Volume 28, No. 1 May 1998

ASTIN BULLETIN

A Journal of the International Actuarial Association

CONTENTS

Ceuterick

EDITORIAL POLICY

ASTIN BULLETIN started in 1958 as a journal providing an outlet for actuarial studies in non-life insurance Since then a well-established non-life methodology has resulted, which is also applicable to other fields of insurance For that reason *ASTIN BULLETIN* has always published papers written from any quantitative point of view – whether actuarial, econometric, engineering, mathematical, statistical, etc – attacking theoretical and applied problems in any field faced with elements of insurance and risk Since the foundation of the AFIR section of IAA, i e since 1988, *ASTIN BULLETIN* has opened its editorial policy to include any papers dealing with financial risk

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

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

MEMBERSHIP

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

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

SUBSCRIPTION AND BACK ISSUES

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

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

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

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

INDEX TO VOLUMES 1-27

The Cumulative Index to Volumes 1-27 is also published for ASTIN by Ceuterick at the above address and is available for the price of BEF 400

EDITORIAL

It is now almost 2 years since I took over from David Wilkie as a co-editor of ASTIN Bulletin. It was indeed a great honour for me to have been appointed to the editorial committee and certainly it is my intention to work hard to maintain and enhance the excellent reputation that ASTIN Bulletin has.

As with my predecessor, my primary responsibility lies on the AFIR side of the journal. It is now, I think, almost 10 years since AFIR was formed. As a consequence the scope of ASTIN Bulletin was broadened to encompass not just non-life research but also actuarial problems dealing with financial risk. As most readers of ASTIN Bulletin will realise we still have a long way to go before we get a good balance in the journal between ASTIN-type and AFIR-type papers

Perhaps this reflects how well the AFIR colloquia have been organised in the past and no doubt the future also. In particular, the success of the meeting depends upon the papers presented and, in consequence, the proceedings produced for the meeting. For many authors the publication of a paper in a colloquium proceedings is a satisfactory endpoint. I would argue that this is not good enough! For every one AFIR member who attends an AFIR colloquium there are ten who do not. It is important that we cater for these members by ensuring that the best of the colloquia papers get through to ASTIN Bulletin. What, then, do I regard as a paper in the core of AFIR? AFIR translates as Actuarial Approach to Financial Risks. Here there is something of a two way flow. On the one hand actuaries have the ability to apply well known actuarial methods to purely financial problems On the other hand actuaries also need to import the best of financial economics into the traditional actuarial problems of risk management (for example, of an insurance company) I would say that this flow of ideas is essential for us to maintain our position as the leaders in this field. Thus actuaries already active in this field need to take on board, and adapt as appropriate, financial economic theory Furthermore our systems of education will also need to adapt to equip the actuaries of tomorrow with the necessary tools to cope with tomorrow's problems. Those who insist that we already have the tools will be left behind

There are a number of areas which I would like to see flourish within the pages of ASTIN Bulletin. asset-liability modelling; securitization of insurance risks; models for long-term financial risk analysis; value at risk, to name but a few. However, I would like to concentrate here on the need for papers which work towards a reunification of the financial economic and traditional actuarial theories. I use the word "reunification" here intentionally, since it is only in the last 20 to 30 years that financial economics (as it might be applied to actuarial problems) has split off and become a major field of study in its own right In the process actuaries were left behind, the majority preferring to stick with their tried-and-tested tools. Before that actuaries could be regarded as being as much at the forefront of financial economic thought as any other group. Indeed, recently I found in one of the earliest volumes of the Journal of the Institute of Actuaries (1855) a paper proving a now-well-known result in stochastic interest. If that doesn't prove that we were once at the forefront of financial economic thought I don't know what else could.

Over the last few years I have watched and become involved in some heated debates over which approach is the right one In my view both approaches are correct and that they are compatible Differences of opinion arise because of misconceptions about what the other approach is attempting to do. On the one hand we have problems which require a fair value or price to be put on a set of liabilities (for example, in setting a premium rate or in defining the liabilities which appear in company accounts). In my view the financial economic approach here is the right one. On the other hand we have, for example, problems of reserving A reserve may be some sort of anticipated present value of future net cashflows often calculated along deterministic lines However, reserves may be based on more sound stochastic principles. For example, reserves may be calculated according to the principles of value at risk. This means determining the level of reserve which will have a 95% probability, say, of being sufficient to take care of the future net cashflows as they arise when these are subject to uncertainty (such as stochastic liabilities and assets, parameter uncertainty and model risk). It is immediately possible to tie the two approaches together by describing a value-at-risk reserve as the fair or market value of the future net cashflows plus a contingency margin for future uncertainty.

Papers which do attempt to pull these approaches together are starting to appear and I very much hope that their authors will choose ASTIN Bulletin as the right home for their work.

ANDREW CAIRNS

THANKS

Attentive readers will have noticed that Harry Reid's name no longer appears on the front cover of ASTIN Bulletin This follows Harry's retirement from the editorship of the journal after fifteen years Harry provided a very valuable link with the industry. As a testimony to his achievements we only need to refer readers to a recent survey by Colquitt (1997) on the significance of actuarial journals in which ASTIN Bulletin ranks very high.

On behalf of the readership we would like to wish Harry a long and happy retirement

All members of ASTIN and AFIR will, by now, have received the cumulative index for volumes 1 to 27 of ASTIN Bulletin. This index would not have been produced without the hard work of Marc Goovaerts and his colleagues for which we are extremely grateful

---

L. L. Colquitt (1977) Relative significance of insurance and actuarial journals and articles: a citation analysis. Journal of Risks and Insurance 64· 505-527.

HEDGING IN FINANCIAL MARKETS[1]

By Martin Baxter

Statistical Laboratory, Cambridge University

Abstract

This (mostly) expository paper describes the importance of hedging to the pricing of modern financial products and how hedging may be achieved even when the traditional Black-Scholes assumptions are absent

Keywords

Derivatives; hedging; option-pricing, superhedging; volatility

1 OVERVIEW

Any market practitioner who sells derivatives on his own account will say that hedging is the key to pricing If a contract is not hedged, one can sell it at any price, even the right one, and still lose money. The price of the contract must be the cost of the hedge, plus margin, and the profit/loss of the deal will depend crucially on the hedge being effective

From the earliest days of the rigorous literature, such as Harrison and Pliska (1981), hedging has been used to derive prices in the absence of arbitrage. Text books for practitioners, such as Chapter 14 of Hull (1997) and Baxter and Rennie (1996) stress the centrality of hedging to securities trading The essence of the case being that hedging allows the derivative writer to minimise his exposure to market risk without reducing his profit, thus allowing him, in the words of one banker, 'to quote a price with a view of making a profit through his intermediation rather than by taking a directional view' (Bogni, 1997).

Hedging may be performed on a wide variety of markets for arbitrary derivative products. In simple cases, an option is hedged by trading in the underlying security (stock, currency or bond), but it is equally possible to construct a hedge for a derivative in terms of simpler derivatives, such as forwards and calls Dupire (1993) has done interesting work in developing this area of option-hedging, which we will study in section 5

On the other hand, it might be attractive to work with a model in which it is not possible to hedge. Such incomplete markets can appear intractable, but, following work by El Karoui et al (1996) are now amenable to the tool of

[1] This paper was delivered at a meeting on 'Financial Mathematics and Derivatives' at the International Centre for Mathematical Sciences in Edinburgh on 21 January 1997

ASTIN BULLETIN, Vol 28 No 1 1998, pp 5-16

superhedging. This technique produces a strategy that dominates the option payoff That is, the hedge will produce at least as much as the contract requires, and may produce a surplus Eliminating down-side market risk (in theory at least) is achieved at the expense of the loss of two-way pricing Finally, the superhedging of derivatives using other options gives even better results, and creates an elegant duality between the option-hedging and the superhedging approaches

2 STATIC HEDGING

We begin with the simplest case

Consider the contract to forward purchase at time T one unit of a stock S for a pre-set price k. Imagine that interest rates are constant at a (continuously compounded) rate r and there are no transaction costs payable nor dividends due from the stock At what price k should we sell the forward contract?

At time T, the contract has the (now certain) value of

$$X = S_T - k,$$

so we might expect its time-zero discounted worth to be

$$\mathbb{E}\left(e^{-rT}X\right) = e^{-rT}\mathbb{E}(S_T) - ke^{-rT},$$

and then the price k required to give the contract nil net present value would be $k = \mathbb{E}(S_T)$ More generally, we might discount equities at a different rate, u, than the cash discount rate r. In that case, the appropriate forward price would be $k = e^{-(u-r)T}\mathbb{E}(S_T)$. Either way, this seems to make some sense if S_T is expected to be large, the forward price should be correspondingly large Paradoxically however, this price is wrong

The actual forward price, in this model, is $k = e^{rT}S_0$ That is, the price is just the current stock price S_0 scaled up by the time value of money over the period The price does not depend at all on whether S_T is expected to be high or low The reason for this is a hedge The contract X can be hedged if we.

- buy one unit of stock for price S_0, and
- borrow ke^{-rT} units of cash

This has initial cost $S_0 - ke^{-rT}$ By time T, the stock has evolved to be worth S_T and the debt has grown to $-k$, giving exactly the same net worth as the forward X. So the initial worth of X is the initial cost of the hedge, which is zero only if $k = e^{rT}S_0$

The hedge is essentially to buy one unit of the stock and wait, so that it is ready to be handed over at time T We are unconcerned whether the stock price rises or falls, or indeed whether it is valued 'correctly' at either time 0 or time T. It is enough for us to have it, because we are now unexposed to market risk, in the form of stock price movements

This example demonstrates a static hedge, which can be put on at the start of the contract and left unchanged till the end

Example: forward borrowing The interest-rate market can be described through the behaviour of zero-coupon discount bonds The T-bond pays one unit of cash at time T, and at time t before then has a value (typically) less than 1, written $P(t, T)$ This allows us to lend £ 1 to a customer from time zero to time T by selling $P^{-1}(0, T)$ units of the T-bond into the market now for a price of £ 1, which we loan to the customer At time T, the customer pays us back $P^{-1}(0, T)$ which we use to meet our maturing T-bond liability We could also accept term deposits from the customer, by changing all the signs and buying T-bonds instead.

Our customer may wish instead to borrow later (from time S to time T), but agree on the price now, at time zero Suppose he wants to borrow £ 1 at time S How much should we demand back at time T?

The answer is, we should get back $P(0, S)/P(0,T)$ and here is the hedge

- sell $P(0, S)/P(0,T)$ units of T-bond, and receive $P(0, S)$ now, and
- buy one unit of S-bond, for cost $P(0, S)$ now.

These initial transactions have zero net cost At time S, we receive £ 1 from our S-bond which we can loan to the customer as agreed. At time T, we receive $P(0, S)/P(0,T)$ from the customer which exactly cancels our maturing T-bond liability

In other words, the forward price to sell the T-bond at time S is

$$F = \frac{P(0, T)}{P(0, S)}$$

Away from the special case of forwards, static hedging can still be beneficial, even if it is not perfect. For instance, a static hedge to approximate a claim X can be made by holding ϕ units of stock and ψ units of the cash bond The expected square error of this hedge (to choose a simple loss function), is

$$\mathbb{E}\left((X - \phi S_T - \psi e^{rT})^2\right)$$

We can minimise this, to begin with, over the cash holding, with the optimal choice of ψ being $\psi = e^{-rT}\mathbb{E}(X - \phi S_T)$, and the minimal value being $E(\phi) = \text{Var}(X - \phi S_T)$ This itself can now be minimised over ϕ at the value

$$\phi = \frac{\text{Cov}(X, S_T)}{\text{Var}(S_T)},$$

with value $E(\phi) = \text{Var(X)}\left(1 - \rho^2\right)$, where ρ is the correlation between X and S_T.

Example In the particular case where S_T is normally distributed as a $N\left(\mu, \sigma^2\right)$ and X is the call payoff $X = (S_T - \mu)^+$, then the optimal $\phi = \frac{1}{2}$, and

$$E(\phi) = \frac{\pi - 2}{2\pi - 2} E(0),$$

a reduction in the error variance of over 73%

3. SIMPLE HEDGING

In a sense, forwards are a special case and their hedge has been known for a long time. In fact, any payoff which is a linear function of the stock price has an exact static hedge. The hedge for the claim $X = aS_T + b$, where a and b are constants, is to hold a units of stock and $e^{-rT}b$ units of cash, with initial value $aS_0 + e^{-rT}b$.

This exact answer for simple claims in general markets also holds for general claims in simple markets For instance, take the single-period market with zero interest rates (so there is a constant cash bond $B_t = 1$) and one risky asset S_t. The stock evolves as shown in figure 1

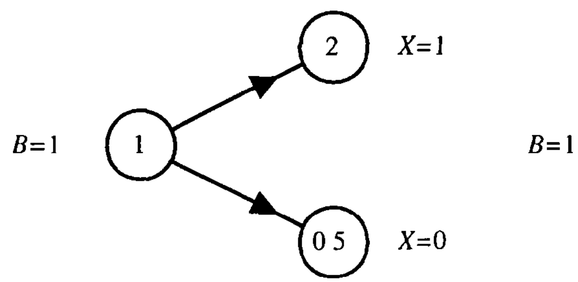

FIGURE 1 Single-period securities market

The stock price either doubles or halves, and cash stays constant at 1 A call option on S_1, struck at £1, with time 1 value of $X = (S_1 - 1)^+$, will pay off £1 if the stock goes up and nothing if it goes down. The paucity of possible values for S_1 enables us to write X as

$$X = \frac{2}{3}S_1 - \frac{1}{3}$$

That is, X has the same payoff as a forward to buy 2/3 units of stock at the price of £0 50 per unit Applying our methods of section 2, we see that the time zero price for X is

$$V = \frac{2}{3}S_0 - \frac{1}{3} = \frac{1}{3}$$

So the price of X is actually 1/3 and the hedge is to

- buy 2/3 units of stock for cost 2/3
- borrow an additional 1/3 units of cash,

which has initial cost of 1/3 and terminal value of X

We could have performed this calculation for any claim X which paid x_u after an up-jump and x_d after a down-jump Such a claim would be worth

$$V = \frac{1}{3}x_u + \frac{2}{3}x_d$$

This has the form of the expected value of X under a probability measure which assigns 1/3 chance to an up-jump and 2/3 chance to a down-jump. This *hedging measure* $(q, 1-q)$ is given by the formula

$$q = \frac{S_0 - s_d}{s_u - s_d}, \text{ or } q = \frac{e^{r\delta t}S_0 - s_d}{s_u - s_d} \text{ if } r \neq 0,$$

where S_1 takes the value s_u after an up-jump, and s_d after a down-jump. To see why this actually is an expectation, see Chapter 2 of Baxter and Rennie (1996)

Although the model is very simple, it can be used as a basic building block of more complex models We can combine many individual branches into a tree (figure 2).

It just takes 10 layers in this tree to produce a final layer containing over 1000 nodes Options can still be priced by working back recursively through the tree from the final layer. See, for example, Chapter 15 of Hull (1997) or Chapter 2 of Baxter and Rennie (1996).

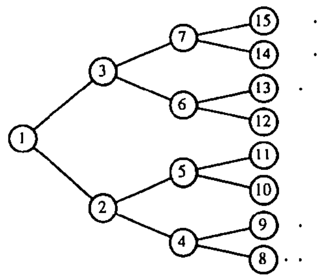

FIGURE 2 Binomial tree

4. BLACK-SCHOLES

The simplest continuous-time model for a stock price is the Black-Scholes model,

$$S_t = S_0 \exp(\sigma W_t + \mu t),$$

where W_t is a Brownian motion, and σ and μ are constants In this model, $\log(S_t/S_0)$ is normally distributed with variance $\sigma^2 t$ and mean μt The variable σ is called the *volatility* of the process

We can also see S as the limit of discrete trees, as in section 3, with current value S_0 evolving to

$$S_{\delta t} = \begin{cases} S_0 \exp(\sigma\sqrt{\delta t} + \mu\delta t) & \text{if up-jump,} \\ S_0 \exp(-\sigma\sqrt{\delta t} + \mu\delta t) & \text{if down-jump} \end{cases}$$

The time increment δt over one step of the tree is going to get ever smaller. Then $\log(S_t/S_0)$ is equal to $\mu t + \sigma\sqrt{\delta t}X(t/\delta t)$, where X_n is a simple symmetric random walk, with X_n distributed as a function of a binomial, 2 Bin(n, 1/2) − n, with zero mean and variance n. By the Central Limit theorem, the distribution of $\log(S_t/S_0)$ converges to that of the Black-Scholes, namely $N(\mu t, \sigma^2 t)$

Choosing q to be the hedging measure

$$q = \frac{e^{r\delta t} - e^{-\sigma\sqrt{\delta t}+\mu\delta t}}{e^{\sigma\sqrt{\delta t}+\mu\delta t} - e^{-\sigma\sqrt{\delta t}+\mu\delta t}} \sim \frac{1}{2}\left(1 - \sqrt{\delta t}\left(\frac{\mu + \frac{1}{2}\sigma^2 - r}{\sigma}\right)\right),$$

then now $\log(S_t/S_0)$ is asymptotically distributed as a normal $N\left(\left(r - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right)$ The hedging price at time zero of any claim X payable at time T will then be $\mathbb{E}_{\mathbb{Q}}(e^{-rT}X)$, where the behaviour of X under $\mathbb{Q}$ is governed by the new asymptotic normal distribution Evaluating this for the European call claim $X = (S_T - k)^+$ gives rise to the celebrated Black and Scholes (1973) call option pricing formula

$$V_0 = e^{-rT}\left(F\Phi\left(\frac{\log\frac{F}{k} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right) - k\Phi\left(\frac{\log\frac{F}{k} - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right)\right),$$

where F is the forward price $F = S_0 e^{rT}$, and Φ is the normal distribution function

Again we do not use this price because it is an expected value of the claim, but because this is the value which lets us hedge In this case the hedge we need at time 0 is $\frac{\partial V}{\partial S}$ For a general option X we can also price and hedge in the same way

In fact the Black-Scholes formula is not just true for the Black-Scholes model. It is enough that the stock S_T and cash bond B_T are jointly log-normally distributed under the hedging measure $\mathbb{Q}$. The formula will then still hold with $\sigma^2 T$ replaced by $\mathrm{Var}(\log(S_T/S_0))$, e^{-rT} replaced by $\mathbb{E}_{\mathbb{Q}}(B_T^{-1})$, and the forward price F equal to $F = S_0/\mathbb{E}_{\mathbb{Q}}(B_T^{-1})$

For a good introduction to Black-Scholes from the actuarial point of view, see the comprehensive review paper by Kemp (1996)

Also Hobson (1996a) reviews the extensions possible from the constant volatility assumptions of the basic Black-Scholes model That paper describes hedging in a stochastic volatility framework, as well as considering discrete-time ARCH and GARCH models, and provides a good introduction to the more advanced techniques

5. OPTION-HEDGING

The above formula does depends crucially on some aspects of the Black-Scholes model, namely that

- volatility is constant (or at least deterministic)
- the market generated by the asset is complete

In practice, these can not be relied upon One solution is to recognise that, say, vanilla call options are so frequently traded as to be liquid assets in their own right As such, they are not priced *per se* by the Black-Scholes formula, but they can themselves be used in hedges to price more complicated products This is, to coin a phrase, option-hedging using options as a hedge for other derivatives, as opposed to the classical hedging of options

As an example, suppose we have a traded stock S_t and traded call options, where $C_t(T, y)$ is the time t price of a call on S_T struck at y. For simplicity, take interest rates to be zero, so that

$$C_t(T, y) = \mathbb{E}_{\mathbb{Q}}((S_T - y)^+ | \mathcal{F}_t)$$

A particular case, originally due to Breeden and Litzenberger (1978), is that of a terminal value payoff $X = f(S_T)$, for some twice-differentiable function f. A simple variant of Taylor's theorem says that

$$f(x) = f(0) + xf'(0) + \int_0^\infty (x-y)^+ f''(y)dy, \quad \text{for al } x \geq 0,$$

which can be proved by integrating by parts. Substituting S_T for x and taking expectations under $\mathbb{Q}$ gives the time t value of the option, V_t, as

$$V_t = f(0) + S_t f'(0) + \int_0^\infty C_t(T, y) f''(y)dy$$

We have calculated not only the price for X, but also a static hedge which is

- hold $f(0)$ units of cash,
- hold $f'(0)$ units of stock, and
- hold $f''(y)dy$ units of the call struck at y.

(In practice, some approximation to the continuous density $f''(y)\,dy$ will be required) We have already seen how linear terms can be statically hedged like a forward Now we see the convex terms being hedged with options.

If interest rates were non-zero, the formula still holds with the single change that we hold $e^{-rT}f(0)$ units of the cash bond, which is worth $e^{-r(T-t)}f(0)$ at time t.

Note that this does not price all options, such as lookbacks or exotics (For example, a put at the maximum price attained by the stock, $X = \sup_{t \leq T} S_t - S_T$, or an down-and-out call which only pays off if the stock never went below a pre-set threshold, $X = (S_T - k)^+ I(\inf_{t \leq T} S_t > c)$.)

The formula's advantages are not only that it is a static hedge, but also that it is completely model independent we make no assumptions about how S_t or $C_t(T, k)$ evolve, or even whether the market is complete But we can still hedge

This example is actually evidence of a deeper idea of Dupire (1993) Given the option prices

$$C_t(T, y) = \mathbb{E}_{\mathbb{Q}}\big((S_T - y)^+ | \mathcal{F}_t\big),$$

their partial derivative with respect to y is

$$\frac{\partial}{\partial y} C_t(T, y) = -\mathbb{Q}\big(S_T > y | \mathcal{F}_t\big),$$

and differentiating once more gives

$$\frac{\partial^2}{\partial y^2} C_t(T, y)dy = \mathbb{Q}(S_T \in dy | \mathcal{F}_t),$$

which is the marginal density of S_T given $\mathcal{F}_t$ Then the time t value of any claim $X = f(S_T)$ will be

$$\int_0^\infty \mathbb{Q}(S_T \in dy | \mathcal{F}_t) f(y) = \int_0^\infty \frac{\partial^2}{\partial y^2} C_t(T, y) f(y)\, dy.$$

Integration by parts transforms this into the Breeden and Litzenberger formula

Convex payoffs

There is a special case of terminal-value options with a convex payoff, which is particularly interesting. For instance, we can use Jensen's inequality (see, for example, 6 6 of Williams, 1991) to show that

$$V_0 = \mathbb{E}_{\mathbb{Q}}(e^{-rT} f(S_T)) \geq e^{-rt} f(F),$$

where F is a forward price $F = e^{rT} S_0 = \mathbb{E}_{\mathbb{Q}}(S_T)$. We can also use the convexity of f once more to show that

$$V_0 \geq e^{-rT} f(F) \geq f(S_0) - (1 - e^{\ rT}) f(0).$$

So that if $f(0) = 0$, for perhaps a call option, the option value V_0 is always worth at least as much as its current intrinsic value $f(S_0)$, and similarly $V_t \geq f(S_t)$ American options, which give the right to the intrinsic value $f(S_t)$ at any time t up to maturity T, have no additional worth for such convex payoffs null at 0.

We can also see how volatility and convexity make prices higher The price of a convex option is increasing in the volatility of the asset. For instance, if

$$S_T^\sigma = F \exp(\sigma Z - \tfrac{1}{2}\sigma^2),$$

where Z is a normal $N(0, 1)$ giving $\mathbb{E}(S_T^\sigma) = \text{F}$, then for $\sigma^2 < \tau^2$,

$$S_T^\tau = S_T^\sigma \exp(\alpha \tilde{Z} - \tfrac{1}{2}\alpha^2) \text{for } \tilde{Z}, \text{ an independent } N(0, 1),$$

where $\alpha^2 = \tau^2 - \sigma^2$ Then again by Jensen's inequality

$$\mathbb{E}(f(S_T^\tau)) = \mathbb{E}(\mathbb{E}(f(S_T^\tau) | S_T^\sigma)) \geq \mathbb{E}(f(S_T^\sigma))$$

Call prices, for instance, increase with the volatility of the asset (as per the Black-Scholes formula), but also in general models This fact, coupled with the Breeden and Litzenberger formula, shows how volatility and convexity work together to give value That is, the option's non-linear terms have worth

$$\int_0^\infty C_t(T, y) f''(y)\, dy,$$

which increases both with the volatility of the asset (which increases all the call prices) and the convexity of f (which increases f'').

Hobson (1996b) unites existing results, using coupling, to show that convex-option prices, even for diffusion models, increase with volatility and that the option value is itself a convex function of the current asset price

A new result generalises the Breeden and Litzenberger formula to higher dimensions Suppose we have a vector of assets S_t in $\mathbb{R}^n$, such that $|S_T|$ is square-integrable and that options on all fixed portfolios (linear combinations) of S_T are traded That is, the call $(\langle\theta, S_T\rangle - y)^+$ is traded, for all vectors θ in $\mathbb{R}^n$ and all real y, and has current price $C_t(T, \theta, y)$

Now for any f in C^{n+3}, which satisfies the boundedness condition that $|\nabla^{n+3}f|$ is integrable over $\mathbb{R}^n$, then f has a Fourier transform $\tilde{f}(\theta)$,

$$\tilde{f}(\theta) = \int_{\mathbb{R}^n} e^{-i\langle\theta, x\rangle}\tilde{f}(x)\,dx,$$

which is bounded by $|\tilde{f}(\theta)| \le c|\theta|^{-(n+3)}$, for some constant c We recall the Fourier inversion formula

$$f(x) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i\langle\theta, x\rangle}\tilde{f}(\theta)\,d\theta.$$

We can also use an adapted version of the existing one-dimensional hedging representation applied to the complex function e^{iz}, evaluated at the portfolio value $z = \langle\theta, S_T\rangle$, thus.

$$e^{i\langle\theta, S_T\rangle} = 1 + i\langle\theta, S_T\rangle - \int_{\mathbb{R}} |y|\left(y^{-1}\langle\theta, S_T\rangle - 1\right)^+ e^{iy}\,dy$$

We can substitute this expression into the Fourier inversion formula above, to deduce that

$$f(S_T) = f(0) + \langle\nabla f(0), S_T\rangle - (2\pi)^{-n}\int_{\mathbb{R}^n}\int_{\mathbb{R}} |y|\left(y^{-1}\langle\theta, S_T\rangle - 1\right)^+ e^{iy}\tilde{f}(\theta)\,dy\,d\theta$$

This expression can be re-expressed, by changing variables to $\phi = y^{-1}\theta$, to give

$$f(S_T) = f(0) + \langle\nabla f(0), S_T\rangle + \int_{\mathbb{R}^n} (\langle\phi, S_T\rangle - 1)^+ F_f(\phi)\,d\phi.$$

where $F_f(\phi) = -(2\pi)^{-n}\, Re\int_{\mathbb{R}}|y|^{n+1}e^{iy}\tilde{f}(y\phi)dy$ Thus the time t price of such a claim $f(S_T)$ is

$$V_t = f(0) + \langle\nabla f(0), S_t\rangle + \int_{\mathbb{R}^n} C_t(T, \phi)F_f(\phi)\,d\phi,$$

where $C_t(T, \phi) = C_t(T, \phi, 1)$ And thus when our claim, $X = f(S_T)$, is the terminal-time evaluation of a smooth function f, then the claim has a static hedge of cash, stocks, and generalised calls. As such functions are dense in the space of all measurable functions f, with $f(S_T)$ integrable, then all such claims can be approximated with static hedges

6. SUPERHEDGING

In incomplete models, where we cannot hedge, and we are pricing exotics (insusceptible to Breeden and Litzenberger), we must try something else Work by El Karoui et al (1996) has brought forward the concept of superhedging.

A clear treatment of the El Karoui results can be found in Hobson (1996b), and Frey (1997) is a good review of the current literature and developments in the superhedging field

Suppose as an example, a stock price behaves under some martingale measure, as a martingale diffusion with volatility σ_t,

$$dS_t = \sigma_t S_t dW_t$$

Suppose further that σ_t is either dependent on a new source of randomness distinct from W_t or simply uncertain – we just do not have a reliable model for it.

Given an upper bound σ_M on the volatility, that is $\sigma_t \leq \sigma_M$, we can bound the price of convex terminal value claims $f(S_T)$. We can even allow σ_M to be non-constant, as long as $\sigma_M = \sigma_M(S_t, t)$ only depends on time and the current stock price. If we hedge *as if* the actual volatility is σ_M then, as the theorem below shows, our hedge's final value will always be at least as large as $f(S_T)$. The claim has been superhedged. So the super-price of the claim is the theoretical price of $f(S_T)$ given the stock's volatility is σ_M. Similarly concave payoffs are superhedged by lower bounds on volatility.

THEOREM (El Karoui et al) *Let $C_t = C(S_t, t)$ be the worth of the convex claim $f(S_T)$ assuming the volatility is σ_M, and let V_t be the worth of the attempted hedge Then $C(x, t)$ is convex in x and the tracking error $e_t = V_t - C_t$ is given by the positive quantity*

$$e_t = \tfrac{1}{2}\int_0^t (\sigma_M^2 - \sigma_u^2) S_u^2 \frac{\partial^2 C}{\partial x^2}(S_u, u)\,du.$$

In the special case where $f(x)$ is the call payoff $(x-k)^+$, then $C_t = C(S_t, t)$ is the Black-Scholes call price, assuming constant volatility σ_M, where $C = C(x, t)$ is

$$C(x, t) = x\Phi\left(\frac{\log\frac{x}{k} + \frac{1}{2}\sigma_M^2(T-t)}{\sigma_M\sqrt{T-t}}\right) - k\Phi\left(\frac{\log\frac{x}{k} - \frac{1}{2}\sigma_M^2(T-t)}{\sigma_M\sqrt{T-t}}\right).$$

and the hedge ϕ_t is equal to $\phi_t = \frac{\partial C}{\partial x}(S_t, t)$, where

$$\frac{\partial C}{\partial x} = \Phi\left(\frac{\log\frac{x}{k} + \frac{1}{2}\sigma_M^2(T-t)}{\sigma_M\sqrt{T-t}}\right)$$

Then the worth of the hedge at time t is $C_0 + \int_0^t \phi_u dS_u$. Crucially, C is convex in x so that the tracking error is always positive as σ_M is an upper bound for σ_t At time T, V_T is the worth of the hedge and C_T is the option worth $(S_T - k)^+$, so a positive tracking error means the option has been superhedged

A pleasing synthesis between the option-hedging of section 5 and super-hedging has been achieved by Paras (1997), a description of which we close with

Following Paras, we let $\mathcal{P}$ be the set of all measures that might model the asset price. These measures might not be equivalent, for instance they might correspond to all possible Markov volatility processes σ_t lying in a band

$$\sigma_m \leq \sigma_t \leq \sigma_M,$$

where σ_m, σ_M can be functions of time and asset price Then the superhedge price of a claim X will be the (supermartingale) process

$$V_t = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}(X | \mathcal{F}_t),$$

where interest rates have been set to zero for simplicity. The superhedging strategy will be to behave as if volatility is

$$\sigma = \begin{cases} \sigma_M \text{ when } \frac{\partial^2 V}{\partial S^2} > 0, \text{ locally convex,} \\ \sigma_m \text{ when } \frac{\partial^2 V}{\partial S^2} > 0, \text{ locally concave} \end{cases}$$

Suppose also that there are currently traded instruments, such as vanilla options, which pay off X_i at time T and are currently worth $C_i(t)$. We might not be able to write X entirely in terms of a combination of the X_i, but we could do the best we could. If we used a hedge of λ_i units of X_i, our valuation for X would be

$$L_t(\lambda) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}\left(X - \sum_i \lambda_i X_i | \mathcal{F}_t \right) + \sum_i \lambda_i C_i(t)$$

As we are completely (super)-hedged for any choice of λ, we could choose λ to minimise $L_t(\lambda)$, and quote the sharpest price possible. As $L_t(\lambda)$ is the supremum of linear functions of λ, it is a convex function of λ, and so susceptible to optimization techniques.

Interestingly, this problem is the Lagrangian dual of the constrained optimization problem which maximises the expectation of X over measures which produce the market price for every X_i That is the problem

$$\sup \mathbb{E}_{\mathbb{P}}(X | \mathcal{F}_t). \text{ subject to } \mathbb{E}_{\mathbb{P}}(X_i \parallel \mathcal{F}_t) = C_i(t). \ \mathbb{P} \in \mathcal{P}$$

This is really just affirms the intuitive observation that measures in $\mathcal{P}$ which do not reflect the current price of traded instruments cannot be the measure we need to price So we have a duality between the best superhedge of the claim over all measures, allowing hedging with traded instruments, and the best superhedge over all measures which price the traded instruments to market

REFERENCES

BAXTER, M W and RENNIE, A J O (1996) *Financial Calculus an introduction to derivative pricing*, Cambridge University Press

BLACK, F and SCHOLES, M (1973) The pricing of options and corporate liabilities, *J Political Econ* , **81**, 637-654

BREEDEN, D T and LITZENBERGER, R H (1978) Price of state contingent claims implicit in options prices, *J Business*, **51**, 621-651

BOGNI, R (1997) Don't forget your tool kit, *Times Higher Ed* , 24/01/97, p28

DUPIRE, B (1993) Arbitrage pricing with stochastic volatility *SORT research paper*, Paribas Capital Markets

EL KAROUI, N , JEANBLANC-PICQUE, M and SHREVE, S (1996) Robustness of the Black and Scholes formula, preprint

FREY, R (1997) Derivative asset analysis in models with level-dependent and stochastic volatility, *CWI Quarterly* (to appear)

HARRISON, J M and PLISKA, S R (1981) Martingales and stochastic integrals in the theory of continuous trading, *Stoch Procs and their applications*, 11, 215-260

HOBSON, D G (1996a) Stochastic Volatility, to appear in *Statistics in Finance*, ed by D Hand and S Jacka, Edward Arnold

HOBSON, D G (1996b) Robust hedging via coupling, *University of Bath preprint*, School of Mathematics 96 04

HULL, J (1997) *Options, futures and other derivative securities*, third edition, Prentice-Hall

KEMP, M H D (1997) Actuaries and derivatives, to appear in *British Actuarial Journal*

PARAS, A (1997) Hedging under uncertain volatility, *SORT research paper*, Paribas Capital Markets

WILLIAMS, D (1991) *Probability with Martingales*, Cambridge University Press

MARTIN BAXTER
Statistical Laboratory
Cambridge University
16 Mill Lane
Cambridge CB2 1SB
England

RISK-MINIMIZING HEDGING STRATEGIES FOR UNIT-LINKED LIFE INSURANCE CONTRACTS

By

THOMAS MØLLER

Laboratory of Actuarial Mathematics
University of Copenhagen

ABSTRACT

A unit-linked life insurance contract is a contract where the insurance benefits depend on the price of some specific traded stocks We consider a model describing the uncertainty of the financial market and a portfolio of insured individuals simultaneously. Due to incompleteness the insurance claims cannot be hedged completely by trading stocks and bonds only, leaving some risk to the insurer. The theory of risk-minimization is briefly reviewed and applied after a change of measure. Risk-minimizing trading strategies and the associated intrinsic risk processes are determined for different types of unit-linked contracts By extending the model to the situation where certain reinsurance contracts on the insured lives are traded, the direct insurer can eliminate the risk completely The corresponding self-financing strategies are determined.

KEYWORDS

Incomplete market, Martingale representation, Minimal martingale measure, Intrinsic risk, Reinsurance.

1 INTRODUCTION

Traditional actuarial analysis of life insurance contracts focuses on calculation of expected values of various discounted random cashflows; the fundamental principle of equivalence states that discounted premiums and benefits should balance on average for any contract. The corresponding premium is called the equivalence premium. Similarly, at any time during the insurance period, the prospective reserve is defined as the conditional expected value of all discounted future benefits less premiums, given the available information. The development of the reserve is described by

Thiele's differential equation, which originally dealt with constant deterministic interest and deterministic benefits, but has been widely generalized, see e.g. Norberg (1995) and Norberg and Møller (1996).

