	4.65
	79.10	*16.70*	*4.20*	*0.00*	*20.40*	*8.40*	*0.50*
$\theta_2 = 1.0$	75.95	19.00	4.45	0.60	23.50	7.70	4.30
	80.30	*18.70*	*1.00*	*0.00*	*19.60*	*1.80*	*0.60*
$\theta_2 = 2.0$	73.35	23.20	0.05	3.40	26.45	2.70	5.00
	77.20	*22.40*	*0.00*	*0.40*	*22.80*	*0.30*	*0.90*
LN					LEIG	G	W
$\sigma = 0.5$	48.80	28.80	18.95	3.45	50.00	24.60	8.60
	34.40	*48.40*	*16.80*	*0.40*	*34.90*	*17.40*	*2.70*
$\sigma = 1.0$	48.35	39.15	11.10	1.40	48.75	12.95	10.75
	36.40	*59.70*	*3.90*	*0.00*	*36.40*	*3.90*	*2.30*
$\sigma = 2.0$	44.20	45.75	0.25	9.80	44.40	3.75	10.35
	39.30	*58.40*	*0.00*	*2.30*	*39.30*	*0.30*	*2.30*
G					LEIG	LN	W
$\gamma = 0.5$	0.85	1.05	72.10	26.00	2.30	3.80	27.80
	0.00	*0.00*	*77.40*	*22.60*	*0.00*	*0.30*	*22.60*
$\gamma = 1.0$	3.05	5.60	42.15	49.20	6.15	8.90	52.55
	0.10	*1.20*	*51.70*	*47.00*	*0.60*	*1.40*	*47.60*
$\gamma = 2.0$	7.50	10.45	46.65	35.40	12.35	17.65	35.40
	1.10	*7.20*	*62.30*	*29.40*	*3.50*	*8.70*	*29.40*
W					LEIG	LN	G
$\beta = 0.5$	3.65	6.55	26.65	63.15	6.45	9.90	26.65
	0.30	*2.50*	*16.80*	*80.40*	*1.30*	*2.80*	*16.80*
$\beta = 1.0$	3.10	5.40	45.00	46.50	5.95	9.30	49.60
	0.00	*2.80*	*47.50*	*49.70*	*1.50*	*2.90*	*49.30*
$\beta = 2.0$	2.50	1.80	22.55	73.15	6.55	9.65	26.85
	0.10	*0.00*	*17.40*	*82.50*	*0.90*	*2.70*	*17.50*

Notes: Italicized values refer to $n = 100$.

Table 4 (Contd.)
Percentage of Selections Among Different Groups of
Candidate Models Using the AIC Criterion when $n = 50$ and $n = 100$

	Number of Candidate Models								
Model	3			3			3		
LEIG	LEIG	LN	G	LEIG	LN	W	LEIG	G	W
$\theta_2 = 0.5$	75.20	14.05	10.75	76.45	20.80	2.75	85.75	13.05	1.20
	79.10	*16.70*	*4.20*	*79.60*	*20.20*	*91.60*	*91.60*	*8.40*	*0.00*
$\theta_2 = 1.0$	75.95	19.00	5.05	76.20	21.10	2.70	92.30	7.10	0.60
	80.30	*18.70*	*1.00*	*80.40*	*19.30*	*0.30*	*98.20*	*1.80*	*0.00*
$\theta_2 = 2.0$	73.45	25.05	1.50	73.35	23.20	3.45	95.00	0.05	4.95
	77.20	*22.70*	*0.10*	*77.20*	*22.40*	*0.40*	*99.10*	*0.00*	*0.90*
LN	LEIG	LN	G	LEIG	LN	W	LN	G	W
$\sigma = 0.5$	48.80	28.80	22.40	49.65	42.05	8.30	75.40	21.10	3.50
	34.40	*48.40*	*17.20*	*34.90*	*62.40*	*2.70*	*82.60*	*17.00*	*0.40*
$\sigma = 1.0$	48.35	39.15	12.50	48.45	40.95	10.60	87.05	11.55	1.40
	36.40	*59.70*	*3.90*	*36.40*	*61.30*	*2.30*	*96.10*	*3.90*	*0.00*
$\sigma = 2.0$	44.35	51.95	3.70	44.20	45.75	10.05	89.65	0.25	10.10
	39.30	*60.40*	*0.30*	*39.30*	*58.40*	*2.30*	*97.70*	*0.00*	*2.30*

Notes: Italicized values refer to $n = 100$.

Table 4 (Contd.)

Percentage of Selections Among Different Groups of Candidate Models Using the AIC Criterion when $n = 50$ and $n = 100$

Model	3			3			3		
G	LEIG	LN	G	LEIG	G	W	LN	G	W
$\gamma = 0.5$	0.95	2.90	96.15	1.20	72.10	26.70	1.75	72.15	26.10
	0.00	*0.30*	*99.70*	*0.00*	*77.40*	*22.60*	*0.00*	*77.40*	*22.60*
$\gamma = 1.0$	3.20	5.90	90.90	5.80	43.75	50.45	8.55	42.25	49.20
	0.10	*1.30*	*98.60*	*0.60*	*52.00*	*47.40*	*1.30*	*51.70*	*47.00*
$\gamma = 2.0$	7.50	10.45	82.05	12.35	52.25	35.40	17.65	46.95	35.40
	1.10	*7.20*	*91.70*	*3.50*	*67.10*	*29.40*	*8.30*	*62.30*	*29.40*
W	LEIG	LN	W	LEIG	G	W	LN	G	W
$\beta = 0.5$	3.65	6.55	89.80	6.45	26.65	66.90	9.90	26.65	63.45
	0.30	*2.50*	*97.20*	*1.30*	*16.80*	*81.90*	*2.80*	*16.80*	*80.40*
$\beta = 1.0$	3.10	6.45	90.45	5.45	46.75	47.80	8.20	45.10	46.70
	0.00	*2.90*	*97.10*	*1.30*	*48.40*	*50.30*	*2.80*	*47.50*	*49.70*
$\beta = 2.0$	2.95	6.75	96.30	3.15	23.70	73.15	3.80	23.05	73.15
	0.10	*2.60*	*97.30*	*0.10*	*17.40*	*82.50*	*0.10*	*17.40*	*82.50*

Notes: Italicized values refer to $n = 100$.

Table 5
Percentage of Selections, Using the QK Criterion, Among Different Groups of Candidate Models in the Absence of LEIG When Data are Generated from Log-EIG for $n = 50$ and 100

	Number of Candidate Models									
Model	3			2		2		2	
LEIG	LN	G	W	LN	G	LN	W	G	W
$\theta_2 = 0.5$	87.25	11.70	1.05	87.25	12.75	97.50	2.50	98.95	1.05
	95.80	*4.20*	*0.00*	*95.80*	*4.20*	*99.90*	*0.10*	*100.0*	*0.00*
$\theta_2 = 1.0$	95.05	4.55	0.44	95.05	4.95	97.60	2.40	98.15	1.85
	98.90	*1.10*	*0.0*	*98.9*	*1.10*	*99.70*	*0.30*	*99.90*	*0.10*
$\theta_2 = 2.0$	97.00	0.00	3.00	98.80	1.20	97.00	3.00	0.10	99.90
	99.60	*0.00*	*0.40*	*99.90*	*0.10*	*99.60*	*0.40*	*0.00*	*100.00*

Notes: Italicized values refer to $n = 100$.

Table 6

Percentile Values for Selected Distributions with Fixed Mean and Variance

LEIG(θ_1, θ_2)					LN(μ, σ)				
Mean = 0.91315 and Variance = 0.16616									
$\theta_1 = 1.0; \theta_2 = 0.5$					$\mu = 0.8338; \sigma = 0.4263$				
P_{25}	P_{50}	P_{75}	P_{95}	P_{99}	P_{25}	P_{50}	P_{75}	P_{95}	P_{99}
0.6162	0.8221	1.1154	1.7094	2.2325	0.6254	0.8338	1.1115	1.6810	2.2477
Mean = 1.0 and Variance = 1.0									
$\theta_1 = 1.0; \theta_2 = 1.0$					$\mu = 0.7071; \sigma = 0.8326$				
0.3797	0.6758	1.2441	2.9221	4.9841	0.4033	0.7071	1.2398	2.7813	4.9053
Mean = 2.0 and Variance = 33.0									
$\theta_1 = 1.0; \theta_2 = 2.0$					$\mu = 0.6576; \sigma = 1.4915$				
0.1442	0.4568	1.5477	8.5385	24.8412	0.2405	0.6576	1.7982	7.6454	21.1269

G(α, γ)					W(λ, β)				
Mean = 0.91315 and Variance = 0.16616									
$\alpha = 5.0184; \gamma = 0.1820$					$\lambda = 1.0302; \beta = 2.3846$				
P_{25}	P_{50}	P_{75}	P_{95}	P_{99}	P_{25}	P_{50}	P_{75}	P_{95}	P_{99}
0.6159	0.8535	1.1458	1.6706	2.1172	0.6110	0.8834	1.1814	1.6321	1.9546
Mean = 1.0 and Variance = 1.0									
$\alpha = 1.0; \gamma = 1.0$					$\lambda = 1.0; \beta = 1.0$				
0.2877	0.6932	1.3863	2.9957	4.6052	0.2877	0.6932	1.3863	2.9957	4.6052
Mean = 2.0 and Variance = 33.0									
$\alpha = 0.1212; \gamma = 16.5000$					$\lambda = 0.6955; \beta = 0.4226$				
0.0001	0.0335	1.0001	11.3970	28.7995	0.0365	0.2922	1.5065	9.3292	25.8069

5 Illustrative Examples

We first consider a well-known data set from Hogg and Klugman (1984, p. 128) on hurricane losses. This data set consists of 38 observations on losses that exceeded \$1,000,000 for the period 1949 to 1980 as compiled by the American Insurance Association. With censoring below \$5,000,000, using the remaining 35 observations, Hogg and Klugman concluded that the Weibull distribution fits the data best when compared with the lognormal and Pareto distributions, using the Chi-squared goodness-of-fit test. Our second data set is obtained from Klugman, Panjer, and Willmot (1998, Table 1.1, p. 18). This data set corresponds to insurance liability payments and reflects a real-life problem encountered by the authors. The third data set of 96 individual claims is from Currie (1992, Table 1, p. 3). Currie (1992) used the chi-square goodness-of-fit test and concluded that the Pareto model is the best model for this data set.

For these data sets, the parameter estimates and the computed values of the selection statistics, S_i and AIC, for the competing distributions are reported in Table 7. For data sets one and two, both the statistics, S_i and AIC, selected the log-EIG distribution as the underlying distribution that generated the data. For the third data set, while the lognormal was chosen, the log-EIG was the closest competitor among the other families of distributions considered in this study.

6 Concluding Remarks

In this study we consider a recently introduced lifetime distribution, the log-EIG distribution. We show that it has a heavier tail than the gamma or Weibull distributions over certain parameter space. Further, the log-EIG distribution appears to be distinct from the other commonly used lifetime distributions. The illustrative examples indicate the usefulness of the log-EIG distribution in fitting some insurance loss data. In the simulated samples, we observed that the log-EIG distribution generated a few unusually large observations more frequently than the other competing distributions. This feature makes the log-EIG distribution a potentially useful model for insurance claims where extreme observations are not uncommon, such as catastrophic losses in liability claims. Another area where log-EIG can be potentially useful is in lifetime and reliability modeling.

Table 7
Parameter Estimates and Values of the Selection Statistics for Selected Data Sets

Data Set	LEIG		LN	
	θ_1	θ_2	μ	σ
1	43190.8773	1.8970	21587.3367	2.7043
	QK	AIC	QK	AIC
	-451.2414	906.1591	-452.4587	908.1124
	θ_1	θ_2	μ	σ
2	321370.7325	2.8030	113498.2855	6.0245
	QK	AIC	QK	AIC
	-585.0090	1173.5732	-586.2208	1176.3376
	θ_1	θ_2	μ	σ
3	2037.6606	1.7378	1120.4416	1.9565
	QK	AIC	QK	AIC
	-845.5107	1693.3413	-843.5242	1688.9931

Data Set	G		W	
	α	γ	λ	β
1	154916.4796	0.4985	49891.5848	0.6185
	QK	AIC	QK	AIC
	-457.9213	918.7478	-455.5948	914.3550
	α	γ	λ	β
2	4531919.5872	0.2856	400863.9222	0.4153
	QK	AIC	QK	AIC
	-593.5978	1190.5629	-589.8482	1183.5582
	α	γ	λ	β
3	4778.3910	0.6257	2244.5103	0.7132
	QK	AIC	QK	AIC
	-856.9187	1715.5828	-851.1017	1704.1546

†QK - Quesenberry and Kent Criterion; AIC - Akaike Criterion

The selection criterion employed here is suboptimal invariant and it is applicable for uncensored data. The procedure requires that the unknown shape parameter be replaced by a scale invariant estimate. From the results reported in the simulation study, it is clear that this procedure performs well in identifying the true family of distribution that generates a given set of data.

References

Bain, L.J. and Engelhardt, M. "Probability of Correct Selection of Weibull Versus Gamma Based on Likelihood Ratio." *Communications in Statistics-Theory and Methods* A9 (1980): 375–381.

Burden, R.L. and Faires, J.D. *Numerical Analysis* Seventh Edition. New York: Brooks/Cole Publishing Company, 2001.

Cairns, A.J.G."A Discussion of Parameter and Model Uncertainty in Insurance." *Insurance: Mathematics and Economics* 27, no. 3 (2000): 313–330.

Campbell, J.Y., Lo, A., and MacKinlay, A.C. *The Econometrics of Financial Markets.* Princeton, NJ: Princeton University Press, 1997.

Cox, D.R. "Further Results on Tests of Separate Families of Hypotheses." *Journal of the Royal Statistical Society, Series B* 24 (1962): 406–424.

Currie, I.D. *Loss Distributions.* London, England: Institute of Actuaries and Faculty of Actuaries, 1992.

Dumonceaux, R., Antle, C.E. and Haas, G. "Likelihood Ratio Test for Discrimination Between Two Models with Unknown Location and Scale Parameters." *Technometrics* 15 (1973): 19–27.

Embrechts, P., Klüppelberg, C. and Mikosch, T. *Modeling Extremal Events for Insurance and Finance.* New York, NY: Springer-Verlag, 1997.

Geisser, S. and Eddy, W.F. "A Predictive Approach to Model Selection." *Journal of the American Statistical Association* 74 (1979): 153–160.

Greenwood, J.A. and Durand, D. "Aids for Fitting the Gamma Distribution by Maximum Likelihood." *Technometrics* 2, no. 1 (1960): 55–65.

Guiahi, F. "Fitting Loss Distributions in the Presence of Rating Variables." *Journal of Actuarial Practice* 9 (2001): 97–129.

Gupta, R.D. and Kundu, D. "Discriminating Between Weibull and Generalized Exponential Distributions." *Computational Statistics and Data Analysis* 43 (2003): 179–196.

Hogg, R.V. and Klugman, S.A. *Loss Distributions.* New York, NY: John Wiley & Sons Inc., 1984.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H. and Lee, T. C. *The Theory and Practice of Econometrics,* Second Edition. New York, NY: John Wiley & Sons Inc., 1985.

Klugmann, S.A., Panjer, H.H., and Willmot, G.E. *Loss Models: From Data to Decisions,* Second Edition. New York, NY: John Wiley & Sons Inc., 2004.

Lehmann, E.L. *Testing Statistical Hypotheses.* New York, NY: John Wiley & Sons Inc., 1959.

Littell, R.C., McClave, J.T. and Often, W.W. "Goodness-of-Fit-Tests for the Two-Parameter Weibull Distribution." *Communications in Statistics-Theory and Methods* B8 (1979): 257-269.

Marshall, A.W., Meza, J.C. and Olkin, I. "Accelerated Degradation Tests: Modeling and Analysis." *Journal of Computational and Graphical Statistics* 10, no. 3 (2001): 555-580.

Quesenberry, C.P. and Kent, J. "Selecting Among Probability Distributions Used in Reliability." *Technometrics* 24, no. 1 (1982): 59-65.

Saw, S. L.C., Balasooriya, U. and Tan, K.C. "The Log-EIG Distribution: A New Probability Model for Lifetime Data." *Communications in Statistics-Theory and Methods* 31, no. 11 (2002): 1913-1926.

Schwarz, G. "Estimating the Dimension of a Model." *Annals of Statistics* 6, no. 2 (1978): 461-464.

Scollnik, D.P.M. "Actuarial Modeling with MCMC and BUGS." *North American Actuarial Journal* 5, no. 2 (2001): 96-124.

Scollnik, D.P.M. "Modeling Size-of-Loss Distributions for Exact Data in WinBUGS." *Journal of Actuarial Practice* 10 (2002): 193-218.

Sugiura, N. "Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections." *Communications in Statistics-Simulation and Computation* 7 (1978): 13-26.

Zhang, S. and Jin, J. *Computation of Special Functions.* New York, NY: John Wiley & Sons Inc., 1996.

Journal of Actuarial Practice Vol. 12, 2005

A Modern Approach to Modeling Insurances on Two Lives

Mária Bilíková* and Graham Luffrum†

Abstract‡

The analysis of life insurance contracts on two lives using the traditional deterministic approach has been an important part of actuarial education for the past fifty years or more. Recently there has been a shift from this deterministic approach to one using a more modern stochastic approach involving the future lifetime random variable. In this paper we will look at the problem using multiple-state models. In our view this approach allows a deeper analysis than either the traditional or the random future lifetime ones.

Key words and phrases: *multiple-state models, Kolmogorov equations*

*Mária Bilíková, RNDr., Ph.D., is an associate professor at the Economics University, Bratislava. She graduated in mathematics in 1974 from Comenius University in Bratislava, Slovakia. Dr. Bilíková has published six textbooks and numerous articles. Her research interests are in the application of mathematical models in life insurance. She is president of the board of the Slovak Society of Actuaries.

Dr. Bilíková's address is: Department of Mathematics, Faculty of Economic Informatics, University of Economics, Dolnozemska cesta 1,852 35 Bratislava 5, SLOVAKIA. E-mail: *bilikova@dec.euba.sk*

†Graham Luffrum, M.A., F.I.A., is an external lecturer at the Economics University, Bratislava. He graduated in mathematics from Cambridge University in 1969 and became a Fellow of the Institute of Actuaries in 1974. After working for a large U.K. insurance company, Mr. Luffrum worked for the Institute of Actuaries, teaching its Life Insurance Practice examination. Since 1992 he has been actively involved in the development of actuarial education in the countries of Central and Eastern Europe. His research interests are in the use of stochastic modeling in life insurance.

Mr. Luffrum's address is: Department of Mathematics, Faculty of Economic Informatics, University of Economics, Dolnozemska cesta 1,852 35 Bratislava 5, SLOVAKIA. E-mail: *grahamgl@stonline.sk*

‡The authors would like to thank the two anonymous referees for their helpful comments that have enabled us to improve the content of the paper.

1 Introduction

Insurance for multiple lives is largely confined to those associated with married couples. So, throughout this paper, we consider a married couple consisting of a husband age x and a wife age y at some initial time $t = 0$, which, for notational convenience, is written as (x, y). The time $t = 0$ usually corresponds in practice to the start of an insurance contract. For simplicity we ignore the possibility of divorce.

Traditionally, actuaries have calculated the premiums for joint-life insurance and annuity contracts using the formula for the joint force of mortality, $\mu_{xy}(t)$,

$$\mu_{xy}(t) = \mu_x(t) + \mu_y(t) \tag{1}$$

where $\mu_x(t)$ and $\mu_y(t)$ are the force of mortality for single lives (x) and (y) at ages $x + t$ and $y + t$, respectively; see, for example, Jordan (1967, Chapter 9), Neill (1977, Chapter 7), and Bowers et al. (1997, Chapter 9). The calculation of the premiums for last survivor insurance and annuity contracts then in addition uses the following standard formulas

$$\overline{A}_{\overline{xy}} = \overline{A}_x + \overline{A}_y - \overline{A}_{xy} \quad \text{and} \quad \overline{a}_{\overline{xy}} = \overline{a}_x + \overline{a}_y - \overline{a}_{xy},$$

which relate the last-survivor functions to those for joint-lives and single lives. The deterministic approach of Jordan and Neill and the random approach of Bowers et al. assume that the two lives are statistically independent.

Several authors have studied the impact of dependence between insured lives; see, for example, Carrière and Chan (1986); Carrière (1994); Frees, Carrière, and Valdez (1996); Dhaene and Goovaerts (1997); Frees and Valdez (1998); Denuit and Cornet (1999); and Youn, Shermyakin, and Herman (2002). Of great interest and relevance to us is the paper by Youn, Shermyakin, and Herman (2002), which shows that we can also derive last survivor insurance and last survivor annuity formulas using more general future lifetime random variables.

The object of this paper is to show how we can use multiple-state models to define more precisely the assumptions required for the standard formulas to apply. We also indicate how we might price insurance and annuity contracts where these assumptions do not apply.

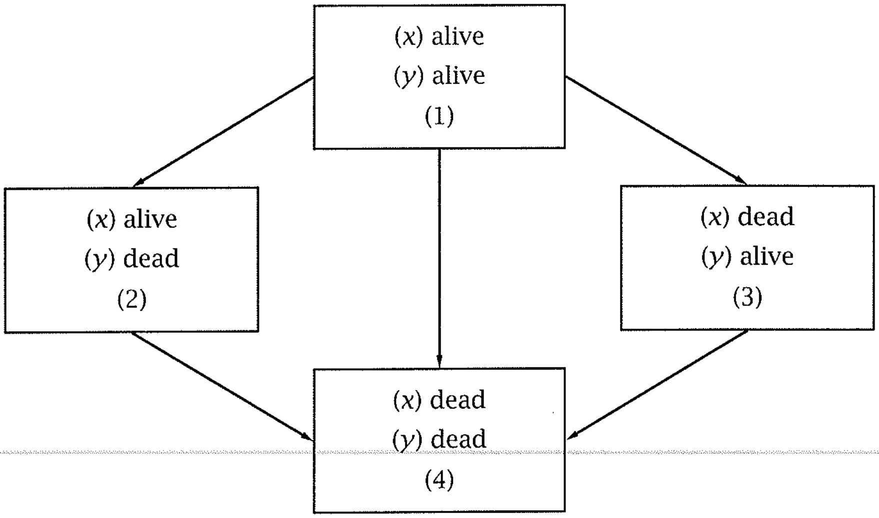

Figure 1: Generalized Mortality Model for Two Lives, (x) and (y)

2 A Model for Two Lives

A well known model for the forces of mortality depending on marital status was proposed by Norberg (1989) as follows:

State 1 = Both husband (x) and wife (y) are alive;
State 2 = Husband (x) is dead and wife (y) is alive;
State 3 = Husband (x) is alive and wife (y) is dead;
State 4 = Both husband (x) and wife (y) are dead.

Norberg regarded the future development of the marital status for the couple as a Markov process. We will generalize Norberg's model by assuming the transition intensities depend on the age at which the previous transition occurred, thus removing the Markov property. We also include transition directly from stage 1 to stage 4. Figure 1 illustrates our generalized model.

The following notations are used for $i, j = 1, 2, 3, 4$:

$p_{xy}^{(ii)}(t,s) =$ Probability that the couple (xy) stays in state i for at least t years (i.e., up to time $t+s$) given they entered state i at time s, for $s,t \geq 0$;

$p_{xy}^{(ij)}(t,s) =$ Probability that the couple (xy) is in state j at time $t+s$ given they entered state i at time s, for $s,t \geq 0$; and

$\mu_{xy}^{(ij)}(t,s)dt =$ Probability that the couple (xy) moves from state i to state j in $(t+s, t+s+dt)$ given they entered state i at time s and remained in state i up to time $t+s$, for $s,t \geq 0$ and infinitesimally small dt.

For convenience we define

$$\mu_{xy}^{(i)}(t,s) = \sum_{\substack{j=1 \\ j\neq i}}^{4} \mu_{xy}^{ij}(t,s),$$

which implies

$$\mu_{xy}^{(i)}(t,s) = \begin{cases} \mu_{xy}^{(12)}(t,s) + \mu_{xy}^{(13)}(t,s) + \mu_{xy}^{(14)}(t,s) & \text{if } i = 1; \\ \mu_{xy}^{(24)}(t,s) & \text{if } i = 2; \\ \mu_{xy}^{(34)}(t,s) & \text{if } i = 3. \end{cases}$$

Our model takes into account the empirical observations that, where there is some connection between the two lives, the mortality of one of the pair depends on whether the other is alive or dead and, if the latter, when death occurred.[1] One unusual feature of our model is the inclusion of transitions from state 1 directly to state 4. This allows for the possibility that the two lives die simultaneously in, for example, a car accident or a plane crash.

It is immediately seen that the basic functions needed for joint-life insurances and annuities are

$$\mu_{xy}(t) \equiv \mu_{xy}^{(1)}(t,0)$$
$${}_tp_{xy} \equiv p_{xy}^{(11)}(t,0) = e^{-\int_0^t \mu_{xy}^{(1)}(r,0)dr}.$$

Although in practice they do not often arise, we can also use our generalized model to price contingent insurance contracts where a payment

[1]See AES Course Notes, Subject D, Unit 8, Institute of Actuaries, London, 1994.

is made on the death of (x) if that occurs before the death of (y) and vice versa and reversionary annuities. Using standard actuarial notation, for example, we have

$$\overline{A}_{\overset{1}{xy}:\overline{n|}} = \int_0^n v^t p_{xy}^{(11)}(t,0)\mu_{xy}^{(13)}(t,0)\mathrm{d}t,$$

which is the net single premium for a contingent insurance that pays \$1 on the death of (x) if that occurs before the death of (y) and within n years.

In a similar manner, the net single premium for an n-year term insurance contract that pays \$1 if (x) and (y) die simultaneously, i.e., on a transition from state 1 to state 4, is given as:

$$\int_0^n v^t p_{xy}^{(11)}(t,0)\mu_{xy}^{(14)}(t,0)\mathrm{d}t$$

for $n \geq 0$. There is no standard actuarial notation for this, and, to the best of the authors' knowledge, no insurance company offers such a contract.

Given the absence of the Markov property, the traditional Chapman-Kolmogorov equations[2] cannot be used for the transition probabilities. As transitions to previous states are not allowed and there are only four states, however, our analysis can be simplified by assuming the first death occurs at time s and the second death at time $s+t$. Thus, the net single premium for a last survivor n-year term insurance is

$$\begin{aligned}\overline{A}_{\overset{1}{\overline{xy}}:\overline{n|}} &= \int_0^n v^s p_{xy}^{(11)}(s,0)\mu_{xy}^{(12)}(s,0)\int_0^{n-s} v^t p_{xy}^{(22)}(t,s)\mu_{xy}^{(24)}(t,s)\mathrm{d}t\,\mathrm{d}s\\ &+ \int_0^n v^s p_{xy}^{(11)}(s,0)\mu_{xy}^{(13)}(s,0)\int_0^{n-s} v^t p_{xy}^{(33)}(t,s)\mu_{xy}^{(34)}(t,s)\mathrm{d}t\,\mathrm{d}s\\ &+ \int_0^n v^s p_{xy}^{(11)}(s,0)\mu_{xy}^{(14)}(s,0)\mathrm{d}s,\end{aligned}$$

and the net single premium for a reversionary annuity is

$$\overline{a}_{x|y} = \int_0^\infty v^s p_{xy}^{(11)}(s,0)\mu_{xy}^{(13)}(s,0)\int_0^\infty \overline{a}_{\overline{t|}}\, p_{xy}^{(33)}(t,s)\mu_{xy}^{(34)}(t,s)\mathrm{d}t\,\mathrm{d}s.$$

[2]See, for example, Cox and Miller (1965, Chapter 4.1) or Taylor and Karlin (1994, Chapter 6.3) for more on Chapman-Kolmogorov equations.