With a unit-linked life insurance contract, benefits depend explicitly on a specified stock index. Typically, the policyholder will receive the maximum of the stock price and some asset value guarantee stipulated in the contract, but other dependencies may be specified These contracts have been analyzed by Brennan and Schwartz (1979), and more recently by e g. Delbaen (1990), Bacinello and Ortu (1993), Aase and Persson (1994) and Nielsen and Sandmann (1995). The last of these authors allow the risk-free interest rate to be stochastic. Various exotic types of contract functions are considered in Ekern and Persson (1996). Aase and Persson (1994) derive a partial differential equation for the value of the reserve of a unit-linked life insurance, which is compared with Thiele's differential equation They also present duplicating strategies that minimize the risk of the insurance company in a sense.

All the papers mentioned consider mortality risk as diversifiable or assume that the insurer is "risk neutral with respect to mortality" and replace the uncertain courses of the insured lives by the expected. In this way, the actual insurance claims, depending on uncertainty within the portfolio of insured lives and the financial markets, are replaced by similar claims which only include the financial uncertainty. These claims are then priced using standard no-arbitrage pricing theory. In the present paper we provide and examine a model where the uncertainty of a portfolio of lives to be insured and a certain financial market are described simultaneously, and consider the problem of hedging the actual claims which depend on both sources of uncertainty.

The insurance company issues life insurance contracts with insurance benefits linked to the price of a specified stock. This stock and one risk-free asset are traded freely on the financial market without transaction costs. We then consider the problem of defining optimal investment strategies. This situation differs from the case of standard life insurance, where the insurance company should try to maximize trading gains in order to compete with other companies on redistributions of bonus With unit-linked contracts, benefits are already linked explicitly to the development of the market, and hence are not influenced by the factual gains generated by the investment strategies of the insurance company. However, by issuing these contracts, the insurer is exposed to a financial risk, and our objective here will be to minimize this risk. In this paper we will measure the risk associated with the contracts using the expected value (under an adjusted measure) of the square of the difference between the insurance benefits to be paid and the gains obtained from investments.

The insurance contracts are characterized as contingent claims in an incomplete model, such that the insurance claims cannot be perfectly duplicated by means of self-financing strategies The theory of risk-minimization for incomplete markets introduced by Föllmer and Sondermann (1986) and developed further by Föllmer and Schweizer (1988) and Schweizer (1991, 1994 and 1995) is reviewed and then applied after a change of measure. With its present formulation, this theory deals with the problem of hedging contingent claims that are payable at a fixed time only. The analysis of more general claims with intermediate payment times would require an extension of the original theory of Follmer and Sondermann (1986), a problem which will be addressed in a forthcoming paper by Moller (1998) Thus, insurance contracts with payments occurring only at fixed times are analyzed within the original setup of Föllmer and Sondermann (1986), whereas some modifications are needed in order to deal with contracts where the sum insured falls due immediately upon the death of the insured. In the present paper, we assume that premiums are paid as single premiums and that all benefits are deferred to the term of the contract. In this way optimal investment strategies minimizing the risk (under the minimal martingale measure) associated with the assigned contracts are determined. Since the model is incomplete, risk cannot be eliminated completely by applying these strategies, leaving some minimum obtainable risk (called the intrinsic risk) to the insurer. This minimum risk process is determined for different types of standard contracts and is taken as a measure of the non-hedgeable risk inherent in the contracts.

In Section 2 we present the combined model and briefly mention some basic results from the theory of mathematical finance. We also introduce the basic types of insurance claims to be analyzed in the paper Section 3 is devoted to a review of the most important concepts of risk-minimization. Unit-linked life insurance contracts by single premium are analyzed in Section 4. Section 5 deals with the situation where reinsurance contracts are traded freely on the market. Finally, some numerical results are presented in Section 6

2 THE MODEL

In this section the two basic elements of the model, the financial market and a portfolio of individuals to be insured, are introduced. We set out by presenting the financial market and reviewing some well-known results from the theory of mathematical finance for complete markets. When extending the model by also including a portfolio of individuals to be insured, the market is no longer complete.

Throughout, we let T denote a fixed, finite time horizon and consider a given probability space $(\Omega, \mathcal{F}, P)$.

2.1. The financial market

We consider a market consisting of only two traded assets: a stock with prices process S and a bond with price process B. At any time t these assets are traded freely at prices S_t and B_t, respectively. The price processes are defined on a probability space $(\Omega, \mathcal{F}, P)$ and are given by the P-dynamics

$$dS_t = \alpha(t, S_t)S_t dt + \sigma(t, S_t)S_t dW_t, \tag{2.1}$$

$$dB_t = r(t, S_t)B_t dt, \tag{2.2}$$

$S_0 > 0$, $B_0 = 1$, where $W = (W_t)_{0 \leq t \leq T}$ is a standard Brownian motion on the time interval $[0, T]$. The filtration $\mathbf{G} = (\mathcal{G}_t)_{0 \leq t \leq T}$ generated by this economy is given by

$$\mathcal{G}_t = \sigma\{(S_u, B_u),\ u \leq t\} = \sigma\{S_u,\ u \leq t\}.$$

A solution to the equation (2.1) exists provided that the functions α and σ satisfy certain regularity conditions, see e.g. Duffie (1996, Appendix E). These conditions are assumed to be fulfilled henceforth. Furthermore, we assume that $\int_0^T r_t dt$ exists and is finite almost surely.

The process α is interpreted as the mean rate of return of S, and σ as the standard deviation of the rate of return. Similarly r is called the short rate of interest and denotes the rate of return of the risk-free asset. The process ν defined by $\nu_t = (\alpha_t - r_t)/\sigma_t$ is known as the market price of risk process associated with S. In addition to the assumptions above, we assume that ν satisfies the integrability conditions from Duffie (1996, Chapter 6). With constant coefficients α, σ and r, all conditions are satisfied, and we have the celebrated Black-Scholes model where S and B are given by

$$S_t = S_0 \exp((\alpha - \tfrac{1}{2}\sigma^2)t + \sigma W_t)\ ,$$
$$B_t = \exp(r\,t).$$

The model above has been thoroughly investigated in the literature of mathematical finance, see e.g. Duffie (1996), Björk (1996) and Lamberton and Lapeyre (1996). Thus some concepts and results from the theory of finance, needed repeatedly in the sequel, will be quoted without explicit reference. Also Aase and Persson (1994) give a brief survey of this theory.

Recall that two measures P and P^* are said to be equivalent if, for each set $A \in \mathcal{F}$, we have that $P(A) = 0$ if and only if $P^*(A) = 0$. By definition, the probability measure P^* defined by

$$\frac{dP^*}{dP} = \exp\left(-\int_0^T \left(\frac{\alpha_u - r_u}{\sigma_u}\right) dW_u - \frac{1}{2}\int_0^T \left(\frac{\alpha_u - r_u}{\sigma_u}\right)^2 du\right) \equiv U_T \tag{2.3}$$

is equivalent to P. It can be verified that the discounted price process S^*, defined by

$$S_t^* = S_t/B_t = S_0 \exp\left(\int_0^t (\alpha_u - r_u)\, du + \int_0^t \sigma_u dW_u\right), \tag{2.4}$$

is a P^*-martingale. Thus P^* is called an *equivalent martingale measure* In the above model, the martingale measure is unique.

A *trading strategy* or *portfolio strategy* is an adapted process $\varphi = (\xi, \eta)$ satisfying some integrability conditions (a precise definition will be given in Section 3). At any time $t \in [0, T]$, ξ_t and η_t represent, respectively, the number of shares and the number of bonds held in the portfolio. The *value process* $\hat{V}^{\varphi}$ associated with φ is defined by

$$\hat{V}_t^{\varphi} = \xi_t S_t + \eta_t B_t \,, \tag{2.5}$$

and the strategy is said to be *self-financing* if

$$\hat{V}_t^{\varphi} = \hat{V}_0^{\varphi} + \int_0^t \xi_u dS_u + \int_0^t \eta_u dB_u \,, \tag{2.6}$$

for all $0 \leq t \leq T$ According to (2.6), any change in the value of the portfolio is generated by changes in the underlying price processes S and B. A *contingent claim* with maturity T is a random variable X that is $\mathcal{G}_T$-measurable and P^*-square integrable In particular, X is called a *simple claim* whenever $X = g(S_T)$, for some function g $\mathbf{R}_+ \to \mathbf{R}$. We say that a contingent claim X can be perfectly duplicated if there exists a self-financing portfolio φ such that $\hat{V}_T^{\varphi} = X$ P-a.s In this case the claim is called *attainable*. If all contingent claims are attainable, then the market is said to be *complete*; otherwise the market is referred to as *incomplete*. A self-financing strategy φ is an *arbitrage* if $\hat{V}_0^{\varphi} < 0$ and $\hat{V}_T^{\varphi} \geq 0$ or if $\hat{V}_0^{\varphi} \leq 0$, $\hat{V}_T^{\varphi} \geq 0$ P-a.s. and $\hat{V}_T^{\varphi} > 0$ with positive probability It is well-known that the market defined by (2.1)-(2.2) and filtration $\mathbf{G}$ is complete and free or arbitrage under the above mentioned assumptions

Note that if $\varphi = (\xi, \eta)$ is self-financing and duplicates the claim X, then we have the following representation from (2.5) and (2.6):

$$X = \xi_0 S_0 + \eta_0 B_0 + \int_0^T \xi_u dS_u + \int_0^T \eta_u dB_u \tag{2.7}$$

The arbitrage-free price process $(F(t, S_t))_{0 \leq t \leq T}$ associated with a simple claim specifying the payment $g(S_T)$ at time T can now be characterized by the partial differential equation (PDE)

$$-r(t,s)F(t,s) + F_t(t,s) + r(t,s)sF_s(t,s) + \frac{1}{2}\sigma(t,s)^2 s^2 F_{ss}(t,s) = 0, \tag{2.8}$$

with boundary value $F(T, s) = g(s)$ Here, exemplifying a general notational convention adopted throughout, $F_s(t, s)$ denotes the partial derivative of $F(t, s)$ with respect to s, $F_{ss}(t, s)$ denotes the second order partial derivative with respect to s, and so on.

The arbitrage-free price process associated with the claim $g(S_T)$ is also given in terms of the unique equivalent martingale measure by

$$F(t, S_t) = E^*\left[\exp\left(-\int_t^T r_u\,du\right)g(S_T)\middle|\mathcal{G}_t\right]. \tag{2.9}$$

(Throughout E^* denotes expectation with respect to P^*). Thus, the price is determined by discounting the T-payment with the asset B and then calculating the conditional expectation under the martingale measure P^*.

2.2. The insurance portfolio

In this paragraph we will introduce a model to describe the lifetimes in a group of individuals. For simplicity, we assume that the lifetimes are mutually independent and identically distributed. The i.i.d. assumption implies that the individuals are selected from a cohort of equal age x, say, and we denote by l_x the number of persons in the group. Mathematically, this is described by representing the individual remaining lifetimes as a sequence $T_1, \ldots, T_{l_x}$ of i.i.d. non-negative random variables defined on $(\Omega, \mathcal{F}, P)$. Assuming that the distribution of T_i is absolutely continuous with hazard rate function μ_{x+t}, the survival function is

$${}_tp_x = P(T_i > t) = \exp\left(-\int_0^t \mu_{x+\tau}\,d\tau\right).$$

Now define a univariate process $N = (N_t)_{0 \leq t \leq T}$ counting the number of deaths in the group;

$$N_t = \sum_{i=1}^{l_x} I(T_i \leq t),$$

and denote by $\mathbf{H} = (\mathcal{H}_t)_{0 \leq t \leq T}$ the natural filtration generated by N, i.e. $\mathcal{H}_t = \sigma\{N_u, u \leq t\}$. By definition, N is cadlag (right-continuous with left-limits) and, since the lifetimes T_i are i.i.d., the counting process N is an $\mathbf{H}$-Markov process. The (stochastic) intensity process λ of the counting process N can be informally defined by

$$E[dN_t \mid \mathcal{H}_{t-}] = (l_x - N_{t-})\mu_{x+t}dt \equiv \lambda_t dt,$$

the hazard rate function μ_{x+t} times the number of individuals under exposure just before time t. The *compensated counting process* M defined by

$$M_t = N_t - \int_0^t \lambda_u du \tag{2.10}$$

is an $\mathbf{H}$-martingale.

2.3. The combined model

Now introduce the filtration $\mathbf{F} = (\mathcal{F}_t)_{0 \le t \le T}$ generated by the economy and the insurance portfolio, that is

$$\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t .$$

We assume throughout that $\mathcal{G}_T$ and $\mathcal{H}_T$ are independent and take

$$\mathcal{F} = \mathcal{G}_T \vee \sigma\{I(T_i \le u),\ 0 \le u \le T,\ i = 1, .., l_x\} .$$

At time 0 the insurance company issues an insurance contract for each of the l_x individuals. These contracts specify payments of benefits and premiums that are contingent on the remaining lifetime of the policyholder, and are linked to the development on the financial market. During the period [0, T] the company is allowed to trade the assets B and S freely (without transaction costs, taxes and short sales restrictions) based on the complete information $\mathbf{F}$ Furthermore, we allow for continuous rebalancing of the portfolio of stocks and bonds in order to hedge against the insurance claims

In the following, we present the two basic forms of insurance contracts to be analyzed in this paper: the *pure endowment* and the *term insurance*. With a *pure endowment* contract, the sum insured is to be paid at the term T if the insured is then still alive. The sum is of the form $g(S_T)$ for some continuous function g stipulated in the contract, thus depending on the price of the risky asset at time T. Some specific functions will be considered as examples, e g. $g(s) = s$ and $g(s) = \max(s, K)$ which are known from the literature as *pure unit-linked* and *unit-linked with guarantee* insurance policies, see Aase and Persson (1994). For each insured person the obligation of the insurance company is given by the *present value*

$$H_i = I(T_i > T)g(S_T)B_T^{-1} = I(T_i > T)g(S_T)e^{-\int_0^T r_u du} \tag{2.11}$$

Here we have adopted widely accepted actuarial usage of the term *present value*, it is taken to be the payments discounted using the bond price process described by (2.2) Thus, the present value is an $\mathcal{F}_T$-measurable random variable. This usage may be at variance with the economical one, where present value typically refers to an $\mathcal{F}_0$-measurable value. The entire portfolio generates the discounted claim

$$H = g(S_T)B_T^{-1}\sum_{i=1}^{l_x} I(T_i > T) = g(S_T)B_T^{-1}(l_x - N_T), \tag{2.12}$$

where $(l_x - N_T)$ is the number of survivors at the end of the insurance period It should be noted that the undiscounted insurance claim HB_T taken from (2.12) is a function of S_T and N_T only. Insurance claims that are payable at time T and are functions of S_T and N_T only will be called *simple T-claims*, whereas more general insurance claims payable at time T are denoted *(general) T-claims*

The *term insurance* states that the sum insured is due immediately upon death before time T. In this case, we consider a time dependent contract function $g_t = g(t, S_t)$. By the definition of the contract, payments can occur at any time during $[0, T]$ and obligations generated by such contracts do not form T-claims without introducing special assumptions. A simple way of transforming the obligations into a (general) T-claim is to assume that all payments are deferred to the term of the contract and are accumulated with the risk-free rate of interest r. With this specific construction, the heirs of a policyholder who died at time t would receive the benefit $g(t, S_t)B_T B_t^{-1}$ at time T. The deferred payments could as well be accumulated differently, for example by using some deterministic first order interest rate δ or by investing $g(t, S_t)$ according to a predefined strategy. These ways of modifying the contracts by deferring the benefits might seem most reasonable for contracts with short time horizons, say one year Although time horizons associated with traditional life insurance contracts are typically much longer, we will assume that the benefits are actually deferred to the end of the insurance period. The insurer's liabilities in respect of a portfolio of term insurance contracts with payments that are deferred and accumulated using the riskless asset B are now described by the discounted general T-claim

$$H_T = B_T^{-1} \sum_{i=1}^{l_x} g(T_i, S_{T_i}) B_{T_i}^{-1} B_T I(T_i \leq T) = \sum_{i=1}^{l_x} \int_0^T g(u, S_u) B_u^{-1} dI(T_i \leq u),$$

which can be rewritten as an integral with respect to the counting process N:

$$H_T = \int_0^T g(u, S_u) B_u^{-1} dN_u \tag{2.13}$$

Various other insurance contracts can be obtained as combinations of the pure endowment and the term insurance. For example, with the *endowment insurance*, the sum insured is payable at the time of death of the insured persons or maturity, whichever comes first. The present value of this claim is a sum of (2.12) and (2 13). Throughout, we assume that premiums are paid as single premiums at time 0. Thus, the present value of all premiums is simply $\pi^s = l_x \cdot \pi_1$, where π_1 is the single premium paid by the insured.

In Section 2 1 it was pointed out that in the complete market every contingent claim can be represented as an integral with respect to the price processes S and B, see (2.7). As we will show later, this property is not preserved when the model consists of the assets (B, S) and filtration $\mathbf{F}$ Intuitively, this follows from the fact that the claims (2 12)-(2 13) are not generated by the price processes (B, S) alone since the uncertainty concerning the insured lives contributes essentially to the final outcome of the claims

We end this section by discussing choice of martingale measure in the combined model. For any **H**-predictable process h, such that $h > -1$, define a likelihood process L by

$$dL_t = L_{t-}h_t dM_t, \tag{2.14}$$

and initial conditional $L_0 = 1$ Provided that $E^P[L_T]$, a new probability measure $\hat{P}$ can be defined by

$$\frac{d\hat{P}}{dP} = U_T \cdot L_T, \tag{2 15}$$

where U_T is given by (2.3). Using the definition of the measure $\hat{P}$ and the independence between N and (B, S) under P we see that S^* defined by (2.4) is also a $\hat{P}$-martingale: for $u < t$ we have

$$\hat{E}[S^*_t|\mathcal{F}_u] = \frac{E[S^*_t U_T L_T|\mathcal{F}_u]}{E[U_T L_T|\mathcal{F}_u]} = \frac{E[S^*_t U_T|\mathcal{F}_u]\ \ E[L_T|\mathcal{F}_u]}{E[U_T|\mathcal{F}_u]\ \ E[L_T|\mathcal{F}_u]} = E^*[S^*_t|\mathcal{F}_u] = S^*_u,$$

using that S^* is a P^*-martingale, and so each $\hat{P}$ is an equivalent martingale measure. Due to this non-uniqueness of the equivalent martingale measure, contracts cannot in general be priced uniquely by no-arbitrage pricing theory alone Actually, all prices

$$\pi(\hat{P}) = E^{\hat{P}}[H]$$

for the claims (2 12)-(2.13) obtained by admissible choices of h are consistent with absence of arbitrage. Furthermore, (B, S) and N are independent under $\hat{P}$ and, by the Girsanov theorem, the process M^h defined by

$$M^h_t = N_t - \int_0^t \lambda_u(1 + h_u)du$$

is an $(\mathbf{F}, \hat{P})$-martingale. The term L_T in (2.15) essentially changes the hazard rate in the model to $\mu_{x+t}(1 + h_t)$. In particular, the measure P^* defined by (2.3) can be obtained from (2.15) with $h \equiv 0$ Note that the change of measure form P to P^* does not affect the distribution of N and that M is an $(\mathbf{F}, P^*)$-martingale.

Throughout this paper we will apply the specific martingale measure P^* defined by (2.3) which is also known as the *minimal martingale measure*, cf Schweizer (1991, 1995) This particular measure is normally applied to pricing of unit-linked contracts, the motivation being the insurer's risk neutrality with respect to mortality, see e.g. Aase and Persson (1994). Thus, we consider the probability space $(\Omega, \mathcal{F}, P^*)$ endowed with the filtration **F**. Note that **F** is equivalently generated by the P^*-martingales S^* and M:

$$\mathcal{F}_t = \sigma\{(S^*_u, M_u),\ 0 \le u \le t\}\ .$$

In the analysis below, we could equally well apply any of the martingale measures $\hat{P}$ defined by (2.15) for admissible choices of h In this case we would obtain similar results with the hazard rate function μ replaced by $(1+h)\mu$ and M replaced by M^h. However, there do exist martingale measures which do not preserve independence between (B, S) and N, and such choices of martingale measures would certainly complicate calculations in Section 4 greatly.

3. A REVIEW OF RISK-MINIMIZATION

In the previous section, a model describing a financial market and an insurance portfolio was introduced. It was pointed out that this market is incomplete in the sense that contingent claims cannot in general be perfectly duplicated by means of self-financing strategies. In this section, we briefly review some results on the theory of risk-minimization, dealing with incomplete as well as complete markets.

Föllmer and Sondermann (1986) extended the established theory for complete markets to the case of an incomplete market. By introducing the concept of *mean-self-financing* strategies they obtained optimal strategies in the sense of minimization of a certain squared error process. In Follmer and Schweizer (1988) a discrete time multiperiod model was examined within this set-up, and they obtained recursion formulas describing the optimal strategies The theory has been further developed by Schweizer (1991, 1994). Föllmer and Sondermann (1986) originally considered the case where the original probability measure P is in fact a martingale measure. Schweizer (1991) introduced the concept of *local risk-minimization* for price processes which are only semimartingales and this criterion was similar to performing risk-minimization using the minimal martingale measure P^*

Recall the space $(\Omega, \mathcal{F}, P^*)$, filtration $\mathbf{F}$ and the $(\mathbf{F}, P^*)$-martingales S^* and M. The deflated value process V^φ is defined by

$$V_t^\varphi = \hat{V}_t^\varphi B_t^{-1} = \xi_t S_t^* + \eta_t , \tag{3.1}$$

where $\hat{V}^\varphi$ is given by (2 5) From Follmer and Sondermann (1986) and Schweizer (1994) we have a slightly modified definition of strategies and the value process. Introducing the space $\mathcal{L}^2(P_S^*)$ of $\mathbf{F}$-predictable square-integrable processes ξ satisfying

$$\mathrm{E}^*\left[\int_0^T \xi_u^2 d\langle S^*\rangle_u\right] < \infty,$$

they state·

Definition 3.1 *An* **F**-*strategy is any process* $\varphi = (\xi, \eta)$ *with* $\xi \in \mathcal{L}^2(P^*_S)$ *and* η **F**-*adapted such that the (deflated) value process* V^φ *is cadlag and* $V^\varphi_t \in \mathcal{L}^2(P^*)$ *for all* t.
The cost process C^φ *associated with the strategy* φ *is defined by*

$$C^\varphi_t = V^\varphi_t - \int_0^t \xi_u dS^*_u, \tag{3.2}$$

and the risk process R^φ *of* φ *is defined by*

$$R^\varphi_t = \mathrm{E}^*\left[(C^\varphi_T - C^\varphi_t)^2 | \mathcal{F}_t\right]. \tag{3.3}$$

In this definition, the notion *risk process* is attached to the conditioned expected squared value of future costs. This usage differs from the traditional actuarial one, where "risk process" would typically denote the cash flow of premiums and benefits

The cost C^φ is the value of the portfolio less the accumulated income from the asset S. The total costs C^φ_t incurred in $[0, t]$ decompose into the costs incurred during $(0, t]$ and an initial cost $C^\varphi_0 = V^\varphi_0$, which typically is greater than zero. A strategy is said to be *mean-self-financing* if the cost process $C^\varphi = (C^\varphi_t)_{0 \le t \le T}$ is an ($\mathbf{F}$, P^*)-martingale. Furthermore, it should be noted that the strategy $\varphi = (\xi, \eta)$ is *self-financing* if and only if

$$V^\varphi_t = V^\varphi_0 + \int_0^t \xi_u dS^*_u,$$

that is, if and only if $C^\varphi_t = C^\varphi_0 = V^\varphi_0$ P^*-a.s.

Let us now turn to the problem of characterizing the optimal strategies. We consider a general contingent claim specifying the $\mathcal{F}_T$-payment H at time T and focus on *admissible* strategies φ satisfying

$$V^\varphi_T = H \text{ a.s.}$$

By means of admissible strategies, the hedger is able to generate the contingent claim, but only at some cost defined by C^φ_T. In particular, for attainable claims, $C^\varphi_T = C^\varphi_0 = V^\varphi_0$ is known at time 0.

As a first result, admissible strategies minimizing the mean squared error R^φ_0 defined by (3 3) are determined. For any admissible φ we have

$$C^\varphi_T = V^\varphi_T - \int_0^T \xi_u dS^*_u = H - \int_0^T \xi_u dS^*_u, \tag{3.4}$$

hence

$$R^\varphi_0 = \mathrm{E}^*\left[(C^\varphi_T - C^\varphi_0)^2\right] = \mathrm{E}^*\left[\left(H - \int_0^T \xi_u dS^*_u - C^\varphi_0\right)^2\right], \tag{3.5}$$

and so R_0^φ is minimized for $C_0^\varphi = \mathrm{E}^*[H]$ $(= \mathrm{E}^*[C_T^\varphi])$. Thus, we should choose ξ so as to minimize the variance

$$\mathrm{E}^*\left[(C_T^\varphi - \mathrm{E}^*[C_T^\varphi])^2\right] \tag{3.6}$$

This criterion does not yield a unique strategy, but it characterizes an entire class of strategies all minimizing the mean squared error (3.5). The non-uniqueness of the optimal admissible strategy is a natural consequence of the simple criterion of minimizing (3.5), which involves only the value of the cost process C^φ at time T, given by (3.4) Furthermore, note that $H = \xi_T S_T^* + \eta_T$, which does not depend on $(\eta_t)_{0\leq t<T}$. Thus, we should not expect the minimization criterion associated with the squared error (3.5) to impose any constraints on the number of bonds held in the time interval (0, T).

The construction of the strategies is based on an application of the Galtchouk-Kunita-Watanabe decomposition, see Föllmer and Sondermann (1986). Defining the *intrincic value process* V^* by

$$V_t^* = \mathrm{E}^*[H|\mathcal{F}_t]\,,$$

and noting that V^* is an $(\mathbf{F}, P^*)$-martingale, the Galtchouk-Kunita-Watanabe decomposition theorem allows us to write V_t^* uniquely in the form

$$V_t^* = \mathrm{E}^*[H] + \int_0^t \xi_u^H dS_u^* + L_t^H, \tag{3.7}$$

where $L^H = (L_t^H)_{0\leq t\leq T}$ is a zero-mean $(\mathbf{F}, P^*)$-martingale, L^H and S^* are orthogonal, and ξ^H is a predictable process in $\mathcal{L}^2(P_S^*)$. By applying the orthogonality of the martingales L^H and S^*, and using $V_T^* = H$, Föllmer and Sondermann (1986, Theorem 1) prove.

Theorem 3.2 (Föllmer and Sondermann) *An admissible strategy* $\varphi = (\xi, \eta)$ *has minimal variance*

$$\mathrm{E}^*\left[(C_T^\varphi - \mathrm{E}^*[C_T^\varphi])^2\right] = \mathrm{E}^*\left[(L_T^H)^2\right]$$

if and only if $\xi = \xi^H$.

Note that if, furthermore, the number of bonds held at time 0 is determined such that the initial value of the portfolio equals $\mathrm{E}^*[H]$, i.e

$$\eta_0 = \mathrm{E}^*[H] - \xi_0 S_0^*,$$

then $R_0^\varphi = \mathrm{E}^*\left[(C_T^\varphi - \mathrm{E}^*[C_T^\varphi])^2\right]$. Thus, the variance is interpreted as the minimal obtainable risk.

A more precise result is obtained by looking for admissible strategies, that is $V_T^\varphi = H$, minimizing the *remaining risk*, defined by R_t^φ at any time t. Such strategies are said to be *risk-minimizing*. Now fix some admissible strategy φ When considering the remaining risk R_t^φ at some point in time t, only admissible strategies $\tilde{\varphi}$ coinciding with φ in the interval $[0, t)$ should be compared. This condition ensures, that the cost processes are given by the same value $C_t^\varphi = C_t^{\tilde{\varphi}}$ at the time of consideration In this case the strategy $\tilde{\varphi}$ is said to be an *admissible continuation* of φ at time t, see Föllmer and Sondermann (1986) for more details. The risk-minimizing strategy, minimizing the risk process $(R_t^\varphi)_{0\le t\le T}$ is determined by Föllmer and Sondermann (1986, Theorem 2).

Theorem 3.3 (Föllmer and Sondermann) *There exists a unique admissible risk-minimizing strategy* $\varphi = (\xi, \eta)$ *given by*

$$(\xi_t, \eta_t) = (\xi_t^H, V_t^* - \xi_t^H S_t^*),\ 0 \le t \le T$$

The associated risk process is given by $R_t^\varphi = \mathrm{E}^*\left[(L_T^H - L_t^H)^2 | \mathcal{F}_t\right]$

The risk process associated with the risk-minimizing strategy is also called the *intrinsic risk process*

4. Unit-linked contracts with single premium

In this section, we apply the technique of risk-minimization in the investigation of the insurance contracts introduced in Section 2. An important step will be the construction of the decomposition (3.7) of the present values (2.12)-(2.13). Having determined this, risk-minimizing strategies and the intrinsic risk process associated with the pure endowment and the deferred term insurance contract can be determined by Theorems 3.2 and 3.3.

From the classical actuarial theory it is known that in the case of fixed premiums and sum insured, the "relative risk" associated with the portfolio decreases as the size l_x of the portfolio increases. More precisely, this means that the ratio between the standard deviation of the present value of all payments and the size of the portfolio l_x will converge to 0 as l_x is increased In the present set-up, we cannot expect such results since the payments associated with different insurance contracts are now linked to the same asset and hence are no longer stochastically independent. However the initial intrincis risk R_0 can be taken as a measure of the risk associated with the non-hedgeable part of the claims, and we will accordingly examine the ratio $\sqrt{R_0}/l_x$

4.1. The pure endowment

Consider the claim with present value H in (2.12);

$$H = g(S_T)B_T^{-1}(l_x - N_T), \tag{4.1}$$

and define the (deflated) intrinsic value process $V^* = (V_t^*)_{0\leq t\leq T}$ by

$$V_t^* = \mathrm{E}^*[H|\mathcal{F}_t]\ ,$$

for all $t \in [0, T]$. Due to the stochastic independence between N and (B, S) under P^*, we get

$$V_t^* = \mathrm{E}^*[(l_x - N_T)|\mathcal{F}_t]B_t^{-1}\mathrm{E}^*\left[g(S_T)B_tB_T^{-1}|\mathcal{F}_t\right]\ . \tag{4.2}$$

Here, the first factor is easily determined as

$$\begin{aligned}\mathrm{E}^*[(l_x - N_T)|\mathcal{F}_t] &= \mathrm{E}^*\left[\sum_{i=1}^{l_x} I(T_i > T)\Big|\mathcal{F}_t\right] = \sum_{i\, T_i>t} \mathrm{E}^*[I(T_i > T)|T_i > t]\\ &= \sum_{i\, T_i>t} {}_{T-t}p_{x+t} = (l_x - N_t)_{T-t}p_{x+t},\end{aligned}$$

that is, at any time t the expected number of individuals alive at the time of maturity T is simply the number of survivors at time t multiplied by the probability ${}_{T-t}p_{x+t}$ of survival to T for an individual, conditional on his/her survival to t. The second factor in (4.2) corresponds to the representation (2.9) of the unique arbitrage-free price process associated with the simple T-claim $g(S_T)$ in the complete model with filtration $\mathbf{G}$. In the present model, the insured lives are included in the filtration $\mathbf{F}$, and arbitrage-free prices are in general not unique. However, as N and (B, S) are stochastically independent, the conditional distribution of (B, S) given $\mathcal{F}_t$ does not depend on information concerning the insured lives $\mathcal{H}_t$ and thus

$$\mathrm{E}^*\left[g(S_T)B_tB_T^{-1}|\mathcal{F}_t\right] = \mathrm{E}^*\left[g(S_T)B_tB_T^{-1}|\mathcal{G}_t\right] = F^g(t, S_t),$$

where the function $F^g(t, s)$ satisfies the same second order PDE as in the complete case (2.8). Consequently, we arrive at the expression

$$V_t^* = (l_x - N_t)_{T-t}p_{x+t}B_t^{-1}F^g(t, S_t). \tag{4.3}$$

The process V^* can be interpreted as the market value process associated with the entire portfolio of pure endowment contracts, using the pricing rule P^*. In particular, the initial value $V_0^* = l_x {}_Tp_x F^g(0, S_0)$ is a natural candidate for the single premium for the entire portfolio. This specific choice of single premium would be in accordance with the well established actuarial principle of equivalence (stating that premiums and benefits should balance on average), but exercised under the martingale measure P^*

Applying the Itô formula to (4.3), we get

$$V_t^* = V_0^* + \int_0^t (l_x - N_{u-}) B_u^{-1} F^g(u, S_u)_{T-u}p_{x+u}\mu_{x+u} du$$
$$+ \int_0^t (l_x - N_{u-})_{T-u}p_{x+u} d(B_u^{-1} F^g(u, S_u)) + \sum_{0<u\leq t} (V_u^* - V_{u-}^*).$$

To determine the integral involving $d(B_t^{-1} F^g(t, S_t))$, recall the definition of the deflated price process $S_t^* = S_t B_t^{-1}$, implying that

$$dS_t = S_t^* dB_t + B_t dS_t^* = S_t r_t dt + B_t dS_t^*.$$

Using the Itô-formula and the PDE (2.8), it is seen that

$$d(B_t^{-1} F^g(t, S_t)) = -r(t, S_t) B_t^{-1} F^g(t, S_t) dt$$
$$+ B_t^{-1} \left(F_t^g(t, S_t) dt + F_s^g(t, S_t) dS_t + \frac{1}{2} F_{ss}^g(t, S_t) \sigma(t, S_t)^2 S_t^2 dt \right)$$
$$= F_s^g(t, S_t) dS_t^*$$

Also, since

$$\sum_{0<u\leq t} (V_u^* - V_{u-}^*) = -\int_0^t B_u^{-1} F^g(u, S_u)_{T-u}p_{x+u}\, dN_u,$$

we obtain.

Lemma 4.1 *For the contingent claim H in (4.1) the process V^** defined by $V_t^* = \mathrm{E}^*[H|\mathcal{F}_t]$ *has the decomposition*

$$V_t^* = V_0^* + \int_0^t \xi_u^H dS_u^* + \int_0^t \nu_u^H dM_u,$$

where (ξ^H, ν^H) are given by

$$\xi_t^H = (l_x - N_{t-})_{T-t}p_{x+t} F_s^g(t, S_t), \tag{4.4}$$
$$\nu_t^H = -B_t^{-1} F^g(t, S_t)_{T-t}p_{x+t}, \quad 0 \leq t \leq T. \tag{4 5}$$

Admissible strategies minimizing the variance

$$\mathrm{E}^*\left[(C_T^\varphi - \mathrm{E}^*[C_T^\varphi])^2\right] \tag{4.6}$$

can now be characterized by applying Theorem 3.2 and Lemma 4.1. By use of the Fubini theorem, the associated minimum obtainable variance is rewritten as

$$\begin{aligned}
\mathrm{E}^*\left[\left(\int_0^T \nu_u^H dM_u\right)^2\right] &= \mathrm{E}^*\left[\int_0^T (\nu_u^H)^2 d\langle M\rangle_u\right] \\
&= \mathrm{E}^*\left[\int_0^T \left(B_u^{-1}F^g(u,S_u)\,{}_{T-u}p_{x+u}\right)^2 \lambda_u du\right] \\
&= \int_0^T \mathrm{E}^*\left[\left(B_u^{-1}F^g(u,S_u)\right)^2\right] {}_{T-u}p_{x+u}^2 \mathrm{E}^*[(l_x - N_u)\mu_{x+u}]\, du \\
&= \int_0^T \mathrm{E}^*\left[\left(B_u^{-1}F^g(u,S_u)\right)^2\right] {}_{T-u}p_{x+u}^2\, l_x\, {}_up_x\, \mu_{x+u}\, du \\
&= l_x\, {}_Tp_x \int_0^T \mathrm{E}^*\left[\left(B_u^{-1}F^g(u,S_u)\right)^2\right] {}_{T-u}p_{x+u}\, \mu_{x+u}\, du \qquad (4\ 7)
\end{aligned}$$

Thus we have obtained

Theorem 4.2 *Consider the pure endowment given by the contingent claim H in (4 1). Admissible strategies φ^* minimizing the variance (4 6) are determined by*

$$\begin{aligned}
\xi_t^* &= (l_x - N_{t-})\ {}_{T-t}p_{x+t}F_x^g(t,S_t),\ 0 \le t \le T, \\
\eta_T^* &= H - \xi_T^* S_T^* .
\end{aligned}$$

The minimal variance is given by (4.7)

The insurance company is able to reduce the total risk associated with the portfolio of unit-linked insurance contracts to the "intrinsic risk" R_0^φ, by following a strategy according to Theorem 4 2 which also satisfies $C_0^\varphi = \mathrm{E}^*[H]$ In particular, it is seen that R_0^φ is proportional to l_x, implying that the ratio between $\sqrt{R_0^\varphi}$ and l_x converges to 0 as l_x converges to infinity

Before determining the unique risk-minimizing strategy, we present one specific strategy from Theorem 4 2, see Follmer and Sondermann (1986, Example 1).

Example 4.3 We shall present one strategy φ that does not require any extra investments during the time interval (0, T). It is self-financing on (0, T), followed by a possible extra payment at time T Define the strategy by

$$\xi_t = \xi_t^H,\ 0 \le t \le T, \qquad (4.8)$$

$$\eta_t = \mathrm{E}^*[H] + \int_0^t \xi_u dS_u^* - \xi_t S_t^*,\ 0 \le t < T, \qquad (4.9)$$

and $\eta_T = H - \xi_T S_T^*$. By definition, this strategy is self-financing on the interval $(0, T)$. Substituting the decomposition of H from Lemma 4.1 into the expression of η_T, we get

$$\eta_T = H - \xi_T S_T^* = \mathrm{E}^*[H] + \int_0^T \xi_u^H dS_u^* + \int_0^T \nu_u^H dM_u - \xi_T S_T^*$$

Likewise we have from (4.9) that

$$\eta_{T-} = \mathrm{E}^*[H] + \int_0^{T-} \xi_u dS_u^* - \xi_{T-} S_{T-}^* = \mathrm{E}^*[H] + \int_0^T \xi_u dS_u^* - \xi_T S_T^*,$$

which proves that

$$\eta_T - \eta_{T-} = \int_0^T \nu_u^H dM_u = L_T^H$$

Thus, the loss L_T^H is an extra payment/investment to be made at time T in order to satisfy the condition of admissibility.

The variance-minimizing trading strategy in Example 4.3 represents a very simple dynamic portfolio strategy from the point of view of the insurer. According to this strategy he is to make an initial investment at time 0 in stocks and bonds. During the time interval $(0, T)$ this portfolio is then adjusted continuously without any additional inflow or outflow of capital as defined by the equations (4.8)-(4.9) At the term T the insurance company now provides the difference L_T^H between the claim H and the value V_{T-}^φ of the portfolio However, there are reasons why this strategy should not be applied. Indeed, it does minimize the variance or the initial intrinsic risk, but at any time t during the insurance period the value V_t^φ of the portfolio will in general not equal the conditional expected present value of the claim V_t^*. Since this difference may be substantial due to adverse development within the insurance portfolio, one should at least require that the value of the portfolio equals V_t^* in order to enhance the solvency of the insurer This additional requirement, in addition with the minimal variance criterion, is actually sufficient to determine the unique risk-minimizing strategy φ. The associated intrinsic risk process is described in Theorem 3 3, and we get

$$\begin{aligned}\mathrm{E}^*\left[(L_T^H - L_t^H)^2 \mid \mathcal{F}_t\right] &= \mathrm{E}^*\left[\left(\int_t^T \nu_u^H dM_u\right)^2 |\mathcal{F}_t\right] = \mathrm{E}^*\left[\int_t^T (\nu_u^H)^2 \lambda_u du|\mathcal{F}_t\right] \\ &= \int_t^T \mathrm{E}^*\left[(\nu_u^H)^2|\mathcal{F}_t\right]\mathrm{E}^*\left[(l_x - N_u)\mu_{x+u}|\mathcal{F}_t\right] du \\ &= (l_x - N_t)\int_t^T \mathrm{E}^*\left[(\nu_u^H)^2|\mathcal{F}_t\right] {}_{u-t}p_{x+t}\, \mu_{x+u}\, du. \qquad (4.10)\end{aligned}$$

From Theorem 3.3 we now have:

Theorem 4.4 *For the pure endowment given by the contingent claim (4.1) the unique admissible risk-minimizing strategy is given by*

$$\xi_t^* = (l_x - N_{t-})\ {}_{T-t}p_{x+t} F_s^g(t, S_t),$$
$$\eta_t^* = (l_x - N_t)\ {}_{T-t}p_{x+t} B_t^{-1} F^g(t, S_t) - \xi_t^* S_t^*,\ 0 \leq t \leq T$$

The intrinsic risk process R^{φ^*} is given by (4.10)

In the model the insurance company is allowed to trade the assets S and B continuously, thus being able to hedge all contingent claims involving these assets only. This eliminates a part of the total uncertainty, leaving only the uncertainty of "not knowing how many of the insured persons will die in the insurance period". The latter is described by the martingale M, which generates the insurer's loss L^H.

$$dL_t^H = \nu_t^H dM_t = -B_t^{-1} F^g(t, S_t)\ {}_{T-t}p_{x+t}(dN_t - \lambda_t dt). \qquad (4.11)$$

The insurer adjusts his trading strategy according to the conditional expected number of insured persons surviving the insurance period. During the infinitesimal time interval $[t, t + dt)$ the insurer will experience the gain dM_t multiplied by the term $B_t^{-1} F^g(t, S_t)\ {}_{T-t}p_{x+t}$, the latter denoting the price at time t of one security with payment $g(S_T)$ at time T contingent on the survival of some individual That is, a death will produce an immediate gain for the insurer due to the downwards adjustment of the expected number of survivors, whereas no deaths will cause a small loss The expression (4.11) for the loss is similar to the one obtained by Norberg (1992) for general payment streams, using a quite different approach. With this terminology, the term $(\nu_t^H B_t)$ is recognized as the *sum at risk* at time t.

We now turn to some examples in the case of constant deterministic short rate of interest, constant drift term α, and volatility parameter σ on S. We will investigate three different contract functions: pure unit-linked, where $g(s) = s$; unit-linked with guarantee, where $g(s) = \max(s, K)$; and the case of deterministic benefits, $g(s) = K$

Example 4.5 Consider a standard Black-Scholes market, where all coefficients r, α and σ are constant. Let the contract function be of the simple form $g(s) = s$, i e. the insured is to be paid the value of the stock at the maturity date. In this case, the process $(F^g(t, S_t))_{0 \leq t \leq T}$ is easily determined as

$$F^g(t, S_t) = \mathrm{E}^*\left[e^{-r(T-t)} S_T | \mathcal{F}_t\right] = S_t,$$

implying that $F^g_S(t, S_t) = 1$. The intrinsic value process is

$$V^*_t = (l_x - N_t)\ {}_{T-t}p_{x+t}\ e^{-rt} S_t = (l_x - N_t)\ {}_{T-t}p_{x+t}\ S^*_t,$$

and in particular $V^*_0 = l_x\ {}_Tp_x\ S^*_0$. From Theorem 4.4 we have the unique risk-minimizing strategy

$$(\xi_t, \eta_t) = ((l_x - N_{t-})\ {}_{T-t}p_{x+t}, -\Delta N_t\ {}_{T-t}p_{x+t} S^*_t)\ , \tag{4.12}$$

where $\Delta N_t = N_t - N_{t-}$. Finally, we have the aggregated loss

$$L^H_T = -\int_0^T S^*_u\ {}_{T-u}p_{x+u}\ dM_u,$$

and the intrinsic risk process

$$\begin{aligned} R^\varphi_t &= (l_x - N_t)\ {}_{T-t}p_{x+t} \int_t^T \mathrm{E}^*\left[(S^*_u)^2 \middle| \mathcal{F}_t\right]\ {}_{T-u}p_{x+u}\ \mu_{x+u}\ du \\ &= (l_x - N_t)\ {}_{T-t}p_{x+t} (S^*_t)^2 \int_t^T e^{\sigma^2(u-t)}\ {}_{T-u}p_{x+u}\ \mu_{x+u}\ du \end{aligned}$$

The risk-minimizing strategy given by (4.12) is easy to interpret: at any time t the insurance company should hold a number of stocks, corresponding to the expected number of surviving individuals Since the number of stocks is controlled by a predictable process ξ, some adjustments are made each time a death occur within the portfolio in order to ensure that $V^*_t = V^\varphi_t$ for all t. This is described by the adapted process η, which denotes the amount to be cashed by the insurance company in connection with the observed death.

Example 4.6 Now consider the contract function $g(s) = \max(s, K)$, where K is some non-negative constant. Note, that $K = 0$ is just the case treated above in Example 4.5 As in the previous example, prices are described by a standard Black-Scholes market.

Writing the contract function $\max(s, K)$ on the form $K + (s - K)^+$, the process $(F^g(t, S_t))_{0 \le t \le T}$ can be evaluated by means of the well-known Black-Scholes formula

$$\begin{aligned} F^g(t, S_t) &= \mathrm{E}^*\left[e^{-r(T-t)}(K + (S_T - K)^+) \middle| \mathcal{F}_t\right] \\ &= Ke^{-r(T-t)} + \left(S_t\Phi(z_t) - Ke^{-r(T-t)}\Phi\left(z_t - \sigma\sqrt{T-t}\right)\right) \\ &= Ke^{-r(T-t)}\Phi\left(-z_t + \sigma\sqrt{T-t}\right) + S_t\Phi(z_t), \end{aligned} \tag{4.13}$$

where Φ is the standard normal distribution function and

$$z_t = \frac{\log(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\ .$$

In particular, the first order partial derivative is $F^g_s(t, S_t) = \Phi(z_t)$. Thus, the risk-minimizing strategy is given by

$$\xi_t = (l_x - N_{t-})\ {}_{T-t}p_{x+t}\Phi(z_t), \tag{4.14}$$

$$\begin{aligned}\eta_t &= (l_x - N_t)\ {}_{T-t}p_{x+t}\ e^{-rt}F^g(t, S_t) - (l_x - N_{t-})\ {}_{T-t}p_{x+t}\Phi(z_t)S^*_t \\ &= (l_x - N_t)\ {}_{T-t}p_{x+t}\ Ke^{-rT}\Phi\Big(-z_t + \sigma\sqrt{T-t}\Big) \\ &\quad - \Delta N_t\ {}_{T-t}p_{x+t}\Phi(z_t)S^*_t,\end{aligned} \tag{4.15}$$

and the intrinsic risk process R^φ is now given by

$$R^\varphi_t = (l_x - N_t)\ {}_{T-t}p_{x+t}\int_t^T \mathrm{E}^*\Big[(e^{-ru}F^g(u, S_u))^2\big|\mathcal{F}_t\Big]\ {}_{T-u}p_{x+u}\ \mu_{x+u}\,du,$$

with F^g defined by (4.13).

Example 4.7 As a last example, consider the case of deterministic benefits, that is $g(S_T) = K$ for some non-negative K. Here, the risk-minimizing strategy is given by

$$(\xi_t, \eta_t) = \Big(0,\ (l_x - N_t)\ {}_{T-t}p_{x+t}Ke^{-rT}\Big)\ , \tag{4.16}$$

and the intrinsic risk process is

$$\begin{aligned}R^\varphi_t &= (l_x - N_t)\ {}_{T-t}p_{x+t}\int_t^T K^2e^{-2rT}\ {}_{T-u}p_{x+u}\ \mu_{x+u}\,du \\ &= (l_x - N_t)\ {}_{T-t}p_{x+t}(1 - {}_{T-t}\,p_{x+t})K^2e^{-2rT}.\end{aligned}$$

In Example 4.5-4.7, we have determined risk-minimizing strategies for three different contract functions, in the setting of a standard Black-Scholes market. The strategies are associated with an entire portfolio l_x; single-life strategies are obtained by specializing to $l_x = 1$. For example, the strategy (4.14)-(4.15) for a single life becomes

$$\xi_t = I(T_1 \geq t)\ {}_{T-t}p_{x+t}\Phi(z_t), \tag{4.17}$$

$$\begin{aligned}\eta_t &= I(T_1 > t)\ {}_{T-t}p_{x+t}Ke^{-rT}\Phi\Big(-z_t + \sigma\sqrt{T-t}\Big) \\ &\quad - I(T_1 = t)\ {}_{T-t}p_{x+t}\Phi(z_t)S^*_t,\end{aligned} \tag{4.18}$$

and the intrinsic value process is

$$V^*_t = I(T_1 > t)\ {}_{T-t}p_{x+t}\Big(Ke^{-rT}\Phi\Big(-z_t + \sigma\sqrt{T-t}\Big) + S^*_t\Phi(z_t)\Big)$$

The process V^* is in a sense similar to a traditional prospective reserve. First, an indicator function appears, which guarantees that the reserve is only different from zero as long as the policyholder is still alive. The rest of the terms are interpreted as the conditional expected present value of the insurance benefit, given the policyholder is alive at t. Provided that the policyholder survives to the maturity date, that is $T_1 > T$, the risk-minimizing strategy (4 17)-(4 18) for a single life reduces to the strategy

$$(\xi_t^0, \eta_t^0) = \left({}_{T-t}p_{x+t}\Phi(z_t),\ {}_{T-t}p_{x+t}Ke^{-rT}\Phi\left(-z_t + \sigma\sqrt{T-t}\right)\right),$$

which is exactly equal to the corresponding duplicating strategy obtained by Aase and Persson (1994). The result (4.17)-(4.18) is to be interpreted as follows: As long as the policyholder is alive, the insurance company should hold a portfolio, where the number of stocks is determined as the probability ${}_{T-t}p_{x+t}$ of survial to T conditioned on survival to t times the factor $\Phi(z_t)$; the latter is recognized as the hedge from the Black-Scholes formula of a European Call Option. If the policyholder dies before the maturity date T, the insurer immediately cashes the reserve, as is apparent in the definition of η. These interpretations are easily carried over to the situation where the insurance portfolio consists of more than one individual In this case, the numbers of stocks and bods held are adjusted in accordance with the conditional expected number of survivors to T, that is $(l_x - N_t)_{T-t}p_{x+t}$ Thus, the risk-minimizing strategies reflect the actual development in the insurance portfolio, and bring to the surface the uncertainty associated with the insured lives. For example, we obtain expressions for the intrinsic risk processes, which serve as characterizations of the non-hedgeable risk inherent in a portfolio of unit-linked contracts. In Section 6 we present some numerical results in the set-up of Examples 4.5 and 4.6 obtained by Monte Carlo simulation.

4.2. Term insurance

Now consider the term insurance with single premium π^s paid at time 0. The payments generated by this contract are described by the discounted claim

$$H_T = \int_0^T g(u, S_u)B_u^{-1}dN_u \tag{4.19}$$

An important step is the construction of the decomposition for the intrinsic value process for H_T First of all, observe that

$$V_t^* = \mathrm{E}^*[H_T|\mathcal{F}_t] = \int_0^t g(u, S_u)B_u^{-1}dN_u + \mathrm{E}^*\left[\int_t^T g(u, S_u)B_u^{-1}dN_u|\mathcal{F}_t\right]$$
$$= \int_0^t g(u, S_u)B_u^{-1}dN_u + \int_t^T B_t^{-1}F^{g_u}(t, S_t)(l_x - N_t)\ {}_{u-t}p_{x+t}\,\mu_{x+u}\,du,$$

where

$$F^{g_u}(t,S_t) = \mathrm{E}^*\left[e^{-\int_t^u r_\tau d\tau} g(u,S_u) \mid \mathcal{G}_t\right]$$

is the unique arbitrage-free price at time t of the simple u-claim $g(u,S_u)$ in the complete model with filtration $\mathbf{G}$. Secondly, by calculations similar to the ones in the previous section, we see that

$$d\left(B_t^{-1}F^{g_u}(t,S_t)\right) = F_s^{g_u}(t,S_t)dS_t^*$$

Using the general Itô formula and the Fubini Theorem for Itô processes, see Ikeda and Watanabe (1981), V^* can now be rewritten as

$$\begin{aligned}
V_t^* = V_0^* &+ \int_0^t \left(-B_\tau^{-1}F^{g_\tau}(\tau,S_\tau)\mu_{x+\tau}(l_x - N_\tau)\right) d\tau \\
&+ \int_0^t \left(g(\tau,S_\tau)B_\tau^{-1} - \int_\tau^T B_\tau^{-1}F^{g_u}(\tau,S_\tau)\ {}_{u-\tau}p_{x+\tau}\mu_{x+u}du\right) dN_\tau \\
&+ \int_0^t \left(\int_\tau^T B_\tau^{-1}F^{g_u}(\tau,S_\tau)\ {}_{u-\tau}p_{x+\tau}\mu_{x+u}du\right)(l_x - N_{\tau-})\mu_{x+\tau}d\tau \\
&+ \int_0^t \left((l_x - N_{\tau-})\int_\tau^T F_s^{g_u}(\tau,S_\tau)\ {}_{u-\tau}p_{x+\tau}\mu_{x+u}du\right) dS_\tau^*.
\end{aligned}$$

Upon gathering terms, and using $F^{g_t}(t,S_t) = g(t,S_t)$, we obtain a decomposition corresponding to Lemma 4 1:

Lemma 4.8 *For the claim H_T in (4.19) the process V^** defined by $V_t^* = \mathrm{E}^*[H_T|\mathcal{F}_t]$ *has the decomposition*

$$V_t^* = V_0^* + \int_0^t \xi_u^H dS_u^* + \int_0^t \nu_u^H dM_u,$$

where (ξ^H, ν^H) are given by

$$\xi_t^H = (l_x - N_{t-})\int_t^T {}_{u-t}p_{x+t}\,\mu_{x+u}\,F_s^{g_u}(t,S_t)du, \tag{4.20}$$

$$\nu_t^H = g(t,S_t)B_t^{-1} - \int_t^T F^{g_u}(t,S_t)B_t^{-1}\,{}_{u-t}p_{x+t}\,\mu_{x+u}\,du \tag{4.21}$$

Using Theorem 3.3 we have now proved:

Theorem 4.9 *For the term insurance given by the contingent claim (4.19) the unique admissible risk-minimizing strategy is given by*

$$\xi_t^* = (l_x - N_{t-}) \int_t^T F_x^{g_u}(t, S_t)\, {}_{u-t}p_{x+t}\, \mu_{x+u}\, du,$$

$$\eta_t^* = \int_0^t g(u, S_u) B_u^{-1} dN_u + (l_x - N_t) \int_t^T B_t^{-1} F^{g_u}(t, S_t)\, {}_{u-t}p_{x+t}\, \mu_{x+u}\, du$$
$$- \xi_t^* S_t^*, \ 0 \leq t \leq T.$$

The intrinsic risk process R^{φ^*} is given by

$$R_t^{\varphi^*} = (l_x - N_t) \int_t^T \mathrm{E}^*\left[(\nu_u^H)^2 | \mathcal{F}_t\right] {}_{u-t}p_{x+t}\, \mu_{x+u}\, du,$$

where ν^H is taken from (4.21).

To give the resulting portfolio an interpretation, note that $\varphi = (\xi, \eta)$ is determined such that

$$V_t^{\varphi} = \int_0^t g(u, S_u) B_u^{-1} dN_u + \mathrm{E}^*\left[\int_t^T g(u, S_u) B_u^{-1} dN_u | \mathcal{F}_t\right].$$

Thus, V_t^{φ} is determined as the sum of the benefits set aside to deaths already occurred and the expected discounted value of payments associated with future deaths

As in the case of the pure endowment, the term ν_t^H denotes the immediate loss due to the death of one of the insured persons. Here, the insurer has to set aside the sum insured $g(t, S_t)$ immediately upon a death within the portfolio at time t. In connection with the incurred death, the insurance company adjusts its expectations regarding the further development of the insurance portfolio. Since the number of survivors has been reduced by one, the insurer now reduces his reserves by the amount

$$\int_t^T F^{g_u}(t, S_t) B_t^{-1}\, {}_{u-t}p_{x+t}\, \mu_{x+u}\, du,$$

which is the expected discounted value of future payments conditional on survival to time t.

Example 4.10 Consider a unit-linked term insurance contract with guarantee in the case of a standard Black-Scholes market. Let the contract function be on the form $g(u, s) = \max(s, Ke^{\delta u})$, that is the guarantee is adjusted in accordance with some constant force of inflation δ The functions $F^{g_u}(t,s)$ are determined by

$$F^{g_u}(t,S_t) = Ke^{\delta u}e^{-r(u-t)}\Phi\left(-z_t^{(u)} + \sigma\sqrt{u-t}\right) + S_t\Phi(z_t^{(u)}), \tag{4.22}$$

with

$$z_t^{(u)} = \frac{\log(S_t/Ke^{\delta u}) + (r+\sigma^2/2)(u-t)}{\sigma\sqrt{u-t}}.$$

Using Theorem 4.9 we find the risk-minimizing strategy

$$\xi_t = (l_x - N_{t-})\int_t^T {}_{u-t}p_{x+t}\,\mu_{x+u}\Phi(z_t^{(u)})du,$$

$$\eta_t = (l_x - N_t)\int_t^T {}_{u-t}p_{x+t}\,\mu_{x+u}Ke^{-(r-\delta)u}\Phi(-z_t^{(u)} + \sigma\sqrt{u-t})\,du$$

$$+ \int_0^t g(u,S_u)B_u^{-1}dN_u - \Delta N_t\int_t^T {}_{u-t}p_{x+t}\,\mu_{x+u}\Phi(z_t^{(u)})S_t^*du.$$

The intrinsic risk process is also determined by that theorem upon inserting the functions F^{g_u} from (4 22) in (4.21).

5. Extending the financial market

In the previous sections we have analyzed a model where the financial market consists of two assets only, namely a risk-free asset B (the bond) and a risky asset S (the stock). That model, which also describes the development of a given portfolio of insured lives, is incomplete. We considered two different basic types of insurance products, and in both cases risk-minimizing strategies were constructed and the corresponding intrinsic risk processes were determined. Due to incompleteness, the risk could not be eliminated completely and thus some uncertainty regarding the course of the insured lives in the portfolio (the intrinsic risk) remains with the insurance company.

The present section is devoted to a brief investigation of the situation where the financial market is extended by a third tradeable asset that is related to the specific insured lives. As in Section 4, focus will be on the pure endowment, but all results can be repeated for the term insurance and the endowment insurance as well. Furthermore we restrict the analysis to the case where the risk-free interest rate r is assumed to be constant.

In addition to the assets (B, S) with prices processes defined by (2 1) and (2.2), respectively, we introduce an asset with price process $Z = (Z_t)_{0\le t\le T}$, where

$$Z_t = (l_x - N_t)\,{}_{T-t}p_{x+t}\,e^{-r(T-t)}. \tag{5.1}$$

The initial value $Z_0 = l_x \, {}_Tp_x \, e^{-rT}$ is equal to the price at time 0 of l_x standard pure endowment contracts with sum insured 1 calculated on a valuation basis consisting of the mortality hazard function μ_x and the risk-free interest rate r. Assuming that premiums are paid as a single premium at time $0, Z_t$ represents, at any time $0 \leq t \leq T$, the traditional prospective reserve for the portfolio. This reserve is calculated as the conditional expected value of future benefits, given the current number of survivors $(l_x - N_t)$. The introduction of this extra investment possibility is motivated by the existence of reinsurance markets, where the direct insurer is able to reduce his total risk by selling some part of the insurance portfolio. Trading on the reinsurance markets will typically be controlled by certain restrictions such as short-selling constraints and upper limits for the amount reinsured. However, in the present formulation we do not impose any restrictions on the trading of any of the three assets

As an example, let us now consider an insurer facing the contingent claim arising from the portfolio of pure endowment unit-linked contracts with sum insured $g(S_T)$ for the portfolio, that is

$$H = (l_x - N_T) B_T^{-1} g(S_T), \tag{5 2}$$

and assume that the insurer is allowed to trade continuously on the extended market (B, S, Z) Note that the asset Z depends on the uncertainty from the insured lives only and evolves independently of the other assets (B, S). The insurance claim H, however, depends on both sources of uncertainty.

Define the deflated price processes S^* and Z^* by $S^* = S/B$ and $Z^* = Z/B$, respectively. In this new setup a trading strategy is a sufficiently integrable process $\varphi = (\xi, \vartheta, \eta)$, where ξ and ϑ are $\mathbf{F}$-predictable and η is $\mathbf{F}$-adapted. At any time t, ϑ_t, ξ_t and η_t are the number of units held of standard pure endowment contracts, stocks, and bonds respectively, and the (discounted) value process V^φ is now given by

$$V_t^\varphi = \xi_t S_t^* + \vartheta_t Z_t^* + \eta_t$$

We set out by verifying that the measure P^* defined by (2 3) is a martingale measure for S^* and Z^*. It already follows from the calculations in Section 4 that S^* is an $(\mathbf{F}, P^*)$-martingale, and the process Z^* is obviously also an $(\mathbf{F}, P^*)$-martingale, since

$$(l_x - N_t) \, {}_{T-t}p_{x+t} = \mathrm{E}^*[(l_x - N_T)|\mathcal{F}_t] \, .$$

From the decomposition for the intrinsic value process V^* for (5.2) and a similar representation result for Z^* with respect to M, we obtain

$$V_t^* = V_0^* + \int_0^t \xi_u^H dS_u^* + \int_0^t \vartheta_u^H dZ_u^*,$$

with

$$(\xi_t^H, \vartheta_t^H) = \Big((l_x - N_{t-})\,{}_{T-t}p_{x+t}\,F_s^g(t, S_t),\; e^{r(T-t)}F^g(t, S_t) \Big). \qquad (5.3)$$

The intrinsic value process V^* has now been rewritten as a sum of two integrals with respect to the price processes S^* and Z^* This implies that the contingent claim H associated with the pure endowment can be replicated by means of self-financing strategies in terms of the three assets (B, S, Z). We can summarize this result by

Theorem 5.1 *Consider the pure endowment with present value (5.2) and assume that standard pure endowment contracts with sum insured 1 are traded freely on a financial market with constant short rate of interest. A self-financing admissible (risk-minimizing) strategy φ^* is given by*

$$\xi_t^* = (l_x - N_{t-})\,{}_{T-t}p_{x+t}F_s^g(t, S_t), \qquad (5.4)$$

$$\vartheta_t^* = e^{r(T-t)}F^g(t, S_t), \qquad (5.5)$$

$$\eta_t^* = V_t^* - \xi_t^* S_t^* - \vartheta_t^* Z_t^*, \quad 0 \le t \le T \qquad (5.6)$$

Furthermore, the intrinsic risk process R^{φ^*} is identically 0.

The insurer is now able to eliminate the risk associated with the insurance claims completely by following a strategy in accordance with Theorem 5 1 According to this result, the insurer should not only adjust the portfolio of stocks and bonds continuously – also the portfolio of reinsurance contracts should be continuously rebalanced. By some simple calculations involving (5.4) and (5.5), formula (5.6) can be rewritten as

$$\eta_t^* = -(l_x - N_{t-})_{T-t}p_{x+t}F_s^g(t, S_t)S_t^* = -\xi_t^* S_t^*$$

Furthermore, φ^* satisfies $V_t^* = \vartheta_t^* Z_t^*$. Thus, the self-financing (and risk-minimizing) strategy consists of a number ϑ^* of shares of standard pure endowment contracts on the portfolio of insured lives, which is adjusted such that the value $\vartheta_t^* Z_t$ exactly equals the intrinsic value process V_t^* at any time $t \in [0, T]$ When allowing trading of reinsurance contracts, the criterion of risk-minimization simply states that all risk should be surrendered to the reinsurer. Furthermore, the number of stocks ξ^* to be held is the same as in the situation where standard insurance contracts are not traded. By the above calculations, we see that this position is financed by an equivalent short position η^* in the risk-free asset, that is, $\eta_t^* = -\xi_t^* S_t^*$.

We end this section by mentioning that P^* would not be a martingale measure for Z^* had we defined the price process $Z = (Z_t)_{0 \le t \le T}$ by

$$Z_t = (l_x - N_t)_{T-t}p_{x+t}e^{-\delta(T-t)}.$$

Here, the risk-free interest rate r has been replaced by some first order interest rate $\delta \neq r$. In this case, a martingale measure $\hat{P}$ for (Z^*, S^*) could be defined by (2.15) with $h_t = (\delta - r)/\mu_{x+t}$, provided that $h_t > -1$ for all t. This, in turn, would impose unique arbitrage-free prices for the unit-linked contracts that differ from those computed using the minimal martingale measure P^*

6. Numerical results

We round off by presenting some Monte Carlo simulation results We consider the pure endowment where the sum insured is due at the maturity date if the insured is then still alive. Premiums are assumed to be paid as a single premium at time 0 The contract functions from Example 4.5-4 6 will then be examined by evaluating the initial value of the intrinsic risk process V_0^*, the initial intrinsic risk R_0 and the risk-increase associated with some simple (piecewise constant) strategies. Since these quantities are proportional to the size of the portfolio l_x, recall e.g. (4.3) and (4.10), we consider an insurance portfolio consisting of only one individual, that is, we take $l_x = 1$. Furthermore we take the age of the policyholder to be $x = 45$ upon issue of the contract, and fix the term of the contract to be $T = 15$ years. We use the Gompertz-Makeham hazard function as mortality law of the policyholder

$$\mu_{x+t} = 0\ 0005 + 0.000075858 \quad 1.09144^{x+t}, \qquad t \geq 0,$$

which is used in the Danish 1982 technical basis for men. With this mortality law, the conditional probability ${}_{15}p_{45}$ of surviving another 15 years given survival to age 45 is 0 8796 The basic financial market is standard Black-Scholes with parameters $\sigma = 0\ 25$ and $r = 0.06$, that is, the deterministic risk-free interest is 6% and the volatility of the stock is 25% Furthermore, we take $S_0 = 1$ and $B_0 = 1$ The importance of the volatility parameter is illustrated by considering, in addition, the case of small market volatility ($\sigma = 0\ 15$) and large market volatility ($\sigma = 0\ 35$).

The value at time 0 of the intrinsic value process V^*, given by

$$V_0^* = l_x \; {}_Tp_x F^g(0, S_0), \tag{6.1}$$

is evaluated by simply inserting the parameters (r, σ) and $S_0 = 1$ in the function F^g determined in Example 4 5 and 4.6. Results are listed in Table 1 for different choices of guarantees, the pure unit-linked insurance corresponds to guarantee $K = 0$ The initial intrinsic risk R_0 is given by

$$R_0 = \mathrm{E}^*\left[l_x \; {}_Tp_x \int_0^T \left(e^{-ru} F^g(u, S_u)\right)^2 \; {}_{T-u}p_{x+u} \, \mu_{x+u} \, du\right], \tag{6 2}$$

and since we have no explicit expression for the expected value of $(F^g(u, S_u))^2$, we apply Monte Carlo simulation combined with numerical integration in order to evaluate (6.2)

The price process for the stock S under P^*

$$S_t = e^{(r-\frac{1}{2}\sigma^2)t+\sigma W_t} \tag{6.3}$$

can be simulated by simply simulating a standard Brownian motion and inserting this in (6.3). Let $n = 100$ be the number of time intervals per time unit (one year) and denote by $\Delta t = 1/n$ the mesh of this partition. Also let M denote the number of paths of S to be simulated and let $\varepsilon_j^{(m)}$, $m = 1, .., M$, $j = 1, ..., T \cdot n$ be a sequence of simulated independent standard normal variables. The simulated versions $\hat{S}^{(m)}$ of (6.3) are determined as

$$\hat{S}_k^{(m)} = \exp\left((r-\frac{1}{2}\sigma^2)k\cdot\Delta t + \sum_{j=1}^{k}\sigma\sqrt{\Delta t}\,\varepsilon_j^{(m)}\right), \quad k=1, .., T\cdot n, m=1, .., M,$$

where $\hat{S}_k^{(m)}$ has same distribution as $S_{k\Delta t}$. The initial risk R_0 is now approximated numerically by applying Monte Carlo simulation for the integral (6.2) which is discretized using the so-called summed Simpson rule, see e.g. Schwarz (1989). In all computations we apply the step size $\Delta t = 1/100$. In Table 1 we have also presented the estimate for R_0 and the standard error on this estimate based on $M = 300000$ simulated paths for $\sigma = 0.15$ and 0.25 and $M = 500000$ for $\sigma = 0.35$.

TABLE I

THE INITIAL INTRINSIC VALUES AND RISKS ASSOCIATED WITH UNIT-LINKED PURE ENDOWMENT CONTRACTS FOR VARIOUS CHOICES OF GUARANTEE AND VOLATILITY

	Guarantee (K)	V_0^*	R_0	*(std.dev.)*	$\sqrt{R_0}/V_0^*$
$\sigma = 0.15$	0	0.8796	0.131		0.411
	0.5 exp(rT)	0.8996	0.134	(0.0002)	0.407
	exp(rT)	1.0807	0.173	(0.0002)	0.385
(M = 300000)	2 exp(rT)	1.7993	0.446	(0.0001)	0.371
$\sigma = 0.25$	0	0.8796	0.194	--	0.501
	0.5 exp(rT)	0.9580	0.205	(0.001)	0.474
	exp(rT)	1.2066	0.261	(0.001)	0.422
(M = 300000)	2 exp(rT)	1.9161	0.538	(0.001)	0.383
$\sigma = 0.35$	0	0.8796	0.365		0.687
	0.5 exp(rT)	1.0255	0.380	(0.005)	0.608
	exp(rT)	1.3213	0.449	(0.005)	0.513
(M = 500000)	2 exp(rT)	2.0511	0.743	(0.005)	0.423

The unrestricted risk-minimizing strategies are not applicable in practice, since they are based on the assumption of continuously adjustable portfolios. However, the expressions can be used as a guide in practical portfolio administration. For example, the insurer could apply a piecewise constant strategy on the form

$$\hat{\xi}_t = \sum_{j=1}^{J} I(t \in (t_{j-1}, t_j])\xi_{t_{j-1}}, \tag{6.4}$$

where ξ denotes the unrestricted risk-minimizing strategy determined in Section 4. Thus, the portfolio of stocks is adjusted at fixed times $0 = t_0 < t_1 < \dots < t_{J-1} < t_J = T$, as an approximation to the continuously adjustable risk-minimizing strategy. Here, we have chosen $t_j = j$ and $t_j = j/12$, which implies trading once a year and once a month, respectively. In Table 2, we have listed the risk-increase associated with the piecewise constant strategies (6.4), obtained by evaluating the expression

$$\sum_{j=1}^{J} \mathrm{E}^* \left[\int_{t_{j-1}}^{t_j} (\xi_u - \xi_{t_{j-1}})^2 \sigma^2 S_u^{*2} du \right].$$

In Møller (1996) optimal simple strategies are derived by means of some heuristic calculations.