3 Practical Simplifications

One problem with our generalized model is that, in practice, there will be a lack of adequate data to provide estimates of all the transition intensities required by this model. As a result, our model may be impractical to implement. We therefore need to introduce a set of simplifying assumptions that are intended to facilitate estimating these intensities. We will prove that one consequence of our assumptions is that the generalized model will yield results that are consistent with those produced by the independence assumption of the traditional or random future lifetime approaches. To this end we let $T(x)$ and $T(y)$ denote the random future lifetime of (x) and (y), respectively.

Assumption 1. *The events* $\{T(x) \in (t, t+\delta t)\}$ *and* $\{T(y) \in (t, t+\delta t)\}$ *are independent for all* $t \geq 0$ *and* δt *is infinitesimally small.*

Assumption 1 implies that

$$\mathbb{P}\left[T(x) \in (t, t+\delta t\} \bigcap T(y) \in (t, t+\delta t)\right] \\ = \mathbb{P}[T(x) \in (t, t+\delta t)]\,\mathbb{P}[T(y) \in (t, t+\delta t)].$$

Using the transition probabilities we have

$$p_{xy}^{(11)}(t,0)\mu_{xy}^{(14)}(t,0)\delta t = {}_tp_x\mu_x^{(*)}(t)\delta t \times {}_tp_y\mu_y^{(*)}(t)\delta t \qquad (2)$$

where ${}_tp_x$ and $\mu_x^{(*)}(t)$ are the marginal survival function and force of mortality of $T(x)$ and ${}_tp_y$ and $\mu_y^{(*)}(t)$ are the marginal survival function and force of mortality of $T(y)$. If we assume that x and y are less than the oldest possible age ω, then these marginal survival function and force of mortality will be positive and finite for $t < \min(\omega - x, \omega - y)$. Dividing both sides of equation (2) by δt and then let $\delta t \to 0$, we obtain

$$\mu_{xy}^{(14)}(t,0) \equiv 0 \quad \text{for} \quad t \geq 0,$$

i.e., transition from state 1 to state 4 is not possible.

For pricing joint-life insurances and annuities we can then use the simplification

$$\mu_{xy}^{(1)}(t,0) = \mu_x^{(*)}(t) + \mu_y^{(*)}(t) \qquad (3)$$

which implies

$$p_{xy}^{(11)}(t,0) \equiv {}_tp_x \times {}_tp_y. \qquad (4)$$

Thus, we get the traditional actuarial independence assumption of Jordan (1967), Neill (1977), and Bowers et al. (1997).

Assumption 2. *For all $t \geq 0$ and δt is infinitesimally small, the probability that (x) or (y) dies in time period $(t, t + \delta t)$ does not depend on whether the other is alive or dead at t, i.e.,*

$$\mathbb{P}\left[T(x) \in (t, t + \delta t) | \{T(y) \leq t \bigcup T(y) > t\}\right] = \mathbb{P}[T(x) \in (t, t + \delta t)]$$
$$\mathbb{P}\left[T(y) \in (t, t + \delta t) | \{T(x) \leq t \bigcup T(x) > t\}\right] = \mathbb{P}[T(y) \in (t, t + \delta t)].$$

On the basis of Assumption 2 we can state that, for $0 \leq s \leq t$,

$$\mu_{xy}^{(13)}(t, 0) \equiv \mu_{xy}^{(24)}(t - s, s) \equiv \mu_x^{(**)}(t) \tag{5}$$

and

$$\mu_{xy}^{(12)}(t + s, 0) \equiv \mu_{xy}^{(34)}(t, s) \equiv \mu_y^{(**)}(t). \tag{6}$$

Equations (5) and (6) constitute the independence assumption of Youn, Shermyakin, and Herman (2002).

If Assumptions 1 and 2 jointly apply, it is clear from the description of our generalized model that we must have

$$\mu_x^{(*)}(t) \equiv \mu_x^{(**)}(t) \quad \text{and} \quad \mu_y^{(*)}(t) \equiv \mu_y^{(**)}(t).$$

We assert that Assumptions 1 and 2 are both necessary if we are to have independence of the two lives. Assumption 1 clearly does not imply Assumption 2; see Youn, Shermyakin, and Herman (2002). Assuming that only Assumption 2 applies does not lead to any contradictions. Assumptions 1 and 2 are sufficient in themselves to derive the simple formulas used in practice. They require the estimation of transition intensities $\mu_x^{(**)}(t)$ and $\mu_y^{(**)}(t)$, which relate to joint lives, i.e., lives that have taken out a contract jointly.

In situations where there are not adequate experience data on joint lives, it is usual to make use also of the following assumption:

Assumption 3. *The mortality of each of the individual lives (x) and (y) in the pair (x, y) is identical to that of the single lives (x) and (y), respectively.*

Assumptions 3 is a different type of assumption from Assumptions 1 and 2, which are concerned with the relationship between the mortality of the two joint lives. Assumption 3 simply equates the numerical values of the transition intensities for each of the joint lives to those for individual lives, where we give the phrase "individual lives" the meaning commonly applied to it in life insurance mortality investigations. This allows us to replace the joint life intensities by those relating to individual lives, i.e.,

$$\mu_x^{(**)}(t) \equiv \mu_x(t) \quad \text{and} \quad \mu_y^{(**)}(t) \equiv \mu_y(t). \tag{7}$$

In practice reliable estimates of the transition intensities for individual lives are almost always available.

Applying all three assumptions[3] to our generalized model we obtain Norberg's (1989) Markov model with transition intensities depending only on current age. The Chapman-Kolmogorov forward equations are

$$\begin{aligned}
\frac{\partial p_{xy}^{(11)}(t,0)}{\partial t} &= -\left(\mu_x(t) + \mu_y(t)\right) p_{xy}^{(11)}(t,0) \\
\frac{\partial p_{xy}^{(12)}(t,0)}{\partial t} &= p_{xy}^{(11)}(t,0)\mu_y(t) - p_{xy}^{(11)}(t,0)\mu_x(t) \\
\frac{\partial p_{xy}^{(13)}(t,0)}{\partial t} &= p_{xy}^{(11)}(t,0)\mu_x(t) - p_{xy}^{(13)}(t,0)\mu_y(t)
\end{aligned}$$

which yield the solutions

$$\begin{aligned}
p_{xy}^{(11)}(t,0) &= e^{-\int_0^t (\mu_x(s)+\mu_y(s))ds} = {}_t p_{xy} \\
p_{xy}^{(12)}(t,0) &= e^{-\int_0^t \mu_x(s)ds}\left(1 - e^{-\int_0^t \mu_y(s)ds}\right) = {}_t p_x {}_t q_y \\
p_{xy}^{(13)}(t,0) &= e^{-\int_0^t \mu_y(s)ds}\left(1 - e^{-\int_0^t \mu_x(s)ds}\right) = {}_t p_y {}_t q_x.
\end{aligned}$$

It follows that the net single premium of the last survivor annuity is:

$$\begin{aligned}
\overline{a}_{\overline{xy}} &= \int_0^\infty v^t \left[p_{xy}^{(11)}(t,0) + p_{xy}^{(12)}(t,0) + p_{xy}^{(13)}(t,0)\right] dt \\
&= \int_0^\infty v^t \left[{}_t p_x + {}_t p_y - {}_t p_{xy}\right] dt \\
&= \overline{a}_x + \overline{a}_y - \overline{a}_{xy}.
\end{aligned}$$

[3]If we do not make use of Assumption 3, the derivation is the same, but with the joint life intensities replacing those relating to individual lives.

4 Closing Comments

We have indicated how we can price life insurances involving two lives using a generalized multi-state model. By introducing a set of clearly defined assumptions we have shown that using our model we can also derive the standard formulas traditionally used for pricing joint-life and last-survivor contracts. These assumptions are unrealistic, however. Thus, if they are used in practice, care must be taken in deciding whether any premiums calculated using these formulas are adequate.

References

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. *Actuarial Mathematics,* Second Edition. Schaumburg, Ill.: Society of Actuaries, 1997.

Carrière, J.F. "Dependent Decrement Theory." *Transactions of the Society of Actuaries* 46 (1994): 45–74.

Carrière, J.F. and Chan, L.K. "The Bounds of Bivariate Distributions That Limit the Value of Last-Survivor Annuities." *Transactions of the Society of Actuaries* 38 (1986): 51–74.

Cox, D.R. and Miller, H.D. *The Theory of Stochastic Processes.* London: Chapman and Hall, 1965.

Denuit, M. and Cornet, A. "Multilife Premium Calculation with Dependent Future Lifetimes." *Journal of Actuarial Practice* 7 (1999): 147–180.

Dhaene, J. and Goovaerts, M.J. "On the Dependency of Risks in the Individual Life Model." *Insurance: Mathematics and Economics* 19 (1997): 243-253.

Frees, E.W., Carrière, J.F. and Valdez, E. "Annuity Valuation with Dependent Mortality." *The Journal of Risk and Insurance* 63 (1996): 229-261.

Frees, E.W. and Valdez, E.A. "Understanding Relationships Using Copulae." *North American Actuarial Journal* 2 (1998): 1–25.

Jordan, C.W. *Life Contingencies,* Second Edition. Chicago, Ill.: Society of Actuaries, 1967.

Neill, A. *Life Contingencies.* London: Heinemann, 1977.

Norberg, R. "Actuarial Analysis of Dependent Lives." *Bulletin de l'Association Suisse des Actuaires* (1989): 243–254.

Taylor, H.M. and Karlin, S. *An Introduction to Stochastic Modeling.* San Diego, CA: Academic Press, 1994.

Youn, H., Shemyakin, A. and Herman, E. "A Re-examination of the Joint Mortality Functions." *North American Actuarial Journal* 6, no. 1 (2002): 166–170.

Journal of Actuarial Practice Vol. 12, 2005

On the Pricing of Top and Drop Excess of Loss Covers

Jean-François Walhin* and Michel Denuit†

Abstract

A top and drop cover is a treaty that can be found on the retrocession market. It offers capacity that can be used either to protect a top layer or a working layer. The former is called a "top" and the latter is called a "drop." Using the traditional collective risk model, we demonstrate the use of a multivariate version of Panjer's algorithm to price this cover. We also compare the premium obtained within the exact model with the premiums obtained either with the Fréchet bounds or with the wrong assumption of independence.

Key words and phrases: *multivariate Panjer's algorithm, excess of loss pricing, dependence, correlation order, stop-loss order, comonotonic risks, Fréchet bounds, supermodular order*

1 Introduction

The traditional collective risk model assumes that an insurance portfolio produces a random number of N positive claims in a year. The claim sizes, $X_1, X_2, \ldots, X_N$, are assumed to be independent and identically distributed positive random variables. The annual aggregate claims S is then given by

*Jean-François Walhin, Ph.D., is R&D manager at Secura and visiting professor at Université Catholique de Louvain. He is a civil engineer and an actuary (fellow of ARAB-KVBA). He obtained his Ph.D. in sciences from the Université Catholique de Louvain.

Dr. Walhin's address is: Secura; Avenue des Nerviens, 9-31; B-1040 Bruxelles; BELGIUM. E-mail: *jfw@secura-re.com*

†Michel Denuit, Ph.D., is professor of statistics and actuarial science at Université Catholique de Louvain. He is a mathematician and an actuary (fellow of ARAB-KVBA) and holds a Ph.D. in statistics from the Université Libre de Bruxelles.

Dr. Denuit's address is: Institut des Sciences Actuarielles, Université Catholique de Louvain, Rue des Wallons, 6; B-1348 Louvain-la-Neuve; BELGIUM. E-mail: *denuit@stat.ucl.ac.be*

$$S = X_1 + \cdots + X_N.$$

When N belongs to the $(a, b, 0)$ class of counting distributions,[1] i.e., when the probabilities associated with N satisfy

$$\frac{\mathbb{P}[N = n]}{\mathbb{P}[N = n-1]} = a + \frac{b}{n} \quad \text{for} \quad n \geq 1,$$

and the X_is are discrete, then it is easy to obtain the distribution of S using the recursive algorithm due to Panjer (1981).

The assumption of mutual independence of claim sizes in the collective risk model, which makes sense in many situations, offers the advantage of mathematical simplicity. There are situations, however, where the independence assumption needs to be relaxed. Some authors have addressed the problem by imposing upper and lower bounds on the results when some form of stochastic dependence is observed (see Dhaene et al., 2001), while others have attempted to model the dependence (e.g., Frees and Valdez, 1998).

This paper extends the collective risk model to include dependent claims. We distinguish between two models:

Model 1, which considers independent occurrences of the random couple (X, Y), i.e., (X_1, Y_1), (X_2, Y_2),..., (X_N, Y_N), with the X_is and Y_is all independent of the counting random variable N. The bivariate aggregate claim is then defined as

$$(S, T) = (\sum_{i=1}^{N} X_i, \sum_{i=1}^{N} Y_i).$$

The dependence between S and T originates from them sharing the same claim number N as well as from possible correlations between the components of the (X_i, Y_i)s. Sundt (1999) proposed a multivariate extension of Panjer's algorithm, allowing for practical calculations within this multivariate collective risk model.

Model 2, which considers the N independent claim sizes $X_1, X_2, \ldots, X_N$, and M independent claim sizes $Y_1, Y_2, \ldots, Y_M$. We assume a mutual independence between the X_is and the Y_js, as well as with the counting variables N and M. However, N and M may be dependent. The bivariate aggregate claims is then defined as

[1] The $(a, b, 0)$ class of counting distributions contains Poisson ($a = 0$), negative binomial ($a > 0$), and binomial ($a < 0$) distributions.

$$(S,T) = (\sum_{i=1}^{N} X_i, \sum_{i=1}^{M} Y_i).$$

Note that the dependence between S and T now originates only from the dependence between N and M because the claim sizes are mutually independent. Some types of dependence between N and M may be modeled using the trivariate reduction method or by mixing the bivariate Poisson distribution. Walhin and Paris (2000b) and Walhin and Paris (2001) provide some sophisticated bivariate counting models allowing for these calculations.

In another departure from the collective risk model, we distinguish between two types of claims (the extension to more types of claims is trivial): (i) small claims and (ii) large claims. We assume the behavior of small claims may differ significantly from the behavior of large claims. In our models we assume the X_is and the Y_is represent the size of the large and small claims, respectively, while N_l and N_s are the annual number of large and small claims, respectively. The common cdf of the X_is is a limited Pareto distributed with nonnegative parameters A_l, B_l, and α_l, while that of the Y_is is a limited Pareto distribution with nonnegative parameters A_s, B_s, and α_s. Because we assume only two types of claims (large and small), then $B_s = A_l$. The numbers of claims N_l and N_s are assumed to be Poisson distributed with mean λ_l and λ_s, respectively.

A random variable X has a limited Pareto distribution with parameters A, B, and α (which, for notational convenience, can be written as $X \sim \mathrm{Par}(A, B, \alpha)$) if its cdf, F_X, can be written as:

$$F_X(x) = \mathbb{P}[X \leq x] = \begin{cases} 0 & \text{if } x < A \\ \dfrac{A^{-\alpha} - x^{-\alpha}}{A^{-\alpha} - B^{-\alpha}} & \text{if } A \leq x < B \\ 1 & \text{if } x \geq B. \end{cases} \qquad (1)$$

Throughout the rest of this paper, we assume mutual independence between the random variables N_l, N_s, X_i, and Y_i. We consider also the following values for the Poisson and limited Pareto distribution parameters as shown in Table 1.

In this paper, we assume Model 1 holds, as this will allow us to derive specific solutions. We propose a new application of the multivariate extension of the Panjer's algorithm to price the so-called *top and drop* cover. This reinsurance treaty, used primarily for retrocession, includes a top layer and a working layer. There is an obvious stochastic dependence in the model as large claims (affecting the top layer)

Table 1
Parameters

	λ	α	A	B
Large Claims	0.3	0.9	400	1000
Small Claims	2.5	1.4	20	400

necessarily hit the working layer. To use the multivariate extension of the Panjer's algorithm, we discretize the claims size distributions, thus making the derived solutions approximations only. We will compare these solutions to those based on the incorrectly assumed independence hypothesis between S and T, as well as to some upper and lower bounds. These comparisons will be done with theoretical or empirical results.

The rest of the paper is organized as follows. Section 2 provides a brief review of excess of loss reinsurance, and describes two types of top and drop covers within a relatively general collective risk model. Section 3 recalls the multivariate Panjer's algorithm. Section 4 reviews some necessary results on stochastic orderings. Section 5 provides the numerical results and compares them with the case where independence would be incorrectly assumed and with the corresponding Fréchet bounds.

2 Top and Drop Covers

Excess of loss reinsurance is a means to share risks between the ceding insurer (the cedent) and and the reinsurer. The cedent always remains liable for the part of the claim below a given attachment point or deductible P, while the reinsurer offers some capacity between P and the limit $P + L$. So we can write the liability of the excess of loss reinsurer for each claim X_i as

$$R_i = \min(L, \max(0, X_i - P)).$$

In the collective risk model, the aggregate liability of the reinsurer is

$$S_R = R_1 + \cdots + R_N.$$

The reinsurance capacity L may be subject to k reinstatements. If $k = 0$, it means that there is no reinstatement and the reinsurer's liability for the whole period (usually one year) is limited to L, regardless of the number of occurrences. Otherwise, the aggregate capacity is $(k + 1)L$. Keep in mind the reinsurer's liability in any occurrence is limited to L, i.e., the aggregate liability of the reinsurer is $\min((k + 1)L, S_R)$. In practice, reinstatements can be paid or free. In the present paper we will only discuss the situation where the reinstatements are free.

An annual aggregate deductible (AAD) will reduce the aggregate claims of the reinsurer. A higher AAD should reduce the reinsurance premium. For the general case where there are k reinstatements and an AAD, the annual liability of the reinsurer is $\min((k + 1)L, \max(0, S_R - \text{AAD}))$.

It is interesting to see how reinsurance can introduce dependencies in some treaties: for example, the ECOMOR-type treaties involving order statistics (see Thépaut, 1950) or the exotic excess of loss treaty described in Walhin (2002) where some layers inure to the benefit of other layers. Walhin (2002) used a multivariate version of Panjer's algorithm to price that treaty. In Walhin and Paris (2000a) this multivariate version of the Panjer's algorithm is used to study the retained risk of the cedent when it buys excess of loss reinsurance with paid reinstatements. We now describe two treaties.

Treaty 1: Recently Secura has been given the opportunity to examine the following excess of loss cover: in reinsurers' jargon (see below for a translation into formulas), the characteristics of this treaty were

- 200 in excess of 800 (written as 200 **XS** 800)

  AND / OR

- 200 **XS** 200 in the aggregate for each loss exceeding 20 (losses to be aggregated from ground up but with a maximum of 100 each and every loss occurrence).
- No reinstatement granted, i.e., the maximal annual amount to be paid by the reinsurer is 200.

The aim of this treaty is to cover a top layer (200 **XS** 800) that has a very low probability of being hit and, simultaneously, a potential high frequency of small claims.

In mathematical terms, the characteristics of this reinsurance cover can be summarized as follows:

$X_i^{\mathrm{RT}} = \min(200, \max(0, X_i - 800))$, which is the reinsurer's liability for the top part of large claims;

$$X_i^{\mathrm{RD}} = \min(100, X_i I_{X_i \geq 20}), \text{ which is the reinsurer's liability for the drop part of large claims;}$$

$$Y_i^{\mathrm{RD}} = \min(100, Y_i I_{Y_i \geq 20}), \text{ which is the reinsurer's liability for the drop part of small claims;}$$

$$S = X_1^{\mathrm{RT}} + \cdots + X_{N_l}^{\mathrm{RT}}, \text{ which is the reinsurer's aggregate liability for the top part of large claims;}$$

$$T = X_1^{\mathrm{RD}} + \cdots + X_{N_l}^{\mathrm{RD}}, \text{ which is the reinsurer's aggregate liability for the drop part of large claims;}$$

$$U = Y_1^{\mathrm{RD}} + \cdots + Y_{N_s}^{\mathrm{RD}}, \text{ which is the reinsurer's aggregate liability for the drop part of small claims; and}$$

$$\text{Cover} = \min(200, S + \max(0, T + U - 200))$$

where I_A is the indicator function, i.e., $I_A = 1$ if A is true, $I_A = 0$ otherwise. Note that the choice made for B_s implies that the small claims Y_i do not trigger the top cover.

Treaty 2: Another example of top and drop cover is described below:

- 200 **XS** 800

  AND / OR

- 200 **XS** 200 with a global annual aggregate deductible of 400 and unlimited free reinstatements.
- The aim of the treaty is clearly to cover an extra reinstatement on the low layer (which typically would be protected by a classical 200 **XS** 200 with one reinstatement) and/or a top layer (200 **XS** 800).

The reinsurance cover can be described as follows:

$$X_i^{\mathrm{RT}} = \min(200, \max(0, X_i - 800)), \text{ which is the reinsurer's liability for the top part of large claims;}$$

$$X_i^{\mathrm{RD}} = \min(200, \max(0, X_i - 200)), \text{ which is the reinsurer's liability for the drop part of large claims;}$$

$$Y_i^{\mathrm{RD}} = \min(200, \max(0, Y_i - 200)), \text{ which is the reinsurer's liability for the drop part of small claims; and}$$

$$\text{Cover} = \max(0, S + T + U - 400)$$

with S, T, and U described as in Treaty 1.

As a consequence of our choice of distributions for small and large claims, we can simplify the model in two ways:

1)

$$T = \begin{cases} 100N_l & \text{Treaty 1} \\ 200N_l & \text{Treaty 2} \end{cases}$$

which leads to

$$\text{Cover} = \begin{cases} \min(200, S + \max(0, 100N_l + U - 200)) & \text{Treaty 1} \\ \max(0, 200N_l + U - 400) & \text{Treaty 2.} \end{cases}$$

2) Using two independent compound Poisson distributions with limited Pareto distributions for the small and large claims is equivalent to a single compound Poisson with i.i.d. claim sizes that are mixtures of limited Pareto distributions. The new number of claims random variable is $N = N_l + N_s$, which is Poisson with mean $\lambda_l + \lambda_s$, and the new claim sizes are Z_i, which is a mixture of limited Pareto distributions with cdf $F_Z(x)$ given by

$$F_Z(x) = \mathbb{P}[Z_i \le x] = \begin{cases} 0 & \text{if } x < A_s \\ \frac{\lambda_s}{\lambda_s+\lambda_l} \frac{A_s^{-\alpha_s} - x^{-\alpha_s}}{A_s^{-\alpha_s} - A_l^{-\alpha_s}} & \text{if } A_s \le x < B_s = A_l \\ \frac{\lambda_l}{\lambda_s+\lambda_l} \frac{A_l^{-\alpha_l} - x^{-\alpha_l}}{A_l^{-\alpha_l} - B_l^{-\alpha_l}} & \text{if } B_s = A_l \le x < B_l \\ 1 & \text{if } x \ge B_l. \end{cases}$$

We obtain for Treaty 1:

$$\begin{aligned} Z_i^{\mathrm{RT}} &= \min(200, \max(0, Z_i - 800)) \\ Z_i^{\mathrm{RD}} &= \min(100, Z_i I_{Z_i \ge 20}) \\ S &= Z_1^{\mathrm{RT}} + \cdots + Z_N^{\mathrm{RT}} \\ T &= Z_1^{\mathrm{RD}} + \cdots + Z_N^{\mathrm{RD}} \\ \text{Cover} &= \min(200, S + \max(0, T - 200)), \end{aligned}$$

and for Treaty 2:

$$\begin{aligned}
Z_i^{\mathrm{RT}} &= \min(200, \max(0, Z_i - 800))\\
Z_i^{\mathrm{RD}} &= \min(200, \max(0, Z_i - 200))\\
S &= Z_1^{\mathrm{RT}} + \cdots + Z_N^{\mathrm{RT}}\\
T &= Z_1^{\mathrm{RD}} + \cdots + Z_N^{\mathrm{RD}}\\
\text{Cover} &= \max(0, S + T - 400).
\end{aligned}$$

Though Model 1 yields treaties that can be simplified as above, we will not use these simplifications; rather we use the general formulation in the rest of this paper.

In both treaties, S and T are correlated. We have

- S and T are random sums of non-negative random variables with identical number N of terms.
- The summands, X_i^{RT} and X_i^{RD} are themselves correlated.

This means that even the computation of the pure reinsurance premium $\mathbb{E}[\text{Cover}]$ requires the joint distribution of (S, T). As explained in the introduction, it is possible to obtain this joint distribution by using the multivariate version of the Panjer's algorithm as is explained below.

3 The Multivariate Version of Panjer's Algorithm

Panjer's type algorithms require lattice distributions. Therefore we must first discretize claim amounts. The local one moment matching method (see Gerber, 1982) is a good choice in the sense that it conserves the first moment and is stop-loss conservative, i.e., for any retention, the stop-loss premium calculated with the discretized distribution will be higher than the stop-loss premium calculated with the original distribution. Furthermore, in the case of the limited Pareto distribution ($X \sim \mathrm{Par}(A, B, \alpha)$), it is not difficult to obtain a closed-form of the corresponding lattice distribution. Let us choose a span h and a positive integer m such that $mh = B - A$. It is easy to demonstrate that the probabilities of the lattice version of X, denoted as X_{dis}, with probability function are given by:

$$\begin{aligned}
&f_{X_{dis}}(A + jh)\\
&\qquad = \frac{2(A + jh)^{1-\alpha} - (A + (j-1)h)^{1-\alpha} - (A + (j+1)h)^{1-\alpha}}{h(1-\alpha)(A^{-\alpha} - B^{-\alpha})}
\end{aligned}$$

for $j = 1, \ldots, m-1$, with

$$f_{X_{dis}}(A) = 1 - \frac{\frac{(A+h)^{1-\alpha}}{1-\alpha} - \frac{A^{1-\alpha}}{1-\alpha} - B^{-\alpha}h}{h(A^{-\alpha} - B^{-\alpha})},$$

$$\text{and} \quad f_{X_{dis}}(B) = 1 - f_{X_{dis}}(A) - f_{X_{dis}}(A+h) - \cdots - f_{X_{dis}}(B-h).$$

Now let us turn to the joint distribution of the bivariate random vector

$$(S,T) = (\sum_{i=1}^{N} X_i^{\mathrm{RT}}, \sum_{i=1}^{N} X_i^{\mathrm{RD}}),$$

where $(X_i^{\mathrm{RT}}, X_i^{\mathrm{RD}})$ are independent copies of the lattice random couple $(X^{\mathrm{RT}}, X^{\mathrm{RD}})$. As N is Poisson distributed, Sundt's (1999) multivariate version of the Panjer's algorithm yields

$$f_{S,T}(0,0) = \Psi_N(f_{X^{\mathrm{RT}},X^{\mathrm{RD}}}(0,0)),$$

$$f_{S,T}(s,t) = \sum_{x,y}^{s,t} \frac{\lambda x}{s} f_{S,T}(s-x, t-y) f_{X^{\mathrm{RT}},Y^{\mathrm{RD}}}(x,y) \quad , \quad s \geq 1,$$

$$f_{S,T}(s,t) = \sum_{x,y}^{s,t} \frac{\lambda y}{t} f_{S,T}(s-x, t-y) f_{X^{\mathrm{RT}},X^{\mathrm{RD}}}(x,y) \quad , \quad t \geq 1,$$

where we use the notation

$$\sum_{x,y}^{s,t} g(x,y) = \sum_{x=0}^{s} \sum_{y=0}^{t} g(x,y) - g(0,0),$$

for any function g and $\Psi_N(u) = \mathbb{E}[u^N] = \exp(\lambda(u-1))$.

4 Some Elements of Stochastic Orderings

In this section, we extensively refer to the seminal paper of Dhaene and Goovaerts (1996) on dependency of risks applied in actuarial science. Some results appear more generally in probability theory, and we will extensively refer to the textbook of Müller and Stoyan (2002).

Stop-Loss Order Stop-loss order allows the actuary to order the risks according to their stop-loss premiums.

Definition 1. *A risk X is said to be smaller in the stop-loss order than a risk Y (written $X \leq_{sl} Y$) whenever one of the following equivalent statements holds true:*

1. $\mathbb{E}[\max(0, X - d)] \leq \mathbb{E}[\max(0, Y - d)]$ *for any nonnegative deductible d; or*
2. $\mathbb{E}[u(X)] \leq \mathbb{E}[v(Y)]$ *for all increasing convex functions u and v, provided these expectations exist.*

The ranking $X \leq_{sl} Y$ implies the stop-loss premiums for X are uniformly smaller than those for Y.

PH-Transform Premium Principle We are interested in calculating premiums with the PH-transform premium principle, introduced by Wang (1996). According to this premium principle, the amount $\Pi_\rho(X)$ charged to cover the risk X is given by

$$\Pi_\rho(X) = \int_0^\infty (1 - F_X(x))^\rho dx,$$

where $0 \leq \rho \leq 1$. In particular when $\rho = 1$, the PH premium reduces to the pure premium. Wang (1996) proved that

$$X \leq_{sl} Y \Rightarrow \Pi_\rho(X) \leq \Pi_\rho(Y) \tag{2}$$

which shows that the PH principle is in accordance with the stop-loss order.

Fréchet Space The concept of Fréchet space emerges when dealing with dependence; it offers the appropriate framework to deal with correlated random variables.

Definition 2. *The bivariate Fréchet space $\mathfrak{R}(F_1, F_2)$ is the class of all bivariate distributions with given marginal cdfs F_1 and F_2.*

For the purpose of this paper, we will consider $\mathfrak{R}(F_1, F_2)$ as a set of random couples.

Correlation Order The correlation order offers a powerful tool to compare the elements of a given Fréchet space.

Definition 3. *If (X_1, X_2) and (Y_1, Y_2) are elements of $\mathfrak{R}(F_1, F_2)$, we say that (X_1, X_2) is less correlated than (Y_1, Y_2), written $(X_1, X_2) \leq_c (Y_1, Y_2)$, if*

$$\mathbb{C}ov(f(X_1), g(X_2)) \leq \mathbb{C}ov(f(Y_1), g(Y_2)),$$

for all non-decreasing functions f and g for which the covariances exist.

The intuitive meaning of a ranking $(X_1, X_2) \leq_c (Y_1, Y_2)$ is that (X_1, X_2) is "less positively dependent" than (Y_1, Y_2).

The correlation order enjoys a number of convenient mathematical properties, some of which are reviewed below. These properties of correlation order are found in Müller and Stoyan (2002):

P1 Let (X_1, X_2) and (Y_1, Y_2) be elements of $\mathfrak{R}(F_1, F_2)$, then the following statements are equivalent:

(i) $(X_1, X_2) \leq_c (Y_1, Y_2)$, and

(ii) $F_{X_1,X_2}(x_1, x_2) \leq F_{Y_1,Y_2}(x_1, x_2) \quad , \quad \forall x_1, x_2 \geq 0.$

P2 Let (U_1, U_2) and (V_1, V_2) be elements of $\mathfrak{R}(F_1, F_2)$, and let (R_1, R_2) be a random vector independent of both (U_1, U_2) and (V_1, V_2). It follows that

$$(U_1, U_2) \leq_c (V_1, V_2) \Rightarrow (U_1 + R_1, U_2 + R_2) \leq_c (V_1 + R_1, V_2 + R_2).$$

P3 Suppose (W, X) and (Y, Z) are elements of $\mathfrak{R}(F_1, F_2)$. Let (W_i, X_i) and (Y_i, Z_i) be independent copies of (W, X) and (Y, Z), respectively, such that $(W, X) \leq_c (Y, Z)$, and let N be a nonnegative counting random variable independent of (W, X) and (Y, Z). It follows that $(S_W, S_X) \leq_c (S_Y, S_Z)$ where $(S_W, S_X) = (\sum_{i=1}^N W_i, \sum_{i=1}^N X_i)$ and $(S_Y, S_Z) = (\sum_{i=1}^N Y_i, \sum_{i=1}^N Z_i)$.

P4 $(W, X) \leq_c (Y, Z)$ implies $(f(W), g(X)) \leq_c (f(Y), g(Z))$ for all increasing functions f and g.

P5 $(W, X) \leq_c (Y, Z)$ implies $W + X \leq_{sl} Y + Z$, i.e., correlation order implies stop-loss order of the sum of the elements.

Using properties P4 and P5, we immediately obtain the following result:

Result 1. *Let (X_1, X_2) and (Y_1, Y_2) be two elements of $\mathfrak{R}(F_1, F_2)$. Then $(X_1, X_2) \leq_c (Y_1, Y_2)$ implies*

$$\max(0, X_1 - a) + \max(0, X_2 - b) \leq_{sl} \max(0, Y_1 - a) + \max(0, Y_2 - b)$$

for all $a, b \geq 0$.

Positive Quadrant Dependence An important concept used in later in this paper is the concept of positive quadrant dependence.

Definition 4. *Let $(X_1, X_2) \in \mathfrak{R}(F_1, F_2)$, and let $(X_1^{\perp}, X_2^{\perp})$ be the independent version of (X_1, X_2), i.e., $(X_1^{\perp}, X_2^{\perp}) \in \mathfrak{R}(F_1, F_2)$ and $X_1^{\perp}$ is independent of $X_2^{\perp}$. Then (X_1, X_2) is said to be positively quadrant dependent if one of the following equivalent statements holds:*

(i) $F_{X_1}(x_1)F_{X_2}(x_2) \leq F_{X_1,X_2}(x_1, x_2), \quad x_1, x_2 \geq 0.$

(ii) $(X_1^{\perp}, X_2^{\perp}) \leq_c (X_1, X_2).$

(iii) $\mathbb{C}ov(f(X_1), g(X_2)) \geq 0$ *for all non-decreasing functions f and g.*

See, e.g., Dhaene and Goovaerts (1996) for a proof of these equivalences.

The following result will be useful for the applications in reinsurance.

Result 2. *Let X_i and Y_i be independent copies of the non-negative random variables X and Y. Let us assume that X, Y, and N are mutually independent and define*

$$S = X_1 + \cdots + X_N \text{ and } T = Y_1 + \cdots + Y_N.$$

Then (S, T) is positively quadrant dependent.

Proof: f and g are non-decreasing functions. By the decomposition formula of the covariance, we have

$$\begin{aligned}\mathbb{C}ov(f(S), g(T)) = {} & \mathbb{E}(\mathbb{C}ov(f(S), g(T)|N)) \\ & + \mathbb{C}ov(\mathbb{E}(f(S)|N), \mathbb{E}(g(T)|N)).\end{aligned}$$

The first term of the right part of the equality vanishes because the covariance between independent random variables is 0. For the second term, it is clear that the expectations are increasing functions of N (because the summands in S and T are assumed to be positive) and therefore the second term can be rewritten as $\mathbb{C}ov(u(N), v(N))$, where u and v are non-decreasing functions. This covariance is clearly non-negative, which closes the proof.

Comonotonicity The concept of comonotonicity generalizes perfect correlation. Comonotonic random variables are functionally (and not necessarily linearly) dependent. For a reference in actuarial science, see e.g., Wang and Dhaene (1998).

Definition 5. *Two risks X and Y are said to be comonotonic if*

(i) Their joint cdf satisfies $F_{X,Y}(x, y) = \min(F_X(x), F_Y(y))$ for any $x, y \geq 0$, or, equivalently,

(ii) There exists a random variable Z and non-decreasing functions u and v on $\mathbb{R}$ such that (X, Y) is distributed as $(u(Z), v(Z))$.

By construction, the couples (X_i^{RT}, X_i^{RD}) are comonotonic.

Fréchet's Theorem An interesting result that is related to the concept of comonotonicity is the following theorem, due to Fréchet (1951) and Hoeffding (1940). It gives the extremal elements of any Fréchet space with respect to $\leq_c$.

Theorem 1 (Fréchet). *Let $(X_1, X_2) \in \mathfrak{R}(F_1, F_2)$, then*

$$(F_1^{-1}(U), F_2^{-1}(1-U)) \leq_c (X_1, X_2) \leq_c (F_1^{-1}(U), F_2^{-1}(U))$$

with U uniformly distributed over (0,1), or, equivalently, in terms of distribution functions, the inequalities

$$\max[F_1(x_1)+F_2(x_2)-1; 0] \leq F_{X_1,X_2}(x_1, x_2) \leq \min[F_1(x_1); F_2(x_2)]$$

hold for any $x_1, x_2 \in \mathbb{R}$.

5 Numerical Results

As mentioned in Section 3, all continuous random variables are discretized using the local one moment matching method with a discretization step $h = 10$. Thus, all random variables in this section are the discrete version of the original random variable.

5.1 Treaty 1

Table 2 shows some interesting characteristics of the claims. Note that the Pearson's correlation coefficient between S^{RT} and T^{RD} is estimated at 0.35. The pure premium for this cover is

$$\begin{aligned}\mathbb{E}[\text{Cover}] &= \sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\sum_{u=0}^{\infty} f_{S,T}(s,t) f_U(u) \min(200, s + \max(0, t + u - 200)) \\ &= 20.519.\end{aligned}$$

Table 2
Means and Variances for Treaty 1

	X^{RD}	X^{RT}	Y^{RD}	S^{RT}	T^{RD}	U^{RD}
Mean	100	16.14	42.87	30	4.84	107.17
Variance	0	1817	631.72	3000	623.4	6173.89

We easily can obtain upper and lower bounds. Let

$$F_{S^{min},T^{min}}(x_1,x_2) = \max[F_S(x_1) + F_T(x_2) - 1; 0], \quad \text{and}$$
$$F_{S^{max},T^{max}}(x_1,x_2) = \min[F_S(x_1), F_T(x_2)].$$

Using Theorem 1 in connection with P1 we have

$$(S^{min}, T^{min}) \leq_c (S,T) \leq_c (S^{max}, T^{max}).$$

Using P2, we have

$$(S^{min}, T^{min} + U) \leq_c (S, T+U) \leq_c (S^{max}, T^{max} + U).$$

Using Result 1, we have

$$\begin{aligned}\mathbb{E}[\max(0; S^{min} &+ \max(0; T^{min} + U - 200) - 200)] \\ &\leq \mathbb{E}[\max(0; S + \max(0; T + U - 200) - 200)] \\ &\leq \mathbb{E}[\max(0; S^{max} + \max(0; T^{max} + U - 200) - 200)],\end{aligned}$$

which is equivalent to

$$\begin{aligned}\mathbb{E}[\min(200; S^{max} &+ \max(0; T^{max} + U - 200))] \\ &\leq \mathbb{E}[\min(200; S + \max(0; T + U - 200))] \\ &\leq \mathbb{E}[\min(200; S^{min} + \max(0; T^{min} + U - 200))].\end{aligned}$$

The numerical bounds are $19.469 \leq 20.519 \leq 21.279$.

It is possible to improve the upper bound. Let $(X^{\mathrm{RT},\perp}, X^{\mathrm{RD},\perp})$ be the independent version of $(X^{\mathrm{RT}}, X^{\mathrm{RD}})$. We define

$$S^{\perp} = X_1^{\mathrm{RT},\perp} + \cdots + X_N^{\mathrm{RT},\perp},$$
$$T^{\perp} = X_1^{\mathrm{RD},\perp} + \cdots + X_N^{\mathrm{RD},\perp}.$$

Clearly X^{RT} and X^{RD} are comonotonic random variables. We then have

$$\begin{aligned} F_{X^{\mathrm{RT},\perp},X^{\mathrm{RD},\perp}}(x_1,x_2) &\le \min(F_{X^{\mathrm{RT}}}(x_1),F_{X^{\mathrm{RD}}}(x_2)) \\ &= F_{X^{\mathrm{RT}},X^{\mathrm{RD}}}(x_1,x_2), \quad \forall x_1,x_2 \ge 0. \end{aligned}$$

Therefore, $(X^{\mathrm{RT},\perp},X^{\mathrm{RD},\perp}) \le_c (X^{\mathrm{RT}},X^{\mathrm{RD}})$. Using P3 we have $(S^\perp,T^\perp) \le_c (S,T)$. $S^\perp$ and $T^\perp$ are dependent, however, because they involve the same number of summands. Therefore, let us define the independent versions of $(S^\perp,T^\perp)$ as $(S^{\perp\perp},T^{\perp\perp})$. Using Result 2 we have $(S^{\perp\perp},T^{\perp\perp}) \le_c (S^\perp,T^\perp)$. By transitivity, we then obtain $(S^{\perp\perp},T^{\perp\perp}) \le_c (S,T)$. Using P2 we have $(S^{\perp\perp},T^{\perp\perp}+U) \le_c (S,T+U)$. Using Result 1 we have

$$\begin{aligned} &\mathbb{E}[\max(0;S^{\perp\perp}+\max(0;T^{\perp\perp}+U-200)-200)] \\ &\qquad \le \mathbb{E}[\max(0;S+\max(0;T+U-200)-200)], \end{aligned}$$

which is equivalent to

$$\begin{aligned} &\mathbb{E}[\min(200;S+\max(0;T+U-200))] \\ &\qquad \le \mathbb{E}[\min(200;S^{\perp\perp}+\max(0;T^{\perp\perp}+U-200))]. \end{aligned}$$

Numerically, we have $20.519 \le 21.131$. A summary of the results is shown in Table 3.

Table 3

Pure Premiums for Treaty 1

Fréchet Lower Bound	19.469
Exact Result	20.519
Independent Case	21.131
Fréchet Upper Bound	21.279

It is also interesting to analyze other moments of the cover, or premiums obtained by the PH-transform premium principle. They are given in Table 4.

Unfortunately, the other moments, as well as the premiums obtained with the PH-transform premium principle, are not ordered anymore.

5.2 Treaty 2

Some preliminary statistics are displayed in Table 5. The correlation between S^{RT} and T^{RD} is 0.35. The pure premium for our cover is:

Table 4
Moments and PH-Transform Premium Principle for Treaty 1

	Exact	Independent
$\mathbb{E}[\text{Cover}]$	20.519	21.131
$\mathbb{E}[\text{Cover}^2]$	265.04	261.32
$\mathbb{E}[\text{Cover}^3]$	4124.3	3915.8
$\mathbb{E}[\text{Cover}^4]$	70331.0	64760.0

ρ	Π_ρ (Exact)	Π_ρ (Independent)
1.00	20.519	21.131
0.75	34.898	35.420
0.50	60.786	61.034
0.25	108.71	108.55

$$\mathbb{E}\text{Cover} = \sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\sum_{u=0}^{\infty} f_{S,T}(s,t) f_U(u) \max(0, s+t+u-400) = 2.252.$$

Upper and lower bounds can be obtained using Theorem 1 in connection with P1 to give

$$(S^{min}, T^{min}) \leq_c (S,T) \leq_c (S^{max}, T^{max}).$$

Using P2, we have

$$(S^{min}, T^{min} + U) \leq_c (S, T+U) \leq_c (S^{max}, T^{max} + U).$$

Using Result 1, we have

Table 5
Means and Variances for Treaty 2

	X^{RD}	X^{RT}	Y^{RD}	S^{RT}	T^{RD}	U^{RD}
Means	200	16.14	1.83	600	4.84	4.57
Variances	0	1817.63	206.31	12000	623.4	524.15

$$\begin{aligned}\mathbb{E}[\max(0; S^{min} + T^{min} + U - 400)] \\ &\le \mathbb{E}[\max(0; S + T + U - 400)] \\ &\le \mathbb{E}[\max(0; S^{max} + T^{max} + U - 400)].\end{aligned}$$

These numerical bounds are $0.952 \le 2.252 \le 5.471$.

Though we are now able to improve the lower bound, the process is, unfortunately, less interesting than in the previous example. $X^{\mathrm{RT},\perp}$ and $X^{\mathrm{RD},\perp}$ are the independent versions of X^{RT} and X^{RD}; that is $X^{\mathrm{RT},\perp}$ and $X^{\mathrm{RD},\perp}$ are independent and have the same distributions as X^{RT} and X^{RD}, respectively. Let us define

$$\begin{aligned} S^\perp &= X_1^{\mathrm{RT},\perp} + \cdots + X_N^{\mathrm{RT},\perp}, \\ T^\perp &= X_1^{\mathrm{RD},\perp} + \cdots + X_N^{\mathrm{RD},\perp}. \end{aligned}$$

It is clear, from their construction, that X^{RT} and X^{RD} are comonotonic random variables. We then have

$$\begin{aligned} F_{X^{\mathrm{RT},\perp}, X^{\mathrm{RD},\perp}}(x_1, x_2) &\le \min(F_{X^{\mathrm{RT}}}(x_1), F_{X^{\mathrm{RD}}}(x_2)) \\ &= F_{X^{\mathrm{RT}}, X^{\mathrm{RD}}}(x_1, x_2) \end{aligned}$$

for any $x_1, x_2 \ge 0$. Therefore,

$$(X^{\mathrm{RT},\perp}, X^{\mathrm{RD},\perp}) \le_c (X^{\mathrm{RT}}, X^{\mathrm{RD}}).$$

Using P3, we have $(S^\perp, T^\perp) \le_c (S, T)$. As $S^\perp$ and $T^\perp$ are dependent, however, there is little interest in working with this random vector.

We define the independent versions of $S^\perp$ and $T^\perp$ as $S^{\perp\perp}$ and $T^{\perp\perp}$, respectively. Using Result 2, $(S^{\perp\perp}, T^{\perp\perp}) \le_c (S^\perp, T^\perp)$. By transitivity, we obtain $(S^{\perp\perp}, T^{\perp\perp}) \le_c (S, T)$. Using P2, we have $(S^{\perp\perp}, T^{\perp\perp} + U) \le_c (S, T + U)$. Using Result 1, we have

$$\mathbb{E}[\max(0; S^{\perp\perp} + T^{\perp\perp} + U - 400)] \le \mathbb{E}[\max(0; S + T + U - 400)].$$

Numerically, we have $1.153 \le 2.252$. These results are summarized in Table 6. Contrary to Treaty 1, we are now able to compare other moments and premiums obtained by the PH-transform premium principle. Numerical results are summarized in Table 7. In this case the characterization of the stop-loss order and equation (2) are directly applicable because we have $\mathrm{Cover}^\perp \le_{sl}$ Cover.

Table 6
Treaty 2: Pure Premium

Fréchet Lower Bound	0.952
Independent Case	1.153
Exact Result	2.252
Fréchet Upper Bound	5.471

Table 7
Moments and PH-Transform
Premium Principle for Treaty 2

	Exact	Independent
$\mathbb{E}[\text{Cover}]$	2.252	1.152
$\mathbb{E}[\text{Cover}^2]$	486.9	242.3
$\mathbb{E}[\text{Cover}^3]$	140198.0	61874.0
$\mathbb{E}[\text{Cover}^4]$	51084848.0	19062223.0

ρ	Π_ρ (Exact)	Π_ρ (Independent)
1.00	2.252	1.152
0.75	7.815	4.611
0.50	30.22	20.60
0.25	170.04	143.41

5.3 Treaty 2bis

Typically, a reinsurer will not offer an unlimited cover, at least for property business. Therefore, the cover of Treaty 2 should be limited in practice and could read

$$\text{Cover} = \min(400, \max(0, S + T + U - 400)),$$

which we call Treaty 2bis.

Pricing this realistic cover thus requires exact computations because of our inability to show that the derived bounds remain valid. We obtain the following results in Table 8. For this example, the bounds remain valid. This result is probably due to the very low probability of exhausting the cover, a fact that is confirmed by observing that the pure premium is the same (at least with three decimal digits) in both cases.

Table 8
Pure premiums for Treaty 2bis

Fréchet Lower Bound	0.952
Independent Case	1.153
Exact Result	2.252
Fréchet Upper Bound	5.471

6 Conclusion

By assuming the reinsurance cover is a function of a comonotonic random vector, we have shown how it is possible to obtain bounds for the pure premium. In particular, we observed that for Treaty 1, the wrong hypothesis of independence provides an upper bound for the pure premium of the treaty. This happens when the cover of the treaty is limited and when the comonotonic random variables are expressed as an excess of the same underlying random variable. Unfortunately, we have found in one case that the other moments of the cover are no longer ordered, i.e., even if we can prove that the first moment under the wrong hypothesis of independence is larger than the first moment under the exact hypothesis of independence, this property is not true for higher moments. In addition, we do not have a theoretical result on these orders.

In a second example we show that the wrong hypothesis of independence was not conservative, which shows the following consequence of not working with the exact model when it is known: if you work with the wrong model, you compute wrong premiums, which are too low when compared to exact premiums. Furthermore, the upper and lower Fréchet bounds may be quite far from the exact result, as shown in Treaty 2.

The theoretical results derived in this paper were based on a two dimensional paradigm. However these results can be extended to higher dimensions by using the supermodular ordering. Further research is being pursued in this area.

References

Ambagaspitiya, R.S. "On the Distribution of a Sum of Correlated Aggregate Claims." *Insurance: Mathematics and Economics* 23 (1998): 15–19.

Dhaene, J. and Goovaerts, M.J. "Dependency of Risks and Stop-Loss Order." *Astin Bulletin* 26 (1996): 201–212.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. "The Concept of Comonotonicity in Actuarial Science and Finance: Theory." *Insurance: Mathematics and Economics* 31 (2002): 3–33.

Fréchet, M. "Sur les Tableaux de Corrélation dont les Marges sont Données." *Ann. Univ. Lyon Sect. A.*, Séries 3, 14 (1951): 53–77.

Frees, E.W. and Valdez, E.A. "Understanding Relationships Using Copulas." *North American Actuarial Journal* 2 (1998): 1–15.

Gerber, H.U. "On the Numerical Evaluation of the Distribution of Aggregate Claims and its Stop-Loss Premiums." *Insurance: Mathematics and Economics* 1 (1982): 13–18.

Hoeffding, W. "Masstabinvariante Korrelations-Theorie." *Schr. Math. Inst. Univ. Berlin* 5 (1940): 181–233.

Müller, A. and Stoyan, D. *Comparison Methods for Stochastic Models and Risks.* New York, NY: Wiley Series in Probability and Statistics, 2002.

Panjer, H.H. "Recursive Evaluation of a Family of Compound Distributions." *Astin Bulletin* 12 (1981): 22–26.

Sundt, B. "On Multivariate Panjer's Recursions." *Astin Bulletin* 29 (1999): 29–46.

Thépaut, A. "Le Traité d'Excédent du Coût Moyen Relatif." *Bulletin Trimestriel de l'Institut des Actuaires Français* 192 (1950): 273–343.

Walhin, J.F. and Paris, J. "The Effect of Excess of Loss Reinsurance with Reinstatements on the Cedent's Portfolio." *Blätter der Deutsche Gesellschaft für Versicherungsmathematik* 24 (2000a): 615–627.

Walhin, J.F. and Paris, J. "Recursive Formulae for Some Bivariate Counting Distributions Obtained by the Trivariare Reduction Method." *Astin Bulletin* 30 (2000b): 141–155.

Walhin, J.F. and Paris, J. "The Mixed Bivariate Hodman Distribution." *Astin Bulletin* 31 (2001): 123–138.

Walhin, J.F. "Some Comments on the Pricing of Exotic Excess of Loss Treaties." *Journal of Actuarial Practice* 10 (2002): 175–191.

Wang, S. "Premium Calculation by Transforming the Layer Premium Density." *Astin Bulletin* 26 (1996): 71-92.

Wang, S. and Dhaene, J. "Comonotonicity, Correlation Order and Premium Principles." *Insurance: Mathematics and Economics* 22 (1998): 235-242.

Journal of Actuarial Practice Vol. 12, 2005

An Application of Control Theory to the Individual Aggregate Cost Method

Alexandros A. Zimbidis* and Steven Haberman†

Abstract

The paper investigates the individual aggregate cost method (also known as the individual spread-gain method), which is normally applicable in small pension funds or fully contributory schemes, using a control theoretical framework. We construct the difference equations describing the mechanisms of the respective funding method and then calculate the optimal control path of the contribution rate assuming (first) a stochastic and (second) a deterministic pattern for the future investment rates of return. For the first case, the optimal solution is achieved through a linear approximation and using stochastic optimization techniques. It is proved that the contribution rate is (optimally) controlled through the control of the valuation rate (which is determined incorporating a certain feedback mechanism of the past contribution rate). The optimal solution for the deterministic case is obtained using standard calculus and the method of Lagrange multipliers.

Key words and phrases: *individual aggregate funding, linear approximation, optimal stochastic control, Lagrange multipliers*

*Alexandros A. Zimbidis Ph.D., F.H.A.S., is a lecturer at Athens University of Economics and Business, Greece. He received his B.Sc. (first class) in mathematics from the University of Athens and his masters and Ph.D. in actuarial science from City University, London. He worked as a director in a large multinational insurance company. Dr. Zimbidis is a Fellow of the Hellenic Actuarial Society and is the chair of the President of the National Actuarial Authority in Greece. His research interests include applications of stochastic analysis and control theory to insurance and financial systems.

Dr. Zimbidis's address is: Athens University of Economics and Business, Department of Statistics, 76 Patission St., Athens 10434, GREECE. E-mail: *aaz@aueb.gr*

†Steven Haberman M.A., Ph.D., D.Sc., F.I.A., is professor and deputy dean at Cass Business School, City University. He has worked at Prudential Assurance and the Government Actuary's Department in London. He has published extensively in the fields of actuarial science and applied statistics. His current research interests include modeling pension plans (defined benefit and defined contribution); mortality forecasting, longevity risk and annuities; modeling guarantees in life insurance; non-life insurance premiums in competitive markets.

Dr. Haberman's address is: Cass Business School, City University, 106 Bunhill Row, London, EC1Y 8TZ, UNITED KINGDOM. E-mail: *s.haberman@city.ac.uk*

1 Introduction

Over the past fifteen years or so, control theory has been used to study different types of insurance systems. Researchers have found that this specific theory provides a powerful research framework to analyze the evolution of insurance and pension systems, as well as to determine optimal strategies for determining, for example, the level of insurance premiums, the level of pension fund contributions, or asset allocation. Several authors have used control theory for the investigation of the properties of different pension funding methods; e.g., Benjamin (1989), Zimbidis and Haberman (1993), Haberman and Sung (1994), Loades (1998), Owadally and Haberman (1999, 2004), Cairns (2000), and Taylor (2002).