TABLE 2

THE RISK INCREASE ASSOCIATED WITH SIMPLE STRATEGIES WITH YEARLY AND MONTHLY ADJUSTMENTS FOR UNIT-LINKED PURE ENDOWMENT CONTRACTS

	K	R_0	*Yearly*	*(std. dev.)*	*Monthly*	*(std. dev.)*
$\sigma = 0.15$	0	0.131	0.0015	–	0.00012	–
	0.5 exp(*rT*)	0.134	0.0014	$(1.5 \cdot 10^{-6})$	0.00012	$(1.3 \cdot 10^{-7})$
	exp(*rT*)	0.173	0.0011	$(1.6 \cdot 10^{-6})$	0.00009	$(1.3 \cdot 10^{-7})$
(M = 1000000)	2 exp(*rT*)	0.446	0.0004	$(1.4 \cdot 10^{-6})$	0.00003	$(1.1 \cdot 10^{-7})$
$\sigma = 0.25$	0	0.194	0.0060		0.00051	–
	0.5 exp(*rT*)	0.205	0.0058	$(1.9 \cdot 10^{-5})$	0.00050	$(1.6 \cdot 10^{-6})$
	exp(*rT*)	0.261	0.0051	$(1.9 \cdot 10^{-5})$	0.00044	$(1.6 \cdot 10^{-6})$
(M = 1000000)	2 exp(*rT*)	0.538	0.0040	$(1.9 \cdot 10^{-5})$	0.00034	$(1.6 \cdot 10^{-6})$
$\sigma = 0.35$	0	0.365	0.0225	–	0.00187	–
	0.5 exp(*rT*)	0.380	0.0218	$(3.1 \cdot 10^{-4})$	0.00186	$(2.6 \cdot 10^{-5})$
	exp(*rT*)	0.449	0.0209	$(3.1 \cdot 10^{-4})$	0.00178	$(2.6 \cdot 10^{-5})$
(M = 1000000)	2 exp(*rT*)	0.743	0.0193	$(3.1 \cdot 10^{-4})$	0.00160	$(2.6 \cdot 10^{-5})$

With volatility parameter $\sigma = 0.25$, the ratio between the square root of the initial intrinsic risk $\sqrt{R_0}$ and the intrinsic value process V_0^* is 0.5 for the pure unit-linked life insurance, see Table 1. By increasing the size l_x of the portfolio to 100, say, the corresponding ratio is reduced by the factor $\sqrt{100}/100 = 0.1$ to 0.05. As mentioned in the previous sections, V_0^* can be interpreted as a natural candidate for the single premium. In non-life insurance premiums are often increased by adding a safety loading, typically twice the standard deviation of the total liability. This procedure would lead to a safety loading about $2 \cdot 5\%$, that is 10% when $l_x = 100$. Furthermore, it is noted that the minimal risk associated with the simple strategy (6.4) with trading once per year is only 0.006 higher than the minimum obtainable risk $R_0 = 0.194$. This corresponds to an increase of 3.1%. Thus, the uncertainty associated with the death of the policyholders seems to be by far the most important.

The results obtained for the unit-linked contract with guarantee different from 0 indicate lower values of the ratio between the square root of the minimal obtainable risk R_0 and the intrinsic value process V_0^* than in the pure unit-linked case. Furthermore, the ratio seems to be decreasing as a function of the guaranteed amount. Also the relative risk increase associated with simple strategies is smaller than the corresponding results for the pure unit-linked life insurance. These properties could be partly explained by considering the exact form of the sum insured, described by the underlying derivative

$$\max(S_T, K) = K + (S_T - K)^+$$

Obviously, the probability of the European Call Option $(S_T - K)^+$ being in the money will converge to zero as K converges to infinity. In this way the relative uncertainty associated with the sum insured should decrease when the guaranteed amount increases.

Table 1 also gives indications of the consequences of possible misspecification of the volatility parameter σ. It is seen that all quantities listed here seem to be non-decreasing functions of the volatility. In particular, calculation of premiums based on the initial intrinsic value V_0^* only would neglect the increase in the ratio $\sqrt{R_0}/V_0^*$ as σ increases. Thus, this principle could result in premiums which are not adequate to cover the insurer's liabilities to the insured.

ACKNOWLEDGEMENT

The paper is based on my master's thesis at the University of Copenhagen. I am grateful to my supervisor Ragnar Norberg for proposing the problem and for helpful comments and fruitful discussions.

REFERENCES

AASE, K K and PERSSON, S A (1994) Pricing of Unit-linked Life insurance Policies, *Scandinavian Actuarial Journal* **1**, 26-52

BACINELLO, A R and ORTU, F (1993) Pricing equity-linked life insurance with endogenous minimum guarantees, *Insurance Mathematics and Economics* **12**, 245-257

BJORK, T (1996) *Arbitrage Theory in Continuous Time*, Stockholm School of Economics

BRENNAN, M J and SCHWARTZ, E S (1979) *Pricing and investment strategies for equity-linked life insurance*, Monograph no 7/The S S Huebner Foundation for insurance education, Wharton School, University of Pennsylvania, Philadelphia

DELBAEN, F (1990) Equity linked Policies, *Bulletin Association Royal Actuaires Belges*, 33-52

DUFFIE, D (1996) *Dynamic Asset Pricing Theory*, 2nd ed, Princeton University Press, Princeton, New Jersey

EKERN, S E and PERSSON, S-A (1996) Exotic Unit-Linked Life Insurance Contracts, *The Geneva Papers on Risk and Insurance Theory* **21**(1), 35-63

FOLLMER, H and SONDERMANN, D (1986) Hedging of non-redundant Contingent Claims, in *Contributions to Mathematical Economics*, eds W Hildenbrand and A Mas-Colell, 205-223

FOLLMER, H and SCHWEIZER, M (1988) Hedging by sequential regression An introduction to the mathematics of option trading, *The ASTIN Bulletin* **18**(2), 147-160

IKEDA, N and WATANABE, S (1981) *Stochastic Differential Equations and Diffusion Processes*, North-Holland

LAMBERTON, D and LAPEYRE, B (1996) *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall

MØLLER, T (1996) Risk-minimizing strategies for unit-linked life insurance products, Master's thesis, University of Copenhagen

MØLLER, T (1998) Risk-minimizing hedging strategies for general payment processes, Working paper, Laboratory of Actuarial Mathematics, University of Copenhagen (Forthcoming)

NIELSEN, J and SANDMANN, K (1995) Equity-linked Life Insurance A Model with Stochastic Interest Rates, *Insurance Mathematics and Economics* **16**, 225-253

NORBERG, R (1992) Hattendorff's Theorm and Thiele's Differential Equation Generalized, *Scandinavian Actuarial Journal* **1**, 2-14

NORBERG, R (1995) Differential equations for moments of present values in life insurance, *Insurance Mathematics and Economics* **17**, 171-180

NORBERG, R and MØLLER, C M (1996) Thiele's Differential Equation with Stochastic Interest of Diffusion Type, *Scandinavian Actuarial Journal* **1**, 37-49

SCHWARTZ, H R (1989) *Numerical Analysis A Comprehensive Introduction*, John Wiley & Sons Ltd

SCHWEIZER, M (1991) Option hedging for semimartingales, *Stochastic Processes and their Applications* **37**, 339-363

SCHWEIZER, M (1994) Risk-minimizing hedging strategies under restricted information, *Mathematical Finance* **4**, 327-342

SCHWEIZER, M (1995) On the Minimal Martingale Measure and the Follmer-Schweizer Decomposition, *Stochastic Analysis and Applications* **13**, 573-599

THOMAS MØLLER
Laboratory of Actuarial Mathematics
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
E-mail tmoller@math.ku.dk

WITHDRAWAL BENEFITS UNDER A DEPENDENT DOUBLE DECREMENT MODEL

BY JACQUES F. CARRIERE

Dept of Mathematical Sciences
University of Alberta

ABSTRACT

This article presents an explicit formula for the value of a withdrawal benefit when the times of death and withdrawal are dependent. The derivation is based on an actuarial equivalence principle. As a special case, we show that in the fully continuous case, the withdrawal benefit is the reserve when the decrements are independent We also present a definition of antiselection and prove that the withdrawal benefit will be smaller under antiselection.

KEYWORDS

Dependent decrement theory, withdrawal benefits, antiselection, the equivalence principle, varying life insurance.

1. INTRODUCTION

In some markets, like the United States, life insurance products have a withdrawal benefit when the policy is terminated. This article will examine the implications of dependent probabilities of withdrawal and death on withdrawal benefits for life insurance in discrete time. Specifically, we will give an explicit expression of the withdrawal benefit under a dependent decrement model thus allowing us to characterize the withdrawal benefit under antiselection

In the book, Actuarial Mathematics (1986), the authors state that "if the withdrawal benefit in a double decrement model whole life insurance, fully continuous payment basis, is the reserve under the single decrement model whole life insurance, the premium and reserves under the double decrement model are equal to the premium and reserves under the single decrement model " This incredible result is not always true. The reason that the reserve is not always equal to the withdrawal benefit was given by D.R. Schuette (reported by Nesbitt (1964)), who found the withdrawal benefit is not the reserve in the discrete model because "the probability of withdrawal depends on the force of mortality." Thus, this article delves into the issue of dependent mortality and withdrawal in a discrete model For an introduction to the mathematics of dependent decrement theory, consult Carriere (1994).

2. The Single-Decrement Model

In this section, we present the classical single-decrement model for life insurance pricing and reserving Let T_d denote the time and death for some life aged x. Next, let

$$S^d(t) = Pr(T_d > t), t \geq 0, \tag{1}$$

be the survival function of T_d. Throughout the discussion we will assume that this survival function is absolutely continuous with a density denoted as $f^d(t)$ and a force of mortality equal to $\mu^d(t) = f^d(t)/S^d(t)$. Thus $S^d(t) = \exp\{-\int_0^t \mu^d(z)dz\}$. Now consider the probability that the life survived to time $t+s$ given that it survived to time t. This survival function is denoted as ${}_sp_t^d$ and it is equal to.

$$ {}_sp_t^d = Pr(T_d > t+s | T_d > t) = \frac{S^d(t+s)}{S^d(t)} = \exp\left\{-\int_t^{t+s} \mu^d(z)dz\right\}. \tag{2}$$

It will be convenient to define $T_d(t)$ as the random variable induced by ${}_sp_t^d$ so that ${}_sp_t^d = Pr(T_d(t) > s)$ Note that $T_d = T_d(0)$ and ${}_tp_0^d = S^d(t)$. Moreover, if $T_d > t$ then $T_d(t) = T_d - t$, otherwise $T_d(t)$ is undefined. It is instructive to note that if the expectation $E[g(T_d(t))]$ exists for some function $g(s)$, then

$$E[g(T_d(t))] = E[g(T_d - t)|T_d > t]. \tag{3}$$

This last fact will be used repeatedly.

Usually we will assume that premiums and death benefits are paid at the discrete times $t = k/m$ where $k = 0, 1, 2, \ldots$ and $m > 0$ Therefore, it is convenient to define the discrete random variable

$$K_d^m = \frac{\lfloor mT_d \rfloor}{m}, \tag{4}$$

where $\lfloor . \rfloor$ is the floor function In other words, $\lfloor y \rfloor$ is the integer part of y. Thus, if $y \geq 0$ then $\lfloor y \rfloor = k$ if and only if $k \leq y < k+1$ Note that T_d is simply equal to K_d^m when $m = \infty$ and so any discussion about the continuous model is subsumed within the discrete model.

In this article, we assume that the life insurance has a varying death benefit equal to $b(t)$, if death occurs at time t. Typically, $b(t) = 1$ for all t

Next, let

$$v(t) = \exp\left\{-\int_0^t \delta_z dz\right\} \tag{5}$$

denote the interest discount function Traditionally, actuaries have assumed that $\delta_z = \delta$ is constant thus implying that $v(t) = e^{-\delta t}$ Using the equivalence principle and the functions $v(t)$ and $b(t)$, we define the net single premium at time t for the future benefits from the life insurance as

$$\begin{aligned} A_d^m(t) &= E[b(K_d^m + 1/m)v(K_d^m + 1/m)/v(t)|T_d > t] \\ &= E[b(\lfloor mT_d + 1\rfloor/m)v(\lfloor mT_d + 1\rfloor/m)/v(t)|T_d > t] \\ &= E[b(\lfloor m(T_d(t) + t) + 1\rfloor/m)v(\lfloor m(T_d(t) + t) + 1\rfloor/m)/v(t)] \end{aligned} \tag{6}$$

Note that the last equality in equation (6) follows from equation (3). Now, let us focus on the valuation of the premium payments. Assume that payments of $1/m$ are made at the times $t = 0, 1/m, 2/m, .$ Then the present value at time $t = 0$ of all the payments made in the period $[r,s)$ will be denoted as $a[r,s)$ and calculated as·

$$a[r,s) = \frac{1}{m}\sum_{k=0}^{\infty} v(k/m)1(r \le (k/m) < s) \tag{7}$$

In this definition, $1(e)$ is an indicator function that is equal to 1 if the event e is true and 0, otherwise. It is instructive, to verify that $a[r,s) = a[0,s) - a[0,r)$ Using this annuity-certain formula and the equivalence principle, we find that the net single premium for the future payments from the life annuity at time t is.

$$\begin{aligned} a_d^m(t) &= E[a[t, K_d^m + 1/m)/v(t)|T_d > t] \\ &= E[a[t, \lfloor mT_d + 1\rfloor/m)/v(t)|T_d > t] \\ &= E[a[t, \lfloor m(T_d(t) + t) + 1\rfloor/m/v(t)]. \end{aligned} \tag{8}$$

Under the single decrement model, the net level premium for the life insurance is denoted as P_d^m and it is equal to.

$$P_d^m = A_d^m(0)/a_d^m(0), \tag{9}$$

under the equivalence principle. Thus, we can define the *link* function as.

$$\mathcal{L}(r,s) = b(s)\, v(s)/v(r) - P_d^m a[r,s)/v(r), \tag{10}$$

This link function will be useful when the withdrawal benefit is derived for the double-decrement model in the next section. This link function can also be used to define the *prospective loss* at time t, which is·

$$L^m(t) = \mathcal{L}(t, K_d^m + 1/m) \tag{11}$$

Note that $E[L^m(0)] = 0$. Finally, we find that the *prospective reserve* at any time t is:

$$\begin{aligned} V^m(t) &= E[L^m(t) | T_d > t] = E[\mathcal{L}(t, \lfloor mT_d + 1 \rfloor / m) | T_d > t] \\ &= E[\mathcal{L}(t, \lfloor m(T_d(t) + t) + 1 \rfloor / m)] = A_d^m(t) - P_d^m a_d^m(t) \end{aligned} \tag{12}$$

Note that the random variable $\lfloor m(T_d(t) + t) + 1 \rfloor / m$ has a central role. Thus it will be convenient to define

$$\mathcal{K}_d^m(t) = \lfloor m(T_d(t) + t) + 1 \rfloor / m. \tag{13}$$

With this notation, we can write

$$\begin{aligned} A_d^m(t) &= E[b(\mathcal{K}_d^m(t)) v(\mathcal{K}_d^m(t)) / v(t)], \\ a_d^m(t) &= E[a[t, \mathcal{K}_d^m(t)) / v(t)], \\ V^m(t) &= E[\mathcal{L}(t, \mathcal{K}_d^m(t))] \end{aligned} \tag{14}$$

Let $S^d(s|t,m)$ denote the survival function of $\mathcal{K}_d^m(t)$. Let us derive this function. Consider the fact that if $y \geq 0$ and $x \geq 0$, then $\lfloor y \rfloor + 1 > x$ if and only if $y \geq \lfloor x \rfloor$. Using this result we find that $\mathcal{K}_d^m(t) > s$ if and only if $T_d(t) \geq (\lfloor ms \rfloor / m) - t$. Therefore,

$$S^d(s|t,m) = \exp\left\{ -\int_t^{\lfloor ms \rfloor / m} \mu^d(z)\, dz \right\}. \tag{15}$$

3 The Double-Decrement Model

In this section, we present the probabilistic structure for a dependent double-decrement model. This will allow us to derive an expression for the withdrawal benefit, $W_{d|w}^m(t)$, that represents the benefit that is returned to the policyholder at time $\lfloor mt + 1 \rfloor / m$ when withdrawal occurs at time t

Let T_w denote the time at withdrawal from a life insurance contract for a life aged x, where $f^w(t)$ is the density, $S^w(t) = \int_t^\infty f^w(z)\, dz$ is the survival function and $\mu^w(t) = f^w(t)/S^w(t)$ is the force We will find it useful to define the discrete random variable

$$K_w^m = \frac{\lfloor mT_w \rfloor}{m}. \tag{16}$$

Generally, we assume that T_d and T_w are not stochastically independent Therefore, let us consider the conditional density of T_d given that $T_w = t$, which is denoted as $f^{d|w}(t_d|t)$. Also, let $S^{d|w}(t_d|t) = \int_{t_d}^\infty f^{d|w}(z|t)\, dz$ denote the conditional survival function of T_d given that $T_w = t$. Hence, the conditional force of mortality is

$$\mu^{d|w}(t_d|t) = \frac{f^{d|w}(t_d|t)}{S^{d|w}(t_d|t)}. \tag{17}$$

Thus $S^{d|w}(t_d|t) = \exp\{-\int_0^{t_d} \mu^{d|w}(z|t)\,dz\}$. In the case of independence, we get $f^{d|w}(t_d|t) = f^d(t_d)$, $S^{d|w}(t_d|t) = S^d(t_d)$, and $\mu^{d|w}(t_d|t) = \mu^d(t_d)$. It is important to note that the ensuing discussion and results assume that we know the density $f^{d|w}(t_d|t)$. However, estimating this density is not a trivial exercise because we can only observe the minimum of the random variables T_d and T_w.

Now consider the probability that the life survived to time $t + s$ given that it survived to time t and withdrawal occurred at time t This survival function is denoted as ${}_sp_t^{d|w}$ and it is equal to.

$$\begin{aligned} {}_sp_t^{d|w} &= Pr(T_d > t+s|T_d > t, T_w = t) = \frac{Pr(T_d > t+s, T_d > t|T_w = t)}{Pr(T_d > t|T_w = t)} \\ &= \frac{S^{d|w}(t+s|t)}{S^{d|w}(t|t)} = \exp\left\{-\int_t^{t+s} \mu^{d|w}(z|t)\,dz\right\}. \end{aligned} \tag{18}$$

It will be convenient to define $T_{d|w}(t)$ as the random variable induced by ${}_sp_t^{d|w}$ so that ${}_sp_t^{d|w} = Pr(T_{d|w}(t) > s)$. Note that ${}_tp_0^d = S^{d|w}(t|0)$. We let $T_{d|w}(0) = T_{d|w}$. It is instructive to note that if the expectation $E[g(T_{d|w}(t))]$ exists for some function $g(s)$, then

$$E[g(T_{d|w}(t))] = E[g(T_d - t)|T_d > t, T_w = t] \tag{19}$$

This last fact will be used repeatedly In the definition of the withdrawal benefit, the random variable $\lfloor m(T_{d|w}(t) + t) + 1\rfloor/m$ has a central role Thus it will be convenient to define

$$K_{d|w}^m(t) = \lfloor m(T_{d|w}(t) + t) + 1\rfloor/m. \tag{20}$$

With this notation, we can write

$$\begin{aligned} A_{d|w}^m(t) &= E[b(K_{d|w}^m(t)\,v(K_{d|w}^m(t))/v(t)], \\ a_{d|w}^m(t) &= E[a[t, K_{d|w}^m(t))/v(t)] \end{aligned} \tag{21}$$

Let $S^{d|w}(s|t,m)$ denote the survival function of $K_{d|w}^m(t)$. Note that $K_{d|w}^m(t) > s$ if and only if $T_{d|w}(t) \geq \lfloor ms\rfloor/m - t$. Therefore,

$$S^{d|w}(s|t,m) = \exp\left\{-\int_t^{\lfloor ms\rfloor/m} \mu^{d|w}(z)\,dz\right\}. \tag{22}$$

We are now ready to state our first theorem.

Theorem 3.1. *Let $\mathcal{L}(t,s)$ be the link function If the equivalence principle holds, then under a double-decrement model where the premiums are equal to P_d^m, the withdrawal benefit function is*

$$\begin{aligned} W_{d|w}^m(t) &= E[\mathcal{L}(\lfloor mt+1\rfloor/m, K_{d|w}^m(t))] \\ &= E[\mathcal{L}(\lfloor mt+1\rfloor/m, \lfloor m(T_{d|w}(t)+t)+1\rfloor/m)] \\ &= E[\mathcal{L}(\lfloor mt+1\rfloor/m, \lfloor mT_d+1\rfloor/m)|T_d>t, T_w=t] \\ &= E[\mathcal{L}(\lfloor mt+1\rfloor/m, K_d^m+1/m)|T_d>t, T_w=t] \end{aligned} \tag{23}$$

Proof. If $W_{d|w}^m(t)$ is the withdrawal benefit, then under the equivalence principle,

$$P_d^m E[a[0,K_d^m+1/m)1(T_d\le T_w)+a[0,K_w^m+1/m)1(T_d>T_w)]=$$
$$E[b(K_d^m+1/m)\,v(K_d^m+1/m)1(T_d\le T_w)+W_{d|w}^m(T_w)\,v(K_w^m+1/m)1(T_d>T_w)].$$

Therefore,

$$\begin{aligned} &P_d^m E[a[0,K_d^m+1/m)]+ \\ &P_d^m E[[a[0,K_w^m+1/m)-a[0,K_d^m+1/m)]1(T_d>T_w)] = \\ &E[b(K_d^m+1/m)\;v(K_d^m+1/m)]+ \\ &E[[W_{d|w}^m(T_w)\;v(K_w^m+1/m)-b(K_d^m+1/m)\;v(K_d^m+1/m)]1(T_d>T_w)] \end{aligned}$$

But

$$P_d^m\times E[a[0,K_d^m+1/m)] = E[b(K_d^m+1/m)\;v(K_d^m+1/m)].$$

For simplicity, let $Y=1(T_d>T_w)\;v(K_w^m+1/m)$. Then

$$\begin{aligned} &E[YW_{d|w}^m(T_w)] = \\ &E[Yb(K_d^m+1/m)\;v(K_d^m+1/m)/v(K_w^m+1/m)]- \\ &E[YP_d^m(a[0,K_d^m+1/m)-a[0,K_w^m+1/m))/v(K_w^m+1/m)]. \end{aligned}$$

Note that $a[0,K_d^m+1/m)-a[0,K_w^m+1/m)=a[K_w^m+1/m,K_d^m+1/m)$, hence the right-hand side of the last equation is equal to

$$E[Y\mathcal{L}(K_w^m+1/m,K_d^m+1/m)]=E[YE[\mathcal{L}(K_w^m+1/m,K_d^m+1/m)|T_d>T_w,T_w]]=$$
$$E[YE[\mathcal{L}\lfloor mt+1\rfloor/m,K_d^m+1/m)|T_d>t,T_w=t]|_{t=T_w}]=E[YW_{d|w}^m(T_w)]$$

Hence, the result is proved □

Now, let us compare the withdrawal benefit $W_{d|w}^m(t)$, as defined in equation (23), with the reserve formula $V^m(t)$, shown in equation (12). Clearly, they are different, even when T_d and T_w are stochastically independent. In the case of independence, $T_d(t)$ and $T_{d|w}(t)$ are identically distributed Let $W_d^m(t)$ denote the withdrawal benefit under the independent decrement model, then

$$W_d^m(t) = E[\mathcal{L}(\lfloor mt+1 \rfloor/m, \mathcal{K}_d^m(t))]. \tag{24}$$

Thus

$$\lim_{m\to\infty} V^m(t) = \lim_{m\to\infty} W_{d|w}^m(t) = E[\mathcal{L}(t, T_d(t)+t)].$$

In other words, the withdrawal benefit is equal to the reserve in the continuous model, thus confirming a well-known fact

4. Withdrawal Benefits under Antiselection

In this section, we give a definition of antiselection and we show that the withdrawal benefit under antiselection is smaller than the benefit under the single-decrement model, as expected. We are now ready to give a definition of *antiselection* We say that life insurance is subject to antiselection at withdrawal, if

$$\mu^{d|w}(t_d|t_w) < \mu^d(t_d)\ \forall t_d \geq t_w. \tag{25}$$

If we reverse this inequality, then we have antiselection for life annuities. Using our definition of antiselection, we immediately find that

$$S^{d|w}(s|t,m) \geq S^d(s|t,m) \tag{26}$$

for alls $s \geq 0$.

First, we discuss the implications of antiselection to the valuation of the life insurance. Assume that $g(s) = b(s)v(s)$ is an absolutely continuous function with $g'(s) \leq 0$ so that

$$g(s) = g(0) + \int_0^s g'(z)\,dz$$

Actually, this is a weak assumption because the assumption is obviously true when $b(s) = 1$ and $v(s) = \exp(-\delta\, s)$

Lemma 4.1. *Suppose that $g(s) = b(s)\,v(s)$ for $s \geq 0$ is absolutely continuous and $g'(s) \leq 0$ If $g(s)$ is integrable with respect to the cumulative distribution functions $1 - S^{d|w}(s|t,m)$ and $1 - S^d(s|t,m)$, then under the antiselection condition $\mu^{d|w}(t_d|t_w) < \mu^d(t_d)$ and the equivalence principle, we get*

$$A_{d|w}^m(t) \leq A_d^m(t)$$

Proof. First, note that

$$\begin{aligned}
v(t)A_d^m(t) &= E[b(\mathcal{K}_d^m(t))\, v(\mathcal{K}_d^m(t))] \\
&= \int_{[0,\infty)} g(s)d(1 - S^d(s|t,m)) \\
&= \int_{[0,\infty)} \left[g(0) + \int_0^s g'(z)\,dz \right] d(1 - S^d(s|t,m)) \\
&= g(0) + \int_0^\infty \int_{(z,\infty)} g'(z)\; d(1 - S^d(s|t,m))\; dz \\
&= g(0) + \int_0^\infty g'(z) S^d(z|t,m)\; dz
\end{aligned}$$

Next, under antiselection $S^{d|w}(z|t,m) \geq S^d(z|t,m)$ and so $g'(z)S^{d|w}(z|t,m) \leq g'(z)S^d(z|t,m)$. Integrating both sides of the inequality yields the result. □

Next, we present a lemma on the implications of antiselection to the valuation of life annuities.

Lemma 4.2. *Under the equivalence principle and the antiselection condition* $\mu^{d|w}(t_d|t_w) < \mu^d(t_d)$, *we get*

$$a_{d|w}^m(t) \geq a_d^m(t).$$

Proof. First, note that

$$\begin{aligned}
v(t)a_d^m(t) &= E[a[t, \mathcal{K}_d^m(t))] \\
&= E\Big[\frac{1}{m}\sum_{k=0}^{\infty} v(k/m)1(t \leq k/m < \mathcal{K}_d^m(t))\Big] \\
&= \frac{1}{m}\sum_{k=0}^{\infty} v(k/m)1(t \leq k/m)E[1(k/m < \mathcal{K}_d^m(t))] \\
&= \frac{1}{m}\sum_{k=0}^{\infty} v(k/m)1(t \leq k/m)S^d(k/m|t,m).
\end{aligned}$$

But under antiselection $S^{d|w}(s|t,m) \geq S^d(s|t,m)$. Summing both sides of the inequality yields the result. □

Applying Lemma 4.1 and 4.2, we immediately find that under antiselection the withdrawal benefit under the classical independent decrement model is too large. We summarize this result with the following theorem.

Theorem 4.3. *Under the conditions in Lemma 4.1 and 4.2, we get*

$$W_{d|u}^{m}(t) \leq W_{d}^{m}(t). \tag{27}$$

REFERENCES

BOWERS, N L , GERBER, H U , HICKMAN, J C , JONES, D A and NESBITT, C J (1986) *Actuarial Mathematics* Society of Actuaries Schaumburg, Illinois

CARRIERE, J F (1994) "Dependent decrement theory" *Transactions Society of Actuaries*, XLVI, 45-73

NESBITT, C J (1964) Discussion of "A Statistical Approach to Premiums and Reserves in Multiple Decrement Theory" *Transactions Society of Actuaries*, XVI, Part 1, 149-153

PROFESSOR JACQUES CARRIERE
Dept of Mathematical Sciences
University of Alberta
Edmonton, Alberta
Canada T6G 2G1

MODELING AND COMPARING DEPENDENCIES IN MULTIVARIATE RISK PORTFOLIOS

By

NICOLE BÄUERLE

Abteilung Mathematik VII
Operations Research
Universität Ulm

AND

ALFRED MÜLLER

Institut für Wirtschaftstheorie
und Operations Research
Universität Karlsruhe

ABSTRACT

In this paper we investigate multivariate risk portfolios, where the risks are dependent. By providing some natural models for risk portfolios with the same marginal distributions we are able to compare two portfolios with different dependence structure with respect to their stop-loss premiums. In particular, some comparison results for portfolios with two-point distributions are obtained. The analysis is based on the concept of the so-called supermodular ordering. We also give some numerical results which indicate that dependencies in risk portfolios can have a severe impact on the stop-loss premium. In fact, we show that the effect of dependencies can grow beyond any given bound.

KEYWORDS

Dependent risks; Stop-loss premium, Supermodular order; Stop-loss order; Majorization, Comonotonicity, Exchangeable Bernoulli random variables

1. INTRODUCTION

In traditional risk theory for means of tractability, individual risks are usually assumed to be independent. Recent research has shown, however, that a positive dependence between risks leads to underestimation of the stop-loss premium for the aggregated loss. To the best of our knowledge, Heilmann (1986) and Hürlimann (1993) have been the first authors, who demonstrated the impact of dependencies on stop-loss premiums More recently, Dhaene and Goovaerts (1996) investigated the effect of bivariate dependencies on the related stop-loss premium and gave an upper bound by determining the riskiest portfolio Dhaene and Goovaerts (1997) made a first attempt to treat multivariate dependencies They considered a special life insurance portfolio with two-point distributions. Their results were generalized by Müller (1997) who characterized the riskiest portfolio under all portfolios with equal marginals for arbitrary distributions. Wang (1997) suggested a set of tools for concrete modeling of dependencies in risk portfolios using the information given by the correlation coefficients.

In this paper we now propose some natural models for multivariate risk portfolios with different degree of dependence and same marginal distributions The assumption about equal marginals is crucial here since our focus lies on comparing dependencies only The results can of course be extended to unequal marginals by adding stochastic dominance. The models are defined in such a way that it is possible to compare two portfolios from the same class of models with respect to their stop-loss premiums. More precisely, we consider the classical individual model from risk theory, where the aggregate claim amount of a portfolio in a period is given by

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

where X_i is the random claim amount caused by policy i, $i = 1, .., n$ Throughout the paper we assume that the random variables X_i are non-negative with finite expectation. In a first model (model 3 1 in section 3) we assume that the risks can be divided into several groups, where each risk of a group is influenced by a global risk factor, a group specific risk factor and an individual risk factor. We show how the group structure of the portfolio affects the stop-loss premium and determine the safest and riskiest portfolio in this model class. On that occasion, we use the notion of majorization in order to compare the group structures

In a second model (model 3.2 in section 3) we compare two portfolios, where both are subject to the same economic/physical environment, but the second portfolio contains an additional global risk factor which influences the risks of this portfolio in the same direction. Again, the marginal distributions are assumed to be equal for both portfolios. It can be proved that the stop-loss premium in the second scenario is greater than in the first one. This result is used later on to construct a portfolio, where the risks have two-point distributions and the portfolio can be characterized by a dependence parameter $\rho \in [0, 1]$. The construction is such that increasing ρ leads to a higher correlation in the portfolio and the two extreme cases

$\rho = 0$ and $\rho = 1$ correspond to independence and comonotonicity respectively. We show that the stop-loss premium is increasing in the dependence parameter ρ.

In another model we compare portfolios which are given by exchangeable Bernoulli random variables Here it can be shown that stop-loss order of the mixing distribution implies more riskiness for the aggregate claims. Moreover, in this setting, we prove that the ratio of the stop-loss premium in the riskiest scenario divided by the stop-loss premium of an arbitrary portfolio is increasing in the retention level.

Our models are very general and cover most of the specific parametric models considered by Wang (1997) There is one main difference between Wang's paper and this one We mainly investigate, how dependencies affect the riskiness of portfolios, whereas Wang focuses on algorithms for simulation and efficient computation of concrete parametric models for correlated risks. Thus the two papers are complementary in so far as his algorithms for simulation can be easily adapted to our models.

Most of the comparison results we provide in this paper are based on the so-called supermodular ordering. This concept has recently proven to be valuable for comparing dependencies in random vectors in a wide range of applied probability models For details see Bauerle (1997a), Shaked and Shanthikumar (1997) and the references therein.

At the end of the paper we give a numerical example for model 3.1, which shows that dependencies can have a severe effect on the stop-loss premium. In particular we demonstrate that whenever the retention level exceeds the expected aggregate claim amount, the effect of dependence can be arbitrary worse

The paper is organized as follows section 2 contains some basic definitions and results about stochastic orderings and dependence which we will use in the sequel. Section 3 covers model 3.1 and 3.2 and section 4 is dedicated to the special case of risks with two-point distributions. The numerical results are summarized in section 5.

2. Stochastic Orderings and Dependence

Let us first fix the notation. A portfolio of risks is a random vector $X = (X_1, \ldots, X_n)$ of n individual risks, where an individual risk X_i, $1 \le i \le n$ is a non-negative (univariate) random variable with a finite mean. For arbitrary univariate random variables Y we denote the distribution function by $F_Y(t) = P(Y \le t)$, $t \in \mathbb{R}$ and $\bar{F}_Y(t) := P(Y > t) = 1 - F_Y(t)$ shall be the corresponding survival function. We will also frequently use the stop-loss transform $\pi_Y(t) := E(Y - t)^+ = \int_t^\infty \bar{F}_Y(x)dx$, $t \in \mathbb{R}$. For a random vector $X = (X_1, \ldots, X_n)$ we similarly define the distribution function

$$F_X(t) = P(X \le t) = P(X_1 \le t_1, \ldots, X_n \le t_n), \quad t = (t_1, \ldots, t_n) \in \mathbb{R}^n$$

and the survival function

$$\bar{F}_X(t) := P(X > t) = P(X_1 > t_1, \ldots, X_n > t_n), \quad t = (t_1, \ldots, t_n) \in \mathbb{R}^n$$

Note that for multivariate distributions in general $\bar{F}_X(t) \neq 1 - F_X(t)$. If two random variables or vectors X and Y have the same distribution, we will write $X \stackrel{d}{=} Y$. $X \sim F$ should be read as: X has the distribution F.

Now we will introduce some stochastic order relations, which are well-known concepts for comparing risks.

Definition 2.1 *Let X, Y be real random variables with finite means*
a) *We say that X precedes Y in stochastic order, written $X \leq_{st} Y$, if $\bar{F}_X(t) \geq \bar{F}_Y(t)$ for all $t \in \mathbb{R}$.*
b) *X precedes Y in stop-loss order, written $X \leq_{sl} Y$, if $\pi_X(t) \leq \pi_Y(t)$ for all $t \in \mathbb{R}$.*

Remarks
a) If $X \preceq Y$, where $\preceq$ may be any stochastic order relation, then we will also write $F_X \preceq F_Y$ whenever it is convenient.
b) If we have a family F_θ, $\theta \in \Theta \subset \mathbb{R}$ of distributions, then we say that F_θ is stochastically increasing in θ, if $F_\theta \leq_{st} F_{\theta'}$ for $\theta < \theta'$.
c) Stop-loss order means, that the stop-loss reinsurance premium for the risk Y is higher than that for X for any retention t.