This paper focuses on the application of control theory to the individual aggregate cost method, which is also known as the individual spread-gain method. The individual aggregate cost method is normally applicable to two broad categories of pension funds: (i) small pension funds where the number of members participating in the plan is so small or the membership is so heterogeneous that the average contribution rate obtained by the aggregate (or other) funding method is not reliable or sufficient; or (ii) pension funds (whether small or large) where the individual members contribute the major share of the total annual contributions; see McGill et al. (1989).

Consider an employee who was hired at age e and will retire at the normal retirement age r, i.e., the employee is expected to give $m = r - e$ years of service. If the employee is currently age x at time n, where $x = n + e$, then under the individual aggregate cost method, an individual's contribution rate at the beginning of year $n+1$ for $n = 0, 1, 2, \ldots, m-1$, is C_n, where:

$$C_n = \begin{cases} \dfrac{\mathrm{PVTB}_n - F_n}{{}^s\ddot{a}_{x:\overline{m-n}|}} \\[2ex] \dfrac{\mathrm{PVTB}_n - F_n}{\ddot{a}_{x:\overline{m-n}|}} \end{cases} \tag{1}$$

for $n = 0, 1, 2, \ldots, m-1$ while $C_n = 0$ for $n = m, m+1, \ldots$, F_n is the accumulated fund assets at time n, PVTB_n is the actuarial present value of total retirement benefits earned at time n, ${}^s\ddot{a}_{x:\overline{m-n}|}$ is the actuarial present value of a life annuity with payments increasing according to a salary scale, and $\ddot{a}_{x:\overline{m-n}|}$ is the actuarial present value of a life annuity with level payments. All actuarial present values are assumed to be discounted at the valuation rate of interest of i.

We assume each plan member has his/her own separate account that changes due to the employee's own contributions, investment returns on the fund, and expenses associated with managing the fund. For example, in Greece the fund value for each member, F_n, is calculated by crediting the employee's own contributions, debiting management and other expenses either as a flat amount for each person or as a standard percentage based upon the employee's own contributions, and finally crediting investment proceeds proportionally according to the prior fund value F_{n-1}. A full description of the individual aggregate cost method may be found in McGill et al. (1989) and Winklevoss (1993).

2 Description of The Model

We now present some of the assumptions used throughout this paper.

A.1: The pension plan is a defined benefit plan with normal retirement age r.

A.2: The plan uses the individual aggregate funding cost method for plan valuations.

A.3: We consider a plan member who was hired at age e and has a future working lifetime of m years, where $m = r - e$.

A.4: There are no pre-retirement mortality, disability, or other decrements.[1]

A.5: The normal retirement benefit is $B/12$ per month.

A.6: The unit of currency used is such that the product of the annuity factor and the annual retirement benefit is equal to one monetary unit, i.e.,

$$B\ddot{a}_r^{(12)} = 1.$$

A.7: As the normal retirement benefit is independent of salary, normal cost is calculated on the basis of a level-dollar amount.

A.8: The contribution rate for the plan year $[n, n+1)$ is C_n monetary units paid at time n and is equal to the plan's normal cost.

A.9: The total funding period of m years is divided in two sub-periods: $[0, T)$ and $[T, m)$. In the first sub-period (up to

[1]This assumption may be justified because, in fully contributory plans (where this specific method is normally applicable), the ancillary non-retirement (death, disability, or other) benefits are normally equal to (or approximately equal to) the accrued liability at the date of decrement, resulting in no gain or loss to the plan.

time T) the investment return for the year $[n-1, n)$, j_n, is considered to be a stationary stochastic process where $\mathbb{E}(j_n) = j > 0$ and $\mathbb{V}\mathrm{ar}(j_n) = \sigma^2 > 0$, and the j_ns are mutually independent. In the second sub-period (after time T) the annual investment returns, j_n, are deterministic for $n = T, t+1, \ldots, m-1$.

A.10: The valuation rate for the plan year $[n, n+1)$, i_n, is determined at the end of the previous year (i.e., at time n) and is based on the information and experience available at n, the i_n is used to determine the contribution rate C_n for the plan year $[n, n+1)$.

Under the traditional approach to determining the contribution rate for the individual aggregate cost method (as formulated in equation (1)), we assume a constant valuation rate of interest i for each year n and an initial fund value F_0, which is normally zero. The contributions rates $(C_n, n = 0, 1, 2, \ldots, m-1)$ determined by equation (1) vary because of fluctuations in the investment returns on the accumulated assets, F_n. In order to reduce the fluctuations in the C_n, $n = 0, 1, 2, \ldots, m-1$, we control the contribution rates by adapting the valuation rate of interest. This is justified because the valuation rate of a pension fund is highly correlated to the long-term interest rates (see Wilkie, 1995 or Ang and Sherris, 1997) so that the estimation of these rates will influence the determination of the valuation rate.

Our approach follows the standard practice, which is commonly called the life-style investment strategy.[2] Following assumption A.9, at the beginning of the first sub-period, we choose a high risk high expected return investment policy. Once in the second sub-period, the assets are switched to assets with a lower risk lower expected return investment policy in order to secure the benefits of the member at the date of retirement. The value of T is determined by the pension fund manager, and we suggest that T should be close to m in order to be able to obtain an investment product with guaranteed rates for the second period.

In the first sub-period, where the returns follow a stochastic process, we control the contribution rate by adopting a control feedback mechanism for the annual valuation rates i_n. In the second sub-period, where the returns follow a deterministic process, we control the contribution rate directly by calculating the optimal path that produces the promised retirement benefit.

[2]See Vigna and Haberman (2001) for a discussion in the context of defined contribution pension schemes.

Because there are no pre-retirement decrements (assumption A. 4) and the retirement benefit is independent of salary, we obtain the following system of difference equations:

$$C_n = \frac{v_{(i_n)}^{m-n} B \ddot{a}_r^{(12)} - F_n}{\ddot{a}_{\overline{m-n}|\, i_n}}, \tag{2}$$

which is the equation for the normal cost, and

$$F_n = (F_{n-1} + C_{n-1})(1 + j_n) \tag{3}$$

where $v_{(i_n)}^{m-n} = (1 + i_n)^{-(m-n)}$ is the discount factor. Equation (2) can be rewritten as $F_n = v_{(i_n)}^{m-n} - C_n \ddot{a}_{\overline{m-n}|\, i_n}$ and substituted into equation (3) to yield, after some elementary algebra,

$$C_n = (1 + j_n) \frac{\ddot{a}_{\overline{m-n}|\, i_{n-1}}}{\ddot{a}_{\overline{m-n}|\, i_n}} C_{n-1} + \frac{v_{(i_n)}^{m-n}}{\ddot{a}_{\overline{m-n}|\, i_n}} - (1 + j_n) \frac{v_{(i_{n-1})}^{m-n+1}}{\ddot{a}_{\overline{m-n}|\, i_n}}. \tag{4}$$

It is clear from equation (4) that the system is non-linear in the valuation rates i_n and i_{n-1} and linear in j_n and C_{n-1}. The state variable of the system is the contribution rate C_n while the input variable is the actual rate of investment return j_n. The valuation rate of interest i_n that appears in equation (4) is the source of non-linearity and operates as the control variable, which attempts to balance the system.

In a steady state, where $j_n = j$, a constant for $n = 0, 1, 2, \ldots$, and, if the valuation rate also is equal to j (i.e., $i_n = j$), then we obtain from equation (4)

$$C_{m-1} = C_{m-2} = \cdots = C_0. \tag{5}$$

The initial contribution (C_0) then is calculated using equation (2) and assuming $F_0 = 0$, i.e.,

$$C_0 = \frac{v_{(j)}^m}{\ddot{a}_{\overline{m}|\, j}} = \frac{1}{\ddot{s}_{\overline{m}|\, j}}. \tag{6}$$

Then the fund value at each time point n is equal to the respective accumulation of contributions, i.e.,

$$F_n = C_0 \ddot{s}_{\overline{n}|\, j} = \frac{\ddot{s}_{\overline{n}|\, j}}{\ddot{s}_{\overline{m}|\, j}} \tag{7}$$

for $n = 1, \ldots, m$. As the recursive equation (4) is non-linear, we propose to employ linear approximation techniques in order to solve the problem.

3 Mean and Variance of Contribution Rate

Before going further with the control theoretical analysis of the model and the linear approximation, we investigate the mean and variance of the contribution rates under the traditional approach where the valuation rate of interest is assumed to be constant for each year and equal to the expected rate of investment return, i.e.,

$$i_0 = i_n = j = \mathbb{E}(j_n) \tag{8}$$

for $n = 1, \ldots, m-1$. Substituting equation (8) in equation (4) we obtain

$$C_n = \frac{1 + j_n}{1 + j} C_{n-1} + \frac{j - j_n}{1 + j} f_n(j) \tag{9}$$

where

$$f_n(j) = \frac{1}{\ddot{s}_{\overline{n}|j}}. \tag{10}$$

In order to facilitate our calculations, we introduce the filtration H_n, which represents all the available information generated by the entire funding process (the annual investment rates, the decisions for the contribution rates, etc.) up to and including time n. We also use two well known results from the theory of conditional probabilities:

$$\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)] \quad \text{and} \tag{11}$$

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{V}\mathrm{ar}[\mathbb{E}(X|Y)] + \mathbb{E}[\mathbb{V}\mathrm{ar}(X|Y)] \tag{12}$$

where X, Y are random variables.

From equation (9) and conditioning on H_{n-1} give

$$\begin{aligned}
\mathbb{E}(C_n) &= \mathbb{E}(\mathbb{E}(C_n|H_{n-1})) \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{1+j_n}{1+j}C_{n-1} + \frac{j-j_n}{1+j}f_n(j)|H_{n-1}\right]\right] \\
&= \mathbb{E}\left[\frac{1+\mathbb{E}(j_n)}{1+j}C_{n-1} + \frac{j-\mathbb{E}(j_n)}{1+j}f_n(j)\right] \\
&= \mathbb{E}(C_{n-1})
\end{aligned}$$

for $n = 1,\ldots,m-1$. Consequently, we determine that the expected contribution rate is constant over time n, i.e.,

$$\mathbb{E}(C_n) = C_0 \quad \text{for} \quad n = 1,\ldots,m-1. \tag{13}$$

To obtain the variance we note that, using equation (9) and conditioning on H_{n-1} yields

$$\mathbb{V}\mathrm{ar}\,(C_n) = \mathbb{V}\mathrm{ar}\,[\mathbb{E}\,(C_n|H_{n-1})] + \mathbb{E}\,[\mathbb{V}\mathrm{ar}\,(C_n|H_{n-1})]$$

where $\mathbb{V}\mathrm{ar}\,[\mathbb{E}\,(C_n|H_{n-1})] = \mathbb{V}\mathrm{ar}\,(C_{n-1})$ and

$$\begin{aligned}
&\mathbb{E}\,[\mathbb{V}\mathrm{ar}\,(C_n|H_{n-1})] \\
&\quad = \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left(\frac{1+j_n}{1+j}C_{n-1} + \frac{j-j_n}{1+j}f_n(j)|H_{n-1}\right)\right] \\
&\quad = \mathbb{E}\left[\frac{C_{n-1}^2}{(1+j)^2}\mathbb{V}\mathrm{ar}(j_n) + \frac{f_n^2(j) - 2f_n(j)C_{n-1}}{(1+j)^2}\mathbb{V}\mathrm{ar}(j_n)\right] \\
&\quad = \sigma^2\left[\frac{\mathbb{E}(C_{n-1}^2)}{(1+j)^2} + \frac{f_n^2(j) - 2f_n(j)\mathbb{E}(C_{n-1})}{(1+j)^2}\right] \\
&\quad = \sigma^2\left[\frac{\mathbb{V}\mathrm{ar}(C_{n-1}) + [\mathbb{E}(C_{n-1})]^2}{(1+j)^2} + \frac{f_n^2(j) - 2f_n(j)\mathbb{E}(C_{n-1})}{(1+j)^2}\right] \\
&\quad = \left(\frac{\sigma}{1+j}\right)^2 \mathbb{V}\mathrm{ar}\,(C_{n-1}) + \left(\frac{\sigma}{1+j}\right)^2\left(C_0^2 + f_n^2(j) - 2f_n(j)C_0\right)
\end{aligned}$$

for $n = 1,\ldots,m-1$. Hence

$$\mathbb{V}\mathrm{ar}\,(C_n) = \left[1 + \left(\frac{\sigma}{1+j}\right)^2\right]\mathbb{V}\mathrm{ar}\,(C_{n-1}) + \frac{\sigma^2}{(1+j)^2}\left(C_0 - f_n(j)\right)^2.$$

For notational convenience, let $\psi_n = \mathbb{V}\mathrm{ar}\,(C_n)$ and

$$A_n = \frac{\sigma^2}{(1+j)^2}\left[C_0 - f_n(j)\right]^2 = \frac{\sigma^2}{(1+j)^2}\left[\frac{1}{\ddot{s}_{\overline{m}|j}} - \frac{1}{\ddot{s}_{\overline{m-n}|j}}\right]^2.$$

Thus we have the following difference equation for the variance:

$$\psi_n = \left[1 + \left(\frac{\sigma}{1+j}\right)^2\right]\psi_{n-1} + A_n \tag{14}$$

for $n = 1, 2, \ldots, m-1$, which has the solution

$$\psi_n = \left[1 + \left(\frac{\sigma}{1+j}\right)^2\right]^n \psi_0 + \sum_{k=0}^{n-1}\left[1 + \left(\frac{\sigma}{1+j}\right)^2\right]^k A_{n-k}. \tag{15}$$

As $1 + (\sigma/(1+j))^2 > 1$ and $A_{n-k} > 0$, it is clear from equation (15) that ψ_n increases as n increases up to m. We also observe that the rate of increase in ψ_n depends on the ratio $\sigma/(1+j)$.

In order to restrict the magnitude of the ψ_ns, we now consider a control theory approach based on a variable valuation rate of interest.

4 Optimal Control Strategy

4.1 The Objective Function

As described in assumption A.9, we split the total funding period into two sub-periods. In the first sub-period we apply a stochastic model for the investment rate of return, while in the second sub-period we apply a deterministic model. In the first sub-period, we use the valuation rate of interest as a control mechanism, while in the second sub-period we directly determine the contribution rates. In other words, we either control the valuation rate of interest or the contribution rate.

Our main objective in the control problem is the minimization of the contribution rate risk, which is defined as the total mean square deviation of contribution levels from their target values.[3] Following Vandebroek (1990) and Haberman and Sung (1994), we adopt a weighted quadratic objective function of minimizing total mean square deviations of contribution levels from their target values (because our basic

[3] According to Haberman and Sung (1994), the contribution rate risk is one of the two main risks with which a pension plan is confronted, while the other basic risk is the solvency risk.

aim is to reduce the fluctuations of the contribution rate) and also total mean square deviations of valuation rates from their target levels.[4]

Let C_{st} and i_{st} denote the contribution rate and valuation rate, respectively, for the steady state of the system (see equation (20)), and let β be a weighting factor, where $0 \le \beta \le 1$. The choice of the factor β reflects the preference of the pension scheme manager between the contribution rate and the valuation interest rate and which of them may exhibit more or less fluctuations. That is, if the manager chooses β to be close to zero that means the manager prefers a more stable contribution rate and pays almost no attention to possible large fluctuations in the valuation interest rate. For the first sub-period with the stochastic investment rates, the objective function under the quadratic performance criterion has the following form:

$$O_1 = \min_{C_k, i_k} \mathbb{E}\left[(1-\beta)\sum_{k=1}^{T-1}(C_k - C_{st})^2 + \beta\sum_{k=1}^{T-1}(i_k - i_{st})^2\right]. \tag{16}$$

For the second sub-period (from (T) up to (m)) the objective function includes no stochastic elements (hence no expectation operator) and has the following form:

$$O_2 = \min_{C_k} \sum_{k=T}^{m-1} (C_k - C_{st})^2 . \tag{17}$$

It is easy to argue that for $\beta = 1$, the new (controlled) model almost corresponds to the traditional approach of the individual aggregate cost method.

4.2 Optimal Control During the First Sub-Period

The difference equation (4) may be linearized in the neighborhood of a certain steady state, defined by:

$$(j_n, i_{n-1}, i_n, C_n) = (j_{st}, i_{st}, i_{st}, C_{st}) \tag{18}$$

In the steady state, the valuation rate of interest is equal to the actual investment rate of return, and normally we choose j_{st} to equal the mean of the j_n, i.e.,

[4] A quadratic objective function has the advantage of leading to mathematically tractable results but we acknowledge that it has the inherent disadvantage of treating deviations below and above the target in an equivalent manner. The use of semi-variance type measures would allow more flexibility in this direction but at the expense of tractability.

$$j_{st} = i_{st} = \mathbb{E}\left(j_n\right) = j. \tag{19}$$

The respective contribution rate is given by

$$C_{st} = \frac{1}{\ddot{s}_{\overline{m}|j}}. \tag{20}$$

By considering infinitesimally small (∇) changes[5] about the steady state for C_n, C_{n-1}, i_n, i_{n-1}, and j_n, i.e., $C_n = C_{st} + \nabla C_n$, $C_{n-1} = C_{st} + \nabla C_{n-1}$, $i_n = j + \nabla i_n$, $i_{n-1} = j + \nabla i_{n-1}$, and $j_n = j + \nabla j_n$, we obtain the following linear approximation (after using equation (19))

$$\nabla C_n = \nabla C_{n-1} + \xi_n \nabla j_n + \varphi_n \nabla i_{n-1} + \zeta_n \nabla i_n + \text{nonlinear terms} \tag{21}$$

where

$$\begin{aligned}
\xi_n &= \frac{1}{1+j}\left(\frac{1}{\ddot{s}_{\overline{m}|j}} - \frac{1}{\ddot{s}_{\overline{m-n}|j}}\right) \\
\varphi_n &= \frac{m-n+1}{1+j}\frac{1}{\ddot{s}_{\overline{m-n}|j}} + \frac{m-n}{j}\frac{1}{\ddot{s}_{\overline{m}|j}}\frac{1}{\ddot{s}_{\overline{m-n}|j}} - \frac{1}{j}\frac{1}{\ddot{s}_{\overline{m}|j}} \\
\zeta_n &= -\frac{m-n}{1+j}\frac{1}{\ddot{s}_{\overline{m-n}|j}} - \frac{m-n}{j}\frac{1}{\ddot{s}_{\overline{m}|j}}\frac{1}{\ddot{s}_{\overline{m-n}|j}} + \frac{1}{j(1+j)}\frac{1}{\ddot{s}_{\overline{m-n}|j}}
\end{aligned}$$

Note that $\varphi_n = -(\xi_n + \zeta_n)$; hence equation (21) may be rewritten as

$$\nabla C_n = \nabla C_{n-1} + \xi_n \nabla j_n - (\xi_n + \zeta_n)\nabla i_{n-1} + \zeta_n \nabla i_n. \tag{22}$$

At this point it is important to briefly describe the solution to a general linear dynamic difference equation of the form

$$\mathbf{x}_n = \mathbf{A}_n \mathbf{x}_{n-1} + \mathbf{B}_n \mathbf{u}_n + \mathbf{e}_n \tag{23}$$

for $n = 1, 2, \ldots, N$, where $\mathbf{x}_n \in \mathbb{R}^n$ is the state variable, $\mathbf{u}_n \in \mathbb{R}^k$ is the control variable, $\mathbf{A}_n \in \mathbb{R}^{r\times r}$ and $\mathbf{B}_n \in \mathbb{R}^{r\times k}$ are known non-random matrices, and $\mathbf{e}_n \in \mathbb{R}^r$ is a random vector with $\mathbb{E}(\mathbf{e}_n) = 0$ and finite covariance matrix. In addition, we assume $\mathbf{e}_n$ is independent of $\mathbf{x}_n$ and $\mathbf{u}_n$.

The problem is to search for the optimal control $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{N-1}$ that minimizes the following expectation:

---

[5]Here ∇ is the backward difference operator, i.e., for any function $f(x)$, $\nabla f(x) = f(x+1) - f(x)$ and $\nabla^{n+1} f(x) = \nabla^n f(x+1) - \nabla^n f(x)$ for $n = 1, 2, \ldots$.

$$\mathop{\mathbb{E}}_{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_N}\left\{\sum_{n=1}^{N}\mathbf{x}_n^\top\mathbf{K}_n\mathbf{x}_n+\sum_{n=1}^{N-1}\mathbf{u}_n^\top\mathbf{R}_n\mathbf{u}_n\right\} \tag{24}$$

where the symbol $^\top$ denotes the transpose operator, $\mathbf{K}_n$ is a symmetric positive semi-definite matrix, and $\mathbf{R}_n$ is a symmetric positive definite matrix. Following Aoki (1989, pp. 131–148) and Bertsekas (1976, pp. 70–80), the optimal solution, given the initial condition $\mathbf{x}_0$, is described by the following equations:

$$\mathbf{u}_n = \mathbf{M}_n\mathbf{x}_{n-1} \tag{25}$$

where $\mathbf{H}_N = \mathbf{K}_N$, and, for $n = N-1, N-2, \ldots, 1, 0$, we have

$$\mathbf{M}_n = -\left(\mathbf{B}_n^\top\mathbf{H}_{n+1}\mathbf{B}_n+\mathbf{R}_n\right)^{-1}\left(\mathbf{B}_n^\top\mathbf{H}_{n+1}\mathbf{A}_n\right) \tag{26}$$

$$\mathbf{H}_n = A_n^\top\left[\mathbf{H}_{n+1}-\mathbf{H}_{n+1}\mathbf{B}_n\left(\mathbf{B}_n^\top\mathbf{H}_{n+1}\mathbf{B}_n+\mathbf{R}_n\right)^{-1}\mathbf{B}_n^\top\mathbf{H}_{n+1}\right]\mathbf{A}_n+\mathbf{K}_n. \tag{27}$$

In order to fit the last equation with the format of the linear system, which appears in equation (23), we write

$$\nabla C_n = \nabla C_{n-1} + \begin{pmatrix}\zeta_n & 0\\ 0 & -(\xi_n+\zeta_n)\end{pmatrix}\begin{pmatrix}\nabla i_n\\ \nabla i_{n-1}\end{pmatrix}+\xi_n\nabla j_n, \tag{28}$$

in other words $\mathbf{x}_n = \nabla C_n$ and $\mathbf{A}_n = 1$, $\mathbf{e}_n = \xi_n\nabla j_n$ are scalars, and

$$\mathbf{B}_n = \begin{pmatrix}\zeta_n & 0\\ 0 & -(\xi_n+\zeta_n)\end{pmatrix} \quad\text{and}\quad \mathbf{u}_n = \begin{pmatrix}\nabla i_n\\ \nabla i_{n-1}\end{pmatrix}.$$

The objective function is

$$\min_{\{\nabla^2 i_1,\nabla^2 i_2,\ldots,\nabla^2 i_{T-1}\}}\mathbb{E}\left[(1-\beta)\sum_{k=1}^{T-1}\nabla C_k^2+\beta\sum_{k=1}^{T-1}\nabla^2 i_k^2\right], \tag{29}$$

which can be rewritten in matrix form as

$$\mathop{\mathbb{E}}_{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_N}\left\{\sum_{n=1}^{N}\mathbf{x}_n^\top\mathbf{K}_n\mathbf{x}_n+\sum_{n=1}^{N}\mathbf{u}_n^\top\mathbf{R}_n\mathbf{u}_n\right\} \tag{30}$$

where $N = T-1$, $\mathbf{K}_n = (1-\beta)$, a scalar, and

$$\mathbf{R}_n = \beta \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The optimal control for equation (28), which minimizes the objective function (30), is then given by

$$\begin{pmatrix} \nabla i_n \\ \nabla i_{n-1} \end{pmatrix} = M_n \nabla C_{n-1} \tag{31}$$

where M_n is calculated according to equations (25) to (27).

If $\beta = 0$ then we obtain the special case where we pay no attention to the development of the valuation rate and we are fully interested to the development of the contribution rate. Equation (31) becomes

$$\nabla^2 i_n = \nabla i_n - \nabla i_{n-1} = -\frac{1}{\zeta_n} \nabla C_{n-1}. \tag{32}$$

Hence, the valuation rate of interest should be controlled using a feedback mechanism of the state variable (contribution rate). As the contribution increases, the proposed valuation rate of interest decreases.

Substituting the feedback mechanism of equation (32) into equation (22) yields

$$\begin{aligned} \nabla C_n &= \nabla C_{n-1} + \zeta_n \left[-\frac{1}{\zeta_n} \nabla C_{n-1} \right] + \xi_n (\nabla j_n - \nabla i_n) \\ &= \xi_n (\nabla j_n - \nabla i_n). \end{aligned} \tag{33}$$

4.3 Optimal Control During the Second Sub-Period

Having controlled the system for the first sub-period through an optimal path under a stochastic pattern of investment rates of return, we arrive at the time point T with a fund value of F_T. During the second sub-period the rates of return $j_{T+1}, j_{T+2}, \ldots, j_m$ are assumed to follow a deterministic process. Our problem now is to guide the fund value from F_T to $F_m = 1$ while minimizing the objective function:

$$\min_{\{C_T, C_{T+1}, \ldots, C_{m-1}\}} \sum_{k=T}^{m-1} (C_k - C_{st})^2. \tag{34}$$

Combining equation (3) and the requirement $F_m = 1$ yields the following constraint:

$$F_m = F_T \prod_{k=T+1}^{m} (1 + j_k) + \sum_{k=T}^{m-1} C_k \prod_{l=k+1}^{m} (1 + j_l) = 1. \tag{35}$$

Using Lagrange multipliers our problem is translated into the minimization of the Lagrangian function, $\Lambda\left(C_T, C_{T+1}, \ldots, C_{m-1}, \rho\right)$ with respect to $C_T, C_{T+1}, \ldots, C_{m-1}$ and ρ:

$$\begin{aligned} \Lambda(C_T, \ldots, C_{m-1}, \rho) &= \sum_{k=T}^{m-1} (C_k - C_{st})^2 \\ &+ \rho \left\{ F_T \prod_{k=T+1}^{m} (1 + j_k) + \sum_{k=T}^{m-1} C_k \prod_{l=k+1}^{m} (1 + j_l) - 1 \right\}. \end{aligned} \tag{36}$$

We find the minimum of Λ by equating the partial derivatives with respect to $C_T, C_{T+1}, \ldots, C_{m-1}$, ρ to zero. It is then straightforward (although tedious) to solve the resulting system of equations to give:

$$\rho = 2 \frac{\left[F_T \prod_{k=T+1}^{m} (1 + j_k) + C_{st} \sum_{k=T}^{m+1} \prod_{l=k+1}^{m} (1 + j_l) - 1 \right]}{\sum_{k=T}^{m-1} \prod_{l=k+1}^{m} (1 + j_l)^2} \tag{37}$$

$$C_k = C_{st} - \frac{\rho}{2} \prod_{l=k+1}^{m} (1 + j_l) \tag{38}$$

for $k = T, \ldots, m-1$. The case where $j_{T+1} = j_{T+2} = \ldots = j_m = j_*$, where j_* is the risk free rate (normally $j > j_*$) leads to

$$C_k = C_{st} - \frac{F_T \left(1 + j_*\right)^{m-T} + C_{st} \ddot{s}_{\overline{m-T}|j} - 1}{\ddot{s}_{\overline{m-n}|j_\phi}} \left(1 + j_*\right)^{m-k} \tag{39}$$

where $j_\phi = (1 + j_*)^2 - 1$.

5 The Mean and Variance of C_n with $\beta = 0$

In order to obtain a direct comparison between the traditional and control approach, we calculate the mean and variance of the contribution rate under the traditional approach, using the linearized difference equation (22). Under the traditional approach

$$\nabla i_0 = \nabla i_n = \mathbb{E}\left(\nabla j_n\right) \quad \text{for} \quad n = 0, 1, \ldots. \tag{40}$$

Hence, the traditional contribution rate, ∇C_n^{trad}, given in equation (22) becomes

$$\begin{aligned}\nabla C_n^{\text{trad}} &= \nabla C_{n-1}^{\text{trad}} + \xi_n\left(\nabla j_n - \nabla i_{n-1}\right)\\ &= \nabla C_0^{\text{trad}} + \xi_1\left(\nabla j_1 - \nabla i_0\right) + \cdots + \xi_n\left(\nabla j_n - \nabla i_{n-1}\right).\end{aligned} \tag{41}$$

Taking the expectation of both sides of equation (41) we obtain

$$\mathbb{E}\left[\nabla C_n^{\text{trad}}\right] = \mathbb{E}\left[\nabla C_0^{\text{trad}} + \xi_1\left(\nabla j_1 - \nabla i_0\right) + \cdots + \xi_n\left(\nabla j_n - \nabla i_{n-1}\right)\right] = 0$$

by using equation (40) and $\mathbb{E}(\nabla C_0^{\text{trad}}) = 0$. The last condition holds as the initial condition C_0^{trad} is constant so that ∇C_0^{trad} is equal to zero. Hence,

$$\mathbb{E}\left(C_n^{\text{trad}}\right) = \mathbb{E}\left(C_{st}^{\text{trad}} + \nabla C_n^{\text{trad}}\right) = \mathbb{E}\left(C_{st}^{\text{trad}}\right) = C_{st} = C_0.$$

The result is the same as in Section 3 where we used the full non-linear equation (9) for C_n.

Equation (41) also can be used to obtain the variance of the contribution rate under the traditional approach as follows:

$$\begin{aligned}\mathbb{V}\text{ar}\left[\nabla C_n^{\text{trad}}\right] &= \mathbb{V}\text{ar}\left[\nabla C_0^{\text{trad}} + \xi_1\left(\nabla j_1 - \nabla i_0\right) + \ldots + \xi_n\left(\nabla j_n - \nabla i_{n-1}\right)\right]\\ &= \mathbb{V}\text{ar}\left[\nabla C_0^{\text{trad}}\right] + \xi_1^2\sigma^2 + \xi_2^2\sigma^2 + \ldots + \xi_n^2\sigma^2\\ &= \sigma^2 \sum_{k=1}^{n} \xi_k^2\end{aligned} \tag{42}$$

because $\mathbb{V}\text{ar}\left(\nabla C_0^{\text{trad}}\right) = 0$.

Let C_n^{ctrl} denote the contribution rate under the control approach. We use equation (33) for ∇C_n^{ctrl}, i.e.,

$$\nabla C_n^{\text{ctrl}} = \xi_n\left(\nabla j_n - \nabla i_{n-1}\right) = \xi_n(j_n - i_{n-1}). \tag{43}$$

Proceeding as before,

$$\mathbb{E}(\nabla C_n^{\text{ctrl}}) = \mathbb{E}\left(\mathbb{E}\left(\xi_n(j_n - i_{n-1})\,|H_{n-1}\right)\right) = \xi_n\left(j - \mathbb{E}\left(i_{n-1}\right)\right) \tag{44}$$

while the variance is given by

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\nabla C_n^{\mathrm{ctrl}}\right) &= \mathbb{V}\mathrm{ar}\left(\mathbb{E}\left(\nabla C_n^{\mathrm{ctrl}}|H_{n-1}\right)\right) + \mathbb{E}\left(\mathbb{V}\mathrm{ar}\left(\nabla C_n^{\mathrm{ctrl}}|H_{n-1}\right)\right)\\ &= \mathbb{V}\mathrm{ar}\left(\xi_n(j - i_{n-1})\right) + \mathbb{E}\left(\xi_n^2\sigma^2\right)\\ &= \xi_n^2 \cdot \mathbb{V}\mathrm{ar}\left(i_{n-1}\right) + \sigma^2\xi_n^2. \end{aligned} \tag{45}$$

From equations (44) and (45) we observe that the mean and the variance of the contribution rate depend upon the mean and the variance of the valuation rate of interest.

Recall the control law for the valuation rate of interest, i.e., equation (32). Substituting the expression for $\nabla C_{n-1}^{\mathrm{ctrl}}$ from equation (43) in equation (32) we obtain

$$i_n = i_{n-1} + \frac{\xi_{n-1}}{\zeta_n} i_{n-2} - \frac{\xi_{n-1}}{\zeta_n} j_{n-1}, \tag{46}$$

which is a difference equation of time-varying format with initial conditions $i_0 = i_1 = j$. We now directly can obtain a recursive relationship for the means by taking the expectations of both sides of equation (46), i.e.,

$$\mathbb{E}(i_n) = \mathbb{E}(i_{n-1}) + \frac{\xi_{n-1}}{\zeta_n}\mathbb{E}(i_{n-2}) - \frac{\xi_{n-1}}{\zeta_n}\mathbb{E}(j_{n-1}).$$

Using the initial conditions $i_0 = i_1 = j$ and $\mathbb{E}(j_{n-1}) = j$, we obtain (by induction) that

$$\mathbb{E}(i_n) = j \quad \text{for} \quad n = 0, 1, 2, \ldots, m - 1. \tag{47}$$

Hence, combining equation (47) and equation (44) we obtain

$$\mathbb{E}(\nabla C_n^{\mathrm{ctrl}}) = \mathbb{E}(\nabla C_n^{\mathrm{trad}}) = 0 \tag{48}$$

for any $n = 0, 1, \ldots, m-1$, which is the same result as for the traditional approach.

Considering the difference equation (46) and the initial conditions $i_0 = i_1 = j$, we obtain (by induction) the following relationship for the variance of the valuation rate:

$$\mathbb{V}\mathrm{ar}(i_n) = \sigma^2\varphi_n \tag{49}$$

where

$$\varphi_n = \sum_{k=2}^{n} \left(\frac{\xi_{k-1}}{\zeta_k}\right)^2 + \sum_{k_1,k_2} \left(\frac{\xi_{k_1-1}}{\zeta_{k_1}}\right)^2 \left(\frac{\xi_{k_2-1}}{\zeta_{k_2}}\right)^2 + \sum_{k_1,k_2,k_3} \left(\frac{\xi_{k_1-1}}{\zeta_{k_1}}\right)^2 \left(\frac{\xi_{k_2-1}}{\zeta_{k_2}}\right)^2 \left(\frac{\xi_{k_3-1}}{\zeta_{k_3}}\right)^2 + \dots \tag{50}$$

for $k_r = 2, 3, \dots, n$ and $1 + k_r < k_{r+1}$ for $r = 1, 2, \dots$. Substituting equation (49) into equation (45) we obtain

$$\mathbb{V}\mathrm{ar}\left(\nabla C_n^{\mathrm{ctrl}}\right) = (1 + \varphi_n)\xi_n^2\sigma^2. \tag{51}$$

Compare the variances under the traditional and the control approaches in the first sub-period; we would expect to see

$$\mathbb{V}\mathrm{ar}\left(\nabla C_n^{\mathrm{trad}}\right) > \mathbb{V}\mathrm{ar}\left(\nabla C_n^{\mathrm{ctrl}}\right) \tag{52}$$

because $\xi_n < \zeta_{k+1}$ for any $n < m$ and $k < n$.

6 Numerical Example

Consider an employee age 25 who will retire at age 65, i.e., $m = 40$. We assume $T = 36$, i.e., it might possible for the fund manager to find in the market a 4-year guaranteed interest rate deposit account; $i = j = 4\%$, which we assume reflects the level of long-term rates in the market; j_n is log-normally distributed[6] with parameters $\mu = -3.2492$ and $\sigma = 0.2462$ for $n = 1, 2, \dots$, i.e., $\mathbb{E}(j_n) = 4\%$ and $\mathbb{V}\mathrm{ar}(j_n) = 0.0001$.

We perform 3,000 simulations for each of three different values of beta ($\beta = 0.0$, 0.5 , 1.0) and then calculate $\mathbb{E}(C_n^{\mathrm{trad}})$, $\mathbb{E}(C_n^{\mathrm{ctrl}})$, $\mathbb{E}(F_n^{\mathrm{trad}})$, and $\mathbb{E}(F_n^{\mathrm{ctrl}})$, and the standard deviations $\sigma(C_n^{\mathrm{trad}})$, $\sigma(C_n^{\mathrm{ctrl}})$, $\sigma(F_n^{\mathrm{trad}})$, and $\sigma(F_n^{\mathrm{ctrl}})$ for the contribution rate and the fund levels under the traditional and the control approach, respectively. Results are provided in Tables 1 and 2.

[6]The assumption of log-normality for investment returns is a simple though realistic approximation to observations of actual investment rates; see, for example, Baxter and Rennie (1996).

Table 1
Standard Deviations of Contribution Rates Under
Control (C_n^{ctrl}) and Traditional (C_n^{ctrl}) Approaches and Various Values of β

	$\beta = 0.00$		$\beta = 0.50$		$\beta = 1.00$		$\beta = 0.00$	$\beta = 0.50$	$\beta = 1.00$
n	C_n^{ctrl}	C_n^{trad}	C_n^{ctrl}	C_n^{trad}	C_n^{ctrl}	C_n^{trad}	$C_n^{\text{ctrl}}/C_n^{\text{trad}}$	$C_n^{\text{ctrl}}/C_n^{\text{trad}}$	$C_n^{\text{ctrl}}/C_n^{\text{trad}}$
1	0.00	0.01	0.00	0.01	0.00	0.01	0%	0%	0%
2	0.01	0.01	0.01	0.01	0.01	0.01	91%	90%	90%
3	0.02	0.02	0.02	0.02	0.02	0.02	82%	91%	96%
4	0.02	0.03	0.03	0.03	0.03	0.03	75%	87%	97%
5	0.03	0.04	0.04	0.04	0.04	0.04	70%	84%	99%
6	0.04	0.06	0.04	0.06	0.05	0.06	67%	81%	99%
7	0.05	0.07	0.06	0.07	0.07	0.07	64%	78%	98%
8	0.05	0.09	0.07	0.09	0.09	0.09	61%	75%	98%
9	0.06	0.11	0.08	0.11	0.11	0.11	56%	72%	98%
10	0.07	0.13	0.09	0.13	0.13	0.13	54%	70%	97%
11	0.08	0.16	0.10	0.16	0.15	0.15	53%	67%	97%
12	0.09	0.18	0.11	0.18	0.17	0.18	51%	63%	96%
13	0.10	0.22	0.13	0.21	0.20	0.21	47%	62%	96%
14	0.12	0.25	0.14	0.24	0.23	0.24	46%	60%	95%
15	0.13	0.27	0.16	0.27	0.26	0.28	48%	60%	95%
20	0.23	0.49	0.26	0.49	0.46	0.50	46%	53%	91%
25	0.38	0.84	0.45	0.86	0.76	0.87	45%	52%	87%
30	0.71	1.50	0.83	1.56	1.23	1.52	48%	53%	81%
35	1.83	3.14	2.03	3.24	2.11	3.07	58%	63%	69%

Table 2
Standard Deviations (and Expectations for $\beta = 0.50$ only) of Fund Levels Under Control (F_n^{ctrl}) and Traditional (F_n^{ctrl}) Approaches and Various Values of β

	Standard Deviations										
	$\beta = 0.00$		$\beta = 0.50$		$\beta = 1.00$		Ratios $F_n^{\text{ctrl}}/F_n^{\text{trad}}$			$\beta = 0.50$	
n	F_n^{ctrl}	F_n^{trad}	F_n^{ctrl}	F_n^{trad}	F_n^{ctrl}	F_n^{trad}	$\beta = 0.00$	$\beta = 0.50$	$\beta = 1.00$	$\mathbb{E}(F_n^{\text{ctrl}})$	$\mathbb{E}(F_n^{\text{trad}})$
1	0.1	0.1	0.1	0.1	0.1	0.1	100%	100%	100%	0.01	0.01
2	0.2	0.2	0.2	0.2	0.2	0.2	101%	101%	101%	0.02	0.02
3	0.4	0.4	0.4	0.4	0.4	0.4	101%	101%	101%	0.03	0.03
4	0.6	0.6	0.6	0.6	0.6	0.6	101%	101%	101%	0.04	0.04
5	0.8	0.8	0.8	0.8	0.8	0.8	102%	101%	100%	0.06	0.06
6	1.1	1.1	1.1	1.1	1.1	1.1	103%	101%	100%	0.07	0.07
7	1.4	1.4	1.4	1.3	1.4	1.3	104%	101%	100%	0.08	0.08
8	1.7	1.6	1.7	1.6	1.7	1.7	105%	102%	100%	0.10	0.10
9	2.1	2.0	2.0	2.0	2.0	2.0	106%	103%	100%	0.11	0.11
10	2.6	2.4	2.4	2.3	2.3	2.3	107%	103%	100%	0.13	0.13
11	3.0	2.8	2.9	2.8	2.7	2.7	108%	104%	100%	0.14	0.14
12	3.4	3.1	3.2	3.0	3.1	3.1	110%	105%	100%	0.16	0.16
13	4.1	3.7	3.7	3.5	3.6	3.6	111%	106%	100%	0.18	0.18
14	4.7	4.2	4.2	3.9	4.1	4.1	113%	108%	101%	0.19	0.19
15	5.1	4.5	4.7	4.4	4.6	4.5	116%	109%	101%	0.21	0.21
20	8.7	6.9	8.2	7.0	7.2	7.1	125%	117%	102%	0.32	0.32
25	13.5	9.7	12.7	10.0	10.4	10.0	139%	127%	104%	0.44	0.44
30	19.8	12.6	18.7	13.2	14.0	12.8	156%	142%	109%	0.60	0.59
35	26.7	14.5	25.0	15.0	17.3	14.2	184%	167%	122%	0.78	0.78

As we observe (and also anticipated by expressions (13) and (49)) the mean of the contribution rates under the two approaches (controlled or traditional) remain almost constant and close to the initial rate C_{st} (i.e., $\mathbb{E}(C_n^{\text{trad}}) = \mathbb{E}(C_n^{\text{ctrl}}) = 0.01$ when $\beta = 0.50$). As regards the expectations of the contribution rates, we present the simulation results only for the value of $\beta = 0.5$ because the results for the other values of β are almost identical to that of $\beta = 0.5$.

The standard deviation of the contribution rate under the controlled approach exhibits a slightly increasing pattern (as anticipated by expression (51)) but always remains (as anticipated by expression (52)) below the standard deviation of the contribution rate under the traditional approach which exhibits a steeper increasing pattern (as anticipated by expressions (15) and (42)). The proportional difference between the controlled and traditional approach decreases as the β parameter increases toward unity. Actually, under the extreme value of $\beta = 1$ the controlled approach is almost the same as the traditional approach. (See the first columns of Table 1.) With respect to the standard deviations of the contribution rates, we present the results for all the three simulated values of the beta factor ($\beta = 0.0$, 0.5, and 1.0).

Additionally, we also may observe the expectation of the fund level under the two approaches (controlled or traditional) that remains almost the same for any value of n. (See the last two columns of Table 2.) With respect to the mean fund level, we again present the simulation results only for the value of $\beta = 0.5$ because the results for the other values of β are almost identical.

The standard deviation of the fund level under the traditional approach exhibits a slightly increasing pattern but always remains below the standard deviation of the fund level under the controlled approach which exhibits a steeper increasing pattern (the opposite pattern of the contribution rate). The proportional difference between the traditional and controlled approach decreases as the β parameter increases toward unity. Actually, under the extreme value of $\beta = 1$ the controlled approach is almost the same as the traditional approach. (See the first columns of Table 2) As regards the standard deviations of the fund level, we present the results for all the three simulated values of the beta factor ($\beta = 0.0$, 0.5, and 1.0).

It is clear from the results above that both the traditional and the controlled approach succeed in achieving (in expected value terms) the target value of the fund in a very similar way. The controlled approach also succeeds in reducing the variance of the contribution rate but this advantage is balanced with the disadvantage of a higher variance for the fund value. It is also interesting to identify the important role of the weighting factor β which may act as a regulator. The extreme values of the weighing factor β produce two extreme versions of the model.

When $\beta = 0$, we obtain the absolute controlled version of the model with the minimum value for the standard deviation of the contribution rate and the maximum value for the standard deviation of the fund value. When $\beta = 1$, we obtain almost the traditional approach with the maximum value for the standard deviation of the contribution rate and the minimum value for the standard deviation of the fund value. Hence, the choice of β may balance the levels of standard deviations between the fund value and the contribution rate.

7 Summary and Areas for Further Research

The central concept of our paper is the consideration of the valuation rate of interest as a free control variable (as proposed by Benjamin, 1989). This concept may be deemed an attractive one if one wants to determine the actual position of a pension fund. Unfortunately, however, it may pose practical problems with legislative or other regulatory restrictions. The optimal path for the contribution rate (according to our objective function) is then determined by controlling the pattern of the valuation rate of interest through a feedback mechanism. Actually, our model process permits the actuary to adjust the initial valuation rate (which may be based on projections of long-term rates) to reflect the recent investment experience.

The model is solved using a standard linear approximation procedure for the basic equation of the system. The important result is provided by equation (31) where the valuation rate is optimally controlled through a feedback mechanism of the state variable (which is the contribution rate). Under this optimal control law, we observe that the expected contribution rate remains the same (as for the traditional approach) for the whole funding period, while the variance of the contribution rate exhibits a slightly increasing pattern. This increase in variance is less than the increase in variance under the traditional approach. Unfortunately, this advantage of the controlled approach is counterbalanced with the higher fluctuations of its fund levels over time.

It is also interesting to identify the regulatory role of the weighting factor β which under the extreme values $\beta = 0$ and $\beta = 1$ produces the absolute controlled and traditional version of the model. Hence, the specific approach illustrates that the traditional form of the individual aggregate cost method may be seen in a wider context as a special case (for $\beta = 0$) of a controlled model. In this new controlled model, we can design in advance the desired level of the variances for the contribution rate and the fund value by selecting the appropriate value for the weighting factor of the objective function named β.

The model may be extended further by relaxing some assumptions. For example, we can make the pension benefit dependent on final salary and assume fluctuations in the interest rates available to purchase the retirement annuity at the time of retirement.

It is clear from the model investigated above that control theory may be applicable to the individual aggregate cost method by providing a system with an improved performance relative to the traditional version of this specific method.

References

Ang, A. and Sherris, M. "Interest Rate Risk Management: Developments in Interest Rate Term Structure Modeling for Risk Management and Valuation of Interest-Rate-Dependent Cash Flows." *North American Actuarial Journal* 1 (1997): 1-21.

Aoki, M. *Optimization of Stochastic Systems,* 2nd Edition. New York, NY: Academic Press, 1989.

Benjamin, S. "Driving the Pension Fund." *Journal of the Institute of Actuaries* 116 (1989): 717-735.

Baxter, M. and Rennie, A. *Financial Calculus.* Cambridge, United Kingdom: Cambridge University Press, 1996.

Bertsekas, D.P. *Dynamic Programming and Stochastic Control.* Cambridge, United Kingdom: Academic Press, 1976.

Cairns, A. "Some Notes on the Dynamic and Optimal Control of Stochastic Pension Fund Models in Continuous Time." *ASTIN Bulletin* 30 (2000): 19-55.

Haberman, S. and Sung, J.H. "Dynamic Approaches to Pension Funding." *Insurance: Mathematics and Economics* 15 (1994): 151-162.

Loades, D.H. "Elementary Engineering Control and Pension Funding." *Transactions of the 26th International Congress of Actuaries* 5 (1998): 239-266.

McGill, D.M., Brown, K.N., Haley, J.J. and Schieber, S.J. *Fundamentals of Private Pension Funds.* Philadelphia, PA: Pension Research Council, University of Pennsylvania, 1989.

Owadally, M.I. and Haberman, S. "Pension Fund Dynamics and Gains/Losses Due to Random Rates of Return." *North American Actuarial Journal* 3, no. 3 (1999): 105-117.

Owadally, M.I. and Haberman, S. (2004). "Efficient Gain and Loss Amortization and Optimal Funding of Pension Plans." *North American Actuarial Journal* 8, no. 1 (2004): 21-36.

Taylor, G.C. "Stochastic Control of Funding Systems." *Insurance: Mathematics and Economics* 30 (2002): 323-350.

Vandebroek M. "Pension Funding and Optimal Control." *Mitteilungen der Vereinigung Schweizerischer Versichrungsmathematiker* (1990): 313-325.

Vigna, E. and Haberman, S. "Optimal Investment Strategy For Defined Contribution Pension Schemes." *Insurance: Mathematics and Economics* 28 (2001): 233-262.

Wilkie, A.D. "More on a Stochastic Asset Model for Actuarial Use." *British Actuarial Journal* 1 (1995): 777-945..

Winklevoss, H.E. *Pension Mathematics with Numerical Illustrations.* Philadelphia, PA: Pension Research Council, University of Pennsylvania, 1993.

Zimbidis, A. and Haberman, S. "Delay, Feedback and Variability of Pension Contributions and Fund Levels." *Insurance: Mathematics and Economics* 13 (1993): 271-285.

Journal of Actuarial Practice Vol. 12, 2005

Reputation Pricing: A Model for Valuing Future Life Insurance Policies

Rami Yosef*

Abstract†

The reputation of a life insurer is used to develop a model for determining the value of future life insurance policies. An $M/G/\infty$ process is used to describe the sales and terminations (due to death or maturity) of future policies. The intensity of the arrival process is assumed to depend on the company's reputation. Explicit expressions are derived for the actuarial reserves and expected profits of these future policies.

Key words and phrases: *future policyholders, expected profits, expected reserve,* $M/G/\infty$ *queue*

1 Introduction

When investors are interested in purchasing an insurance company, they usually seek an expert appraisal of the value of the company from actuaries, accountants, and other financial professionals. The insurance company may have diverse business interests, including different lines of products sold. As it is common for an insurance company to group similar insurance policies into portfolios (i.e., blocks of policies), the appraised value of the company should reflect the value of each

*Rami Yosef, Ph.D., is a assistant professor of actuarial science at Ben-Gurion University. He obtained his Ph.D. in statistics and actuarial science from the University of Haifa, Israel. In addition to his academic research, Dr. Yosef has provided expert actuarial opinions to various public and private entities in Israel since 1997.

Dr. Yosef's address is: School of Management, Department of Business Administration, Ben-Gurion University of the Negev, Beer-Sheva 84105, ISRAEL. E-mail: *ramiyo@som.bgu.ac.il*

†The author acknowledges the helpful comments of the anonymous referees, the editor, and Shulamith T. Gross.

portfolio[1] including any intangible assets[2] associated with each portfolio.

The value of a life insurance portfolio consists of two components:

(i) The value of the active portfolio, which consists of the life insurance contracts that currently exist and remain active in the portfolio and for which actuarial reserves are not equal to zero. The value of the active portfolio depends on the assets and the aggregate actuarial reserves associated with all of the contracts in the active portfolio.

(ii) The value of the future portfolio is based largely on the insurer's intangible assets: its reputation and its management/marketing strategies for attracting and maintaining new policies. We will assume new policyholders will purchase their insurance from the insurance company based on the strength of the company's reputation. Thus, to determine the value of a future portfolio, assumptions must be made about the insurer's reputation, and its attitude toward new policies.

Economists and accounts long have recognized that one of a firm's intangible assets is its name, or the reputation conveyed by its name. Economists have used game theory to study a firms reputation; see, for example, Kreps and Wilson (1982), Fudenberg and Kreps (1987), Diamond (1989), Fudenberg and Levine (1989), Kreps (1990), Kreps et al. (1992), and Hart (1995). As an example, Kreps (1990) developed a theory of the firm as a bearer of reputation and provides a simple example that demonstrates, using the ideas of the folk theorem in repeated games, how a firm's reputation can become a tradable asset. Game theoretic techniques, however, can be difficult to apply to the problem of valuing the reputation of a life insurance portfolio because of the uncertainties associated with determining the makeup of a future portfolio. Unfortunately, there is no established actuarial theory to assist in valuing an insurer's reputation.[3]

---

[1]In Israel, for example, experts conducting these appraisal valuations most commonly perform a separate evaluation of each portfolio.

[2]A firm's intangible assets or goodwill, which includes the firm's reputation and/or name, are usually hidden in its balance sheet. The actual value of intangible assets is known only when the firm is sold and is obtained by subtracting the value of tangible net assets from the firm's sale price.

[3]In Israel the aggregate actuarial reserves is multiplied by a loading factor to yield a value for a future portfolio that is based on the insurer's reputation. There is no actuarial guidance, however, on how the size of this loading factor is determined. For example, a private investor in Israel recently purchased the successful Israel Phoenix

In the author's opinion, when actuaries or other experts evaluate future (reputation) life insurance portfolios, they can use one of the two approaches described above for their evaluation process. The first approach is to use the insurance company's historical data to project the composition of the portfolio's future insureds. These data contain information on the date of policy inception, age, gender, mortality level, amount of insurance, type of policy, date of exit from portfolio, cause of exit, assumed mortality table, etc. By assuming these data can accurately represent the insureds in the future portfolio, one can anticipate the development of that portfolio by using deterministic methods. The second approach is to use the historical data to develop a stochastic model (such as a queueing model) of the influx and efflux of the portfolio's future insureds. In either approach, one problem will be the choice of mortality table to use. As mortality is continuously improving in most countries, the mortality table used should have built in factors that account for this improvement.

Another aspect mentioned above is the management/marketing strategy employed with respect to new policies. When investors purchase an insurance company, they often continue managing the various insurance portfolios without any restrictions on the sale of policies to future customers. In other words, the investors allow applicants for new life insurance contracts to purchase policies after they have satisfactorily completed the necessary underwriting. This approach, however, may not always be best for the investor. For example, in Israel insurance regulators require that the reserve for a life insurance portfolio be proportional to the number of policyholders insured in the portfolio. If investors do not believe the reserve requirements needed for expanding a portfolio will be available or do not believe it worthwhile to raise this money, then it may be best to restrict the sale of new policies and restrain the growth of the portfolio. In this paper we will assume there are no limits on the number policyholders accepted.

As was mentioned above, there is no established actuarial theory or model for valuing an insurer's reputation. Given an insurer's reputation, however, can we determine the value of one of its future portfolio? In the author's opinion, when actuaries or other experts evaluate future life insurance portfolios, they should use one of two approaches:

1. Use the insurance company's historical data to project the composition of the portfolio's future insureds. These data will contain

Insurance Company. Analysts on Israeli television commented that 35% of the price paid reflected the value of its life insurance portfolio, including the intangible asset based on the Phoenix's reputation.

information on the date of policy inception, age, gender, mortality level, amount of insurance, type of policy, date of exit from portfolio, cause of exit, assumed mortality table, etc. By assuming the data can represent the insureds in the future portfolio accurately, one can anticipate the development of that portfolio by using deterministic methods.