Now we collect some important properties of these orderings, which we will use frequently. They can be found e.g. in Shaked and Shanthikumar (1994) or Goovaerts et al. (1990)

Theorem 2.2
a) *The following conditions are equivalent.*
   1. $X \leq_{st} Y$,
   2. *$Ef(X) \leq Ef(Y)$ for all non-decreasing functions f,*
   3. *There are random variables $\tilde{X} \stackrel{d}{=} X$ and $\tilde{Y} \stackrel{d}{=} Y$ such that $\tilde{X} \leq \tilde{Y}$ almost sure.*
b) *The following conditions are equivalent*
   1. $X \leq_{sl} Y$,
   2. *$Ef(X) \leq Ef(Y)$ for all non-decreasing convex functions f,*
   3. *There are random variables $\tilde{X} \stackrel{d}{=} X$ and $\tilde{Y} \stackrel{d}{=} Y$ such that $E[\tilde{Y}|\tilde{X}] \geq \tilde{X}$ almost sure.*

As stated before, the main topic of this paper is the comparison of the riskiness of portfolios. In order to do so we need notions of stochastic order relations for random vectors. We say that a portfolio $X = (X_1, \ldots, X_n)$ is less risky than a portfolio $Y = (Y_1, \ldots, Y_n)$, if the corresponding aggregate claims $S = \sum_{i=1}^n X_i$ and $S' = \sum_{i=1}^n Y_i$ are stop-loss ordered, i.e $S \leq_{sl} S'$. It will turn out that a sufficient condition for this is given by the so-called *supermodular ordering* or the *symmetric supermodular ordering* These stochastic order relations have recently been considered in applied probability by Bauerle (1997a, b), Bäuerle and Rieder (1997), Shaked and Shanthikumar (1997) and others. In the actuarial literature the supermodular ordering has been introduced by Müller (1997) It is based on the comparison of integrals of (symmetric) supermodular functions, which are defined as follows

Definition 2.3

a) A function f $\mathbb{R}^n \to \mathbb{R}$ is said to be **supermodular**, *if*

$$f(x_1, \ldots, x_i+\varepsilon, \ldots, x_j+\delta, \ldots, x_n) - f(x_1, \ldots, x_i+\varepsilon, \ldots, x_j, \ldots, x_n) \geq f(x_1, \ldots, x_i, \ldots, x_j+\delta, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) \qquad (1)$$

holds for all $x \in \mathbb{R}^n$, $1 \leq i < j \leq n$ and all $\varepsilon, \delta > 0$

b) A function $f . \mathbb{R}^n \to \mathbb{R}$ is called **symmetric**, *if $f(x) = f(\Pi x)$ for all permutations Πx of x.*

An intuitive explanation of the notion of supermodularity can be given as follows: Let $x_1, \ldots, x_n$ be the individual claim amounts of n policy holders and let $f(x_1, \ldots, x_n)$ be the loss for the insurance company caused by these claims Then supermodularity of the function f means that the consequences of an increase of a single claim are the worse, the higher the other claims are.

Symmetric functions do not depend on the order of the variables. This means in our context that the policy holders are indistinguishable

The following properties of supermodular functions are well-known.

Theorem 2.4

a) If f is twice differentiable, then f is supermodular if and only if

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) \geq 0 \text{ for all } x \in \mathbb{R}^n, \ 1 \leq i < j \leq n$$

b) If $g_1, \ldots, g_n$ $\mathbb{R} \to \mathbb{R}$ are increasing functions and f is supermodular, then $f(g_1(\cdot), \ldots, g_n(\cdot))$ is also supermodular.

A proof of this theorem and many examples can be found in Marshall and Olkin (1979, p. 146ff). Now we will introduce the supermodular stochastic order relation.

Definition 2.5

a) A random vector $X = (X_1, \ldots, X_n)$ is said to be smaller than the random vector $Y = (Y_1, \ldots, Y_n)$ in the **supermodular ordering**, *written $X \leq_{sm} Y$, if $Ef(X) \leq Ef(Y)$ for all supermodular functions f such that the expectations exist.*

b) A random vector $X = (X_1, \ldots, X_n)$ is said to be smaller than the random vector $Y = (Y_1, \ldots, Y_n)$ in the **symmetric supermodular ordering**, *written $X \leq_{symsm} Y$, if $Ef(X) \leq Ef(Y)$ for all symmetric supermodular functions f such that the expectations exist.*

Supermodular ordering is a useful tool for comparing dependence structures of random vectors. Since any function $f: \mathbb{R}^n \to \mathbb{R}$ that depends only on one variable (i.e. $f(x_1, ..., x_n) = g(x_i)$ for some $g: \mathbb{R} \to \mathbb{R}$ and some $i \in \{1, ..., n\}$) is supermodular, it follows immediately from the definition that only distributions with the same marginals can be compared by supermodular ordering. Moreover, all functions $f(x) = x_i x_j$, $i \neq j$ are supermodular. Hence $X \leq_{sm} Y$ implies $Corr(X_i, X_j) \leq Corr(Y_i, Y_j)$, $i \neq j$.

The usefulness of these concepts in our setting is shown clearly in the next result.

Theorem 2.6 *Let $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ be random vectors with $X \leq_{sm} Y$ ($X \leq_{symsm} Y$) and let*

$$S = \sum_{i=1}^{n} X_i \quad \text{and} \quad S' = \sum_{i=1}^{n} Y_i$$

Then $S \leq_{sl} S'$.

Proof: For the supermodular ordering this has been shown in Müller (1997, Th. 3.1). The case of symmetric supermodular ordering can be shown along the same lines, as the function $x \to \sum x_i$ is obviously symmetric. □

The Theorem says that stronger dependence in the sense of supermodular ordering leads to more risky portfolios. Next we will construct a special random vector with given marginals, which exhibits a very strong form of dependence. Let U be a random variable uniformly distributed on [0,1] and let $F_1, ..., F_n$ be n marginal distributions. Define $X = (X_1, ..., X_n) = (F_1^{-1}(U), ..., F_n^{-1}(U))$. Using the well-known fact in simulation that $F^{-1}(U) \sim F$, we see that X in fact has the marginal distributions $F_1, ..., F_n$. Since F_i^{-1} is increasing for all i it follows that $X_i(\omega_1) < X_i(\omega_2)$ implies $X_j(\omega_1) \leq X_j(\omega_2)$ for all $j \neq i$. Schmeidler (1986) and Yaari (1987) introduced the notion **comonotonicity** for this property. An easy calculation shows that the distribution function of X is given by $F_X(t) = \min_{i=1}^{n} F_i(t_i)$. Summing up, we can give four equivalent definitions of comonotonicity.

Definition 2.7 *The distribution F with marginal distributions $F_1, ..., F_n$ is called comonotonic, if one of the following four equivalent conditions is fulfilled*

1. $$F(t) = \min_{i=1}^{n} F_i(t_i), \quad t \in \mathbb{R}^n,$$
2. *The random vector $X = (F_1^{-1}(U), ..., F_n^{-1}(U))$, where U is uniformly distributed on* [0,1], *has the distribution F,*
3. *There is a univariate random variable Z and there are increasing functions $f_1, ..., f_n$, such that $X = (f_1(Z), ..., f_n(Z))$ has the distribution F.*
4. *There is a random vector $X \sim F$, such that $X_i(\omega_1) < X_i(\omega_2)$ implies $X_j(\omega_1) \leq X_j(\omega_2)$ for all $j \neq i$.*

The comonotonic distribution F is also called *upper Fréchet bound*, since Fréchet has shown that for any distribution function G with marginals $F_1, \dots, F_n$ we have $G \leq F$. An even stronger result is the so-called Lorentz-inequality. It can be found e.g. as Theorem 5 in Tchen (1980) and can be stated as follows.

Theorem 2.8 *Let X be an arbitrary random vector and let Y be the comonotonic random vector with the same marginals as X. Then $X \leq_{sm} Y$.*

This means that comonotonicity is the strongest possible dependence structure and hence by Theorem 2.6 the corresponding portfolio is the riskiest one under all portfolios with the same marginals.

3. The Models

In this section we consider several possibilites of modeling dependencies in risky portfolios. In our first model we assume that the portfolio consists of different groups, such that there is a strong dependence between the members of one group, but much less dependence between members of different groups. As a typical example where this is very realistic imagine a catastrophe risk like earthquakes or hurricanes, where the groups are specified by geographic regions. There is certainly a strong dependency between the expected losses of people from the same region, but the losses will be nearly independent for people who live far from each other For such situations we suggest the following model. It was introduced by Tong (1989) and was further considered by Bäuerle (1997a).

Model 3.1
Consider a portfolio $X = (X_1, \dots, X_n)$, consisting of n risks $X_1, \dots, X_n$. We assume that the risks can be divided into $r \leq n$ groups according to an n-dimensional vector $k = (k_1, \dots, k_r, 0, \dots, 0)$, $k_\nu \in I\!N$, $\sum_{\nu=1}^{r} k_\nu = n$, where risk X_i is in group ν if and only if $k_1 + \dots + k_{\nu-1} < i \leq k_1 + \dots + k_\nu$. Each of the risks in the portfolio is influenced by three risk factors which will be modeled as independent random variables V, G_ν and Z_i

1. an overall risk factor V which is due to global environmental changes and concerns all of the risks in the portfolio in the same fashion,
2. a group specific risk factor G_ν which influences only the risks in group ν, $1 \leq \nu \leq r$ and has no effect on other risks in the portfolio,
3. an individual risk factor Z_i which reflects the individual share of risk X_i, $1 \leq i \leq n$

Moreover, we assume that there exists a function $g : I\!R^3 \to I\!R$ such that the i-th risk is given by $X_i = g(V, G_\nu, Z_i)$ whenever i is in group ν. Since we associate higher outcomes of a risk factor with higher risk in the portfolio, we suppose that g is increasing. This situation is typical for a lot of insurance portfolios In private health insurance for example, the risk caused by an individual person depends on an overall risk factor which collects environmental aspects (e.g. pollution, greenhouse effect, epidemics), on a group specific factor like profession and on an individual risk factor which summarizes health conditions. In car insurance, the group risk factor could be interpreted as the local area of the policy holder. Assuming this kind of dependence within a portfolio it is now interesting to investigate the effect, the constellation of group sizes has on the aggregate claim of the portfolio, since it is well-known that positive correlations in a risk portfolio increase the payable amount of the insurance company, see e.g. Dhaene and Goovaerts (1996, 1997) or Müller (1997). Obviously it is quite hard to compare two risky portfolios when for example the number and sizes of the groups change. However, in some cases this is possible as we will show in the next theorem. In order to state it, let k and k' be two n-dimensional vectors with

$$k = (k_1, \dots, k_r, \dots, 0, \dots, 0), \quad k' = (k'_1, \dots, k'_l, 0, \dots, 0)$$

$1 \le r$, $l \le n$, $k_i, k'_i \in \mathbb{N}$ for all i and $\sum_{i=1}^n k_i = \sum_{i=1}^n k'_i = n$. Let two n-dimensional risky portfolios X and Y be given by

$$\begin{array}{ll} X_1 = g(Z_1, G_1, V) & Y_1 = g(U_1, G_1, V) \\ \vdots & \vdots \\ X_{k_1} = g(Z_{k_1}, G_1, V) & Y_{k'_1} = g(U_{k'_1}, G_1, V) \\ X_{k_1+1} = g(Z_{k_1+1}, G_2, V) & Y_{k'_1+1} = g(U_{k'_1+1}, G_1, V) \\ \vdots & \vdots \\ X_{k_1+k_2} = g(Z_{k_1+k_2}, G_2, V) & Y_{k'_1+k'_2} = g(U_{k'_1+k'_2}, G_2, V) \\ \vdots & \vdots \\ X_n = g(Z_n, G_r, V) & Y_n = g(U_n, G_l, V) \end{array}$$

where the individual risk factors $Z_1, \dots, Z_n, U_1, \dots, U_n$ are i.i.d. random variables, the group specific risk factors $G_1, \dots, G_{\max\{r,l\}}$ are i.i.d random variables and the environmental risk factor V is a random variable independent of $\{Z_i\}$, $\{U_i\}$ and $\{G_i\}$. $g: \mathbb{R}^3 \to \mathbb{R}$ is an increasing function. Denote $S = \sum_{i=1}^n X_i$ and $S' = \sum_{i=1}^n Y_i$ respectively.

Moreover, we need an appropriate order relation for vectors to compare the group structures k and k'. It turns out that the notion of majorization is best suited for this purpose. The definition is as follows.

Definition 3.1 *Let $x, y \in \mathbb{N}_0^n$ and denote by $x_{[1]} \ge \dots \ge x_{[n]}$ the decreasing rearrangement of x, analogously for y. We say that y majorizes x ($x \prec y$) if and only if*

$$\sum_{i=1}^r x_{[i]} \le \sum_{i=1}^r y_{[i]}, \quad r = 1, \dots, n-1, \quad \textit{and} \quad \sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}.$$

A host of results and applications of this order relation can be found in Marshall and Olkin (1979) Intuitively speaking $k \prec k'$ means that in k' the groups are larger and/or more unequal. Some examples are given in section 5 Now we are able to state the main result for this model

Theorem 3.2 *If $k \prec k'$, we obtain under the assumptions of model 3.1*
a) $X \leq_{symsm} Y$,
b) $S \leq_{sl} S'$.

Proof A complete proof of statement a) can be found in Bauerle (1997a). The main ideas are as follows: in a first step we show that for a sequence $\{G_\nu\}$ of i.i.d random variables and

$$X = (G_1, \quad , G_1, G_2, \ldots, G_2, \quad , G_r, \ldots, G_r)$$
$$Y = (G_1, \ldots, G_1, G_2, \ldots, G_2, \ldots, G_l, \ldots, G_l)$$

where the block of G_i's in X (Y) has length k_i (k_i'), the relation $k \prec k'$ implies that $X \leq_{symsm} Y$. Applying properties of symmetric supermodular functions we obtain a) Part b) then follows from Theorem 2.6. □

In this setting it is easy to determine the riskiest and the safest portfolio with respect to the stop-loss ordering of aggregate claims. In order to do so we only need to determine the minimum and maximum with respect to majorization under all vectors k with $\sum k_i = n$ It is nearly obvious that the minimum is given by $k^s = (1, 1, \ldots, 1)$ and the maximum is given by $k^r = (n, 0, \quad, 0)$. This yields the following result.

Corollary 3.3 *Let* $k^r = (n, 0, \ldots, 0)$ and $k^s = (1, \ldots, 1)$ *be two n-dimensional vectors and denote by S^r and S^s the aggregate claims of the corresponding risk portfolios as in model 3.1 Then we obtain for arbitrary $k \in \mathbb{N}_0^n$ with $\sum_{i=1}^n k_i = n$ and respective aggregate claim S*

$$S^s \leq_{sl} S \leq_{sl} S^r.$$

Hence the riskiest portfolio is given, when there is only one group and the safest portfolio is obtained, when each individual forms his/her own group.

Our model 3.1 is strongly related to the component models introduced in chapter 9 of Wang (1997). As another important class of models he considers *common mixture models*, which we will investigate now.

Model 3.2
The intuition behind this model is as follows The model for X as well as the model for Y is a so called common mixture model This means that there are some external mechanisms, described by random variables, which have influence on all the risks Given these environmental parameters, the individual risks are independent. The parameters can be some state of nature (weather conditions, earthquakes, ...) as well as economic or legal

environments (inflation, court rules etc) which have a common impact on all risks. In contrast to model 3.1 we will now compare the portfolios with respect to the number of external mechanisms which affect them.

The following model for this situation has been considered by Bauerle (1997a) (cf. also Shaked and Tong (1985)): Suppose there are two n-dimensional random vectors X and Y with the structure

$$(X_1, ..., X_n) = (g_1(Z_1, W), .., g_n(Z_n, W)) \tag{2}$$

$$(Y_1, .., Y_n) = (\tilde{g}_1(U_1, V, W), ..., \tilde{g}_n(U_n, V, W)) \tag{3}$$

where $Z_1, ..., Z_n$ are i.i.d. random variables, $U_1, ..., U_n$ are i.i.d random variables and (V, W) is a random vector independent of Z_i and U_i. Moreover, the functions $g_i : \mathbb{R}^2 \to \mathbb{R}$ and $\tilde{g}_i : \mathbb{R}^3 \to \mathbb{R}$ are such that for every fixed w and all $i = 1, ..., n$ we have

$$g_i(Z_i, w) \overset{d}{=} \tilde{g}_i(U_i, V, w), \tag{4}$$

i.e. they have the same distribution.

We will show now, that the portfolio $Y = (Y_1, ..., Y_n)$ is more risky than the portfolio $X = (X_1, ..., X_n)$, if the functions $\tilde{g}_i$ are increasing in the second argument. In fact, let $S := \sum_{i=1}^n X_i$ and $S' := \sum_{i=1}^n Y_i$. Then the following holds.

Theorem 3.4 *If the functions $\tilde{g}_i$ are increasing in the second argument, then*
a) $X \leq_{sm} Y$,
b) $S \leq_{sl} S'$.

Proof: a) can be found as Theorem 3.1 in Bauerle (1997a). Part b) then follows immediately from a) by Theorem 2.6. □

The model for Y contains an additional environmental variable V, which has an influence on $Y_1, .., Y_n$ in the same direction. Hence there is more dependence in Y than in X, since the external mechanism, which has a common influence on all risk, is more important in Y. This will become more explicit in the special case we will treat now.

Let us assume that W is constant. Hence $Y_i = \tilde{g}_i(U_i, V)$ and $X_i = g_i(Z_i)$. This means that $Y_1, ..., Y_n$ are conditionally independent given $V = v$ and the monotonicity of $\tilde{g}_i$ in the second argument means that the conditional distribution of Y_i given $V = v$ is stochastically increasing in v for all $i = 1, .., n$. Moreover, $X_1, .., X_n$ are independent random variables, which by (4) have the same marginal distributions as $Y_1, ..., Y_n$. Summing up, we get the following corollary of Theorem 3.4.

Corollary 3.5 *Let V be any random variable and let $Y = Y_1, ..., Y_n$ be a random vector such that $Y_1, ..., Y_n$ are conditionally independent given $V = v$ and such that the conditional distributions $P(Y_i \in \cdot | V = v)$ are stochastically* *increasing in v for all i = 1, ..., n Moreover, let $X = (X_1, \ldots, X_n)$ be a vector of independent random variables with the same marginal distributions as Y. Then*

$$X \leq_{sm} Y \quad and \quad S = \sum_{i=1}^{n} X_i \leq_{sl} S' = \sum_{i=1}^{n} Y_i.$$

Another application of Theorem 3.4 will be given in the next section. Many more examples can be found in chapter 7 of Wang (1997)

4. Risks with two-point Distributions

Now we consider the important special case of portfolios consisting of risks X_i having a two-point distribution in 0 and α_i with $P(X_i = 0) = p_i$. This occurs e.g. in the individual life model. Dhaene and Goovaerts (1997) determined the riskiest portfolio with given marginals for this case and especially considered portfolios with dependencies only between couples.

The riskiest portfolio has the property that if a policy holder with a low mortality dies, then all policy holder with higher mortality also die with probability 1 We think that this is very unrealistic. It would be desirable to have a parametric model with a dependence parameter ρ, which continuously varies between independence and maximal dependence as described above.

We investigate here two such models, one for the case of indistinguishable individuals and one for the case that the probability for no claim differs between the individuals.

Indistinguishable individuals

We say that the individuals in a portfolio are indistinguishable, if the joint distribution of the random vector of their risks is not affected by permutations of the risks In probability theory a sequence of such random variables is said to be exchangeable (or interchangeable), see e.g. Feller (1966, p. 228ff) or Chow and Teicher (1978) Of course this implies that all risks have the same marginal distribution, i e. there is a $p \in (0, 1)$ and some $\alpha > 0$ such that $P(X_i = 0) = p = 1 - P(X_i = \alpha)$ for all $i = 1, \ldots, n$ Without loss of generality we can assume $\alpha = 1$, so that the random variables $X_1, X_2, \ldots$ form a sequence of exchangeable Bernoulli variables.

Therefore let us assume that S_n is the total claim amount of a portfolio of n risks, which stem from a sequence of exchangeable Bernoulli variables. A well-known theorem of de Finetti (see e.g Feller (1966, p 228)) states that in this case S_n is a mixture of binomial distributions, i.e.

$$P(S_n = k) = \int_0^1 \binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k} F(d\vartheta)$$

for some mixing distributions F. Thus, the distribution of S_n is completely determined by the mixing distribution F In fact, it is completely determined by the first n moments of F. For a survey on exchangeable Bernoulli variables, including many examples and methods for estimating their parameters we refer to Madsen (1993).

Now we want to show, how the mixing distribution F affects the riskiness of the portfolio S_n We have the following result.

Theorem 4.1 *Let S_n (S_n') be the total claim amount of a portfolio of n risks, which stem from a sequence of exchangeable Bernoulli variables with mixing distribution F (F'). Then $F \leq_{sl} F'$ implies $S_n \leq_{sl} S_n'$.*

Proof This follows directly from Corollary 3.7 in Lefèvre and Utev (1996). □

Remark: From Theorem 4.1 it follows easily that the least risky portfolio of exchangeable Bernoulli variables with given marginals is the one that consists of independent risks and the riskiest portfolio is the one with mixing distribution concentrated on $\{0, 1\}$, which means that the risks are comonotonic In fact, this means that the portfolio consists of identical risks $X = (X_1, X_1, \ldots, X_1)$ and the distribution of the total claim amount $S_n = n \cdot X_1$ is a two-point distribution with $P(S_n = 0) = p = 1 - P(S_n = n)$. If we compare the stop-loss premiums of this portfolio with an arbitrary other portfolio of $bi(1, p)$-distributed risks, then we can strengthen Theorem 4 1 to the following result.

Theorem 4.2 *Let $X = (X_1, \ldots, X_n)$ be a portfolio of $bi(1, p)$-distributed risks with an arbitrary dependence structure and let $Y = (Y_1, \ldots, Y_1)$ be a portfolio of identical risks with the same distribution. Let $\pi_X(t) := E(\sum X_i - t)^+$ be the net stop-loss reinsurance premium of portfolio X and define $\pi_Y(t)$ similarly. Then the ratio $\pi_Y(t)/\pi_X(t)$ is increasing on its range $[0, n)$*

Proof: Since $\sum Y_i = nY_1$ is a two-point distribution on $\{0, n\}$, the function π_Y is affine linear. Since any stop-loss transform is decreasing and convex (see e.g. Muller (1996)) this implies that $g(x) := \pi_X \circ \pi_Y^{-1}(x)$ is a convex function. Differentiation yields that

$$g'(x) = \frac{\pi'_X \circ \pi_Y^{-1}(x)}{\pi'_Y \circ \pi_Y^{-1}(x)}$$

is increasing, and hence $\pi'_X(x)/\pi'_Y(x)$ is decreasing, since π_Y^{-1} is decreasing. This can be written equivalently as

$$\pi'_X(t)\pi'_Y(s) \geq \pi'_X(s)\pi'_Y(t) \quad \text{for all } t < s$$

and hence

$$\int_t^\infty \pi'_X(t)\pi'_Y(s)\,ds \geq \int_t^\infty \pi'_X(s)\pi'_Y(t)\,ds$$
$$\Leftrightarrow \pi'_X(t)\pi_Y(t) \leq \pi_X(t)\pi'_Y(t)$$
$$\Leftrightarrow \frac{\pi_Y(t)}{\pi_X(t)} \text{ is increasing}$$

□

Remark. Computational results indicate that Theorem 4.2 may be true for arbitrary distributions. We are, however, not yet able to give a proof for this conjecture.

Distinguishable individuals.
Now we propose a model where the individuals in the portfolio may have different probabilities for claims and different claim amounts. We want to construct a portfolio of risks X_i with $P(X_i = 0) = p_i$ and $P(X_i = \alpha_i) = q_i = 1 - p_i$ where $0 < p_i < 1$ and $\alpha_i > 0$ are arbitrary. Moreover we want to introduce a dependence parameter $\rho \in [0, 1]$ such that $\rho = 0$ corresponds to independence and $\rho = 1$ corresponds to comonotonicity. A very simple model with this property would be to take some mixture of the independent and the comonotone case. We think, however, that this is not very realistic. We propose some sort of an additive damage model, which is well known in reliability theory. Assume that there are two sources, that cause some normally distributed damage. One source influences all individuals in the same manner, whereas the other source depends on the individual behavior of each individual. A claim of amount α_i occurs, if the sum of these two damages exceeds some level z_i.

The formal construction will be based on model 3.2 with distributions and functions, which assume only two values. We denote by $N(\mu, \sigma^2)$ the univariate normal distribution with mean μ and variance $\sigma^2 > 0$ For convenience we extend the definition to the case $\sigma^2 = 0$, where $N(\mu, 0)$ denotes the one-point distribution in μ. The p-quantile of the standard normal distribution will be denoted by z_p, i.e if $X \sim N(0, 1)$, then $P(X \leq z_p) = p$. Now assume that $0 \leq \sigma^2 < \tau^2 \leq 1$ and consider model 3.2 with $W \sim N(0, \sigma^2)$, $V \sim N(0, \tau^2 - \sigma^2)$, $Z_i \sim N(0, 1 - \sigma^2)$ and $U_i \sim N(0, 1 - \tau^2)$ All random variables shall be independent We define

$$g_i(z, w) = \alpha_i \cdot 1\{z + w \geq z_{p_i}\} = \begin{cases} \alpha_i, & z + w \geq z_{p_i} \\ 0, & \text{else} \end{cases}$$

and

$$\tilde{g}_i(u, v, w) = \alpha_i \cdot 1\{u + v + w \geq z_{p_i}\}$$

Recall that $X_i = g_i(Z_i, W)$ and $Y_i = \tilde{g}_i(U_i, V, W)$ for $i = 1, \ldots, n$. Since $U_i + V \stackrel{d}{=} Z_i \sim N(0, 1 - \sigma^2)$, condition (4) is fulfilled. Moreover, $Z_i + W \stackrel{d}{=} U_i + V + W \sim N(0, 1)$, so that $P(X_i = \alpha_i) = P(Z_i + W \geq z_{p_i}) = q_i$ and $P(X_i = 0) = P(Z_i + W \leq z_{p_i}) = p_i$. Similarly $P(Y_i = 0) = p_i = 1 - P(Y_i = \alpha_i)$. By Theorem 3.4 $X \leq_{sm} Y$ and hence X is less risky than Y.

Now let us write $X(\sigma) = (X_1(\sigma), \ldots, X_n(\sigma))$ for the above defined portfolio X to make the dependency on σ explicit The definition of Y implies that $Y \stackrel{d}{=} X(\tau^2)$ which can be seen by interchanging the roles of Z_i and U_i as well as the one of W and $V + W$. Hence we obtain the following result.

Theorem 4.3 Let $0 \leq \rho < \rho' \leq 1$ *Then* $X(\rho) \leq_{sm} X(\rho')$ *and hence*

$$\sum_{i=1}^{n} X_i(\rho) \leq_{sl} \sum_{i=1}^{n} X_i(\rho').$$

It is easy to see that $X(0)$ is a portfolio of independent risks and $X(1)$ is a portfolio of comonotonic risks, which is the riskiest portfolio under all portfolios with given marginals, as has been shown by Müller (1997) for general distributions and in Dhaene and Goovaerts (1997) for the case of two-point distributions as considered here Now we will show that we can get any positive dependence structure by varying ρ continuously between these two extreme cases. In fact, we have the following result.

Theorem 4.4 *The function* $\rho \rightarrow Corr(X_i(\rho), X_j(\rho))$ *is non-negative and continuously increasing for all* $i, j = 1, \ldots, n$, $i \neq j$.

Proof: The marginal distribution of $X_i(\rho)$ and hence also the variance of $X_i(\rho)$ is independent of ρ for $i = 1, \ldots, n$ Thus we only have to examine the covariance. A straightforward calculation shows that

$$Cov(X_i(\rho), X_j(\rho)) = \alpha_i\alpha_j \quad (P(X_i(\rho) = \alpha_i, X_j(\rho) = \alpha_j) - q_iq_j).$$

Hence it is sufficient to consider the expression

$$P(X_i(\rho) = \alpha_i, X_j(\rho) = \alpha_j) = P(Z_i + W \geq z_{p_i}, Z_j + W \geq z_{p_j}) =: \bar{F}_\rho(z_{p_i}, z_{p_j})$$

where $\bar{F}_\rho$ is the survival function of a bivariate normal distribution with standard normal marginals and correlation coefficient ρ. It follows from Slepian's inequality and its proof as given e.g. in Tong (1980, p. 8ff) that $\rho \rightarrow \bar{F}_\rho$ is increasing and continuous. Hence $\rho \rightarrow Corr(X_i(\rho), X_j(\rho))$ is also increasing and continuous. Non-negativity then follows from the fact that $X(0)$ is a vector of independent random variables □

5. Numerical Example

Let us now illustrate the effect of dependencies in model 3.1 by a numerical example. In order to keep the computation simple, we have chosen $g(x, y, z) = y$. The sequence of random variables $\{G_\nu\}$ is i.i.d. with a two-point distribution on 0 and 4, where the value of 4 occurs with probability 0.06. The portfolio consists of 20 risks. We have computed the relative stop-loss premiums for 8 different scenarios which are given by their group structures k_i, $i = 1, \ldots, 8$ listed in table 1.

TABLE 1

scenario i	k_i
1	(1, 1, 1, , 1, 1, 1)
2	(4, 3, 3, 2, 2, 1, 1, 1, 1, 1, 1)
3	(8, 2, 2, 2, 2, 2, 2)
4	(4, 4, 4, 3, 3, 2)
5	(15, 2, 1, 1, 1)
6	(5, 5, 5, 5)
7	(10, 5, 5)
8	(20)

Scenario 1 corresponds to the safest portfolio with 20 independent risks and scenario 8 is the riskiest portfolio, where the same risk occurs 20 times. In the next table we summarize the ordering relations of these vectors with respect to the majorization ordering.

TABLE 2

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8
k_1	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$
k_2		$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$
k_3			$\prec$	$\not\prec$	$\prec$	$\not\prec$	$\prec$	$\prec$
k_4				$\prec$	$\prec$	$\prec$	$\prec$	$\prec$
k_5					$\prec$	$\not\prec$	$\not\prec$	$\prec$
k_6						$\prec$	$\prec$	$\prec$
k_7							$\prec$	$\prec$

The symbol $\not\prec$ indicates that the vectors cannot be compared The following table now contains the relative stop-loss premiums (divided by the independent case $i = 1$) multiplied by 100 for several retention levels. Note that the expectation of the aggregate claims equals 4 8 and the outcomes range between 0 and 80.

TABLE 3

	scenario							
retention	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8
0	100	100	100	100	100	100	100	100
1	100	105	109	110	111	112	113	116
2	100	113	121	124	126	129	132	139
3	100	124	140	145	150	155	161	173
4	100	144	173	182	191	200	210	233
6	100	174	210	229	272	272	295	347
8	100	270	330	385	537	506	572	717
10	100	327	478	480	830	700	834	1128

Because of Theorem 3.2 we know that given a retention level, the relative stop-loss premium increases in k. Table 3 shows that the increase is moderate if k_i and k_j are in some sense nearby as for example k_6 and k_7. In the cases where we were not able to establish the comparison theoretically like for example for scenario 5 and 6, we find that the order can change when the retention level increases Theorem 4 2 explains the monotonicity of the relative stop-loss premium with respect to the retention in scenario 8. The numerical data suggest that this is also true for the other scenarios. This was already observed by Dhaene and Goovaerts (1996). To our knowledge this is still an open problem.

A very important conclusion that we can draw from the computation is that the increase in the relative stop-loss premium can be dramatic in the presence of positive dependence Even minor occurrence of dependence like in scenario 2 has a severe effect. Moreover, if a portfolio contains positive dependence between the risks, the situation deteriorates in the number of insured risks.

Suppose Y, X_1, .., X_n are i.i.d. random variables (w.l o g we assume that they are concentrated on [0,1]) and we are interested in the stop-loss premiums of the safest portfolio $\pi^n_X(t) = E(\sum_{i=1}^n X_i - nt)^+$ and the riskiest one $\pi^n_Y(t) = E(nY - nt)^+$, where $t \in (0, 1)$ gives the retention percentage. In this setting we obtain

Theorem 5.1 *The ratio $\pi^n_Y(t)/\pi^n_X(t)$ is increasing in the number n of aggregate risks and the limit is equal to $E(Y-t)^+/(EY-t)$ if $t < EY$ and $+\infty$ if $t \geq EY$*

Proof We obtain that

$$\frac{\pi^n_Y(t)}{\pi^n_X(t)} = \frac{E(nY - nt)^+}{E(\sum_{i=1}^n X_i - nt)^+} = \frac{E(Y-t)^+}{E(\frac{1}{n}\sum_{i=1}^n X_i - t)^+}$$

Hence it suffices to prove that $E(\frac{1}{n}\sum_{i=1}^{n} X_i - t)^+$ is decreasing in n. Since $X_1, \ldots, X_n$ are i.i.d. it follows from Theorem 4 in Arnold and Villasenor (1986) that

$$\frac{1}{n+1}\sum_{i=1}^{n+1} X_i \leq_{sl} \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5}$$

and the monotonicity follows.
Since the random variables $X_1, X_2, \ldots$ are independent and identically distributed with a finite mean, the assumptions of the strong law of large numbers are fulfilled Therefore

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i = EX_1 = EY \tag{6}$$

Hence the stated limit follows. □

Remark: Arnold and Villasenor (1986) have shown that for Equation 5 it is sufficient, that $X_1, X_2,$ are exchangeable Hence the monotonicity part of Theorem 5.1 remains true for the more general case of exchangeable random variables, but in that case the limit will be different. In fact, there is also a version of the strong law of large numbers for sequences of exchangeable random variables. It states that in this case

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i = E[X_1|\Theta],$$

where Θ is the random variable, which describes the mixing mechanism in de Finetti's Theorem (cf. Feller (1966) and Chow and Teicher (1978) for more details) Hence in this case we get

$$\lim_{n\to\infty}\frac{\pi_{Y}^{n}(t)}{\pi_{X}^{n}(t)} = \frac{E(Y-t)^+}{E(E[Y|\Theta]-t)^+}$$

From Theorem 5.1 we see that the relative stop-loss premium can be arbitrary high, when the retention exceeds the expected aggregate claim
Altogether we can conclude that the usual assumption of independence in risky portfolios leads to a dangerous underestimation of the risk Hence the adequate modeling of dependent risks will remain an important task for future research.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referees for several helpful comments

REFERENCES

ARNOLD, B C , J A VILLASENOR (1986) Lorenz ordering of means and medians *Statistics and Probability Letters* **4**, 47-49

BAUERLE, N (1997a) Inequalities for stochastic models via supermodular orderings *Communications in Statistics Stochastic Models* **13**, 181-201

BAUERLE, N (1997b) Monotonicity results for MR/GI/1 queues *Journal of Applied Probability* **34**, 514-524

BAUERLE, N , U RIEDER (1997) Comparison results for Markov-modulated recursive models *Probability in the Engineering and Informational Sciences* **11**, 203-217

CHOW, Y S , H TEICHER (1978) *Probability Theory* Springer, New York

DHAENE, J , M J GOOVAERTS (1996) Dependency of risks and stop-loss order *ASTIN Bulletin* **26**, 201-212

DHAENE, J , M J GOOVAERTS (1997) On the dependency of risks in the individual life model *Insurance Mathematics and Economics* **19**, 243-253

FELLER, W (1966) *An introduction to Probability Theory and Its Applications, Volume II* Wiley, New York

GOOVAERTS, M J , R KAAS, A E VAN HEERWAARDEN, T BAUWELINCKX (1990) *Effective Actuarial Methods* Insurance Series vol 3, North Holland

HEILMANN, W -R (1986) On the impact of independence of risks on stop loss transforms *Insurance Mathematics and Economics* **5**, 197-199

HURLIMANN, W (1993) Bivariate distributions with diatomic conditionals and stop-loss transforms of random sums *Statistics and Probability Letters* **17**, 329-335

LEFEVRE, C , S UTEV (1996) Comparing sums of exchangeable Bernoulli random variables *Journal of Applied Probability* **33**, 285-310

MADSEN, R W (1993) Generalized binomial distributions *Communications in Statistics – Theory and Methods* **22**, 3065-3086

MARSHALL, A W , I OLKIN (1979) *Inequalities Theory of Majorization and its Applications* Academic Press, New York

MULLER, A (1996) Orderings of risks A comparative study via stop-loss transforms *Insurance Mathematics and Economics* **17**, 215-222

MULLER, A (1997) Stop-Loss Order for Portfolios of Dependent Risks *Insurance Mathematics and Economics* **21**, 219-223

SCHMEIDLER, D (1986) Integral Representation without additivity Proceedings of the American Mathematical Society **97**, 255-261

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

SHAKED, M , J G SHANTHIKUMAR (1997) Supermodular Stochastic Orders and Positive Dependence of Random Vectors *Journal of Multivariate Analysis* **61**, 86-101

SHAKED, M , Y L TONG (1985) Some Partial Orderings of Exchangeable Random Variables by Positive Dependence *Journal of Multivariate Analysis* **17**, 333-349

TCHEN, A H (1980) Inequalities for Distributions with given marginals *Annals of Probability* **8**, 814-827

TONG, Y L (1980) *Probability Inequalities in multivariate Distributions* Academic Press, New York

TONG, Y L (1989) Inequalities for a class of positively dependent random variables with a common marginal *Annals of Statistics* **17**, 429-435

WANG, S (1997) Aggregation of Correlated Risk Portfolios Models and Algorithms *Preprint*

YAARI, M E (1987) The dual theory of choice under risk *Econometrica* **55**, 95-115

NICOLE BAUERLE
Abteilung Mathematik VII
(Operations Research)
Helmholtzstr. 18
Universitat Ulm
D-89069 Ulm, Germany

ALFRED MULLER
Institut fur Wirtschaftstheorie und
Operations Research
Kaiserstr. 12
Universität Karlsruhe
D-76128 Karlsruhe, Germany

SOME APPLICATIONS OF LÉVY PROCESSES TO STOCHASTIC INVESTMENT MODELS FOR ACTUARIAL USE

BY

TERENCE CHAN

Dept. of Actuarial Mathematics and Statistics
Heriot-Watt University

ABSTRACT

This paper presents a continuous time version of a stochastic investment model originally due to Wilkie. The model is constructed via stochastic differential equations. Explicit distributions are obtained in the case where the SDEs are driven by Brownian motion, which is the continuous time analogue of the time series with white noise residuals considered by Wilkie In addition, the cases where the driving "noise" are stable processes and Gamma processes are considered.