2. Alternatively, use the historical data to develop a stochastic model (such as a queueing model) of the influx and efflux of the portfolio's future insureds.

In either approach, the mortality table used should have built in factors that account for mortality improvement.

The objective of this paper is to present an actuarial model for the evaluation of a future life insurance portfolio. We will propose a dynamic stochastic model of the number of policies in force at any time[4] to describe the evolution of the future life insurance portfolio. The model assumes new policies are issued in a Poisson process and the number of policyholders decreases due to deaths and policy expirations. The rate of new policy issues is assumed to depend on the reputation of the insurer: the better the reputation, the higher the arrival rate. The number of policyholders insured (the in-force process) is allowed to increase without bounds.

Because of the Poisson process assumption, we are implicitly assuming there is an infinite population of potential policyholders. It turns out that our model can efficiently be described as an $M/G/\infty$ queue model where new customers enter the pool of insured parties by a Poisson process (M), each policyholder remains in the portfolio for random period of time that follow a general distribution (G), and the insurance portfolio has infinity capacity (∞). Using this model, we derive an expression for the prospective actuarial reserves of the portfolio t years in the future using each of the two valuation strategies.

2 The Model

Let us consider an insurance portfolio that consists of special fully continuous n-year endowment insurance policies with death benefit B_1

[4]There are not many dynamic models proposed in the actuarial literature. The first one was proposed by Ramsay (1985), who considered a birth-death model of a life insurance portfolio operating in a finite population of potential insureds. Willmot (1990) used techniques from queueing theory to analyze the claim liabilities of an insurance company and provide an example of the application to life insurance portfolio.

and survival benefit B_2, and with premiums paid for h ($h \le n$) years. When this policy is sold to a person age x, the net annual premium, π_x, can be expressed in standard actuarial notation as

$$\pi_x = \frac{B_1\,\overline{A}^{\,1}_{x:\overline{n}|} + B_2\,{}_nE_x}{\overline{a}_{x:\overline{h}|}} \tag{1}$$

for $x = 0, 1, 2, \ldots$. The actuarial functions are calculated using a known standard mortality with survival function ${}_tp_x$. Assuming the policyholder is alive at age $x + t$, the net premium reserve t years after the policy is issued (i.e., at age $x + t$) is $\overline{V}_x(t)$ where

$$\overline{V}_x(t) = \begin{cases} B_1\,\overline{A}^{\,1}_{x+t:\overline{n-t}|} + B_2\,{}_{n-t}E_{x+t} - \pi_x\,\overline{a}_{x+t:\overline{h-t}|} & 0 \le t \le h \\ B_1\,\overline{A}^{\,1}_{x+t:\overline{n-t}|} + B_2\,{}_{n-t}E_{x+t} & h \le t < n \\ B_2 & t = n \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The following assumptions are needed to fully describe our model:

A.1: Each customer who applies for insurance is subject to underwriting (medical and otherwise). If the applicant is deemed insurable, then he or she is sold the special n-year endowment insurance contract described above and becomes a policyholder in the portfolio.

A.2: The mortality of a policyholder age x follows the same known survival function used to determine premiums and reserves, i.e., ${}_tp_x$. Let $T(x)$ be the future lifetime of a typical policyholder age x. Then the time spent in the portfolio is $T_n(x) = \min(T(x), n)$. The cdf of $T_n(x)$ is $G_n(s, x)$ where

$$G_n(s,x) = \mathbb{P}\left[T_n(x) \le s\right] = \begin{cases} {}_sq_x & \text{for } s < n \\ 1 & \text{for } s \ge n. \end{cases} \tag{3}$$

and the resulting survival function is

$$\overline{G}_n(s,x) = 1 - G_n(s,x).$$

A.3: Policyholders leave the portfolio only through death or at the time of the maturity of the policy. There are no policy conversions, lapses, or cancelations.

A.4: The policyholders are mutually independent and indistinguishable, except, possibly, for their age at the issue of their respective policy.

A.5: At $t = 0$, n_x new policies are issued to policyholders age x.

A.6: The future new policyholders age x arrive in the portfolio in a homogeneous Poisson process with rate λ_x.

A.7: The size of λ_x depends on the reputation of the insurer.

A.8: Finally, there are no expenses.

3 The Main Results

Consider a new policyholder age x who joined the portfolio at time y. The net premium reserve at time $t > y$ due to this policyholder, i.e., $(t - y)$ years after joining the portfolio, is $\overline{V}_x(t - y)$. Now suppose that in the time interval $(0, t)$ we are given that k new policyholders arrived in the portfolio with the i^{th} arrival occurring at time $y_{(i)}$, where $0 < y_{(1)} < y_{(2)} < \cdots < y_{(k)} < t$. Then the expected reserve at time t given these k arrivals is

$$\sum_{i=1}^{k} \overline{V}_x(t - y_{(i)})\ {}_{t-y_{(i)}}p_x.$$

The total expected reserve at time t for policies sold to persons age x in $(0, t)$, $R_x(t)$, is thus:

$$\begin{aligned} R_x(t) = \sum_{k=0}^{\infty} \frac{e^{-\lambda_x t}(\lambda_x t)^k}{k!} & \int_{y_{(1)}=0}^{t} \int_{y_{(2)}=y_{(1)}}^{t} \cdots \int_{y_{(k)}=y_{(k-1)}}^{t} \\ & \left[\sum_{i=1}^{k} \overline{V}_x(t - y_{(i)})\ {}_{t-y_{(i)}}p_x\right] f(y_{(1)}, y_{(2)}, \ldots, y_{(k)}) \\ & dy_{(k)} \ldots dy_{(1)}. \end{aligned}$$

From Ross (1996, Theorem 2.3.1), the conditional ordinal arrival times of a homogeneous Poisson process in $(0, t)$, given there are k arrivals, follow the same distribution as that of the order statistics of a random sample of uniform $(0, t)$ variables. Thus, the joint p.d.f. is

$$f(y_{(1)}, y_{(2)}, \ldots, y_{(k)}) = \frac{k!}{t^k} \quad \text{for} \quad 0 < y_{(1)} < y_{(2)} < \ldots < y_{(k)} < t.$$

The multiple integral can now be simplified as follows:

$$\int_{y_{(1)}=0}^{t}\int_{y_{(2)}=y_{(1)}}^{t}\cdots\int_{y_{(k)}=y_{(k-1)}}^{t}\left[\sum_{i=1}^{k}\overline{V}_x(t-y_{(i)})\,{}_{t-y_{(i)}}p_x\right]\frac{k!}{t^k}dy_{(k)}\ldots dy_{(1)}$$
$$=\int_{y_1=0}^{t}\int_{y_2=0}^{t}\cdots\int_{y_k=0}^{t}\left[\sum_{i=1}^{k}\overline{V}_x(t-y_i)\,{}_{t-y_i}p_x\right]\left(\frac{1}{t^k}\right)dy_k\cdots dy_1$$
$$=\frac{k}{t}\int_{y=0}^{t}\overline{V}_x(t-y)\,{}_{t-y}p_x dy.$$

The expression for $R_x(t)$ is now seen to be

$$R_x(t)=\sum_{k=0}^{\infty}\frac{e^{-\lambda_x t}(\lambda_x t)^k}{k!}\frac{k}{t}\int_{y=0}^{t}\overline{V}_x(t-y)\,{}_{t-y}p_x dy$$
$$=\lambda_x\int_{y=0}^{t}\overline{V}_x(t-y)\,{}_{t-y}p_x dy. \tag{4}$$

The total expected reserve at time t for all policies sold $[0,t)$, including those newly in existence at time $t=0$, is

$$R(t)=\sum_x\left[n_x\overline{V}_x(t)+\lambda_x\int_{y=0}^{t}\overline{V}_x(t-y)\,{}_{t-y}p_x dy\right]. \tag{5}$$

The reserve process $R(t)$ represents the liabilities of the insurer to its portfolio of policyholders at time t. This means that an investor who purchases the life insurance portfolio will have a commitment or obligation of amount $R(t)$ to these policyholders. If $A(t)$ represents the amount of assets the portfolio has on hand at time t, then the portfolio's surplus at time t is, $U(t)$, where

$$U(t)=A(t)-R(t). \tag{6}$$

To value the future portfolio we must perform a profit evaluation, which requires knowledge of the future expected rate of profits generated by this portfolio. To this end, we must determine the gross (profit-loaded) premium charged given assumption A.8 (there are no expenses). In practice there are typically three ways to obtain the gross premium rate that allows a profit to the insurer:

1. Use conservative estimates of the various parameters involved in the pricing process. For example, assume a lower interest rate,

higher mortality rates, and lower investment returns than are actually expected. This results in insureds paying a higher premium than they would pay if the best estimates were used.

2. Explicitly specify a profit objective then include that profit as an expense. The gross premium then can be calculated by the actuarial equivalence principle; see, for example, Bowers et al. (1997, Chapter 15). Or,

3. Increase the net premium by a loading factor.

Regardless of the approach used, given the net premium rate π_x and assuming there are no expenses, let $\pi_x^{(g)}$ and $\pi_x^{(p)}$ be the gross premium rate and the expected profit rate, respectively, for policies in the portfolio of policies sold to persons age x, i.e.,

$$\pi_x^{(g)} = \pi_x + \pi_x^{(p)}.$$

To determine the discounted expected portfolio profits we need an expression for the expected number of policyholders expected to be insured at any time t. Let $Q_x(t)$ denote the in-force process, i.e., the number of policyholders who bought their policies at age x at some time y ($y \leq t$) and are still in force at time t with $Q_x(0) = n_x$. (Unlike the models used in traditional risk theory, $Q_x(t)$ is a stochastic (queueing) process.) Thus the expected amount of profits in the time interval $(s, s + \mathrm{d}s)$ generated by the portfolio of policies that were sold to persons age x is $\pi_x^{(p)}\mathbb{E}[Q_x(s) \mid Q_x(0) = n_x]\mathrm{d}s$. If we let $\text{Profit}_x(t)$ be the discounted expected profits in $(0, t)$ from the portfolio of policies that were sold to persons age x, then

$$\text{Profits}_x(t) = \int_0^t \pi_x^{(p)}(1 + i)^{-s}\mathbb{E}[Q_x(s) \mid Q_x(0) = n_x]\mathrm{d}s \tag{7}$$

where i is the valuation rate of interest. The ultimate expected profits from the entire portfolio is

$$\text{Profits} = \sum_x \pi_x^{(p)} \int_0^\infty (1 + i)^{-s}\mathbb{E}[Q_x(s) \mid Q_x(0) = n_x]\mathrm{d}s. \tag{8}$$

Finally, we need an expression for $\mathbb{E}[Q_x(t) \mid Q_x(0) = n_x]$. Clearly $Q_x(t)$ is the number of customers at time t in an $M/G/\infty$ queue with Poisson arrivals at rate λ_x and service time distribution $G_n(s)$ given in equation (3). It is well known (e.g., Ross 1996, p. 70 and Medhi 2003, Chapter 6.10.1) that the distribution of $Q_x(t) \mid Q_x(0) = 0$ is a Poisson distribution with mean $\lambda_x(t)$ given by

$$\lambda_x(t) = \lambda_x \int_0^t \overline{G}_n(s,x)\mathrm{d}s.$$

Thus we can consider an $M/G/\infty$ queue with initial queue length $Q_x(0) = n_x$ as being artificially partitioned into two independent and disjoint sub-queues: a permanently closed queue that consists of the n_x busy servers and a permanently open (but initially empty) $M/G/\infty$ queueing system such that every newly arriving customer can only be served at the open queue. Clearly the number still remaining in the closed queue at time t is binomially distributed with parameters n_x and $\overline{G}_n(t,x)$, which gives a mean of $n_x\overline{G}_n(t,x)$. On the other hand, the expected number of customers at time t in the open queue is $\lambda_x(t)$. Thus the expected number in the queue at time t is $n_x\overline{G}_n(t,x) + \lambda_x(t)$, i.e., we have the following result:

Theorem 1. *If $Q_x(t)$ is the number of customers in an $M/G/\infty$ queue with Poisson arrivals at rate λ_x and independent service times with distribution function $G_n(s)$, then*

$$\mathbb{E}[Q_x(t) \mid Q_x(0) = n_x] = n_x\overline{G}_n(t,x) + \lambda_x \int_0^t \overline{G}_n(s,x)\mathrm{d}s. \qquad (9)$$

4 Summary and Closing Comments

This paper introduces to actuarial pricing a method for evaluating a future life insurance portfolio, which has a growth rate that depends on the reputation of the insurer. When an investor is interested in purchasing such a portfolio, the insurer must be compensated for the reputation of the portfolio. As there is no actuarial theory to assist in valuing an insurer's reputation, the common approach for actuarial practitioners in Israel, for example, is to evaluate the active portfolio which consists of the life insurance contracts that currently exist and remain active in the portfolio. This value is multiplied by a loading factor for which there have been no guidelines for determination.

To correct this state of affairs, we suggest a stochastic model for valuing the future life insurance policies. We specifically use an $M/G/\infty$ process to describe the sales and terminations of future policies and to provide expressions for the total expected reserve and the profit of the future portfolio.

In deriving the expression for the expected reserves in equation (4) we used the marginal distribution of the number of arrivals and the order statistics property of the conditioned arrival times for the homogeneous Poisson process. This technique can be applied to other processes such as the nonhomogeneous Poisson process and the Yule process because they also have the order statistics property; see, for example, Berg and Spizzichino (1999).

Other areas for further research include:

- Considering alternative models. For example, we can consider the case where the insurer intends to limit the size of the portfolio to at most c policyholders so that we have an $M/G/c$ queue with no waiting room. Thus if there are fewer than c policyholders in the portfolio, new contracts are sold (subject to underwriting approval) until there are c policyholders in the portfolio. Once there are c policyholders in the portfolio, then all applications for insurance are denied until there is a death or a policy matures.

- Using stochastic interest rates to determine present values; and

- Use insurance demand function for profit determination. For example, we can assume the demand of insurance decreases as the profit loading increases, i.e., assume that for a given reputation and insurance policy λ_x is a decreasing function of $\pi_x^{(p)}$. This is similar to the work of Kliger and Levikson (1998) and Ramsay (2005).

References

Berg, M. and Spizzichino, F. "Time-lagged Point Processes: The Role Order-Statistics Property." *Mathematical Methods of Operations Research* 51, no. 2 (2000): 301-314.

Diamond, D. "Reputation in Acquisition and Debt Markets." *Journal of Political Economy* 97 (1989): 828-862.

Eick, S. G., Massey, W.A., and Whitt, W. "The Physics of the $M/G/\infty$ Queue." *Operations Research* 41 (1993): 731-742.

Fudenberg , D. and Kreps, D. "Reputation and Simultaneous Opponents." *Review of Economic Studies* 54 (1987): 541-568.

Fudenberg , D. and Levine, D. "Reputation and Equilibrium Selection in Games with a Patient Player." *Econometrica* 57 (1989): 759-778.

Hart, O. *Firms, Contracts and Financial Structure.* Oxford, England: Oxford University Press, 1995.

Kliger, D. and Levikson, B. "Pricing Insurance Contracts-An Economic Viewpoint." *Insurance: Mathematics and Economics* 22 (1998): 243-249.

Kreps, D.M. "Corporate Culture and Economic Theory." In *Perspectives on Positive Political Economy.* (J.E. Alt and K.A. Shepsle, eds.) New York, NY: Cambridge University Press, (1990): 90–143.

Kreps, D.M., Milgrom, P.R., Roberts. J, and Wilson. R.J. "Rational Co-operation in the Finitely Repeated Prisoner's Dilemma." *Journal of Economic Theory* 27, no. 2 (August 1992): 245–252.

Kreps, D.M. and Wilson, R. "Reputation and Imperfect Information." *Journal of Economic Theory* 27 (1982): 253–279.

Medhi, J. *Stochastic Models in Queueing Theory.* New York, NY: Academic Press, 2003.

Ramsay, C.M. "A System of Integro-Differential-Difference Equations in Risk Theory, Using Compound Birth-Death Processes." *Scandinavian Actuarial Journal* (1985): 39–48.

Ramsay, C.M. "Pricing Optional Group Term Insurance: A New Approach Using Reservation Prices." *Insurance: Mathematics and Economics* 36, no. 1 (2005): 37–55.

Ross, S.M. *Stochastic Processes* 2nd Edition. New York, NY: John Wiley and Sons, Inc., 1996.

Serfozo, R.F. "Point Processes." In *Handbooks in OR and MS*, Volume 2. (D.P. Heyman and M.J. Sobel, eds.) Amsterdam, Netherlands: Elsevier Science Publishers, (1990): 1-93.

Whitt, W. "The Renewal-Process Stationary-Excess Operator." *Journal of Applied Probability* 22 (1985): 156–167.

Willmot, G.E. "A Queueing Theoretic Approach to the Analysis of the Claims Payment Process." *Transactions of the Society of Actuaries* (1990): 447-497.

Wolff, R.W. *Stochastic Modeling and Theory of Queues.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1989.

Ultimate Ruin Probability for a Time-Series Risk Model with Dependent Classes of Insurance Business

Lai Mei Wan,* Kam Chuen Yuen,† and Wai Keung Li‡

Abstract§

We consider a discrete-time risk model with m ($m \geq 2$) dependent classes of insurance business. The claim processes of these m classes are assumed to follow a multivariate autoregressive time-series model of order 1. Given this claims model, we explore the probability of ultimate ruin assuming exponentially bounded claims. As an example, we use simulations to study the case where there are two business and the underlying losses are of two types: bivariate exponential and bivariate gamma claim distributions.

Key words and phrases: *adjustment coefficient, bivariate exponential distribution, bivariate gamma distribution, discrete-time risk model, multivariate autoregressive model, time series, ultimate ruin probability*

---

*Lai Mei Wan is a postgraduate student at University of Hong Kong under the supervision of Professors Li and Yuen.

Ms. Wan's address is: Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, HONG KONG. Email address: *kriswan@graduate.hku.hk*

†Kam Chuen Yuen, Ph.D., A.S.A., is an associate professor at University of Hong Kong.

Dr. Yuen's address is: Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, HONG KONG. Email address: *kcyuen@hku.hk*

‡Wai Keung Li is a professor at University of Hong Kong.

Professor Li's address is: Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, HONG KONG. Email address: *hrntlwk@hku.hk*

§The authors are grateful to the referees for their helpful comments and suggestions. This research was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7054/04P). The third author would like to thank the Croucher Foundation for the award of a senior research fellowship.

1 Introduction

For a book of insurance business, it is often assumed that different classes of policies are independent. This assumption, however, may not be justified in many practical situations. For example, a severe car accident may trigger auto-insurance claims as well as medical insurance claims. In recent years, risk models with various dependence structures have been studied by many researchers; for example, Dhaene and Goovaerts (1997), Ambagaspitiya (1998, 1999), Nyrhinen (1998), Wang (1998), Asmussen et al. (1999), Cossette and Marceau (2000), Müller and Pflug (2001), Albrecher and Kantor (2002), Goovaerts and Kaas (2002), Yuen et al. (2002), Picard et al. (2003), and Wu and Yuen (2003).

Actuaries have considered the time-series method as a possible tool to model risk processes. For example, Gerber (1982) investigated the ruin probability by considering the annual gains which form a linear time series. Extensions of his result can be found in Promislow (1991) and Ramsay (1991). Yang and Zhang (2003) studied a risk model with constant interest in which the claim process and the premium process are described by an autoregressive model.

In this paper, we propose a discrete-time risk model with m dependent classes of policies using a time-series approach. Our objective is to investigate the ultimate ruin probability for this model. Specifically, the claim processes of the m classes are described by a multivariate autoregressive model of order 1 (MAR(1)). The MAR(1) model assumes that for each of the m classes, the total claim in a certain period depends not only on the claims occurring in that period, but also on the total claim of its own class and that of other classes in the previous period. Correlation among the claim amounts of the m classes in each period also may be assumed.

Note that Picard et al. (2003) considered a discrete-time model with several interdependent risks in which the claim amounts during successive periods are independent and identically distributed random variables.

The MAR(1) risk model and some basic assumptions are introduced in Section 2. In Section 3, the ultimate ruin probability for the proposed model and its upper bound are investigated. Finally, simulated results in the bivariate case are given in Section 4 to reveal the impact of dependence structure on the ruin probabilities. Some closing comments are given in Section 5.

2 The Model

We now introduce a discrete-time risk model of an insurance portfolio consisting of m dependent classes of insurance policies, where these classes are labeled 1, 2, ..., m. The following assumptions are made:

- Policies are open ended, i.e., they remain in force for an unlimited length of time.
- Each class of policies has its own premiums and claims.
- Premiums are paid at the start of each time period (a period may be a year, quarter, month, etc.) and remain constant throughout the life of the policy.
- The total premium paid in the i^{th} period for the policies in class j is π_j, for $j = 1, 2, \cdots, m$.
- X_{ji} is the total amount of claims **incurred** by the class j policies in the i^{th} period (we only consider exponentially bounded claims).
- We assume that the events causing X_{ji} will cause further claims in the future periods not only in the j^{th} class but also in other classes.
- W_{ji} is the total amount of claims **paid** on behalf of the class j policies in the i^{th} period. It consists of X_{ji} and a linear combination of all the previous claims in all classes (i.e., a linear combination of all X_{hk}s for $h = 1, 2, \cdots, m$ and $k = 1, 2, \cdots, i-1$), and is defined in equation (1).
- If $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{mi})'$ denotes the column vector of the m total incurred claims in period i, we assume that $\{\mathbf{X}_1, \mathbf{X}_2, \cdots\}$ is a sequence of independent and identically distributed non-negative random vectors having finite mean and covariance matrix. And, finally
- If $\mathbf{W}_i = (W_{1i}, W_{2i}, \ldots, W_{mi})'$ denotes the column vector of the m total paid claims in period i, we assume that $\{\mathbf{W}_1, \mathbf{W}_2, \ldots\}$ is a sequence of dependent vectors such that they follow a MAR(1) process, i.e., $\mathbf{W}_i$ is given by

$$\mathbf{W}_i = \mathbf{A}\mathbf{W}_{i-1} + \mathbf{X}_i, \tag{1}$$

where $\mathbf{A}$ is a non-negative constant $m \times m$ matrix. Hence, the components of $\mathbf{W}_i$ are correlated.

The model defined by equation (1) may be useful in describing the dependence of several classes of insurance business in some real situations. For example, a natural disaster or a serious fire accident often causes various types of claims, and some of these claims such as the medical and disability ones may last for many periods of time. The suitability of the MAR(1) model for practical purposes is limited, however, partly because the inherent dependence structure affects the marginal distributions and thus a separate statistical estimation of marginals and the degree of dependence is not possible from the given data.

Let U_n denote the aggregate surplus process of the insurance portfolio at the end of the n^{th} period. As usual, we define the surplus process of class j as

$$U_{jn} = u_j + n\pi_j - \sum_{i=1}^{n} W_{ji},$$

for $n = 1, 2, \ldots$, where u_j is the initial surplus of class j. Thus,

$$U_n = \sum_{j=1}^{m} U_{jn} = u + n\pi - \sum_{i=1}^{n}\sum_{j=1}^{m} W_{ji}, \tag{2}$$

where u and π are the portfolio's aggregate initial reserve and periodic premiums, respectively, i.e.,

$$u = \sum_{j=1}^{m} u_j \quad \text{and} \quad \pi = \sum_{j=1}^{m} \pi_j. \tag{3}$$

For notational convenience, we write $\sum_{j=1}^{m} W_{ji} = \mathbf{1}'_m \mathbf{W}_i$ where $\mathbf{1}_m$ is an m-dimensional column vector of 1.

It is important for the model to be stationary with finite second-order moments. To fulfill this second-order stationarity condition, the eigenvalues of $\mathbf{A}$ must be smaller than 1 in absolute value (see Reinsel 1993). Specifically, all the roots of the characteristic equation of $\mathbf{A}$ (as a function of λ) must be smaller than 1 in absolute value:

$$h(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}) = 0, \tag{4}$$

where $\mathbf{I}$ is an $m \times m$ identity matrix.

Put $\mathbf{X} = \mathbf{X}_1$. Given initial value $\mathbf{W}_0 = \mathbf{w}$, we see from (1)

$$\mathbb{E}(\mathbf{W}_i) = \frac{\mathbf{I} - \mathbf{A}^i}{\mathbf{I} - \mathbf{A}}\mathbb{E}(\mathbf{X}) + \mathbf{A}^i\mathbf{w},$$

which depends on i. Hence, the MAR(1) process is locally non-stationary. As $i \to \infty$, however, its asymptotic mean becomes

$$\mathbb{E}(\mathbf{W}) = \lim_{i\to\infty} \mathbb{E}(\mathbf{W}_i) = (\mathbf{I} - \mathbf{A})^{-1}\mathbb{E}(\mathbf{X}),$$

which is independent of i. Besides, the covariance of W_{ji} and $W_{j,i+l}$ only depends on lag l, where $l = -i, -i+1, \ldots, n-i$, but not on i. Thus, the MAR(1) process is asymptotically stationary. In this paper, the term stationarity is generally used to mean asymptotic stationarity.

The net-profit condition requires that the aggregate premium should be greater than the expected value of the claims in each time period, that is,

$$\pi > \mathbf{1}'_m \left(\frac{\mathbf{I} - \mathbf{A}^i}{\mathbf{I} - \mathbf{A}} \mathbb{E}(\mathbf{X}) + \mathbf{A}^i \mathbf{w} \right),$$

for all i. Here, we assume in the sense of stationarity that

$$\pi > \mathbf{1}'_m \left((\mathbf{I} - \mathbf{A})^{-1} \mathbb{E}(\mathbf{X}) \right), \tag{5}$$

which is a necessary condition for deriving Theorem 1 given below.