KEYWORDS

Lévy process; Brownian motion; stochastic investment model

1. INTRODUCTION

Wilkie (1986) presented an investment model based on time series, which has since been updated and extended in Wilkie (1995) This paper presents some continuous time variants of Wilkie's original model using stochastic differential equations driven by appropriate Lévy processes. There is no single correct continuous time equivalent to the model in Wilkie (1986), the aim of this paper is to suggest some possible ways of constructing the analogous continuous time models and to analyse these mathematically. It seems that whatever one takes to be the "right" continuous time equivalent of the Wilkie model, similar methods to those presented here can be used to analyse it.

One reason one might be interested in a continuous time model is that in a continuous time setting one is free to choose any unit of time and to model the state of the various investment variables at any time, not just at discrete instants. However, the main attraction of continuous time models is their mathematical tractability; whereas the Wilkie model is mainly intended for computer simulations, in the continuous time setting here many questions admit explicit answers which can be obtained in a simple way. Here, we concentrate on obtaining explicit distributions but other questions can undoubtedly be answered

The model introduced in Wilkie (1986) only makes use of Gaussian (white noise) series, for our model the driving noises are allowed to have other distributions.

2. Description of the model

In many ways, the model described here is the most direct and obvious continuous-time version of the model in Wilkie (1986), although some modifications are necessitated by the transition to a continuous time scale. We do not make any special claims about its appropriateness to practical situations beyond pointing out its similarity to the original Wilkie model which has by now gained wide acceptance, at least in the world of insurance. The two main guiding principles behind the construction of the continuous time model presented here are firstly the analogy with the corresponding time series and secondly the similarities between certain features of the Wilkie model and other models which feature widely in different areas of financial modelling, occasionally we shall depart from an exact analogy with the time series to emphasise these similarities because the qualitative features common to all these models are of potentially greater interest Thus, it would be more appropriate to refer to the continuous time model presented here as inspired by the Wilkie model, rather than "the continuous time Wilkie model" The model should be treated as a "first draft" rather than a final version. As with the original Wilkie model, the model here is based on four processes (although these are not exactly the same as the ones in Wilkie (1986)) and we describe each of these in turn.

Let Z_1, Z_2, Z_3 and Z_4 be four independent (not necessarily continuous) processes. Exactly what kind of processes are the Z_i will be discussed later.

1.1. Retail prices index and inflation

Consider first a retail prices index, $Q_t \equiv \exp\{P_t\}$ We use an Ornstein-Uhlenbeck type model for the process P.

$$\begin{aligned} dP_t &= R_t dt \\ dR_t &= -a_1 R_t dt + \phi(t)dt + \sigma_1 dZ_1(t) \end{aligned} \tag{1 1}$$

where $a_1 > 0$, $\sigma_1 \in \mathbb{R}$ and ϕ is a (deterministic) positive periodic function with period $h > 0$. Here the process R plays the role of the continuous force of inflation. A direct translation of Wilkie's model would have $\phi \equiv$ constant, but in passing to continuous time it may be desirable to take into account the seasonal fluctuations in inflation over a year. The period h here corresponds to a year in our units of time (see Remark (ii) below) (To spell things out in a little more detail, supposing $\phi \equiv$ constant, the process R_t in (1 1) corresponds to Wilkie's $\nabla \log Q(t)$, the parameter a_1 corresponds to the parameter Wilkie calls 1 – QA, ϕ corresponds to QMU(1 – QA) in Wilkie's original paper and σ_1 plays the role of QSD).

Because (1.1) is a linear equation, it is easy to solve explicitly, whatever our choice of driving noise Z_1. The general version of Itô's formula for discontinuous semimartingales X states that if f is a continuous function with the necessary derivatives,

$$f(X_t,t) - f(X_0,0) = \int_0^t \dot{f}(X_s,s)ds + \int_0^t f'(X_{s-},s)dX_s + \frac{1}{2}\int_{0+}^t f''(X_{s-},s)d[X]_s$$
$$+ \sum_{0<s\le t} \left(f(X_s,s) - f(X_{s-},s) - f'(X_{s-},s)\Delta X_s - \frac{1}{2}f''(X_{s-},s)(\Delta X_s)^2 \right)$$

where $\Delta X_s = X_s - X_{s-}$, $\dot{f} \equiv \partial f/\partial t$, $f' \equiv \partial f/\partial x$ etc. For this and other aspects of the general theory of stochastic integration with respect to semimartingales, we refer the reader to Protter (1990) and Roger and Williams (1987), which approach the subject in different ways. (Note that X is assumed to be right-continuous and can only have countably many jumps, so the sum above is actually a sum over countably many values of s). Consider now the case that $f(x,t) = e^{at}X_t$ We have $f'' \equiv 0$ and $f'(X_{s-},s)\Delta X_s = f(X_s,s) - f(X_{s-},s)$, so the terms involving the jumps of X in Itô's formula all vanish Therefore applying Itô's formula to $e^{a_1 t}R_t$, we obtain an explicit formula for R_t:

$$R_t = e^{-a_1 t}R_0 + \int_0^t e^{-a_1(t-s)}\phi(s)ds + \int_0^t \sigma_1 e^{-a_1(t-s)}dZ_1(s). \qquad (1\ 2)$$

From (1.2), we can find $P_t = P_0 + \int_0^t R_s ds$ and the resulting double integrals can be handled by interchanging the order of integration (e g. see Lemma 3 1 in the sequel)

1.2. Share yield process

Wilkie (1986) next considers two inter-related processes: an index of share dividends and the dividend yield process. Let Y_t denote the share dividend yield The continuous time analogue of Wilkie's model would be

$$Y_t = Y_* \exp\{X_t + \zeta R_t\},$$
$$\text{where} \quad dX_t = -a_2 X_t dt + b_1 dt + \sigma_2 dZ_2(t) \qquad (1\ 3)$$

(Here, $Y_* = Y_0 e^{-(X_0+\zeta R_0)}$. In the sequel, this notation will be frequently used to denote this kind of "modified initial condition".) Equation (1.3) admits an explicit solution similar to (1.2), namely

$$X_t = X_0 e^{-a_2 t} + b_1\left(\frac{1-e^{-a_2 t}}{a_2}\right) + \int_0^t \sigma_2 e^{-a_2(t-s)} dZ_2(s). \tag{1.4}$$

1.3. Share dividend process

We next turn to the index of share dividends, D_t Our model follows Wilkie in using an exponentially discounted "sum of inflation effects".

$$d(\log D_t) = \left(b_2 + \beta\lambda \int_0^t e^{-\lambda s} R_{t-s} ds + \gamma R_t\right) dt + \eta_2 dZ_2(t) + \eta_3 dZ_3(t). \tag{1.5}$$

In Wilkie's time series model, the noise has a simultaneous as well as a lagged effect which is captured by moving average in the noise. There is no sensible equivalent in the continuous time context for such a moving average. Another feature of the model (inherited from Wilkie) is the mixing of the driving noises for Y_t and D_t.

The share price S_t is related to the dividends and the yield by $S_t = D_t / Y_t$ It is interesting to note that the process S_t satisfies an equation of the form

$$dS_t = c_t S_t dt + S_t(\delta_1 dZ_1(t) + \delta_2 dZ_2(t) + \delta_3 dZ_3(t)),$$

which has exactly the same form as the ubiquitous geometric Brownian motion model of share prices, except that the coefficient c_t here takes a rather complicated form which involves the whole path of the force of inflation R up to time t, as well as the usual constant drift terms

Interchanging the order of integration, it is easy to see that

$$\lambda \int_0^t \int_0^s e^{-\lambda u} R_{s-u} du\, ds = \int_0^t (1 - e^{-\lambda(t-u)}) R_u du ,$$

therefore from (1.5) we have

$$D_t = D_* \exp\left\{\eta_2 Z_2(t) + \eta_3 Z_3(t) + \beta \int_0^t (1 - e^{-\lambda(t-u)}) R_u du + \gamma \int_0^t R_u du + b_2 t\right\}. \tag{1.6}$$

where D_* is a constant determined by D_0 and R_0 in a similar manner to Y_* (see Remark (iv) in §1.5 below).

1.4. Consol yield process

Finally, we have the yield on consols C_t,

$$\begin{aligned} C_t &= \xi\rho \int_0^t e^{-\rho s} R_{t-s} ds + C_* e^{V_t}, \\ dV_t &= -a_4 V_t dt + \sigma_4 dZ_4(t), \quad V_0 = v \end{aligned} \tag{1.7}$$

The equation for V in (1.7) admits an explicit solution for the same form as (1 2)

1.5. General remarks on the model

(i) We do not claim that the method in (1.1) is the most appropriate way to model seasonal effects in inflation – it is *one* simple and obvious way to do it without destroying the most attractive features of the Ornstein-Uhlenbeck process but we could equally plausibly let σ_1 be a periodic function as well and we would still be able to obtain an explicit solution as before

(ii) Some remarks on the time scale of the continuous time processes here and their relationship with their discrete-time counterparts in Wilkie (1986) might be useful Typically these continuous time processes run at a much faster speed than their discrete-time equivalents: for example, if the unit of time in Wilkie (1986) is years, the unit of time here might be centuries, so that $h = 0.01$ would correspond to a year. This is essentially an artifact of the discretization in passing from continuous time to discrete time. If we were to discretize (1.1) in multiples of h using first-order Euler approximation together with the approximation $P_t - P_{t-h} = \int_{t-h}^t R_s ds \approx hR_{t-h}$ and noting that $\phi(t) = \phi(t-h) = \phi =$ constant, we would recover the Wilkie model provided we rescale time by defining $\tilde{R}_t := R_{ht}$. For example, assuming that Z_1 is Brownian motion for simplicity, the first-order Euler discretization of (1.1) is

$$R_t - R_{t-h} = -a_1 h R_{t-h} + \phi h + \sigma_1 (Z_1(t) - Z_1(t-h)),$$

which can be rewritten as

$$\begin{aligned} R_t &= (1 - a_1 h) R_{t-h} + \phi h + \sigma_1 \sqrt{h}\, W_t \\ &= \mu + a(R_{t-h} - \mu) + \sigma_1 \sqrt{h}\, W_t, \end{aligned} \tag{1.8}$$

where we have put $a = 1 - a_1 h$, $\mu = \phi / a_1$ and $W_t = (Z_1(t) - Z_1(t-h))/\sqrt{h}$. Note that W_h, W_{2h}, W_{3h}, ... are i.i.d. standard Gaussian random variables. Defining $\tilde{R}_t := R_{ht}$, we obtain from (1.8) the AR(1) time-series model of Wilkie:

$$\tilde{R}_t = \mu + a(\tilde{R}_{t-1} - \mu) + \tilde{\sigma}_1 W_t, \tag{1.9}$$

where $\tilde{\sigma}_1 = \sigma_1\sqrt{h}$. The calculations at (2.5) and (2.6) below and the subsequent discussion illustrate this point in greater detail Observe that the corresponding parameters in (1.1) are rescaled in the appropriate way with this time change: $\tilde{a}_1 = ha_1$ and $\tilde{\sigma}_1 = \sigma_1\sqrt{h}$ Because the parameters are automatically scaled accordingly once a time scale has been chosen, such comparisons with the discrete time-series are usually irrelevant from a practical point of view; in practice, one would choose a suitable time scale and then fit the model to data directly without reference to any discrete-time model and if one wished to do simulation, one would choose a discretization for its numerical efficiency rather than for its consistency with another discrete-time model. The same comment applies to all the other processes discussed above

(iii) For our choices of Z_1, the process R will have a stationary distribution. Throughout this paper, we assume that the initial condition R_0 is some fixed number as in (1.2) However, it is also possible to let R_0 be a random variable with the stationary distribution, in which case R would be a stationary process. The same can be said of all the other processes which have stationary distributions.

(iv) Because the processes X and R in (1.3) are not spatially homogeneous, the initial values X_0 and R_0 cannot be absorbed into Y_* and so separate parameters for the initial values are needed The same applies to the processes D and C. Also, Wilkie (1986) has an extra drift term of the form $c\,dt$ appearing in the equation for V_t in (1.7) but we have omitted it here because it is clear from the explicit formula for V_t that c can be absorbed into the two parameters ν and C_*, and so serves no additional purpose

1.6. Lévy processes

We are mainly interested in the case where the "noise" processes Z_t are symmetric Lévy processes, that is processes with stationary independent increments ("Symmetric" in this context just means that Z and $-Z$ have the same law.) We end this section by briefly recalling some results about Lévy processes which we shall need in the sequel. Let Z be a (symmetric) Lévy process. Since Z has stationary independent increments, its characteristic function must take the form $\mathbb{E}\left[e^{i\theta Z_t}\right] = e^{-t\psi(\theta)}$ for some function ψ, called the *Lévy exponent* of Z. The Lévy-Khintchine formula says that

$$\psi(\theta)=\frac{\sigma^2}{2}\theta^2+ia\theta+\int_{\{|x|<1\}}(1-e^{i\theta x}-i\theta x)\nu(dx)+\int_{\{|x|\geq 1\}}(1-e^{i\theta x})\nu(dx) \quad (1.10)$$

for $a, \sigma \in \mathbb{R}$ (if Z is symmetric, $a = 0$) and for some σ-finite measure ν on $\mathbb{R}\setminus\{0\}$ satisfying $\int \min(1, x^2)\nu(dx) < \infty$. The measure ν is called the *Lévy measure* of Z. (To put readers on more familiar ground, consider the situation when $\sigma = a = 0$ and suppose that the total mass λ of ν, $\lambda = \int_{\mathbb{R}} \nu(dx)$ is finite. Then the Lévy process Z with such a Lévy measure is just an ordinary compount Poisson process which jumps occurring as a Poisson process of rate λ and whose jump-size distribution is $\lambda^{-1}\nu(dx)$. In the case that the integral of ν diverges near 0, Z will have infinitely many small jumps in a finite time-interval At the other extreme, if $\nu \equiv 0$, there are no jumps so we just have Brownian motion and ψ is the same as the exponent for a normal distribution)

From the Lévy-Khintchine formula we can deduce the exact form Z must take. it turns out that Z must be a linear combination of a Brownian motion (the continuous part) and a pure-jump process independent of the Brownian part. Specifically, let $Q\ (dt, dx)$ be a Poisson measure on $(0, \infty)\times \mathbb{R}\ \{0\}$ with expectation measure $dt \times \nu$ (here dt denotes Lebesgue measure), then (assuming $a = 0$ in (1.10)) we have the Lévy decomposition

$$Z_t = \sigma B_t + J_t + A_t \tag{1.11}$$

where, corresponding to each of the three terms in (1.10) respectively, B is a Brownian motion, J is the pure-jump martingale $J_t = \int_{|x|<1} x(Q((0,t], dx) - t\nu(dx))$ and A is the finite-variation jump process $A_t = \int_{\{|x|\geq 1\}} xQ((0, t], dx)$. The processes B, J and A are independent A more detailed treatment can be found in Protter (1990) and Rogers and Williams (1987) also contains a nice direct construction of (1.11). Because of independence, we lose no generality in treating separately the cases where Z is a Brownian motion and where Z is a pure-jump process We do this in the next two sections

3 Explicit distributions in the Brownian case

If the Z_t are all Brownian motions, all the processes described in the previous section are either Gaussian processes or exponentials of Gaussian processes Since in order to specify the law of a Guassian process one only has to specify the mean and the covariance, the results of this section are essentially trivial.

Recall that for a Brownian motion W, $\int_0^t f(s)dW_s = B\left(\int_0^t f(s)^2 ds\right)$ where B is some other Brownian motion. Applying this result to (1.2) gives

$$R_t = e^{-a_1 t}R_0 + \int_0^t e^{-a_1(t-s)}\phi(s)ds + \sigma_1 e^{-a_1 t}B_1\left(\frac{e^{2a_1 t} - 1}{2a_1}\right) \tag{2.1}$$

where B_1 is a Brownian motion. Hence, R_t has Gaussian distribution with mean

$$\mu_R(t) = e^{-a_1 t} R_0 + \int_0^t e^{-a_1(t-s)} \phi(s) ds \tag{2.2a}$$

and variance

$$v_R(t) = \sigma_1^2 \left(\frac{1 - e^{-2a_1 t}}{2a_1} \right) . \tag{2.2b}$$

(In 2.2a,b) we have used the fact that B_t is Gaussian with mean 0 and variance t). Similar results hold for the other Ornstein-Uhlenbeck type processes X and V introduced in Section 1.

From (1.6) and (1.7), it is clear that the key to finding the distributions of D_t and C_t lies in obtaining the distribution of $\int_0^t f(s) R_s\, ds$ for suitable (deterministic) functions f. Since R is a Gaussian process, so is $t \mapsto \int_0^t f(s) R_s\, ds$ and so all we need to do is work out the mean and variance of $\int_0^t f(s) R_s\, ds$. The mean is trivial: by interchanging the order of integration it is easy to see that the mean is just $\int_0^t f(s) \mu_R(s) ds$ We now turn to the variance Since the mean is irrelevant here, the variance is simply given by

$$\mathbb{E}^0 \left[\left(\int_0^t f(s) H_s ds \right)^2 \right] = \mathbb{E}^0 \left[\int_0^t \int_0^t f(s) f(u) H_s H_u \, du\, ds \right]$$

where we have put

$$H_t = \sigma_1 e^{-a_1 t} B_1 \left(\frac{e^{2a_1 t} - 1}{2a_1} \right)$$

and we use the superscript in $\mathbb{E}^0$ to emphasise that $H_0 = B_1(0) = 0$. Using the covariance of Brownian motion $\mathbb{E}\ (B_s B_u) = \min(s, u)$ and interchanging the order of integration, we get

$$\begin{aligned} &\mathbb{E}^0 \left[\int_0^t \int_0^t f(s) f(u) H_s H_u \, du\, ds \right] \\ &= 2 \int_0^t f(s) \int_0^s f(u) E(H_s H_u) \, du\, ds \\ &= 2\sigma_1^2 \int_0^t f(s) e^{-a_1 s} \int_0^s f(u) e^{-a_1 u} \left(\frac{e^{2a_1 u} - 1}{2a_1} \right) du\, ds \end{aligned} \tag{2.3}$$

Putting $f \equiv 1$ in (2.3) gives the variance of $\int_0^t R_s\, ds$ to be

$$\frac{\sigma_1^2}{a_1} \left(\frac{t}{a_1} + \frac{2e^{-a_1 t}}{a_1^2} - \frac{e^{-2a_1 t}}{2a_1^2} - \frac{3}{2a_1^2} \right) . \tag{2.4}$$

At this point, it may be instructive to compare these results with the analogous ones for the AR(1) time series (1.9) The mean and variance of $\sum_{i=1}^{t} \tilde{R}_i$ has been obtained by Hurlimann (1992) and Wilkie (1995). Keeping to our notation established in (1.9), the mean of the accumulated force of inflation $\sum_{i=1}^{t} \tilde{R}_i$ is

$$\mu t + (\tilde{R}_0 - \mu)\frac{a(1-a^t)}{1-a} \qquad (2.5)$$

while in the continuous model the mean $\int_0^t R_s\,ds$ is (assuming ϕ = const.)

$$\int_0^t \mu_R(s)ds = \mu t + \frac{R_0 - \mu}{a_1}[1 - e^{-a_1 t}], \qquad (2.6)$$

where $\mu = \phi/a_1$ as before. We see immediately that (2.5) and (2.6) have the same form. To check that they in fact agree, recall that to obtain the time-series (1.9) from (1.1), we discretized time into steps of size h Therefore $\sum_{i=1}^{t/h} h\tilde{R}_i$ is precisely the Riemann-sum approximation to $\int_0^t R_s ds$. According to the formula (2.5), the mean of $\sum_{i=1}^{t/h} h\tilde{R}_i$ is

$$\begin{aligned}
&\mu t + (\tilde{R}_0 - \mu)h\frac{a(1-a^{t/h})}{1-a} \\
&= \mu t + (R_0 - \mu)\left(\frac{1 - a_1 h}{a_1}\right)\left[1 - (1 - a_1 h)^{t/h}\right] \\
&\to \mu t + \frac{R_0 - \mu}{a_1}[1 - e^{-a_1 t}]
\end{aligned}$$

as $h \to 0$, which is precisely the mean of $\int_0^t R_s ds$ given by (2.6). Similarly, Hurlimann (1992) gives the variance of $\sum_{i=1}^{t} \tilde{R}_i$ as

$$\frac{\tilde{\sigma}_1^2}{(1-a)^2}\left[t - \frac{2a(1-a^t)}{1-a} + \frac{a^2(1-a^{2t})}{1-a^2}\right],$$

which has the same form as (2.4).

It is just as easy to obtain the distributions of the other processes in our model Putting $f(s) = \rho e^{-\rho(t-s)}$ in (2.3) we get that $\int_0^t e^{-\rho(t-s)} R_s\,ds$ has Gaussian distribution with mean

$$\int_0^t \rho e^{-\rho(t-s)}\mu_R(s)\,ds \qquad (2.7a)$$

and variance

$$\frac{\rho\sigma_1^2}{a_1}\left(\frac{(a_1-\rho)^2 - (a_1+\rho)(\rho e^{-2a_1 t} + a_1 e^{-2\rho t}) + 4a_1\rho e^{-(a_1+\rho)t}}{2(a_1-\rho)^2(a_1+\rho)}\right) \qquad (2.7b)$$

Putting $f = \beta + \gamma - \beta e^{-\lambda(t-s)}$ also gives an explicit expression for the variance of $\beta \int_0^t (1 - e^{-\lambda(t-s)}) R_s \, ds + \gamma \int_0^t R_s \, ds$, although this is too messy to write down here – the formula is simplified somewhat by choosing $\gamma = 0$ and simplified considerably by choosing $\gamma = -\beta$, for this would then reduce to (2.7b). The full covariance structure of the process $t \mapsto \int_0^t f(s) R_s \, ds$ can also be obtained in this way

Armed with these results, we can now state the distributions of interest. We have already found the distributions of R_t and $P_t = \int_0^t R_s \, ds$ (see (2.2) and (2.4)). Applying the results (2 2) to the process X, we get from (1.3) and (1.4) that $\log Y_t$ has Gaussian distribution with mean

$$\log Y_* + X_0 e^{-a_2 t} + b_1 \left(\frac{1 - e^{-a_2 t}}{a_2} \right) + \zeta \mu_R(t)$$

and variance

$$\zeta^2 v_R(t) + \sigma_2^2 \left(\frac{1 - e^{-2a_2 t}}{2a_2} \right) .$$

For the dividend index D, the result (2.3), with $f(s) = \beta + \gamma - \beta e^{-\lambda(t-s)}$, together with the analogous results (2.2) for U give that $\log D_t$ has Gaussian distribution with mean

$$\log D_* + b_2 t + \int_0^t (\beta + \gamma - \beta e^{-\lambda(t-s)}) \mu_R(s) \, ds$$

and variance

$$\begin{aligned}&(\eta_2^2 + \eta_3^2) t \\ &\quad + 2\sigma_1^2 \int_0^t (\beta + \gamma - \beta e^{-\lambda(t-s)}) e^{-a_1 s} \int_0^s (\beta + \gamma - \beta e^{-\lambda(t-u)}) \left(\frac{e^{a_1 u} - e^{-a_1 u}}{2a_1} \right) du \, ds.\end{aligned}$$

Applying (2 2) to V_t shows that it is Gaussian with mean

$$\mu_V(t) = v e^{-a_4 t}$$

and variance

$$v_V(t) = \sigma_4^2 \left(\frac{1 - e^{-2a_4 t}}{2a_4} \right)$$

The distribution of C_t is the convolution of normal and log normal distributions and the results (2.2) and (2 7) show that C_t has mean

$$\xi \int_0^t \rho e^{-\rho(t-s)} \mu_R(s) \, ds + C_* e^{\mu_V(t) + v_V(t)/2}$$

and variance

$$\frac{\rho\xi^2\sigma_1^2}{a_1}\left(\frac{(a_1-\rho)^2-(a_1+\rho)(\rho e^{-2a_1t}+a_1e^{-2\rho t})+4a_1\rho e^{-(a_1+\rho)t}}{2(a_1-\rho)^2(a_1+\rho)}\right)$$
$$+\,C_*^2e^{2\mu_Y(t)+\nu_Y(t)}(e^{\nu_Y(t)}-1).$$

It is also possible to specify the full multivariate structure of R, Y and D using the methods here. Since R, Y and D are either Gaussian or log Gaussian, their joint law is specified once we have the covariances $\mathrm{Cov}(R_t, \log Y_t)$, $\mathrm{Cov}(R_t, \log D_t)$ and $\mathrm{Cov}(\log Y_t, \log D_t)$. For the most part, we only need to know the covariance stucture of the process $t\mapsto R_t$, which is given by

$$\mathbb{E}[R_tR_s]=\mu_R(t)\mu_R(s)+\mathbb{E}[H_sH_u]=\mu_R(t)\mu_R(s)+\sigma_1^2e^{-a_1t}\left(\frac{e^{a_1s}-1}{2a_1}\right)$$

if $s<t$ Thus, for example,

$$\mathbb{E}[R_t\log D_t]=\mu_R(t)(\log D_*+b_2t)+\int_0^t(\beta+\gamma-\beta e^{-\lambda(t-s)})\mathbb{E}[R_tR_s]\,ds$$

and we can then substitute the relevant previous results into the above expression. In addition, we also need the covariance of X_t and $Z_2(t)$, which is given by

$$\mathbb{E}^0[X_tZ_2(t)]=\sigma_2e^{-a_2t}\min\left(\frac{e^{a_2t}-1}{2a_2},\ t\right)$$

using the covariance of Brownian motion. The detailed computations of the covariances are left to the reader.

4. Explicit distributions in the discontinuous case

There have been some suggestions that Gaussian noise terms are not entirely appropriate for these models and that more realistically, the noise should have jumps In this section, we perform the same analysis as in Section 2 on the assumption that the Z_t are symmetric pure-jump Lévy processes.

From the analysis in Section 2, it is clear that once we know what the law of $\int_0^t f(s)Z_s\,ds$ is for fixed t (where $f(s)$ or $f(t,s)$ is a suitable function and Z is a generic Lévy process), we can obtain the necessary explicit distributions. It all turns out to rest on the following simple lemma allowing the interchange of order of integration.

LEMMA 3.1: *Let f and g be Riemann-integrable functions. Then the laws of*

$$\int_0^t f(s)\int_0^s g(u)\,dZ_u\,ds\quad \textit{and}\quad \int_0^t g(u)\int_u^t f(s)\,ds\,dZ_u$$

are the same for each fixed t and the common law is given by

$$\mathbb{E}\left[\exp\left\{-i\theta\int_0^t g(u)\int_u^t f(s)ds\,dZ_u\right\}\right]=\exp\left\{-\int_0^t \psi(\theta g(u)[F(t)-F(u)])du\right\} \tag{3.1}$$

where ψ is given by the Lévy-Khintchine formula (1.10) and $F(u) = \int_0^u f(s)\,ds$

The proof, although not very pretty, uses only well-known standard results in the theory of stochastic integration and Lévy processes and is presented in the Appendix.

Remarks

(i) The above lemma is trivial if Z has finite variation, for then the integral $\int_0^t g(s)\,dZ_s$ exists as an ordinary Riemann-Stieltjes integral. Changing the order of integration as for ordinary integrals, we actually have the much stronger result that

$$\mathbb{P}\ \left(\int_0^t f(s)\int_0^s g(u)\,dZ_u\,ds = \int_0^t g(u)\int_u^t f(s)\,ds\,dZ_u \quad \forall t\right) = 1.$$

When Z has infinite variation, the integral with respect to Z is a "genuine" stochastic integral. In this case, we have to emphasise that Lemma 3.1 holds only for fixed t; the two integrals clearly cannot have the same law as processes since the former is a process of finite variation while the latter has infinite variation.

(ii) Since t is a fixed parameter in the present context, Lemma 3.1 holds equally if we allow f and g to also depend on t, which we need to do for some of the processes considered earlier.

(iii) Note that a simple special case of (3.1) is that

$$\mathbb{E}\left[\exp\left\{-i\theta\int_0^t G(t,u)\,dZ_u\right\}\right] = \exp\left\{-\int_0^t \psi(\theta G(t,u))\,du\right\} \tag{3.2}$$

for any (Riemann-integrable) function G.

Consider now the model described in Section 1 where the Z_t are symmetric Lévy processes with jumps. From the explicit formula (1.2) for R, we see that to find the law of R_t we can apply (3.2) with $G(t,u) = \sigma_1 e^{-a_1(t-u)}$, in which case we obtain

$$\mathbb{E}[e^{-i\theta R_t}] = \exp\left\{-i\theta\mu_R(t) - \int_0^t \psi(\theta\sigma_1 e^{-a_1(t-u)})\,du\right\},$$

where $\mu_R(t)$ is as defined by (2.2a). In a similar way we can obtain the laws of the processes X, U and V introduced in Section 1. For the law of $\int_0^t R_s\,ds$, we can apply Lemma 3.1 with $g(t,u) = \sigma_1 e^{-a_1(t-u)}$, $f \equiv 1$ and for the law of $\int_0^t \rho e^{-\rho(t-s)}R_s\,ds$ we can take $g(t,u) = \sigma_1 e^{-a_1(t-u)}$, $f(t,s) = \rho e^{-\rho(t-s)}$. In this way, we obtain the distributions of R_t, S_t, C_t and $\log D_t$ in a similar manner to Section 2. However, the joint distribution is much more difficult to obtain.

We end this section with a brief word on some specific examples of Lévy processes one might choose to use in these models. We just mention two commonly used Lévy processes. One is the symmetric α-stable process, whose Lévy exponent is $\psi(\theta) = |\theta|^\alpha$ and whose Lévy measure is

$$\nu(dx) = \frac{C_\alpha}{|x|^{1+\alpha}}dx, \quad x \neq 0,$$

where $C_\alpha = \pi^{-1}\Gamma(1+\alpha)\sin(\pi\alpha/2)$. (Here $0 < \alpha < 2$; $\alpha = 2$ corresponds to the Gaussian distribution and $\alpha = 1$ gives the Cauchy distribution). Stable distributions are examples of so-called heavy-tailed distributions. One of the disadvantages of stable processes is that they do not have higher order moments than 1 (for $\alpha \leq 1$ they do not even have a first moment) which may cause awkward problems, for example, when we take exponentials of stable processes as we are frequently doing in these models.

Another commonly used class of Lévy processes which overcomes this problem is the Gamma process. A Lévy process Y is said to be Gamma with parameters $(\alpha, \beta$ where $\alpha, \beta > 0$ if $\mathbb{P}\,(Y_1 \leq x) = \Gamma(\alpha)^{-1}\beta^\alpha \int_0^x y^{\alpha-1}e^{-\beta y}dy$. Hence

$$\mathbb{E}[e^{-i\theta Y_t}] = \left(\frac{\beta}{\beta + i\theta}\right)^{\alpha t} = \exp\left\{-\alpha t \log\left(1 + \frac{i\theta}{\beta}\right)\right\}.$$

Note that such a process is non-decreasing, so to obtain a symmetric process, we simply take two independent copies Y and $\tilde{Y}$ and define $Z = Y - \tilde{Y}$. The process Z is therefore a symmetric Lévy process with Lévy exponent $\psi(\theta) = \alpha\log(1 + \theta^2/\beta^2)$ and Lévy measure $\nu(dx) = \alpha|x|^{-1}e^{-\beta|x|}dx$. Looking at the Lévy decomposition, since $\int_{\{|x|<1\}} |x|\nu(dx) < \infty$, we see that Z has finite variation and since $\int_{\{|x|\geq 1\}} |x|^n\nu(dx) < \infty$, Z_t has finite moments of all orders.

Applying Lemma 3.1 we obtain (replacing $i\theta$ with θ for convenience)

$$\mathbb{E}\left[\exp\left\{-\theta\int_0^t f(t-s)\,dZ_s\right\}\right] = \exp\left\{-\alpha\int_0^t \log\left(1 - \frac{\theta^2 f(t-s)^2}{\beta^2}\right)ds\right\} \quad (3.3)$$

and

$$\mathbb{E}\left[\exp\left\{-\theta\int_0^t g(u)\int_u^t f(s)\,ds\,dZ_u\right\}\right]$$
$$= \exp\left\{-\alpha\int_0^t \log\left(1 - \frac{\theta^2 g(u)^2}{\beta^2}\left[\int_u^t f(s)\,ds\right]^2\right)du\right\} \quad (3.4)$$

Although in general it is not possible to give explicit formulae for the integrals in (3.3) and (3.4) for our choices of f and g as in the preceding two sections, the Laplace transforms (3.3-4) do give relatively simple expressions for the moments, involving integrals which can be readily evaluated by numerical means.

5. Concluding remarks

We have concentrated here on obtaining explicit formulae, both in the case where our SDEs are driven by Brownian motions and in the case where they are driven by symmetric Lévy processes with jumps. Of course, many other questions – which we have not considered – do not admit explicit answers and one must then resort to numerical solutions. It is not our intention here to give a detailed quantitative analysis of numerical simulations of the models presented in the preceding sections, as this could well constitute a paper in its own right. We simply present some examples of numerical simulations to give a feel for what these processes look like In the case of SDEs driven by Brownian motion, great advances have been made in recent years in numerical methods for solving them. For a comprehensive survey of these techniques as well as an extensive bibliography on the subject, we refer the reader to Kloeden and Platen (1992) By contrast, numerical methods for SDEs driven by processes with jumps, such as stable processes, have received far less attention until recently and the literature on this subject is more limited: a systematic treatment in book form can be found in Janicki and Weron (1993).

For simplicity, we present some simulations for the inflation process R_t only since of the four components, this is closest to the time-series model of Wilkie Figure 1 shows three trajectories of the process R_t, in the case where the noise Z_t is Brownian motion. The scaling used is such that the time interval [0, 1] corresponds to a period of 50 years Specifically, in the context of Remark (ii) in Section 1, we have used $h = 0.02$ and in equation (1.2) our choice of ϕ is $\phi(t) = b + c\cos(2\pi t/h)$. Since the picture is only intended to give a qualitative indication of how the process behaves, the actual numerical values on the vertical axes are not of any great importance: the parameter values in Wilkie (1986) are used as a rough guide to the sort of values which might be appropriate for the parameters here – in particular, the parameter values of Wilkie are rescaled in the manner discussed in Remark (ii) of Section 1

Throughout, we have taken the various parameters in our models as given quantities and we have said nothing about the problems of their estimation. There is some discussion of this question in §6.4 and §13.2 of Kloeden and Platen (1992) which is especially relevant to the linear equations which appear repeatedly in our models.

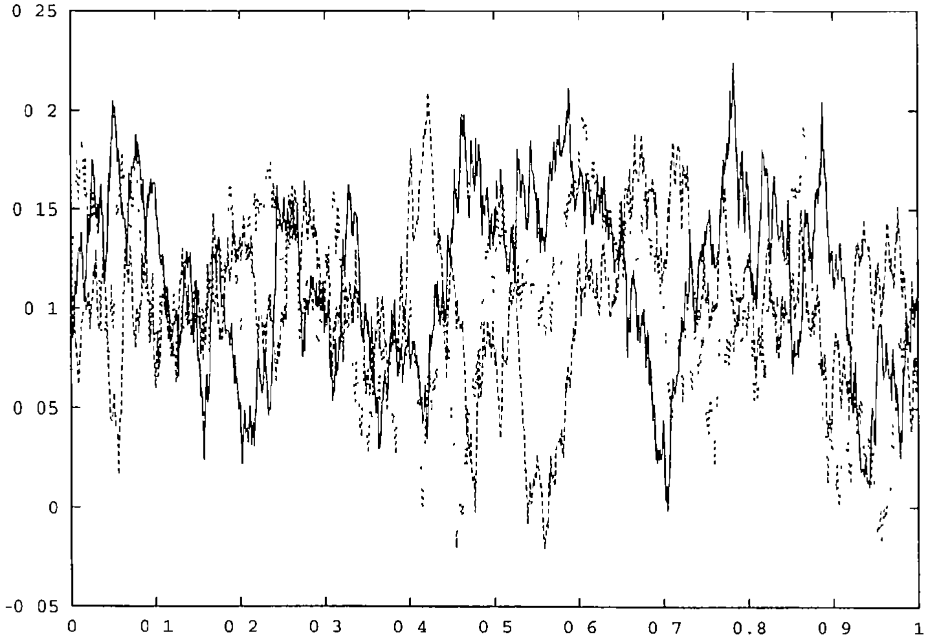

FIGURE 1 SAMPLE PATH REALIZATIONS OF THE FORCE OF INFLATION PROCESS R_t

APPENDIX PROOF OF LEMMA 3 1

Consider first the integral $I(s) = \int_0^s g(u)\,dZ_u$ Take a sequence of partitions $\left(u_k^{(n)}, u_{k+1}^{(n)}\right]$ of the interval [0, t], such that $\sup_k |u_{k+1}^{(n)} - u_k^{(n)}| \to 0$ *as* $n \to \infty$. It is known that, as $n \to \infty$,

$$I_n(s) = \sum_{u_k^{(n)} \le s} g(u_k^{(n)})\left(Z(u_{k+1}^{(n)}) - Z(u_k^{(n)})\right) \to \int_0^s g(u)\,dZ_u$$

in probability uniformly in s over the time interval [0, t] (see Protter (1990)). Therefore, there is a subsequence (n_i) such that $I_{n_i}(s) \to I(s)$ almost surely as $i \to \infty$ and without loss of generality we can assume that $I_n(s) \to I(s)$ almost surely. Next, take a different sequence of successively refining partitions of [0, t] and call this $\left[s_j^{(m)}, s_{j+1}^{(m)}\right)$. Put

$$F_m(u_k^{(n)}, t) = \sum_{s_j^{(m)} \ge u_k^{(n)}} f\left(s_j^{(m)}\right)\left(s_{j+1}^{(m)} - s_j^{(m)}\right)$$

(Of course, $F_m(u_k^{(n)}, t) \to F(t) - F(u_k^{(n)})$ as $n \to \infty$.) We then have

$$\lim_{m\to\infty}\lim_{n\to\infty} \sum_{s_j^{(m)}\le t} f\left(s_j^{(m)}\right)\left(s_{j+1}^{(m)} - s_j^{(m)}\right) \sum_{u_k^{(n)}\le s_j^{(m)}} g\left(u_k^{(n)}\right)\left(Z(u_{k+1}^{(n)}) - Z(u_k^{(n)})\right)$$
$$= \int_0^t f(s) \int_0^s g(u)\, dZ_u\, ds$$

and so for fixed n,

$$\mathbf{E}\left[\exp\left\{-i\theta \sum_{s_j^{(m)}\le t} f\left(s_j^{(m)}\right)\left(s_{j+1}^{(m)} - s_j^{(m)}\right) \sum_{u_k^{(n)}\le s_j^{(m)}} g\left(u_k^{(n)}\right)\left(Z(u_{k+1}^{(n)}) - Z(u_k^{(n)})\right)\right\}\right]$$
$$= \mathbb{E}\left[\exp\left\{-i\theta \sum_{u_k^{(n)}\le t} g\left(u_k^{(n)}\right)\left(Z(u_{k+1}^{(n)}) - Z(u_k^{(n)})\right) \sum_{s_j^{(m)}\ge u_k^{(n)}} f\left(s_j^{(m)}\right)\left(s_{j+1}^{(m)} - s_j^{(m)}\right)\right\}\right]$$
$$= \prod_{u_k^{(n)}\le t} \exp\left\{-\left(u_{k+1}^{(n)} - u_k^{(n)}\right)\psi\left[\theta g(u_k^{(n)}) F_m(u_k^{(n)}, t)\right]\right\}$$
$$= \exp\left\{-\sum_{u_k^{(n)}\le t}\left(u_{k+1}^{(n)} - u_k^{(n)}\right)\psi\left[\theta g(u_k^{(n)}) F_m(u_k^{(n)}, t)\right]\right\}$$
$$\to \exp\left\{-\sum_{u_k^{(n)}\le t}\left(u_{k+1}^{(n)} - u_k^{(n)}\right)\psi\left[\theta g(u_k^{(n)})\left(F(t) - F(u_k^{(n)})\right)\right]\right\} \qquad (A1)$$

as $m \to \infty$. In the above calculation, we have used the stationary independent increments property of Z and the fact that $\mathbf{E}\left[e^{-i\theta(Z_t - Z_s)}\right] = e^{-(t-s)\psi(\theta)}$. Letting $n \to \infty$ in (A1) then gives the right-hand side of (3.1).

For the integral $\int_0^t g(u) \int_u^t f(s)\, ds\, dZ_u$, we know that

$$\sum_k g(u_k^{(n)})\left(F(t) - F(u_k^{(n)})\right)\left(Z(u_{k+1}^{(n)} - Z(u_k^{(n)})\right) \to \int_0^t g(u) \int_u^t f(s)\, ds\, dZ_u$$

almost surely as $n \to \infty$ (passing to a subsequence if necessary) A similar calculation as in (A1) easily yields the identity (3.1).

ACKNOWLEDGEMENT

I am grateful to David Wilkie for his many detailed and helpful comments on an earlier draft of this paper.

REFERENCES

HURLIMANN, W (1992), "Numerical evaluation of the Wilkie inflation model", *Insurance Mathematics & Economics* **11**, 311-314

JANICKI, A and WERON, A (1993), *Simulation and Chaotic Behaviour of α-Stable Stochastic Processes*, Marcel Dekker

KLOEDEN, P E and PLATEN, E (1992), *Numerical Solution of Stochastic Differential Equations*, Springer

PROTTER, P (1990), *Stochastic Integration and Differential Equations*, Springer

ROGERS, L C G and WILLIAMS, D (1987), *Diffusions, Markov Processes and Martingales, Vol 2 Itô Calculus*, Wiley

WILKIE, A D (1986), "A stochastic investment model for actuarial use", *Trans Faculty of Actuaries* **39**, 341-373

WILKIE, A D (1995), "More on a stochastic investment model for actuarial use", *British Actuarial Journal* **1**, 777-964

TERENCE CHAN
Dept of Actuarial Mathematics
and Statistics
Heriot-Watt University
Edinburgh EH 14 4AS U.K

THE COX REGRESSION MODEL FOR CLAIMS DATA IN NON-LIFE INSURANCE

BY

NIELS KEIDING

Department of Biostatistics, University of Copenhagen

AND

CHRISTIAN ANDERSEN and PETER FLEDELIUS

ATP PensionService A/S, Hilleroed, Denmark

ABSTRACT

The Cox regression model is a standard tool in survival analysis for studying the dependence of a hazard rate on covariates (parametrically) and time (nonparametrically). This paper is a case study intended to indicate possible applications to non-life insurance, particularly occurrence of claims and rating

We studied individuals from one Danish county holding policies in auto, property and household insurance simultaneously at some point during the four year period 1988-1991 in one company The hazard of occurrence of claims of each type was studied as function of calendar time, time since the last claim of each type, age of policy holder, urbanization and detailed type of insurance Particular emphasis was given to the technical advantages and disadvantages (particularly the complicated censoring patterns) of considering the nonparametrically underlying time as either calendar time or time since last claim. In the former case the theory is settled, but the results are somewhat complicated The latter choice leads to several issues still under active methodological development. We develop a goodness-of-fit criterion which shows the lack of fit of some models, for which the practical conclusions might otherwise have been useful.

1. INTRODUCTION

Individual rating in non-life insurance may be based on exogenous variables (age of policy holder, urbanization) but in auto insurance various schemes for dynamical individual rating based on endogenous information (previous claim career) are well established. A possible further development of such procedures would be to base rating on endogenous variables for more than

ASTIN BULLETIN, Vol 28, No 1, 1998, pp 95-118

one type of non-life insurance. This would – as all such schemes – require an extensive knowledge base, and to focus ideas we studied the example of household, property and auto insurance. The joint development in time of the occurrences of claims of these three types is conveniently phrased in terms of the theory of event history analysis which has developed rapidly during the last decade, cf. Blossfeld et al (1989) and Blossfeld and Rohwer (1995) for good surveys with social science applications and Andersen et al (1993) for a general treatise with many practical examples, primarily from biostatistics.

In this report we indicate some initial possibilities as well as difficulties in carrying out such a programme Restricting attention to claim occurrence (i.e disregarding claim size) we want to capture the occurrence in time of claims as function of fixed exogenous covariates (age of policy holder, urbanization) and several time variables: calendar time and times since recent claims of each type. There is an active current literature on choice of time scales in statistical models for repeated events, cf. Lawless and Thiagarajah (1996), Lawless (1998) and Oakes (1998).

Our main tool will be versions of the Cox (1972a) regression model for event history data, see Andersen et al. (1993, Chapter VII). In this "semiparametric" model, one time variable is selected as "underlying" and modelled "nonparametrically" while other time variables as well as fixed exogenous covariates are modelled parametrically See Prentice et al. (1981) for an early exposition of alternative time scales in Cox models for repeated events and Oakes (1998) for an excellent concise survey. The Cox model is introduced in Section 3 and two alternative choices of underlying time variable are considered in Section 4 (calendar time) and 5 (time since last claim). Whereas calendar time as underlying time variable leads to a relatively standard application of Cox regression methodology, it will turn out to be rather less standard to consider time since last claim. A brief discussion is contained in Section 6.

The methodology is illustrated on data from a Danish insurance company, introduced in Section 2.

2. DATA

The present case study is based on data from a Danish insurance company. Between 1 January 1988 and 31 December 1991, 15,718 persons across Denmark at least once simultaneously held household, property and auto policies in this company. We study the 1,904 persons from the county of Fyn, in which Odense is by far the largest city. For each person and each type of policy is known

- the start and the end of the policy if within 1988-1991. If there were several policies of the same type within 1988-1991, only the latest was kept in the routine records on which we work.
- age (but not sex) of policy holder
- urbanization

- for household: coverage (amount)
- for auto: coverage
- date and size of claims.

In this study we focused attention on claims that led to payments This means that we removed claims of size 0. We made no other use of claim size.

3. The Cox regression model for event history analysis

For each type $h = 1, 2, 3$ (household, property, auto) and policy holder i the intensity of having a claim at time t is denoted $\lambda_{hi}(t)$. Here t can be calendar time (cf. Section 4) or time since the last claim of a similar type (cf. Section 5), with a special definition necessary if there has not (yet) been such a claim. A third possibility would be that t was time since taking out the policy. We explain later why we do not consider the latter possibility relevant here.

The Cox regression model now postulates that

$$\lambda_{hi}(t) = \alpha_{0h}(t) \exp\left[\beta_h' Z_{hi}(t)\right] Y_{hi}(t)$$

where $\alpha_{0h}(t)$ is a freely varying so-called underlying intensity function common to all policy holders i but specific to insurance type h The indicator $Y_{hi}(t)$ is 1 if policy holder i is at risk to make a claim of type h at time t, 0 otherwise The covariate process $Z_{hi}(t)$ indicates fixed exogenous as well as time-dependent endogenous covariates. The fixed covariates considered are year of birth of policy holder and urbanization of residence, which in practice equals 1 for city (Odense) and 0 for rural (rest of Fyn). The time-dependent covariates indicate duration since last claim of each type (which can and will be parameterized in various ways). Finally the vector β_h contains the regression coefficients on the covariates $Z_{hi}(t)$

Statistical inference in the Cox regression model is primarily based on maximum partial likelihood, which in the generality necessary for this application was surveyed by Andersen et al. (1993, Chapter VII) in the framework of *counting processes*. The regression coefficients β_h are estimated by maximizing the partial likelihood

$$L(\beta_h) = \prod_j \frac{\exp(\beta_h' Z_{hi(j)}(T_{hj}))}{\sum_{i\; Y_{hi}(T_{hj})=1} \exp(\beta_h' Z_{hi}(T_{hj}))}$$

where $T_{h1} < T_{h2} < \ldots$ are the times of claims of type h, policyholder $i(j)$ claiming at time T_{hj}. Large sample results are available to justify the application of the inverse Hessian of the log partial likelihood as approximate covariance matrix for $\hat{\beta}_h$. Because of the time-varying covariates the necessary algorithms are rather elaborate, although we were able to perform all computations on a medium-sized PC using StatUnit (Tjur, 1993). The computations may also be performed in standard packages such as BMDP, SAS or S-plus, or via the Poisson regression approach of Lindsey (1995).

For the *underlying intensity* $\alpha_{0h}(t)$ it is well-established that a natural estimator of the integrated intensity

$$A_{0h}(t) = \int_0^t \alpha_{0h}(u)du$$

is given by the step function (the "Breslow" estimator)

$$\hat{A}_{0h}(t) = \sum_{T_{hj} \leq t} \frac{1}{\sum_{i\ Y_{hi}(T_{hj})=1} \exp(\hat{\beta}_h' Z_{hi}(T_{hj}))}$$

where $T_{h1} < T_{h2} < \ldots$ are the times of claims of type h and $\hat{\beta}_h$ the maximum partial likelihood estimator of β_h.

Unfortunately $\hat{A}_{0h}(t)$ is less than optimal in communicating important features of the structure of $\alpha_{0h}(t)$; it is often desirable to be able to plot an estimate of α_{0h} itself We shall here use *kernel smoothing* (which in the context of estimating the intensity in the multiplicative intensity model for counting processes was incidentally pioneered by the actuary Ramlau-Hansen (1983)). This estimates $\alpha_{0h}(t)$ by

$$\hat{\alpha}_{0h}(t) = \sum_{j\ t-b<T_{hj}<t+b} K\left(\frac{t - T_{hj}}{b}\right) \Delta \hat{A}_{0h}(T_{hj})$$

where b is the *bandwidth*, K a *kernel function*, here restricted to $[-1, 1]$ and $\Delta \hat{A}_{0h}(T_{hj}) = \hat{A}_{0h}(T_{hj}) - \hat{A}_{0h}(T_{h,j-1})$, $T_{h0} = 0$. We choose here the Epanechnikov kernel $K(x) = 0.75(1 - x^2)$. For more documentation, see again Andersen et al. (1993, pp. 483 and 507-509).

Despite its considerable flexibility, the Cox regression model is not assumption-free, the most important assumptions being that of *proportional hazards* and that of *log-linearity* of effect of regressors. There is a well-developed battery of goodness-of-fit procedures available, cf. Andersen et al. (1993, Section VII.3), and several of these methods have been used in the present case-study (never indicating deviation from model assumptions). However, space prevents us from documenting these here.

4 Cox regression of claim intensity
Calendar time as underlying time variable

Our first choice of underlying time scale is *calendar time*, which is always observable and whose association with variations in claim intensity may form an interesting object of study. Technically, the counting process approach elaborated by Andersen et al. (1993, Section III.4) easily allows for entry and exit of policies from observation (the "Aalen filter") in this situation.

However, an important purpose of this study was to ascertain the observability and possible extent of the association of claim intensity to the duration(s) since earlier claim(s), and it is less obvious how to account for these. Because of the relatively limited period of observation (4 years) it was necessary to make several pragmatic choices. First, the dependence on earlier claims was operationalized as dependence on duration since latest claim, and this was achieved by defining the indicator covariates

[1-90]: There has been a claim less than 90 days ago.
[91-180]: The latest claim was between 91 and 180 days ago.
[181-270]: The latest claim was between 181 and 270 days ago.
[271-360] The latest claim was between 271 and 360 days ago.
[> 360] There has been no claim during the past 360 days.

Since the database contains no information on claims before 1988, these covariates would not all be observable early in the period. We therefore decided to use 1988 as run-in year, only for collecting information on earlier claims.

A further problem was the many instances where a new policy was taken out within 1988-1991 In case no claims happened, the above covariates would remain unobservable for 360 days, which forced us to add the covariate

[no inf.]. policy (of this type) was taken out less than 360 days ago and during that time there were no claims.

4.1. Household claims in calendar time

For household claims the relevant covariates were: year of birth of policyholder (categorized in three groups separated by 1 January 1938 and 1 January 1948), urbanization (Odense vs rest of Fyn) and duration since last claim of each type as described above All groups of covariates were of statistical significance and the estimated model had regression coefficients as given in Table 4.1.

It is seen that compared to the "no information" situation when no claim has happened after a recently taken out policy, knowledge of a recent *household* claim during the recent 0-9 months increases the risk of a new household claim by a factor ranging from $e^{0.562} = 1.8$ to $e^{0.808} = 2,2$, i.e , a factor of about 2. On the other hand knowledge of claim-free career of one year decreases the risk by the (statistically insignificant) factor of 0.9.

Past *property* claims have effects according to a similar pattern, although the effects are smaller, except for very recent property claims ($e^{0.629} = 1.9$), some of which may be caused by the same events that caused the household claim. Unfortunately the database cannot identify such cases, which would in principle violate the proportional hazards assumption of the Cox regression model

TABLE 4 1

REGRESSION COEFFICIENTS IN REDUCED COX MODEL FOR HOUSEHOLD CLAIMS

Covariate	*Estimate*	*Standard error*	*P*
Household[no inf]	0	–	–
Household[1-90]	0 562	0 277	0 043
Household[91-180]	0 725	0 275	0 008
Household[181-270]	0 808	0 275	0 003
Household[271-360]	0 206	0 303	0 496
Household[> 360]	−0 105	0 243	0 665
Property[no inf]	0	-	-
Property[1-90]	0 629	0 197	0 001
Property[91-180]	0 178	0 219	0 416
Property[181-270]	0 107	0 225	0 663
Property[271-360]	0 287	0 225	0 202
Property[> 360]	−0 132	0 161	0 413
Auto[no inf]	0	–	–
Auto[1-90]	0 224	0 209	0 284
Auto[91-180]	0 301	0 208	0 148
Auto[181-270]	0 258	0 217	0 234
Auto[271-360]	−0 187	0 260	0 473
Auto[> 360]	−0 144	0 148	0 330
Born[> 1947]	0	–	–
Born[1938-1947]	0 015	0 086	0 860
Born[< 1938]	−0 406	0 100	0 000
Rural	0	–	–
City	0 381	0 076	0 000

Past *auto* claims show overall significance, although the effect of each period is small, generally in a similar pattern as for the other types of insurance.

The *age* pattern has a decreased intensity for older policy-holders (intensity factor $e^{-0\ 406} = 0.7$) while the two younger groups are very similar; finally urbanization generates the expected gradient with an increased risk in the city ($e^{0\ 381} = 1.5$).

The underlying intensity is estimated as described in Section 3, using 3 different bandwidths for illustration, see Fig. 4.1 It is not easy to conclude much from the somewhat irregular pattern except perhaps a slight general decrease. The boundary effects at the start and the end of the studied period are statistical artefacts deriving from the kernel estimation approach.

It may be noticed from Table 4.1 and the following tables that several of the patterns of dependence on time since last claim might be simplified. As an example in Table 4.1, the regression coefficients Auto[1-90], Auto[91-180] and Auto[181-270] look rather similar, as do Auto[271-360] and Auto[> 360]. However, there is no obviously consistent pattern across types of claims and types of risk indicators, so we have refrained from conducting what would in any case be post-hoc attempts at statistical identification of such patterns.

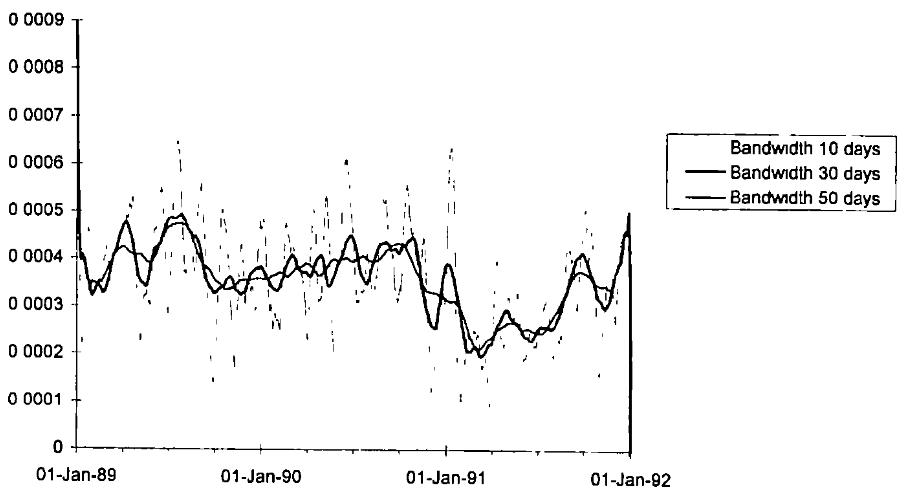

FIGURE 4 1 Kernel smoothed underlying intensities for household claims

4.2. Property claims in calendar time

For property insurance there is a series of optional additional coverage possibilities, which are all included as specific indicator covariates fire, glass, insects, wash basins, pipe, rot

The estimates of the reduced model are given in Table 4.2. Note that urbanization is statistically insignificant and that there is an unusual age pattern, the middle-aged having a somewhat lower risk than the young and the old. In the interpretation of the age effect it is however particularly important to keep in mind the specially selected population each person must have had all three types of policies simultaneously at some point during 1988-1991, this restricts consideration to better situated people.

Of the optional additional coverage, only glass and pipe coverage are retained as risk variables, both clearly increasing the risk. That fire does not appear is related to the fact that almost all policies chose that option. For duration since last claim the general pattern is similar to the earlier one, although one must notice that there is never a significantly lower risk than that of [no inf.], which (as we shall discuss more fully below) will limit the practical applicability of the results.

TABLE 4.2

REGRESSION COEFFICIENTS IN REDUCED COX MODEL FOR PROPERTY CLAIMS

Covariate	*Estimate*	*Standard error*	*P*
Household[no inf.]	0	-	-
Household[1-90]	0.485	0.229	0.034
Household[91-180]	0.302	0.240	0.208
Household[181-270]	0.345	0.240	0.151
Household[271-360]	0.032	0.260	0.902
Household[> 360]	−0.080	0.192	0.676
Property[no inf.]	0	-	
Property[1-90]	0.524	0.206	0.011
Property[91-180]	0.334	0.217	0.124
Property[181-270]	0.206	0.224	0.357
Property[271-360]	0.281	0.224	0.210
Property[> 360]	−0.180	0.181	0.320
Auto[no inf.]	0	-	-
Auto[1-90]	0.501	0.184	0.006
Auto[91-180]	0.262	0.202	0.194
Auto[181-270]	0.182	0.210	0.387
Auto[271-360]	0.267	0.211	0.205
Auto[> 360]	0.026	0.141	0.851
Born[> 1947]	0		-
Born[1938-1947]	−0.196	0.079	0.013
Born[< 1938]	−0.061	0.079	0.438
Glass	0.411	0.140	0.003
Pipe	0.185	0.072	0.010

The underlying intensity is estimated in Fig. 4.2 and shows a dramatic peak in early 1990, apparently traceable to extreme weather conditions.

4.3. Auto claims in calendar time

In addition to the standard covariates, auto claims are expected to depend on whether or not there is auto comprehensive coverage and whether or not a certain "free claim" allowance is included in the policy.

The estimates of the reduced model are given in Table 4.3, where it is immediately noticed that, perhaps contrary to expectation, auto comprehensive coverage does not increase risk of claim for this population of insures. Note also the age pattern, generally unusual for auto insurance with

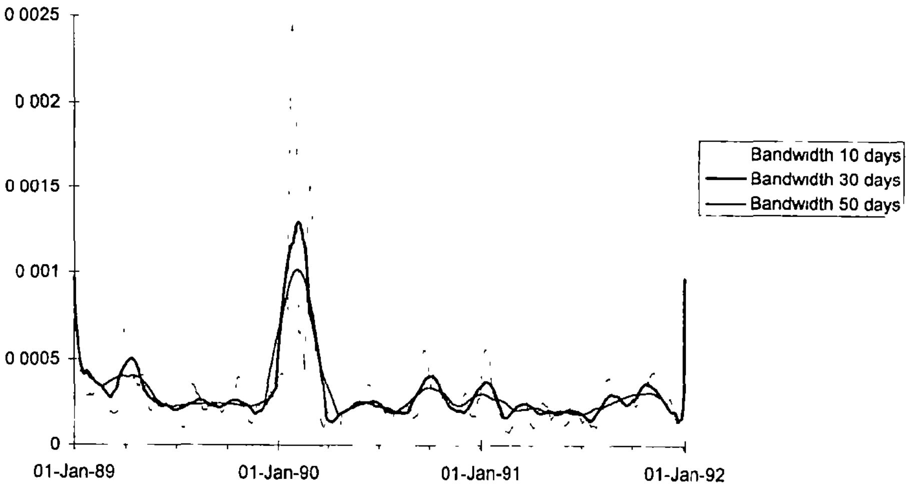

FIGURE 4 2 Kernel smoothed underlying intensity for property claims

maximal risk among the middle-aged policy-holders. (Note that there are no data to account for size of household, and note once again the specially selected population.)

TABLE 4 3

REGRESSION COEFFICIENTS IN REDUCED COX MODEL FOR AUTO CLAIMS

Covariate	*Estimate*	*Standard error*	*P*
Household[no inf]	0	-	-
Household[1-90]	0 388	0 245	0 114
Household[91-180]	0 303	0 251	0 226
Household[181-270]	0 304	0 252	0 227
Household[271-360]	0 493	0 244	0 043
Household[> 360]	0 001	0 193	0 995
Auto[no inf]	0	-	-
Auto[1-90]	0 730	0 259	0 005
Auto[91-180]	0 862	0 257	0 001
Auto[181-270]	0 738	0 264	0 005
Auto[271-360]	0 618	0 273	0 024
Auto[> 360]	0 294	0 231	0 203
Born[> 1947]	0	-	-
Born[1938-1947]	0 100	0 079	0 209
Born[< 1938]	−0 140	0 087	0 106
Free claim	1 048	0 083	0 000

The patterns regarding duration since last claim show no overall effect of recent property claims and some effect (increase) on risk of recent household claim. As expected, recent auto claims considerably increase the risk of a further auto claim, as does the "free claim" option (no penalty in premium scale after a claim)

The underlying intensity (Fig. 4.3) indicates some seasonality with peaks in the winter and the summer, however this pattern is rather irregular.

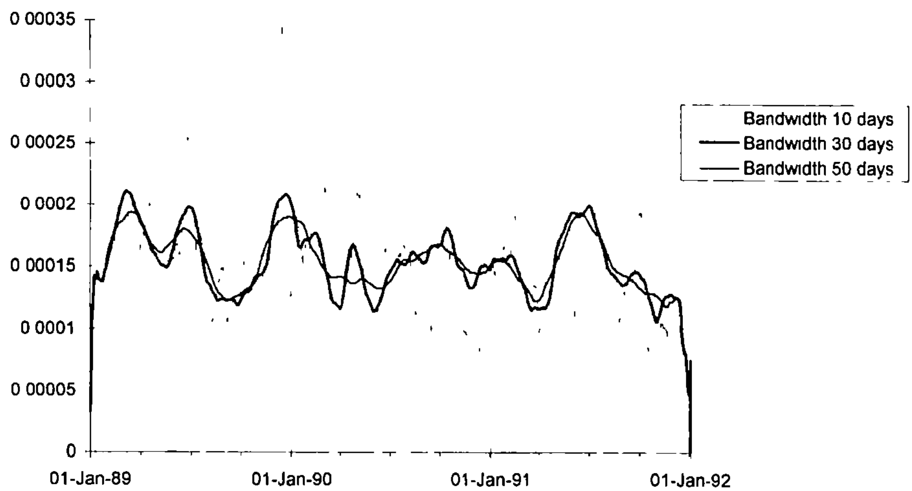

FIGURE 4 3 Kernel smoothed underlying intensity for auto claims

4.4. Preliminary conclusions: calendar time as underlying variable

Two problems are common to all analyses so far. First, the unstable nature of the population of policies during the relatively short observation window of four years make the desired allowance for time since earlier claims difficult to achieve in practice. The general reference category of [no inf.], meaning that a policy of the relevant type was taken out less than a year ago and there have not yet been claims to that policy, in all cases carries a very low risk for new claims of the type under study This relative low-risk behaviour of new policyholders is obviously difficult to integrate into a reward system for faithful customers In this connection it must be emphasized that the routine nature of our database (which may well be typical of such databases) did not allow the distinction between genuinely new policies and "bureaucratical" renewals initiated by the company or the policyholder in order to update conditions.

Secondly, some of our concrete results point to the rather special selection procedure underlying the present database: all policyholders were required to have held all three types simultaneously at least once in 1988-1991 As an example, think of the rather biased selection of young policyholders!

5. COX REGRESSION OF CLAIM INTENSITY: USING DURATION RATHER THAN CALENDAR TIME AS BASIC TIME VARIABLE

In the discussion so far it has become obvious that we need to reason in several time variables: calendar time as well as duration(s) since recent claim(s). At least because of the possibility that there have not yet been any claims, we may also need the time since the policy was taken out. When using the Cox regression model such as introduced in Section 3

$$\lambda_{hi}(t) = \alpha_{0h}(t) \exp\left[\beta_h' Z_{hi}(t)\right] Y_{hi}(t)$$

one may choose one of these time scales as "basic" ($= t$) and handle the other(s) as (time-dependent) covariates $Z_{hi}(t)$. An important criterion for choosing between these possibilities is the additional flexibility in the description offered by the "nonparametric" underlying intensity $\alpha_{0h}(t)$. We actually saw in Section 4 that various indications regarding seasonal patterns appeared in the graphs of Figs. 4.1-3

Another criterion is ease of handling special observation plans. When calendar time is used, the exact time is always known for each policy-holder, in contrast to what is the case for duration since last claim. We discussed the latter problem at the beginning of Section 4, where we constructed time-dependent covariates to account for durations since earlier claims.

However, both prior expectation and our experiences so far point to the importance of time since last claim as decisive time variable, for which the maximal modelling flexibility offered by the nonparametric part of the Cox model would be useful. To discuss an adequate statistical analysis in this time-scale, consider first the simple situation without covariates, which is a renewal process.

5.1. Estimation of renewal processes observed in a fixed time window

Let X_1, X_2, be independent random variables (durations) with distribution functions F_1, $F_2 = F_3 = . = F$, assumed to have finite expectations μ_1 and μ and density functions $f_1 = F_1'$ and $f = F'$. Let $S_n = X_1 + ... + X_n, n = 1, 2,$ .. and the stochastic process (a *renewal process*)

$$N_t = \sum_{n=1}^{\infty} I\{S_n \leq t\},$$

the number of durations since time 0. If $f_1 = (1 - F)/\mu$ the process is *stationary* Observing a renewal process in an interval $[t_1, t_2]$ amounts to observing the renewal times (claims) $T_j \in [t_1, t_2]$ or equivalently $(N_t - N_{t_1}\ \ t \in [t_1, t_2])$. Let T_j be the first renewal after t_1, i.e. $N_{T_j} = N_{t_1} + 1$. Then $T_j - t_1$ is called the *forward recurrence time*, and if the process is stationary, this has *density* function $(1 - F)/\mu$

Observing a renewal process in an observation window $[t_1, t_2]$ involves four different elementary observations

1. Times x_i from one renewal to the next, contributing the density $f(x_i)$ to the likelihood.
2. Times from one renewal to t_2, right-censored observations of F, contributing factors of the form $1 - F(t_2 - T_j)$ to the likelihood
3. Times from t_1 to the first renewal (forward recurrence times), contributing, in the stationary case, factors of the form $(1 - F(T_j - t_1))/\mu$ to the likelihood.
4. Knowledge that no renewal happened in $[t_1, t_2]$, being right-censored observations of the forward recurrence time, contributing in the stationary case a factor

$$\int_{t_2 - t_1}^{\infty} (1 - F(u))du/\mu.$$

In the stationary case the resulting maximum likelihood estimation problem is well understood. Vardi (1982) derived an algorithm (a special case of the EM-algorithm) in a discrete-time version of the problem, and Soon and Woodroofe (1996) gave an elaborate and very well-written discussion in continuous time. McClean and Devine (1995) conditioned on seeing at least one renewal in $[t_1, t_2]$, excluding observations of type 4 and restricting attention to observations of type 3 right-truncated at $t_2 - t_1$, i.e. with density

$$(1 - F(u - t_1))/\left(1 - \int_0^{t_2 - t_1} F(v)\, dv\right)$$

Again an EM-type algorithm is feasible.

In our situation we need to be able to generalize the estimation method from iid variables to the Cox regression model, and we would also prefer to avoid the stationarity condition required for inclusion of the (uncensored and censored) forward recurrence times of type 3 and 4.

This is possible by restricting attention to (uncensored and censored) times since a renewal, that is, observations of type 1 and 2. As discussed repeatedly by Gill (1980, 1983), see also Aalen and Husebye (1991) and Andersen et al. (1993, Example X.1 8), the likelihood based on observations of type 1 and 2 is identical to one based on independent uncensored and censored life times from the renewal distribution F. Therefore the standard estimators (Kaplan-Meier, Nelson-Aalen) from survival analysis are applicable, and their usual large sample properties may be shown (albeit with new proofs) to hold.

The above analysis is sensitive to departures from the assumption of homogeneity between the iid replications of the renewal process. Restricting attention to time since first renewal will be biased (in the direction of short renewal times) if there is unaccounted heterogeneity, as will the re-use of second, third, ... renewals within the time window. As always, incorporation of observed covariates may reduce the unaccounted heterogeneity, but the question is whether this will suffice

5.2. Cox regression of duration since last claim

The Cox (1972a) proportional hazards regression model for survival analysis was implemented by Cox (1972b) in the so-called *modulated renewal processes*, for which the hazard of the renewal distribution is assumed to have a similar semiparametric decomposition. This model has received much less attention than the survival analysis model and its event history analysis generalization (Prentice et al., 1981, Andersen and Gill, 1982, Andersen et al , 1993, Chapter VII), although Kalbfleisch and Prentice (1980) and Oakes and Cui (1994) discussed estimation. Careful mathematical-statistical analysis was provided by Dabrowska et al. (1994) and Dabrowska (1995), who showed that if the covariates depend on no other time variables than the backward recurrence times, then the 'usual' asymptotic results of the Cox partial (or profile) likelihood may be proved.