3 The Probability of Ultimate Ruin

Let the time of ruin T be the smallest time at which equation (2) becomes negative, i.e.,

$$T = \min\{n : U_n < 0 | U_0 = u\}.$$

Then, the probability of ultimate ruin given the initial surplus u, the aggregate premium per period π, and the initial claim $\mathbf{W}_0 = \mathbf{w}$ is given by

$$\psi(u, \pi, \mathbf{w}) = \Pr(T < \infty | U_0 = u, \pi, \mathbf{W}_0 = \mathbf{w}). \tag{6}$$

In order to prove the main result of the paper, we need to make use of the following modified surplus process. Define

$$\epsilon_i = (\mathbf{1}_m + \alpha)'\mathbf{X}_i,$$

where

$$\begin{aligned} \alpha' &= \mathbf{1}'_m \mathbf{A}(\mathbf{I} - \mathbf{A})^{-1} \\ &= (\alpha_1, \alpha_2, \cdots, \alpha_m). \end{aligned}$$

It is obvious that $\{\epsilon_1, \epsilon_2, \cdots\}$ is a sequence of independent and identically distributed random variables with finite mean and variance. The modified surplus process $\hat{U}_n$ is then defined as

$$\begin{aligned}\hat{U}_n &= U_n - \boldsymbol{\alpha}'\mathbf{W}_n \\ &= U_{n-1} + \pi - (\mathbf{1}_m + \boldsymbol{\alpha})'\mathbf{W}_n \\ &= U_{n-1} + \pi - (\mathbf{1}_m + \boldsymbol{\alpha})'(\mathbf{A}\mathbf{W}_{n-1} + \mathbf{X}_n) \\ &= U_{n-1} + \pi - (\mathbf{1}_m + \boldsymbol{\alpha})'\mathbf{A}\mathbf{W}_{n-1} - (\mathbf{1}_m + \boldsymbol{\alpha})'\mathbf{X}_n,\end{aligned}$$

with $\hat{U}_0 = \hat{u} = u - \boldsymbol{\alpha}'\mathbf{w}$. By the definition of $\boldsymbol{\alpha}$, we have

$$\boldsymbol{\alpha}'\mathbf{I} = (\mathbf{1}_m + \boldsymbol{\alpha})'\mathbf{A}.$$

This together with the definition of ϵ_n allow us to rewrite the modified surplus as

$$\begin{aligned}\hat{U}_n &= U_{n-1} + \pi - \boldsymbol{\alpha}'\mathbf{W}_{n-1} - \epsilon_n \\ &= \hat{U}_{n-1} + \pi - \epsilon_n.\end{aligned} \tag{7}$$

It can be shown that the condition (5) is equivalent to

$$\pi > \mathbb{E}(\epsilon_1). \tag{8}$$

The total premium per period can be expressed as $\pi = (1 + \eta)\mathbb{E}(\epsilon_1)$ where $\eta > 0$ is the relative security loading for the modified surplus process. It is intuitively clear from (7) that ruin is certain if η is negative. We now define the adjustment coefficient R as the smallest positive solution of

$$\mathbb{E}[e^{-R(\pi-\epsilon_1)}] = 1.$$

The adjustment coefficient is assumed to exist for all models considered in this paper.

Theorem 1. *For $u \geq 0$,*

$$\psi(u, \pi, \mathbf{w}) = \frac{e^{-R\hat{u}}}{\mathbb{E}[e^{-R\hat{U}_T} | T < \infty]}. \tag{9}$$

To prove Theorem 1, one can make use of equations (7) and (8), and then follow the proof of the one-dimensional case given in Bowers et al. (1997). It should be pointed out that Theorem 1 only holds for exponentially bounded claims. The following corollary is easily established:

Corollary 1. *Given that equation (9) holds, we have*

$$\psi(u, \pi, \mathrm{w}) \leq \psi^{UB}(u, \pi, \mathrm{w}) = e^{-R\hat{u}}.$$

Proof: As all the α_is of $\boldsymbol{\alpha}$ are non-negative, we have $\hat{U}_T \leq U_T < 0$. Therefore, the denominator on the right hand side of (9) is greater than one. This gives us an upper bound $\psi^{UB}(u, \pi, \mathrm{w})$ for the ultimate ruin probability.

4 Simulation Studies: Models and Results

Simulations are used to study the effect of the time-series modeling and the correlation between the current claim amounts on the ruin probabilities in the bivariate case.

4.1 The Models Used

Four discrete-time risk models are used. For notational convenience, we set $W_i = W_{1i}$, $Z_i = W_{2i}$, $X_i = X_{1i}$, and $Y_i = X_{2i}$.

Model 1:

$$\begin{pmatrix} W_i \\ Z_i \end{pmatrix} = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} W_{i-1} \\ Z_{i-1} \end{pmatrix},$$

where a, b, c, and d are non-zero constants, and (X_i, Y_i) follows a bivariate distribution. In this model, the correlation between W_i and Z_i comes from the AR(1) coefficients as well as the correlation of X_i and Y_i.

Model 2:

$$\begin{pmatrix} W_i \\ Z_i \end{pmatrix} = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} + \begin{pmatrix} a' & 0 \\ 0 & d' \end{pmatrix} \begin{pmatrix} W_{i-1} \\ Z_{i-1} \end{pmatrix},$$

where a' and d' are non-zero constants, and (X_i, Y_i) comes from a bivariate distribution. The correlation between W_i and Z_i is solely due to the correlation of X_i and Y_i.

Model 3:

$$\begin{pmatrix} W_i \\ Z_i \end{pmatrix} = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} W_{i-1} \\ Z_{i-1} \end{pmatrix},$$

where a, b, c, and d are non-zero constants, and X_i and Y_i are independent. The correlation between W_i and Z_i comes solely from the AR(1) coefficients.

Model 4:

$$\begin{pmatrix} W_i \\ Z_i \end{pmatrix} = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} + \begin{pmatrix} a' & 0 \\ 0 & d' \end{pmatrix} \begin{pmatrix} W_{i-1} \\ Z_{i-1} \end{pmatrix},$$

where a' and d' are non-zero constants, and X_i and Y_i are independent. In this model, W_i and Z_i are independent.

In order to obtain a consistent comparison across models, X_i and Y_i are set to have equal mean in each of the four models. In order to do a fair comparison, the parameters a, b, c, d, a', and d' are chosen in the way that the asymptotic means of W and Z, $\mathbb{E}[W]$ and $\mathbb{E}[Z]$, in the four models are equal. Thus, we set

$$a' = 1 - \frac{\mathbb{E}(X)}{\mathbb{E}(W)}, \quad \text{and} \quad d' = 1 - \frac{\mathbb{E}(Y)}{\mathbb{E}(Z)}.$$

In our simulation studies, we consider two bivariate distributions for the two types of claims in Models 1 and 2. One is the bivariate exponential distribution while the other is the bivariate gamma distribution. Hence, the claim amounts of the two classes in Models 3 and 4 are generated from the corresponding marginal distributions.

4.2 Bivariate Exponential Distribution

4.2.1 An Overview

Block and Basu (1974) introduced the so-called absolutely continuous bivariate exponential distribution which possesses the loss of memory property. Here, we simply called it the bivariate exponential distribution. Assume that the claim amounts (X, Y) follow the bivariate exponential distribution. With parameters λ_1, λ_2, $\lambda_{12} > 0$, and $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$, the joint distribution function of (X, Y) is defined as

$$\begin{aligned} F(x,y) &= \frac{\lambda}{\lambda_1+\lambda_2} \exp\left(-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)\right) \\ &\quad - \frac{\lambda_{12}}{\lambda_1+\lambda_2} \exp\left(-\lambda \max(x,y)\right), \end{aligned}$$

for $x, y > 0$. Note that λ_{12} is the key parameter determining the correlation between X and Y and that X and Y are independent when $\lambda_{12} = 0$. Some of the statistical properties derived by Block and Basu (1974) (with minor corrections) are as follows:

$$
\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\lambda_1 + \lambda_{12}} + \frac{\lambda_{12}\lambda_2}{\lambda(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_{12})}, \\
\mathbb{E}(Y) &= \frac{1}{\lambda_2 + \lambda_{12}} + \frac{\lambda_{12}\lambda_1}{\lambda(\lambda_1 + \lambda_2)(\lambda_2 + \lambda_{12})}, \\
\mathbb{V}\mathrm{ar}(X) &= \frac{1}{(\lambda_1 + \lambda_{12})^2} + \frac{\lambda_{12}\lambda_2(2\lambda_1\lambda + \lambda_{12}\lambda_2)}{\lambda^2(\lambda_1 + \lambda_2)^2(\lambda_1 + \lambda_{12})^2}, \\
\mathbb{V}\mathrm{ar}(Y) &= \frac{1}{(\lambda_2 + \lambda_{12})^2} + \frac{\lambda_{12}\lambda_1(2\lambda_2\lambda + \lambda_{12}\lambda_1)}{\lambda^2(\lambda_1 + \lambda_2)^2(\lambda_2 + \lambda_{12})^2}, \\
\mathbb{C}\mathrm{ov}(X,Y) &= \frac{(\lambda_1^2 + \lambda_2^2)\lambda_{12}\lambda + \lambda_1\lambda_2\lambda_{12}^2}{\lambda^2(\lambda_1 + \lambda_2)^2(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12})}, \\
\rho(X,Y) &= \lambda_{12}\Big((\lambda_1^2 + \lambda_2^2)\lambda + \lambda_1\lambda_2\lambda_{12}\Big) \\
&\quad \times \sqrt{\Big((\lambda_1 + \lambda_2)^2(\lambda_1 + \lambda_{12})^2 + \lambda_2(\lambda_2 + 2\lambda_1)\lambda^2\Big)} \\
&\quad \times \sqrt{\Big((\lambda_1 + \lambda_2)^2(\lambda_2 + \lambda_{12})^2 + \lambda_1(\lambda_1 + 2\lambda_2)\lambda^2\Big)},
\end{aligned}
$$

where $\rho(X,Y)$ is the correlation coefficient of X and Y. It is easy to see that X and Y are positively correlated. Block and Basu (1974) also derived some useful properties of the bivariate exponential distribution which allow us to generate (X,Y) easily. The properties include:

1. $\min(X,Y)$ follows exponential distribution with mean λ.
2. The difference $G = X - Y$ has distribution function

$$
F(g) = \begin{cases} \frac{\lambda_1}{\lambda_1+\lambda_2} \exp\left((\lambda_2 + \lambda_{12})g\right), & g \le 0, \\ 1 - \frac{\lambda_2}{\lambda_1+\lambda_2} \exp\left(-(\lambda_1 + \lambda_{12})g\right), & g > 0. \end{cases}
$$

3. $\min(X,Y)$ is independent of G.

Based on these properties, one can generate random variables (X,Y) using the following steps:

Step 1: Generate random variables R_1 and R_2 following uniform $(0,1)$ distribution.

Step 2: If $R_2 < \lambda_1/(\lambda_1 + \lambda_2)$, then $G \le 0$ and

$$
G = \frac{1}{\lambda_2 + \lambda_{12}} \ln\Big(\frac{\lambda_1 + \lambda_2}{\lambda_1} R_2\Big),
$$

else go to Step 5.

Step 3: $X = \frac{-\ln(1-R_1)}{\lambda}$ and $Y = X - G$.

Step 4: Go to Step 1 for a new set of X and Y.

Step 5: $G > 0$ and

$$G = \frac{-1}{\lambda_1 + \lambda_{12}} \ln\Big(\frac{\lambda_1 + \lambda_2}{\lambda_2}(1 - R_2)\Big).$$

Step 6: $Y = (-\ln(1 - R_1))/\lambda$ and $X = G + Y$.

Step 7: Go to Step 1 for a new set of X and Y.

4.2.2 Simulation Results

In our simulation studies, we arbitrarily select $a = 0.4$, $b = 0.2$, $c = 0.2$, and $d = 0.4$ for Models 1 and 3. With these parameter values, the solutions of (4) with $m = 2$ are 0.2 and 0.6. For Models 2 and 4, we set $a' = 0.6$ and $d' = 0.6$. Then, both roots of (4) equal 0.6. Therefore, the stationarity condition is satisfied in each of the four models. The parameters of the bivariate exponential distribution are chosen to be $\lambda_1 = \lambda_2 = 0.070466$ and $\lambda_{12} = 0.38486$. Hence, $\mathbb{E}(X) = \mathbb{E}(Y) = 3$ and the asymptotic means of W and Z are $\mathbb{E}(W) = \mathbb{E}(Z) = 7.5$. The correlation coefficient of X and Y is $\rho(X, Y) = 0.3333$. The asymptotic variances $\mathbb{V}\mathrm{ar}(W)$ and $\mathbb{V}\mathrm{ar}(Z)$, the asymptotic covariance $\mathbb{C}\mathrm{ov}(W, Z)$, and the asymptotic correlation coefficient $\rho(W, Z)$ can be calculated using the standard method for a typical stationary MAR(1) model. Further details about the calculation of these values can be found in Reinsel (1993). Their numerical values are summarized in Table 1 and can serve as indicators for the variances, covariances, and correlation coefficients of W_i and Z_i.

Table 1
Asymptotic Variances, Covariances and Correlation Coefficients for Bivariate Exponential distribution

Model	$\mathbb{V}\mathrm{ar}(W)$	$\mathbb{V}\mathrm{ar}(Z)$	$\mathbb{C}\mathrm{ov}(W, Z)$	$\rho(W, Z)$
1	10.0482	10.0482	5.0238	0.5000
2	11.3043	11.3043	3.7677	0.3333
3	9.4203	9.4203	1.8841	0.2000
4	11.3043	11.3043	0.0000	0.0000

The relative security loading η is set to be 0.05, so the constant total premium per period is $\pi = 15.75$. The initial values are $u = 10$ and $w = z = 0$. The number of simulations used for computing the results is 10,000, and the sample size is 100. We first study the finite-time ruin probability which is defined as

$$\psi_N(u, \pi, w, z) = \Pr(T \leq N | U_0 = u, \pi, W_0 = w, Z_0 = z).$$

The results are shown in Table 2 with standard errors in parentheses. It is observed that as N increases, the finite-time ruin probabilities for the four models increase. As $N \to \infty$, with other parameters fixed, the values of $\psi_N(u, \pi, w, z)$ approach the ultimate ruin probability $\psi(u, \pi, w, z)$.

We also compare the results across the four models with the same length of period. The values of the finite-time ruin probability for Model 1 are greater than those for Model 2 simply because Model 1 has a higher degree of dependence which leads to a higher asymptotic correlation coefficient $\rho(W, Z)$. With the same argument, the finite-time ruin probabilities for Model 3 are greater than those for Model 4. Moreover, the finite-time ruin probabilities for Model 1 and Model 2 are higher than those for Model 3 and Model 4, respectively, because the correlation between X and Y introduces additional dependence in the former two models.

From Table 2, we see that the values of $\psi_N(u, \pi, w, z)$ for $N = 1,800$ and $N = 2,000$ are very close. Therefore, the value of $\psi_N(u, \pi, w, z)$ with $N = 1,800$ can be treated as a good approximation of the ultimate ruin probability in the following numerical studies.

Simulation studies are further carried out to investigate how the ultimate ruin probability is affected by the value of the initial surplus u. For $\pi = 15.75$ and $w = z = 0$, the ultimate ruin probabilities for various values of u are summarized in Table 3. With a larger initial surplus, the approximated values of the ultimate ruin probability become smaller. It is also noticed that as the value of u increases, the standard error for the estimated upper bound also increases. This can be easily explained by the form of the upper bound, that is, $\psi^{UB}(u, \pi, w, z) = \exp(-Ru)$. In words, a small deviation of the simulated R from the mean has a relatively much larger effect on the upper bound with a large value of u. The relation between the relative security loading and the ultimate ruin probability also is examined. Table 4 summarizes the results with $u = 10$ and $w = z = 0$. As η increases, the ultimate ruin probabilities decrease very quickly.

Finally, in view of Theorem 1, we discuss the empirical behavior of the adjustment coefficient R as a function of the model parameters. From Tables 3 and 4, we see that R increases as u or π increases. In general, for a given set of (u, π), ψ^{UB} decreases steadily as the degree of dependence decreases from Model 1 to Model 4, and hence R is not too sensitive to the model change (that is, the change in the degree of dependence). If we fix a model and change the correlation between X and Y, empirical evidence also shows that R decreases in a rather uniform manner as the correlation between X and Y increases.

4.3 Bivariate Gamma Distribution

4.3.1 An Overview

Johnson and Kotz (1972) constructed a multivariate gamma distribution from independent random variables $H_0, H_1, \ldots, H_m$ where H_j follows standard gamma distributions with parameters θ_j ($\theta_j > 0$) for $j = 1, 2, \cdots, m$. Here, we only consider the bivariate case. Let $X = H_0 + H_1$ and $Y = H_0 + H_2$. Then, the claim amounts (X, Y) have a bivariate gamma distribution with joint density

$$f(x, y) = \frac{e^{-(x+y)}}{\Gamma(\theta_0)\Gamma(\theta_1)\Gamma(\theta_2)} \int_0^{\min(x,y)} z^{\theta_0-1}(x-z)^{\theta_1-1}(y-z)^{\theta_2-1} e^z dz,$$

with $\mathbb{E}(X) = \mathbb{V}\mathrm{ar}(X) = \theta_0 + \theta_1$, $\mathbb{E}(Y) = \mathbb{V}\mathrm{ar}(Y) = \theta_0 + \theta_2$, $\mathbb{C}\mathrm{ov}(X, Y) = \theta_0$, and

$$\rho(X, Y) = \frac{\theta_0}{\sqrt{(\theta_0 + \theta_1)(\theta_0 + \theta_2)}}.$$

It is clear that X and Y are positively correlated. Hence, the bivariate gamma random variables (X, Y) can be generated using the following steps:

Step 1: Generate H_0, H_1, and H_2 from standard gamma distributions with means θ_0, θ_1, and θ_2, respectively.

Step 2: $X = H_0 + H_1$ and $Y = H_0 + H_2$.

Step 3: Go to Step 1 for a new set of X and Y.

4.3.2 Simulation Results

Similar to the simulation studies in Section 4.2, we set $a = 0.4$, $b = 0.2$, $c = 0.2$, and $d = 0.4$ for Models 1 and 3 and $a' = d' = 0.6$ for Models 2 and 4. These parameter values imply that all the four models satisfy the stationarity condition. The parameters of the bivariate gamma distribution are arbitrarily selected as $\theta_0 = 1$, $\theta_1 = 2$, and $\theta_2 = 2$ so that the means of X and Y and the asymptotic means of W and Z are the same as those in Section 4.2, that is, $\mathbb{E}(X) = \mathbb{E}(Y) = 3$ and $\mathbb{E}(W) = \mathbb{E}(Z) = 7.5$. The correlation coefficient of X and Y is also 0.3333. The asymptotic variances $\mathbb{V}\mathrm{ar}(W)$ and $\mathbb{V}\mathrm{ar}(Z)$, the asymptotic covariance $\mathbb{C}\mathrm{ov}(W, Z)$, and the asymptotic correlation coefficient $\rho(W, Z)$ are shown in Table 5.

Again, we let $\eta = 0.05$, $\pi = 15.75$, $u = 10$, and $w = z = 0$. The number of simulations and the sample size are also 10,000 and 100, respectively. Table 6 presents the finite-time ruin probability $\psi_N(u, \pi, w, z)$ for various values of N. As expected, the observations made in Section 4.2 from Table 2 also hold in this case. The values in Table 2, however, are generally higher than those in Table 6. It is mainly due to the fact that the asymptotic variances $\mathbb{V}\mathrm{ar}(W)$ and $\mathbb{V}\mathrm{ar}(Z)$ are larger in Section 4.2 (although the asymptotic means and the asymptotic correlation coefficients are the same in both sections).

As shown in Table 6, the finite-time ruin probabilities with $N = 1,000$ and $N = 1,500$ are the same. Therefore, we use $N = 1,000$ to obtain approximations of the ultimate ruin probabilities. Table 7 displays the ultimate ruin probabilities for $u = 5, 10, 15, 20, 25, 30, 50$ with $\pi = 15.75$ and $w = z = 0$ while Table 8 shows the ultimate ruin probabilities for eight values of η with $u = 10$ and $w = z = 0$. Not surprisingly, the patterns of Tables 7 and 8 are more or less parallel to those of Tables 3 and 4, respectively. Also, the empirical behavior of R in this case is similar to that in Section 4.2.

Table 2
Finite-Time Ruin Probabilities $\psi_N(u, \pi, w, z)$ and Upper Bounds for Ultimate Ruin Probabilities $\psi^{UB}(u, \pi, w, z)$ with $u = 10$, $\pi = 15.75$ and $w = z = 0$ for Bivariate Exponential Distribution

	Model 1	Model 2	Model 3	Model 4
N	ψ_N	ψ_N	ψ_N	ψ_N
50	0.3328	0.3187	0.2610	0.2394
	(0.0108)	(0.0146)	(0.0098)	(0.0098)
100	0.4148	0.4007	0.3416	0.3217
	(0.0098)	(0.0143)	(0.0080)	(0.0091)
150	0.4518	0.4362	0.3780	0.3592
	(0.0087)	(0.0138)	(0.0069)	(0.0086)
200	0.4708	0.4558	0.3982	0.3794
	(0.0081)	(0.0135)	(0.0066)	(0.0081)
500	0.5046	0.4918	0.4347	0.4154
	(0.0075)	(0.0128)	(0.0060)	(0.0075)
800	0.5105	0.4982	0.4407	0.4213
	(0.0073)	(0.0127)	(0.0060)	(0.0074)
1,000	0.5118	0.4997	0.4420	0.4226
	(0.0073)	(0.0126)	(0.0060)	(0.0074)
1,200	0.5123	0.5003	0.4426	0.4231
	(0.0073)	(0.0126)	(0.0059)	(0.0074)
1,500	0.5134	0.5011	0.4435	0.4241
	(0.0072)	(0.0125)	(0.0059)	(0.0074)
1,600	0.5136	0.5012	0.4436	0.4242
	(0.0072)	(0.0125)	(0.0059)	(0.0074)
1,800	0.5137	0.5013	0.4437	0.4244
	(0.0072)	(0.0125)	(0.0059)	(0.0074)
2,000	0.5137	0.5014	0.4438	0.4244
	(0.0072)	(0.0125)	(0.0059)	(0.0074)
	ψ^{UB}	ψ^{UB}	ψ^{UB}	ψ^{UB}
	0.8914	0.8833	0.8750	0.8717
	(0.0620)	(0.0723)	(0.0672)	(0.0653)

Table 3
Ultimate Ruin Probabilities $\psi(u, \pi, w, z)$ and their Upper Bounds $\psi^{UB}(u, \pi, w, z)$ with $\pi = 15.75$ and $w = z = 0$ for Bivariate Exponential Distribution

u	Model 1		Model 2		Model 3		Model 4	
	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}
5	0.5452	0.9436	0.5323	0.9391	0.4728	0.9347	0.4524	0.9330
	(0.0070)	(0.0329)	(0.0135)	(0.0387)	(0.0057)	(0.0360)	(0.0078)	(0.0350)
10	0.5137	0.8914	0.5013	0.8833	0.4437	0.8750	0.4244	0.8717
	(0.0072)	(0.0620)	(0.0125)	(0.0723)	(0.0059)	(0.0672)	(0.0074)	(0.0653)
30	0.4057	0.7186	0.3963	0.7030	0.3453	0.6817	0.3297	0.6736
	(0.0075)	(0.1486)	(0.0101)	(0.1682)	(0.0063)	(0.1554)	(0.0074)	(0.1518)
50	0.3210	0.5902	0.3137	0.5735	0.2692	0.5432	0.2564	0.5321
	(0.0072)	(0.2013)	(0.0080)	(0.2237)	(0.0063)	(0.2048)	(0.0067)	(0.2007)
70	0.2541	0.4934	0.2490	0.4786	0.2105	0.4422	0.2002	0.4294
	(0.0066)	(0.2333)	(0.0067)	(0.2561)	(0.0059)	(0.2320)	(0.0058)	(0.2280)
90	0.2013	0.4193	0.1973	0.4077	0.1644	0.3673	0.1563	0.3537
	(0.0057)	(0.2526)	(0.0056)	(0.2751)	(0.0052)	(0.2461)	(0.0050)	(0.2427)
110	0.1596	0.3619	0.1567	0.3538	0.1283	0.3107	0.1218	0.2971
	(0.0057)	(0.2640)	(0.0045)	(0.2862)	(0.0049)	(0.2527)	(0.0046)	(0.2500)
150	0.0998	0.2809	0.0989	0.2792	0.0785	0.2335	0.0742	0.2208
	(0.0043)	(0.2746)	(0.0030)	(0.2956)	(0.0038)	(0.2539)	(0.0033)	(0.2534)
200	0.0556	0.2183	0.0556	0.2228	0.0422	0.1762	0.0401	0.1656
	(0.0028)	(0.2780)	(0.0021)	(0.2970)	(0.0024)	(0.2468)	(0.0023)	(0.2495)

Table 4
Ultimate Ruin Probabilities $\psi(u, \pi, w, z)$ and their Upper Bounds $\psi^{UB}(u, \pi, w, z)$ with $u = 10$ and $w = z = 0$ for Bivariate Exponential Distribution

	Model 1		Model 2		Model 3		Model 4	
η	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}
0.01	0.8488	0.9599	0.8347	0.9482	0.8175	0.9485	0.8111	0.9486
	(0.0024)	(0.0465)	(0.0052)	(0.0574)	(0.0031)	(0.0524)	(0.0025)	(0.0509)
0.03	0.6696	0.9282	0.6565	0.9174	0.6124	0.9139	0.5973	0.9111
	(0.0043)	(0.0575)	(0.0094)	(0.0673)	(0.0042)	(0.0639)	(0.0047)	(0.0608)
0.05	0.5137	0.8914	0.5013	0.8833	0.4437	0.8750	0.4244	0.8717
	(0.0072)	(0.0620)	(0.0125)	(0.0723)	(0.0059)	(0.0672)	(0.0074)	(0.0653)
0.08	0.3488	0.8389	0.3362	0.8314	0.2768	0.8184	0.2584	0.8156
	(0.0099)	(0.0648)	(0.0128)	(0.0712)	(0.0084)	(0.0657)	(0.0093)	(0.0671)
0.1	0.2701	0.8065	0.2593	0.7992	0.2032	0.7840	0.1867	0.7811
	(0.0110)	(0.0643)	(0.0120)	(0.0684)	(0.0093)	(0.0638)	(0.0094)	(0.0658)
0.15	0.1442	0.7355	0.1386	0.7297	0.0962	0.7096	0.0849	0.7062
	(0.0108)	(0.0611)	(0.0092)	(0.0631)	(0.0084)	(0.0593)	(0.0075)	(0.0625)
0.2	0.0782	0.6775	0.0762	0.6728	0.0468	0.6487	0.0400	0.6448
	(0.0095)	(0.0596)	(0.0068)	(0.0601)	(0.0066)	(0.0571)	(0.0053)	(0.0608)
0.3	0.0240	0.5883	0.0249	0.5853	0.0116	0.5546	0.0096	0.5503
	(0.0052)	(0.0595)	(0.0035)	(0.0584)	(0.0029)	(0.0570)	(0.0023)	(0.0606)

Table 5
Asymptotic Variances, Covariances and Correlation Coefficients for Bivariate Gamma Distribution

Model	$\mathbb{V}\mathrm{ar}(W)$	$\mathbb{V}\mathrm{ar}(Z)$	$\mathbb{C}\mathrm{ov}(W,Z)$	$\rho(W,Z)$
1	4.1667	4.1667	2.0833	0.5000
2	4.6875	4.6875	1.5625	0.3333
3	3.9062	3.9062	0.7812	0.2000
4	4.6875	4.6875	0.0000	0.0000

Table 6
Finite-Time Ruin Probabilities $\psi_N(u,\pi,w,z)$ and Upper Bounds for Ultimate Probabilities $\psi^{UB}(u,\pi,w,z)$ with $u = 10$, $\pi = 15.75$ and $w = z = 0$ for Bivariate Gamma Distribution

	Model 1	Model 2	Model 3	Model 4
N	ψ_N	ψ_N	ψ_N	ψ_N
50	0.2174	0.1522	0.0964	0.0720
	(0.0053)	(0.0092)	(0.0039)	(0.0026)
100	0.2724	0.2058	0.1344	0.1077
	(0.0051)	(0.0094)	(0.0042)	(0.0031)
150	0.2931	0.2263	0.1489	0.1218
	(0.0052)	(0.0095)	(0.0044)	(0.0030)
200	0.3024	0.2360	0.1556	0.1286
	(0.0052)	(0.0096)	(0.0044)	(0.0031)
500	0.3140	0.2480	0.1618	0.1346
	(0.0052)	(0.0097)	(0.0045)	(0.0031)
800	0.3149	0.2489	0.1621	0.1349
	(0.0051)	(0.0097)	(0.0045)	(0.0031)
1,000	0.3150)	0.2490	0.1622	0.1349
	(0.0051)	(0.0096)	(0.0045)	(0.0031)
1,500	0.3150	0.2490	0.1622	0.1349
	(0.0051)	(0.0096)	(0.0045)	(0.0031)
	ψ^{UB}	ψ^{UB}	ψ^{UB}	ψ^{UB}
	0.7555	0.7508	0.6997	0.6961
	(0.0997)	(0.0912)	(0.0845)	(0.0907)

Table 7
Ultimate Ruin Probabilities $\psi(u,\pi,w,z)$ and their Upper Bounds $\psi^{UB}(u,\pi,w,z)$ with $\pi = 15.75$ and $w = z = 0$ for Bivariate Gamma Distribution

	Model 1		Model 2		Model 3		Model 4	
u	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}
5	0.3643	0.8673	0.2866	0.8649	0.1963	0.8349	0.1632	0.8325
	(0.0058)	(0.0578)	(0.0108)	(0.0529)	(0.0050)	(0.0505)	(0.0033)	(0.0544)
10	0.3150	0.7555	0.2490	0.7508	0.1622	0.6997	0.1349	0.6961
	(0.0051)	(0.0997)	(0.0096)	(0.0912)	(0.0045)	(0.0845)	(0.0031)	(0.0907)
15	0.2729	0.6610	0.2165	0.6541	0.1340	0.5884	0.1118	0.5845
	(0.0050)	(0.1299)	(0.0083)	(0.1185)	(0.0039)	(0.1065)	(0.0028)	(0.1140)
20	0.2366	0.5807	0.1883	0.5720	0.1110	0.4967	0.0925	0.4928
	(0.0046)	(0.1513)	(0.0073)	(0.1375)	(0.0035)	(0.1199)	(0.0024)	(0.1280)
25	0.2056	0.5123	0.1637	0.5019	0.0921	0.4207	0.0764	0.4172
	(0.0042)	(0.1663)	(0.0063)	(0.1503)	(0.0031)	(0.1270)	(0.0021)	(0.1354)
30	0.1785	0.4538	0.1424	0.4419	0.0763	0.3576	0.0632	0.3546
	(0.0037)	(0.1763)	(0.0053)	(0.1584)	(0.0029)	(0.1298)	(0.0018)	(0.1382)
50	0.1011	0.2901	0.0814	0.2745	0.0360	0.1931	0.0292	0.1924
	(0.0027)	(0.1886)	(0.0032)	(0.1640)	(0.0017)	(0.1184)	(0.0014)	(0.1267)

Table 8

Ultimate Ruin Probabilities $\psi(u,\pi,w,z)$ and their Upper Bounds $\psi^{UB}(u,\pi,w,z)$ with $u = 10$ and $w = z = 0$ Bivariate Gamma Distribution

	Model 1		Model 2		Model 3		Model 4	
η	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}	ψ	ψ^{UB}
0.01	0.7516	0.9173	0.7003	0.9142	0.6482	0.9193	0.6215	0.9103
	(0.0036)	(0.0898)	(0.0078)	(0.0868)	(0.0057)	(0.0806)	(0.0031)	(0.0890)
0.03	0.4988	0.8378	0.4286	0.8341	0.3331	0.8084	0.2992	0.8029
	(0.0055)	(0.1029)	(0.0103)	(0.0988)	(0.0052)	(0.0943)	(0.0035)	(0.1023)
0.05	0.3150	0.7555	0.2490	0.7508	0.1622	0.6997	0.1349	0.6961
	(0.0051)	(0.0997)	(0.0096)	(0.0912)	(0.0045)	(0.0845)	(0.0031)	(0.0907)
0.08	0.1587	0.6491	0.1118	0.6451	0.0560	0.5695	0.0410	0.5676
	(0.0048)	(0.0870)	(0.0063)	(0.0770)	(0.0026)	(0.0716)	(0.0019)	(0.0743)
0.1	0.1012	0.5903	0.0667	0.5868	0.0282	0.5002	0.0186	0.4992
	(0.0039)	(0.0805)	(0.0047)	(0.0700)	(0.0018)	(0.0654)	(0.0013)	(0.0661)
0.15	0.0332	0.4750	0.0193	0.4724	0.0053	0.3701	0.0030	0.3707
	(0.0025)	(0.0695)	(0.0024)	(0.0588)	(0.0008)	(0.0556)	(0.0005)	(0.0523)
0.2	0.0110	0.3913	0.0058	0.3892	0.0010	0.2814	0.0004	0.2831
	(0.0014)	(0.0634)	(0.0011)	(0.0534)	(0.0003)	(0.0506)	(0.0002)	(0.0446)
0.3	0.0014	0.2794	0.0006	0.2781	0.0000	0.1725	0.0000	0.1753
	(0.0004)	(0.0584)	(0.0003)	(0.0508)	(0.0000)	(0.0460)	(0.0000)	(0.0371)

Wu, X. and Yuen, K.C. "A Discrete-Time Risk Model with Interaction Between Classes of Business." *Insurance: Mathematics and Economics* 33 (2003): 117–133.

Yang, H. and Zhang, L. "Martingale Method for Ruin Probability in an Autoregressive Model with Constant Interest Rate." *Probability in the Engineering and Informational Sciences* 17 (2003): 183–198.

Yuen, K.C., Guo, J. and Wu, X. "On a Correlated Aggregate Claims Model with Poisson and Erlang Risk Processes." *Insurance: Mathematics and Economics* 31 (2002): 205–214.

Journal of Actuarial Practice Vol. 12, 2005

Optimal Dividend Strategies: Some Economic Interpretations for the Constant Barrier Case

Maite Mármol,* M. Mercè Claramunt,† and Antonio Alegre‡

Abstract§

We consider the surplus process of a non-life insurance portfolio with a dividend component represented by a constant dividend barrier strategy. The optimal dividend barrier is known when individual claim amounts follow an exponential distribution. This result for the optimal dividend barrier is used to develop combinations of the levels of the insurer's initial surplus and of the barrier which, under certain economic and financial criteria, can be regarded as optimal.

Key words and phrases: *optimal dividend strategy, constant barrier, surplus process with dividends, solvency*

*Maite Mármol, Ph.D., is a professor, licensed actuary, and a member of the Catalan Actuarial Association. Dr. Mármol's research interests are in the areas of financial mathematics and risk theory.

Dr. Mármol's address is: Department of Economical, Financial and Actuarial Mathematics, University of Barcelona, Avda. Diagonal, 690. 08034 Barcelona, SPAIN. E-mail: *mmarmol@ub.edu*

†M. Mercè Claramunt, Ph.D., is a professor, licensed actuary, and a member of the Catalan Actuarial Association. Dr. Claramunt's research interest are in the areas of solvency in insurance, risk theory, and rate-making.

Dr. Claramunt's address is: Department of Economical, Financial and Actuarial Mathematics, University of Barcelona, Avda. Diagonal, 690. 08034 Barcelona, SPAIN. E-mail: *mmclaramunt@ub.edu*

‡Antonio Alegre, Ph.D. is Dean of the Faculty of Economics at the University of Barcelona, licensed actuary, and a member of the Catalan Actuarial Association.

Dr. Alegre's address is: Department of Economical, Financial and Actuarial Mathematics, University of Barcelona, Avda. Diagonal, 690. 08034 Barcelona, SPAIN. E-mail: *aalegre@ub.edu*

§The authors wish to thank the anonymous referees for their helpful comments and suggestions that have improved this paper.

1 Introduction

In the classical compound Poisson model of risk theory, an insurance company's surplus can increase without bounds. This is unrealistic, because the company could reinvest its excess surplus in search of even bigger returns or could simply pay them out as dividends to its shareholders. Thus, to make the classic model more realistic, we should include dividend payments.

The question of how much and when to make dividend payments first was studied by De Finneti (1957). He found that the optimal strategy to maximize the expected sum of the discounted dividends must be a barrier strategy, and he showed how the optimal level of the barrier can be determined. Bühlmann (1970, p. 164) proved that the introduction of a constant barrier in the classical model leads to certain ruin.

The problem of finding the optimal dividend-payment strategy assuming a constant barrier was discussed extensively by several other authors. Gerber (1972, 1979) and Bühlmann (1970) analyze the problem in the context of the classical risk model. The random variable representing the present value of dividends also has been analyzed in the discrete case: Claramunt, Mármol, and Alegre (2003) obtain a general solution of its expectation, and Dickson and Waters (2004) obtain higher order moments in the discrete and continuous case. Recently, authors have modified the risk process by considering a Brownian motion risk model; see, for example, Asmussen and Taskar (1997), Paulsen and Gjessing (1997) who include a stochastic interest on reserves, and Gerber and Shiu (2004) who obtain the moments of the present value of dividends. Other forms of the barriers have been considered, for example a linear barrier was studied by Gerber (1981) and Siegl and Tichy (1999) while a non-linear dividend barrier was first introduced by Alegre, Claramunt, and Mármol (2001) and generalized by Albrecher and Kainhofer (2002).

In this paper we study optimal dividend strategies for a non-life insurance portfolio under a compound Poisson model with a constant barrier. We provide combinations of the levels of the initial surplus and levels of dividend barriers that, under certain economic and financial criteria, can be regarded as optimal. For simplicity we assume the individual claim amounts are independent and identically distributed exponential variables, which makes our analysis easier. The analysis of optimal dividend strategies with other individual claim amount distributions can be performed with simulations or discrete risk models.

The paper is organized as follows: Section 2 gives the main characteristics of the model with a constant barrier. Section 3 contains an analysis of the function of the expected present value of dividends, and in Section 4 the optimal combinations are proposed.

2 The Constant Barrier Model

2.1 The Modified Surplus Process

In the classical model of risk theory, the surplus at a given time t, $U(t)$, is defined as

$$U(t) = u + ct - S(t)$$

for $t > 0$ with $U(0) = u$ being the insurer's initial surplus. The term $S(t)$ represents the aggregate claims in $(0, t)$ modeled as a time homogeneous compound Poisson process with rate λ and μ the expected claim amount. The rate at which the premiums are received is $c = \lambda\mu(1 + \theta)$, where $\theta > 0$ is the security loading.

The imposition of a constant dividend barrier $b \geq u$ modifies the behavior of the surplus process because, when the surplus reaches the level b, all premium incomes are paid out as dividends to shareholders, and the modified surplus process remains at level b until the occurrence of the next claim. The modified surplus process, $\bar{U}(t)$ is given by

$$\bar{U}(t) = U(t) - D(t) \tag{1}$$

where $D(t)$ is the aggregate of dividend payments in the interval $(0, t]$, i.e., for infinitesimally small $\mathrm{d}t$,

$$D(t + \mathrm{d}t) - D(t) = \begin{cases} 0 & \text{if } U(t) < b \\ c\mathrm{d}t & \text{if } U(t) \geq b. \end{cases} \tag{2}$$

Ruin is said to occur at time T if $\bar{U}(T) < 0$ and $\bar{U}(t) \geq 0$ for $t < T$ with the understanding that $T = \infty$ if $\bar{U}(t) \geq 0$ for all $t > 0$.

Figure 1 shows a typical sample path of $U(t)$, $\bar{U}(t)$, and $D(t)$ where T_i for $i = 1, 2, \ldots$ denotes the time of occurrence of the i^{th} claim. Notice that whenever the surplus $U(t)$ reaches b, dividends are paid out to the shareholders with intensity c, and the surplus remains on the barrier until the next claim occurs.

Let $W(u, b)$ denote the expected present value of the discounted aggregate dividend payments up to the moment of ruin T, i.e.,

$$W(u, b) = E\left[\int_0^T e^{-\delta t}\mathrm{d}D(t)\right]$$

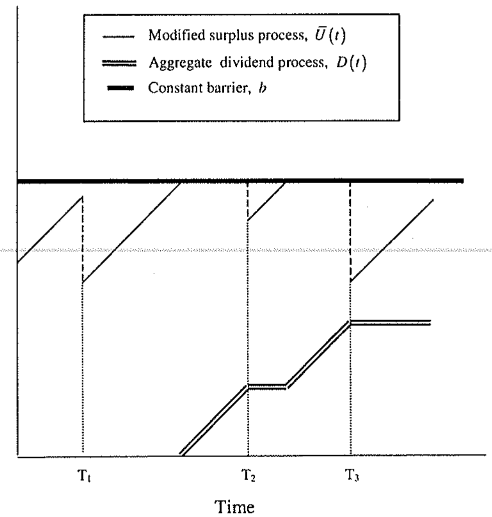

Figure 1: Typical Sample Path of $U(t)$, $\bar{U}(t)$, and $D(t)$

where $\delta \geq 0$ is the force of interest.[1] Bühlmann (1970, p. 173), assuming an exponential distribution for the individual claim amount, obtained an expression for $W(u, b)$. Without loss of generality, we assume the claim size unit is scaled so that the expected claim size is 1. Bühlmann's result can be rewritten as,

$$W(u,b) = \frac{\dfrac{1+r_1}{1+r_2} e^{r_1 u} - e^{r_2 u}}{\dfrac{1+r_1}{1+r_2} r_1 e^{r_1 b} - r_2 e^{r_2 b}} \tag{3}$$

[1]Note that the surplus process is not discounted in order to obtain a tractable model. This is consistent with an economy where the rate of inflation is equal to the rate of return on investment income.

where $r_1 > r_2$ are the roots of

$$\lambda (1+\theta) r^2 - (\delta - \lambda\theta) r - \delta = 0.$$

It is easy to demonstrate $r_1 > 0$ and $r_2 < 0$ and that the following relationships hold between the two roots:

1. $(\delta - \lambda\theta) > 0$ implies $-1 < r_2 < 0 < r_1 < \infty$ and $|r_2| < |r_1|$;
2. $(\delta - \lambda\theta) = 0$ implies $-1 < r_2 < 0 < r_1 < 1$ and $|r_2| = |r_1|$; and
3. $(\delta - \lambda\theta) < 0$ implies $-1 < r_2 < 0 < r_1 < 1$ and $|r_2| > |r_1|$.

Note that $(\lambda\theta - \delta)$ is the difference between the income rate from the security loading and the force of interest used to discount the dividends.

2.2 Some Properties of $W(u, b)$

By considering the other parameters $(\lambda, \theta, \delta, u)$ as fixed, let us find the b^* that maximizes $W(u,b)$. Bühlmann (1970) minimized the denominator of equation (3) to give

$$b^* = \frac{1}{r_2 - r_1} \ln\left(\frac{r_1^2(1+r_1)}{r_2^2(1+r_2)}\right) \qquad -\infty < b^* < \infty. \tag{4}$$

We can observe that b^* doesn't depend on u, so b^* can be less than u and even be negative. When u exceeds b^*, the optimal level of the barrier is $b = u$ (Dickson and Waters, 2004, p. 63).

If $u \le b^*$, then b^* is a maximum point of $W(u,b)$. Interestingly, as b increases to b^* the time that it takes for the surplus to reach the barrier and dividend payments to begin is lengthened; however, $W(u,b)$ increases because the time to ruin is increased thereby allowing dividend payments to be made over a longer period. When b gets beyond b^*, the dividend payments made in the distant future have less impact on the expected present value of the dividends due to the presence of the discount rate.

It is easy to see that

$$\frac{\partial}{\partial u} W(u,b) = \frac{-\frac{1+r_1}{1+r_2} r_1 e^{r_1 u} + r_2 e^{r_2 u}}{-\frac{1+r_1}{1+r_2} r_1 e^{r_1 b} + r_2 e^{r_2 b}} > 0.$$

When $0 \le u \le b$, let u^* denote the optimum value of the initial surplus that maximizes the expected present value of the dividend payments for a given barrier value b, i.e., is $u^* = b$. Thus we must explore the function $W(u^*, b) \equiv W(b, b)$. For convenience we use the notation $W(k) = W(k, k)$, i.e.,

$$W(k) = \frac{-\frac{1+r_1}{1+r_2} e^{r_1 k} + e^{r_2 k}}{-\frac{1+r_1}{1+r_2} r_1 e^{r_1 k} + r_2 e^{r_2 k}}.$$

Note that

$$W(0) = \frac{r_2 - r_1}{r_2(1+r_2) - r_1(1+r_1)} > 0$$

while

$$W'(k) = \frac{\frac{1+r_1}{1+r_2} e^{(r_1+r_2)k} (r_1 - r_2)^2}{\left(-\frac{1+r_1}{1+r_2} r_1 e^{r_1 k} + r_2 e^{r_2 k}\right)^2} > 0,$$

i.e., $W(k)$ is monotonically increasing. The upper bound of $W(k)$ is easily seen to be $W(\infty) = 1/r_1$.

Next we will establish that $W(k)$ has a point of inflection. Let

$$h(k) = -\frac{1+r_1}{1+r_2} e^{r_1 k} + e^{r_2 k}$$

so that $W(k) = h(k)/h'(k)$. Differentiating $W(k)$ twice with respect to k and equating this derivative to zero gives

$$[h'(k)]^2 h''(k) + h(k) h'(k) h'''(k) - 2h(k)[h''(k)]^2 = 0. \tag{5}$$

The solution to equation (5) is $k = k_i$ where

$$k_i = \frac{1}{r_2 - r_1} \ln\left[\left(\frac{-r_1}{r_2}\right)\left(\frac{1+r_1}{1+r_2}\right)\right]. \tag{6}$$

We see that $k = k_i$ is a point of inflection because $W''(k_i) = 0$ and $W'''(k_i) \neq 0$. Thus we have just established the following proposition:

Proposition 1. *For $k > 0$,*

1. *$W(k)$ is positive monotonically increasing;*
2. *$\lim_{k\to\infty} W(k) = 1/r_1$; and*
3. *$W(k)$ has a point of inflection at k_i given in equation (6).*

From equations (6) and (4),

$$k_i - b^* = \frac{1}{r_2 - r_1}\left(\ln \frac{-r_2}{r_1}\right), \tag{7}$$

which leads to the following results:

1. $(\delta - \lambda\theta) > 0$ implies $b^* < k_i < 0$;
2. $(\delta - \lambda\theta) = 0$ implies $b^* = k_i < 0$; and
3. $(\delta - \lambda\theta) < 0$ implies $-\infty < k_i < \infty$ and $k_i < b^*$.

3 Criteria for Choosing k

We now investigate three criteria for choosing k based on: (i) percentiles, (ii) the maximum marginal increase, and (iii) recouping the initial investment.

The Percentile Criterion

From Proposition 1, there exists no value of k that maximizes $W(k)$. As $W(k)$ has an upper limit, $1/r_1$, which is independent of the initial surplus level and the barrier level, an obvious question is what is the value of k that allows us to achieve a specified percentage ($100\alpha\%$) of this limit? Let k_α denote this value, i.e., k_α satisfies $W(k_\alpha) = \alpha/r_1$. It can be proved that

$$k_\alpha = \frac{1}{r_2 - r_1}\ln\left(\frac{(1-\alpha)}{r_1 - \alpha r_2} \times \frac{r_1^2 + r_1}{1 + r_2}\right). \tag{8}$$

In Table 1 we provide some numerical results:

The Maximum Marginal Increase Criterion

Proposition 1 states that the greater the value of the initial reserve and barrier k, the greater the expected present value of the dividends. It is costly, however, for companies to keep increasing the level of k because of the opportunity cost of tying up the company's capital in its surplus. So the question then becomes how large should k be?

Let $M(k)$ denote the marginal rate of increase in the expected present value of the aggregate dividends paid given a barrier at k and initial reserve k, i.e., $M(k) = W'(k)$. This criteria states that investors set k to

Table 1
k_α for Various Values of α
($\delta = 0.03$, $\lambda = 1$, $\theta = 0.5$)

α	k_α	$W(k_\alpha)$
0.1	0.262	1.839
0.2	1.343	3.677
0.3	2.258	5.516
0.4	3.101	7.354
0.5	3.930	9.193
0.6	4.799	11.032
0.7	5.778	12.870
0.8	7.002	14.709
0.9	8.882	16.547

maximize $M(k)$, i.e., k is such that $M'(k) = 0$ and $M''(k) < 0$. In other words the criteria to set $k = k_i$, the point of inflexion of the function $W(k)$, with $W'''(k_i) < 0$. The resulting expression for W is:

$$W(k_i) = \frac{1}{2}\left(\frac{1}{r_1} + \frac{1}{r_2}\right). \tag{9}$$

In Section 2.2 we obtained the values of k_i and b^* according to $(\delta - \lambda\theta)$. If $k_i < b^*$, for $k_i, b^* > 0$, on the combination (k_i, k_i) we can raise the level of the barrier (which involves no extra effort in the initial investment), converting it into (k_i, b^*). We attain a combination that can be regard as optimal for the decision maker, with

$$W(k_i, b^*) = \frac{r_1 + r_2}{r_1 - r_2}\left[\frac{(r_2)^{r_1}}{(-r_1)^{r_2}}\right]^{\frac{1}{r_2 - r_1}}.$$

If $k_i \geq b^*$ we are in a situation in which $k_i, b^* < 0$, and therefore the optimal combination as a function of the values of k_i and b^* is meaningless. We then should focus on the value of k_α , which, fixing the percentage that we consider acceptable to obtain on the maximum of the expected present value of the dividends, α, leads us to choose $u = b = k_\alpha$ as the optimal combination.

Recoup the Initial Investment Criterion

Another way to choose k is for investors to require total recovery of their initial investment of k through future expected dividends, i.e.,

$$W(k) \geq k. \tag{10}$$

Let k_e satisfy the equality $W(k_e) = k_e$. We call k_e the efficiency threshold when the dividends are discounted at a rate δ. At k_e the insurer's rate of return, which we shall represent as $\hat{\theta}$, coincides with the rate δ. It is easy to prove the existence of a unique efficiency threshold, k_e, and that $k > W(k)$ for $k > k_e$ while $k < W(k)$ for $k < k_e$. It follows that the insurer's rate of return is less than δ for $k > k_e$ while it is greater than δ for $k < k_e$. Thus the investors will demand that the insurer set $k < k_e$.

Given that $k < k_e$, it is natural to ask whether there exists a k that maximizes $W(k) - k$. We refer to such a k as k^*, i.e.,

$$k^* = \sup_{0 \leq k \leq k_e} \{W(k) - k\}$$

It is easy to prove that $k^* = b^*$. Thus $k^* > 0$ only when $(\delta - \lambda\theta) < 0$. Under this condition, we therefore can affirm that the optimal value of the expected present value of the dividends according to this criteria is obtained for $k = k^* = b^*$ giving

$$W(k^*) = \frac{r_1 + r_2}{r_1 r_2} = \frac{\lambda\theta}{\delta} - 1.$$

For the case in which $(\delta - \lambda\theta) \geq 0$ leads to $k^* < 0$ and $k_i > k^*$, the maximum difference for $k \geq 0$ will be with a zero initial investment, which is meaningless from an economic standpoint.

Table 2 provides some numerical results as examples of the maximum marginal increase and the recouping of the initial investment criteria presented above. Using $\lambda = 1$, $\delta = 0.03$ and $\delta = 0.05$, and $\theta = 0.2$ and $\theta = 0.5$, we indicate the resulting values of the roots, b^*, the inflection point k_i, the efficiency threshold k_e and the expected present value of dividends for the combinations of u and b.

For example, assuming $\delta = 0.03$ and $\theta = 0.2$, under the maximum marginal increase criteria, we first choose $u = b = k_i = 1.417$, which gives $W(1.417) = 2.833$. Then, without any extra increase in the initial investment, we can raise the level of the barrier in order to increase the expected present value of dividends, $W(1.417, 3.923) = 3.088$. While under the recouping of the initial investment criteria, we have to choose

Table 2
An Example Using the Maximum Marginal Increase and the Recouping of the Initial Investment Criteria with $\lambda = 1$

	$\delta = 0.03$		$\delta = 0.05$	
	$\theta = 0.2$	$\theta = 0.5$	$\theta = 0.2$	$\theta = 0.5$
r_1	0.102	0.054	0.151	0.086
r_2	−0.244	−0.368	−0.276	−0.386
$b^* = k^*$	3.923	7.844	1.740	5.135
k_i	1.417	3.316	0.327	1.963
$W(k_i)$	2.833	7.833	1.500	4.500
$W(k_i, b^*)$	3.088	10.669	1.567	5.595
k_e	8.752	18.350	5.677	11.429
$W(k^*)$	5.667	15.667	3.000	9.000

$u = b^* = k^* = 3.923$ giving the expected present value of dividends as $W(3.923) = 5.667$.

4 Summary

We analyzed the expected present value of dividend payments under a constant dividend barrier, when the aggregate claim amount is assumed to follow a compound Poisson process and the individual claim amount has an exponential distribution. Under these assumptions, we provide some economic/financial criteria for deciding the optimal combination of the initial surplus and the level of the barrier.

An area for further research is to consider other distributions for the individual claim amount, using simulations or discrete approximations. Further research could be done with other models of the risk process such as the Erlang process or Brownian motion (Gerber and Shiu, 2004).

References

Albrecher, H. and Kainhofer, R. "Risk Theory with a Non-Linear Dividend Barrier." *Computing* 68, no. 4 (2002): 289–311.

Alegre, A., Claramunt, M.M., and Mármol, M. "Probabilidad de Ruina y Estrategias de Barrera Bajo un Proceso de Poisson Compuesto." *Cuadernos de Ciencias Economicas y Empresariales* 41 (2001): 75-92.

Asmussen, S. and Taskar, M. "Controlled Diffusion Models for Optimal Dividend Pay-Out." *Insurance: Mathematics and Economics* 20 (1997): 1-15.

Bühlmann, H. *Mathematical Methods in Risk Theory.* Heidelberg, Germany: Springer-Verlag, 1970.

Claramunt, M.M., Mármol, M., and Alegre, A. "A Note on the Expected Present Value of Dividends with a Constant Barrier on the Discrete Time Model." *Mitteilungen der Schweizerischen Aktuarvereinigung* no. 2 (2003): 149-159.

Dickson, D.C.M. and Waters, H.R. "Some Optimal Dividend Problems." *ASTIN Bulletin* 34, no. 1 (2004): 49-71.

de Finetti, B. "Su un Impostazione Alternativa Della Teoria Collectiva del Rischio." *Transactions XV. International Congress of Actuaries* 2 (1957): 433-443.

Gerber, H.U. "Games of Economic Survival with Discrete and Continuous Income Process." *Operations Research* 20, no. 1, (1972): 37-45.

Gerber, H.U. *An Introduction to Mathematical Risk Theory.* Philadelphia, PA: S.S. Huebner Foundation, University of Pennsylvania, 1979

Gerber, H.U. *On the Probability of Ruin in the Presence of a Linear Dividend Barrier.* Scandinavian Actuarial Journal (1981): 105-115.

Gerber, H.U. and Shiu, E.S.W. "Optimal Dividends: Analysis with Brownian Motion." *North American Actuarial Journal* 8, no. 1, (2004): 1-20.

Paulsen, J. and Gjessing, H.K. "Optimal Choice of Dividend Barriers for a Risk Process with Stochastic Return on Investments." *Insurance: Mathematics and Economics* 20 (1997): 215-223.

Siegl, T. and Tichy, F. "A Process with Stochastic Claim Frequency and a Linear Dividend Barrier." *Insurance: Mathematics and Economics* 24 (1999): 51-65.