In the present case we have the additional complication of observing through a fixed (calendar) time window. Inclusion of likelihood factors of types 3 and 4 is then intractable, but if the model were true (in particular, if the observed covariates sufficiently account for individual heterogeneity), valid inference may be drawn from the reduced likelihood based on time since first claim (factors of types 1 and 2)

Finally, we want to incorporate time-dependent covariates not depending on the backward recurrence time only (for example, in the analysis of household claims we want to incorporate times since the last property or auto claim) and the analysis is then no longer covered by Dabrowska's asymptotic results.

As pointed out at the end of the last section, if there is unaccounted heterogeneity the expected bias by restricting attention to time since first renewal will be in the direction of short renewal times, and this will be even worse if times since second, third etc renewal times are also included. We build a goodness-of-fit criterion on this intuition, as follows.

5.3. A goodness-of-fit criterion for the Cox modulated renewal process observed through a fixed time window

We assume that the occurrence of claims of type h for policy holder i at duration t since last claim of that type is governed by a Cox regression model with intensity

$$\lambda_{hi}(t) = \alpha_{0h}(t)\exp\left[\beta_h' Z_{hi}(t)\right] Y_{hi}(t)$$

with interpretation as before. For this model Dabrowska (1995) proved asymptotic results for the 'usual' profile likelihood based inference, under the crucial assumption that the covariates $Z_{hi}(t)$ depend on time only through (the backwards recurrence time) t. (Obviously a full model will require an additional specification of occurrence of the first claim of type h after the policy is taken out.)

The claim occurrences are viewed through a fixed time window, but under the model valid inference may be based on the likelihood composed of the product of contributions from the distribution of time from first to second claim, second to third claim, and so on, the last being right-censored The expected deviation from the model is that time from claim $j = 1$ is longer than times from claims $j = 2, 3,$. We therefore extend the model to the Cox regression model

$$\lambda_{hj}(t) = \alpha_{0hj}(t)\exp\left[\beta_{hj}' Z_{hi}(t)\right] Y_{hi}(t).$$

In practice the regression coefficients β_{hj} and the underlying intensities $\alpha_{0hj}(t)$ after claim j are assumed identical for $j = 2, 3, \ldots$. A good evaluation of the fit of the Cox model can be based on first assessing identity of regression coefficients ($\beta_{h1} = \beta_{h2}$) and then, refitting in a so-called stratified Cox regression model with identical β_{hj} but freely varying $\alpha_{h0j}(t)$ over j, comparing the underlying intensities ($\alpha_{0h1}(t) = \alpha_{0h2}(t)$) after first and after later claims. For the first hypothesis a standard log partial likelihood ratio test may be performed, for the second we use graphical checks as surveyed by Andersen et al. (1993, Section VII. 3) Further development of this goodness-of-fit approach might follow the lines of Andersen et al (1983)

5.4. Household claims by duration since last such claim

The relevant covariates are the same as listed in Section 4 1 except of course that duration since last household claim is now described in the non-parametric part of the Cox model rather than by time-dependent covariates. Table 5.1 shows the final model after elimination of non-significant covariates It is noted that the result is rather simpler than that represented by Table 4 1 since in addition to time since last household claim, also time since last auto claim and age have disappeared.

TABLE 5 1

REGRESSION COEFFICIENTS IN REDUCED COX MODEL FOR HOUSEHOLD CLAIMS

Covariate	*Estimate*	*Standard error*	*P*
Property[no inf]	0	–	–
Property[1-90]	0 659	0 199	0 001
Property[91-180]	0 118	0 243	0 623
Property[181-270]	0 281	0 238	0 237
Property[271-360]	0 211	0 266	0 428
Property[> 360]	–0 140	0 165	0 394
Rural	0	–	–
City	0 251	0 103	0 015

The remaining covariates, time since last property claim and urbanization, have similar effects (particularly for the former) as before, and similar remarks apply.

The underlying intensity is estimated in Fig. 5.1 for the first three years (thereafter the random variation dominates) A clear decrease is seen: the longer the duration since the last household claim, the lower the intensity of a new one.

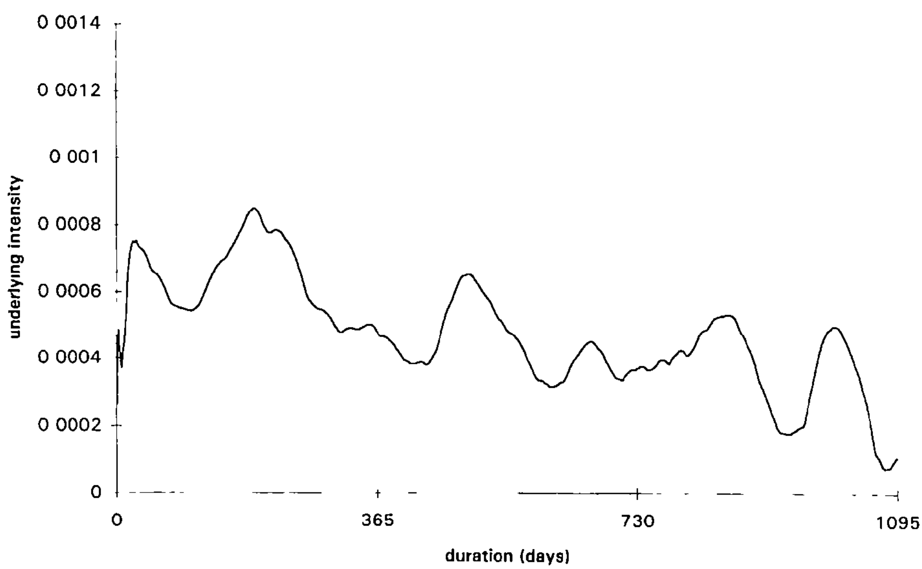

FIGURE 5 1 Kerned smoothed underlying intensity for household claims (bandwidth 50 days)

Fitting the stratified model specified in the previous section to the covariates of Table 5.1 leads to insignificantly different regression parameter estimates after first and after later claims ($\chi^2 = 8.87, f = 6$). To compare the estimates of underlying intensities $\alpha_{h0_j}(t)$ between times since first claim and times since later claims, Fig. 5.2 shows integrated intensity estimates against time, whereas Fig 5.3 shows integrated intensity estimates against one another Both plots indicate good agreements so that the model, and hence the above interpretation, would seem acceptable

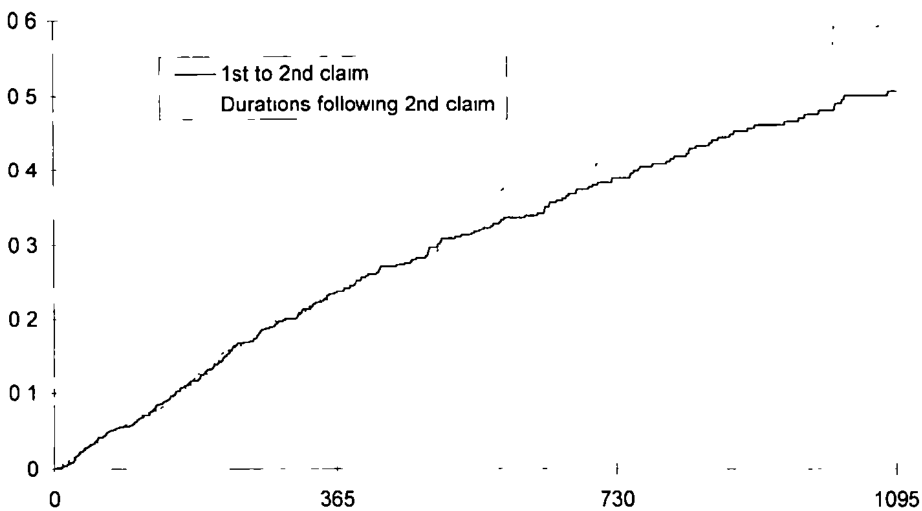

FIGURE 5 2 Estimated integrated underlying intensities for household claims

5.5. Property claims by duration since last such claim

In a similar fashion Table 5.2 shows the final model after elimination of non-significant covariates. (A likelihood ratio test for no effect of time since last household claim gave $P = .01$.)

TABLE 5 2

REGRESSION COEFFICIENTS IN REDUCED COX MODEL FOR PROPERTY CLAIMS

Covariate	*Estimate*	*Standard error*	*P*
Household[no inf]	0	-	-
Household[1-90]	0 198	0 208	0 340
Household[91-180]	0 321	0 213	0 131
Household[181-270]	0 110	0 236	0 634
Household[271-360]	−0 140	0 269	0 602
Household[> 360]	−0 253	0 157	0 106

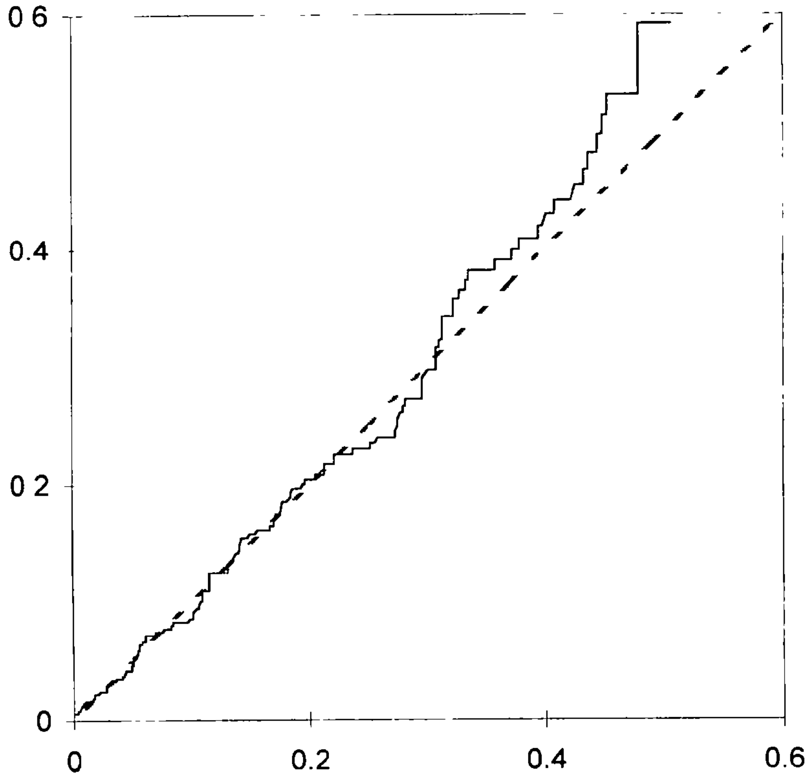

FIGURE 5 3 Estimated integrated underlying intensities for household claims based on durations following second claim plotted against those based on the (possibly right censored) duration from first to second claim

As for household claims, we get a much simpler description in the present time-scale, the only remaining covariate being time since last household claim The effect of this covariate in qualitatively similar to what it was in Table 4.2. The underlying intensity (Fig. 5 4) is decreasing. The gradient between best and worst customers (expressed by range of variation of regression coefficients) is smaller than for household claims, corresponding to common expectation.

For the goodness-of-fit test the identity of regression coefficients was again easily accepted ($\chi^2 = 0.73$, $f = 5$), but here the unfortunate bias in the direction of shorter durations after second and further claims is clearly visible from Figs. 5.5 and 5 6. The model must be judged as not fitting and the above conclusions cannot be sustained

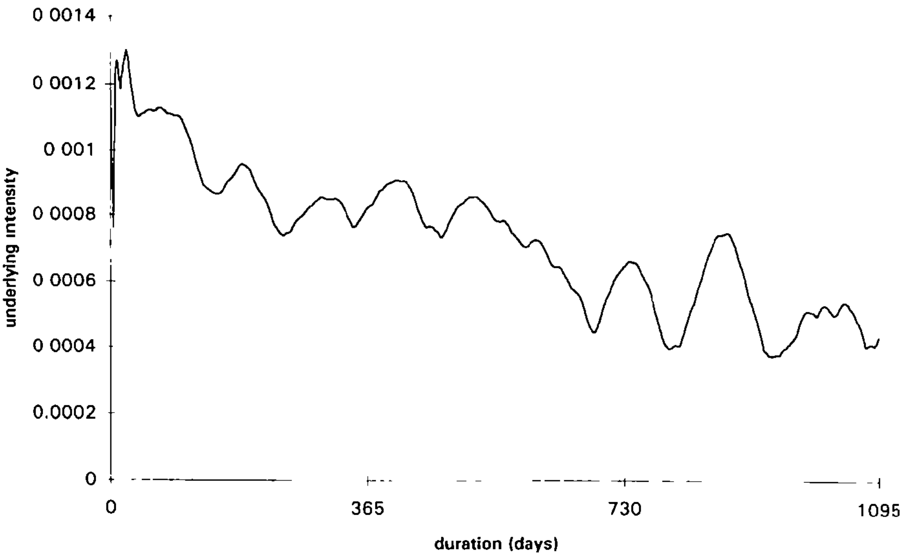

FIGURE 5.4 Kernel smoothed underlying intensity for property claims (bandwidth 50 days)

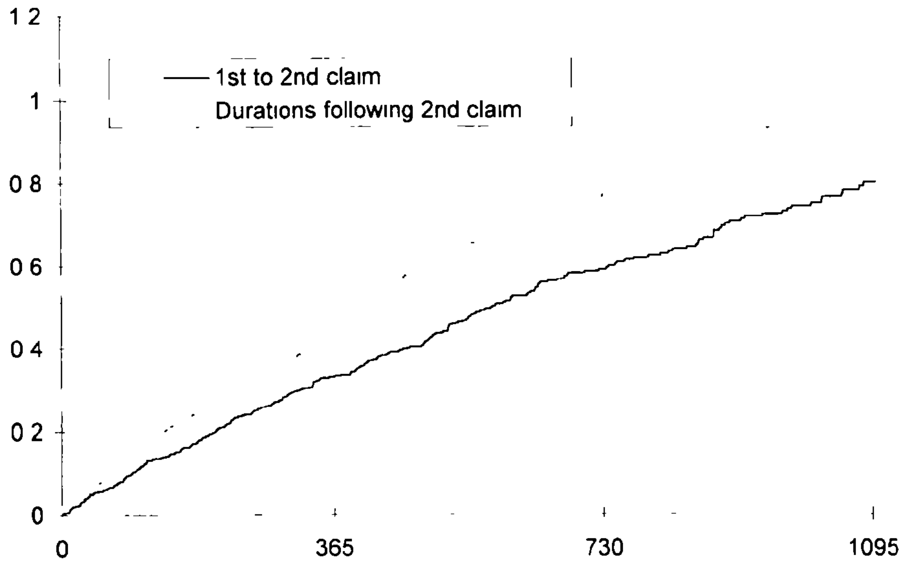

FIGURE 5.5 Estimated integrated underlying intensities for property claims

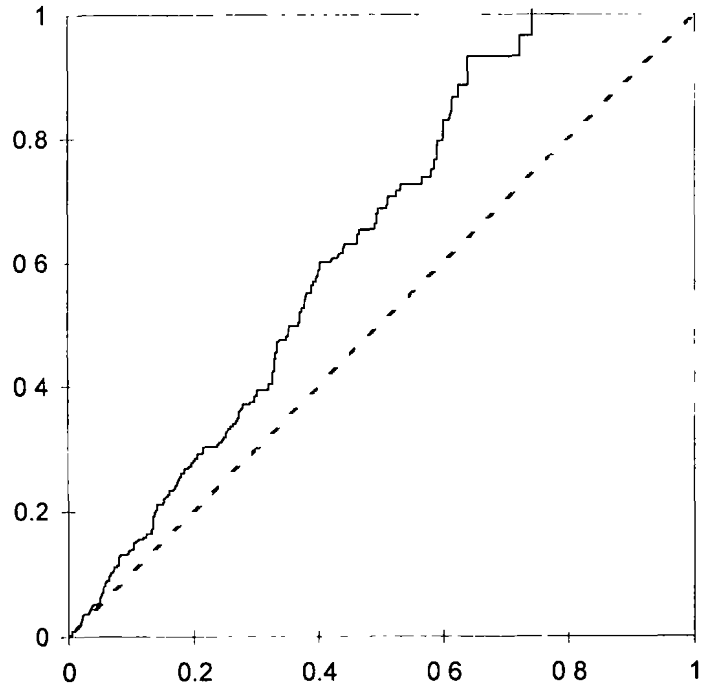

FIGURE 5 6 Estimated integrated underlying intensity for property claims based on durations following second claim plotted against those based on the (possibly right censored) duration from first to second claim

5.6. Auto claims by duration since last such claim

Finally, Table 5.3 documents the result of fitting the Cox regression model to time since last auto claim, using the covariates listed in Section 4, particularly Section 4.3, and eliminating statistically insignificant covariates

TABLE 5 3

Regression coefficients in reduced Cox model for auto claims

Covariate	*Estimate*	*Standard error*	*P*
Household[no inf]	0	–	–
Household[1-90]	0 304	0 205	0 139
Household[91-180]	0 295	0 218	0 175
Household[181-270]	0 053	0 240	0 826
Household[271-360]	0 032	0 251	0 897
Household[> 360]	−0 334	0 155	0 031
Auto comprehensive	−0 405	0 148	0 005
Free claim	0 320	0 121	0 008

Compared to Table 4.3, we necessarily have lost time since last auto claim, but furthermore, age is no longer significant while, most surprisingly, auto comprehensive coverage seems to *decrease* the risk of the next auto claim by a factor of $e^{-405} = 0.67$ We can only interpret the latter phenomenon with reference to a peculiar selection of policyholders who choose comprehensive coverage. The underlying intensity (Fig. 5 7) shows a clear decrease.

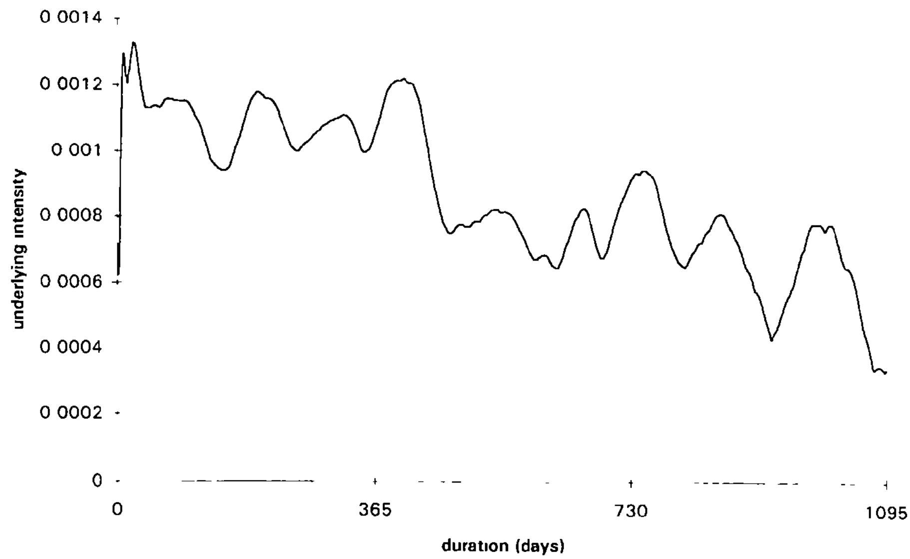

FIGURE 5 7 Kernel smoothed underlying intensity for auto claims (bandwidth 50 days)

The result of the goodness-of-fit test is very similar to that for household insurance above: regression coefficients are easily identical ($\chi^2 = 2\ 26$, $f = 7$), but the expected bias is immediately obvious from Figs. 5.8 and 5.9 The model must thus be considered poorly fitting, and the results cannot be sustained.

5.7. Preliminary conclusions: duration as underlying time variable

The two basic difficulties mentioned in Section 4 4 were not removed by changing to duration as basic time variable. Furthermore, technical problems of estimation (as well as reluctance to postulate stationarity) forced us to omit all durations already running at the start of observation

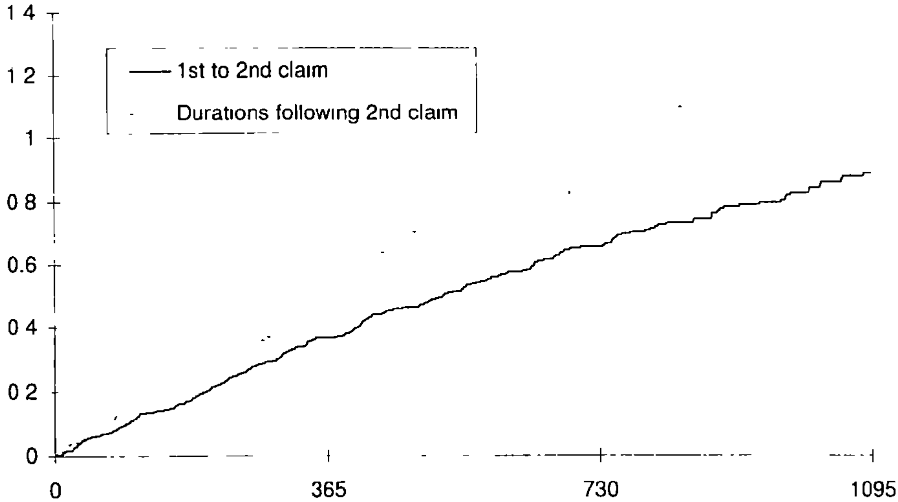

FIGURE 5 8 Estimated integrated underlying intensities for auto claims

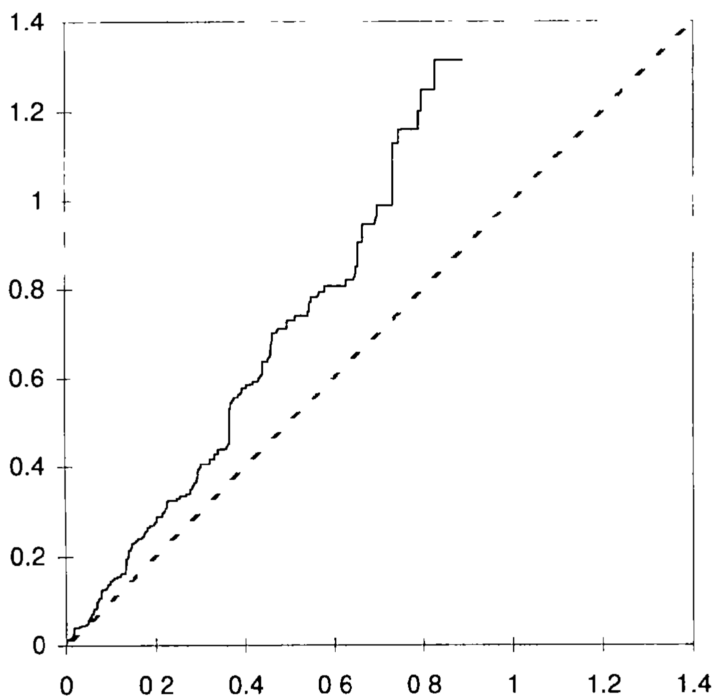

FIGURE 5 9 Estimated integrated underlying intensities for auto claims based on durations following second claim plotted against those based on the (possibly right censored) duration from first to second claim

1 January 1988 or when a new policy was taken out. Even based on these reduced data, we were able to construct a goodness-of-fit criterion that rejected the Cox regression model for property and auto claims, while household claims seemed to be amenable to analysis by this approach.

In any case the analysis performed in this section is in practice restricted to what happens during the first three years after a claim, and it is impossible to extrapolate from here to the situation before the first claim or long after a claim, both of which carry an important weight in practice.

6. Discussion and conclusion

The purpose of this report was to demonstrate some possibilities of recently developed tools in event history analysis in describing routinely collected data on non-life insurance claim histories, with the long-term aim of individualizing rating To simplify matters we ignored claim size but attempted to handle such presumably realistic difficulties as relatively short collection period (4 years), many bureaucratic renewals and the special selection pattern arising from the desire to simultaneously study household, property and auto insurance in the same company.

Our basic tool was an event history generalization of the proportional hazards model due to Cox (1972a) for survival data, see Andersen et al (1993, Chapter VII) for a detailed exposition

A central feature has been the choice of time origin. The primary choice was to use calendar time as underlying time in the Cox regression model, which necessitated a run-in period for assessing time since last claim but otherwise allowed detailed identification of effects of fixed (exogenous) and time-varying (endogenous) covariates, in most but not all cases yielding results in good accordance with expectation.

A more experimental choice was to use time since last claim as underlying time in the Cox regression model, tying to Cox's (1972b) modulated renewal process The mathematical-statistical theory of this model is rather less settled (Dabrowska, 1995). We develop in Section 5 a necessary (but by no means sufficient) goodness-of-fit criterion which, for property and auto claims, is violated even for our restricted data after first claim. Although the use of time since last claim as underlying time variable does have advantages, particularly in leading to much simpler regression models, it will so far have to be considered to be under development. The goodness-of-fit investigation indicated residual unaccounted heterogeneity, for which some kind of frailty modelling (Oakes 1992, 1998, Hougaard 1995, Scheike et al. 1997) might be fruitful.

Several of the difficulties and shortcomings listed in Sections 4.4 and 5.7 refer to the routine nature of the database that we used (and which we believe to be typical). Further attempts at employing such techniques in this context should perhaps make an effort to obtain better tuned databases, to further calibrate and explain the tools before they are released with practical ambitions.

REFERENCES

AALEN, O O and HUSEBYE, E (1991) Statistical analysis of repeated events forming renewal processes *Statistics in Medicine*, **10**, 1227-1240

ANDERSEN, P K , BORGAN, Ø , GILL, R D and KEIDING, N (1993) *Statistical Models Based on Counting Processes*, Springer Verlag, New York

ANDERSEN, P K , CHRISTENSEN, E , FAUERHOLDT, L and SCHLICHTING, P (1983) Evaluating prognoses based on the proportional hazards model *Scand J Statist* , **10**, 141-144

ANDERSEN, P K and GILL, R D (1982) Cox's regressing model for counting processes A large sample study *Ann Statist* , **10**, 1100-1120

BLOSSFELD, H -P , HAMERLE, A and MAYER, K U (1989) *Event History Analysis Statistical Theory and Application in the Social Sciences* Lawrence Erlbaum, Hillsdale, NJ

BLOSSFELD, H -P and ROHWER, G (1995) *Techniques of Event History Modeling* Lawrence Erlbaum, Mahwah, NJ

COX, D R (1972a) Regression models and life tables (with discussion) *J R Statist Soc* , B **34**, 187-220

COX, D R (1972b) The statistical analysis of dependencies in point processes In *Stochastic Point Processes*, Ed P A W Lewis, pp 55-66 John Wiley, New York

DABROWSKA, D M (1995) Estimation of transition probabilities and bootstrap in a semiparametric Markov renewal model *Nonparametric Statistics*, **5**, 237-259

DABROWSKA, D M , SUN, G and HOROWITZ, M M (1994) Cox regression in a Markov renewal model An application to the analysis of bone marrow transplant data *J Amer Statist Association*, **89**, 867-877

GILL, R D (1980) Nonparametric estimation based on censored observations of a Markov renewal process *Z Wahrsch verw Geb* , **53**, 97-116

GILL, R D (1983) Discussion of the papers by Helland and Kurtz *Bull Internat Statist Inst* , **50**, 239-243

HOUGAARD, P (1995) Frailty models for survival data *Lifetime Data Analysis*, **1**, 19-38

KALBFLEISCH, J D and PRENTICE, R L (1980) *The statistical analysis of failure time data* Wiley, New York

LAWLESS, J F (1998) Repeated events *Encyclopedia of Biostatistics*, **5**, 3783-3787

LAWLESS, J F and THIAGARAJAH, K (1996) A point-process model incorporating renewals and time trends, with application to repairable systems *Technometrics*, **38**, 131-138

LINDSEY, J K (1995) Fitting parametric counting processes by using log-linear models *Appl Statist* , **44**, 201-212

MCCLEAN, S and DEVINE, C (1995) A nonparametric maximum likelihood estimator for incomplete renewal data *Biometrika*, **82**, 791-803

OAKES, D (1992) Frailty models for multivariate event times In Klein, J P and Goel, P K (eds) *Survival analysis State of the art*, pp 371-379 Netherlands, Kluwer

OAKES, D (1998) Duration dependence *Encyclopedia of Biostatistics*, **2**, 1248-1252

OAKES, D and CUI, L (1994) On semiparametric inference for modulated renewal processes *Biometrika*, **81**, 83-90

PRENTICE, R L , WILLIAMS, B J and PETERSON, A V (1981) On the regression analysis of multivariate failure time data *Biometrika*, **68**, 373-379

RAMLAU-HANSEN, H (1983) Smoothing counting process intensities by means of kernel functions *Ann Statist* , **11**, 453-466

SCHEIKE, T H , MARTINUSSEN, T and PETERSEN, J H (1997) Retrospective ascertainment of recurrent events An application to time to pregnancy Res Rep 97/14, Dep Biostat , Univ Copenhagen

SOON, G and WOODROOFE, M (1996) Nonparametric estimation and consistency for renewal processes *Journal of Statistical Planning and Inference*, **53**, 171-195
TJUR, T (1993) The StatUnit manual University of Copenhagen, Institute of Mathematical Statistics
VARDI, Y (1982) Nonparametric estimation in renewal processes *Ann Statist*, **10**, 772-785

DR. NIELS KEIDING
Institute of Public Health
Department of Biostatistics
University of Copenhagen
3 Blegdamsvej
DK-2200 Copenhagen N
Denmark
Tel. +45 35 32 79 01
Fax +45 35 32 79 07

CHRISTIAN ANDERSEN
PETER FLEDELIUS
ATP PensionService A/S
Kongens Vaenge 8
DK-3400 Hilleroed
Denmark

ON STOP-LOSS ORDER AND THE DISTORTION PRICING PRINCIPLE

BY WERNER HURLIMANN,

Winterthur

ABSTRACT

A number of more or less well-known, but quite complex, characterizations of stop-loss order are reviewed and proved in an elementary way. Two recent proofs of the stop-loss order preserving property for the distortion pricing principle are invalidated through a simple counterexample A new proof is presented. It is based on the important Hardy-Littlewood transform, which is known to characterize the stop-loss order by reduction to the usual stochastic order, and the dangerousness characterization of stop-loss order under a finite crossing condition Finally, we complete and summarize the main properties of the distortion pricing principle.

KEY WORDS

Pricing theory, distortion function, quantile function, stop-loss order, stochastic order, Hardy-Littlewood transform

1. INTRODUCTION

Since its introduction by Bühlmann (1970), the functional approach to premium calculation in insurance has seen an impressive development. A first general and rather elementary method to generate valuable pricing principles consists of the class of quantile premium calculation principles by Denneberg (1985/90/94) Several recent contributions around this theme have been made in actuarial science and finance, among others Hurlimann (1993), Wang (1995a/b/c, 1996a/b), Wang et al. (1997) and Chateauneuf et al. (1996).

For a given set S of non-negative random variables $X \geq 0$ with finite means, defined on some probability space, and which represent random losses of insurance contracts, a *pricing principle* is a non-negative real function $P \cdot S \rightarrow R$, which depends on the distribution $F_X(x)$ of X, and which is interpreted as price of the insurance risk From an axiomatic point of view, it is well accepted that a pricing principle should satisfy a certain number of desirable properties. Without repeating all well-known interpretations, the following properties are quite reasonable:

(P1) $P[X] \geq E[X]$, for all $X \in S$

(P2) $P[X] \leq \sup[X]$, for all $X \in S$

(P3) $P[aX+b] = aP[X]+b$, for all $a, b, a > 0$, for all $X \in S$

(P4) $P[X+Y] \leq P[X]+P[Y]$, for all $X, Y \in S$ such that $X+Y \in S$

(P5) $P[X] \leq P[Y]$ if $X \leq_{sl} Y$ and $X, Y \in S$

The last property says that the price functional preserves the stop-loss order, or equivalently the increasing convex order (see Kaas et al. (1994) and Shaked and Shanthikumar (1994) for fundamentals). Requiring that the price functional preserves the usual stochastic order $\leq_{st}$ only, is a less stringent property since stochastic order implies stop-loss order. Though the stop-loss ordering preserving property of the Swiss family of premium calculation principles has been known since its actuarial consideration in Bühlmann et al. (1977), the recognition of $\leq_{sl}$ as a sound ordering of risk seems more recent. For example, the order preserving axiom (P5) is considered in Heilmann (1987) but without mention of a specific and accepted partial order, which could be used as selected ordering of risk. Furthermore, the absolute deviation principle and the Gini principle, introduced by Denneberg (1985/90), and which satisfy properties (P1)-(P4), and the weaker stochastic order preserving property, also satisfy (P5), (consequence of our main result in Section 3.2). Previously two quite similar but different proofs of (P5) have been proposed by Wang (1996a/b), but both contain an error (see Section 3.1).

In view of the above discussion, it seems useful to present a short chronological review of some main non-trivial pricing functions, which preserve $\leq_{sl}$, and inspect whether the remaining axioms (P1)-(P4) are satisfied.

The Swiss family is positively homogeneous if, and only if, it is the net principle (see Schmidt (1989), simpler proof by Hurlimann (1997), Example 4.1 (continued), p. 9). The first genuine pricing principles, which satisfy (P1)-(P5), are the absolute deviation principle $P[X] = E[X] + \theta \cdot E[|X - m_X|]$, $0 \leq \theta \leq 1$ (Denneberg (1985/90)) and the Gini principle $P[X] = E[X] + \theta \ \mathrm{Gini}[X]$, $0 \leq \theta \leq 1$ (Denneberg (1990)). These functionals are special cases of the class of distortion pricing principles.

$$P[X] = \int_0^\infty g(\bar{F}_X(x))dx = \int_0^1 F_X^{-1}(1-u)dg(u) = \int_0^1 F_X^{-1}(u)d\gamma(u), \quad (1.1)$$

where $g(x)$ is an increasing concave function such that $g(0) = 0$, $g(1) = 1$, $\bar{F}_X(x) = 1 - F_X(x)$ is the survival function, $\gamma(x) = 1 - g(1-x)$ is the distortion of probabilities in Denneberg's setting, and $F_X^{-1}(u)$ is a quantile function of X. The second equality is obtained through partial integration, and shown by elementary calculus in case $g(x)$ is differentiable. The right-hand side representation has been introduced by Denneberg (1990) and its equivalence with the first integral (up to an alternative appropriate definition of the inverse) has been used by Wang (1996a) (see also Wang et al. (1997)).

Another attractive special case of (1.1) is the PH-transform principle studied by Wang (1995a/95b/96a/96b). Previously to the last example had appeared the Dutch principle (see van Heerwaarden (1991), van Heerwaarden and Kaas (1992), Kaas et al. (1994) and a slight generalization of it (see Hurlimann (1994/95a/95b)). A pricing principle from the Dutch family satisfies (P1)-(P5) if, and only if, it is of the form

$$P[X] = E[X] + \theta \; E\left[(X - E[X])_+\right], \quad 0 \le \theta \le 1. \qquad (1.2)$$

The Dutch family is a special case of the class of so-called "quasi-mean value principles" considered recently by the author. However, only sporadic members of this class define feasible price functionals satisfying ($P1$)-($P5$), of which one may mention the interesting Example 11 1 in Hürlimann (1997a)

A generalization of the class of distortion pricing principles is the class of Choquet pricing principles in Chateauneuf et al. (1996), which is based on the theory of capacities and non-additive measures (exposed in Denneberg (1994)), and breaks with the traditional probabilistic foundations of actuarial science and finance. Finally, let us mention that one misses still feasible price functionals along the economic approach initiated by Buhlmann (1980/84) (see the critical comments by Lemaire (1988)).

In the present paper, we invalidate Wang's proofs of the property ($P5$) for the distortion pricing principle through a simple counterexample, and focus on a new proof of this important property. Using a two-stage limiting argument (dominated convergence theorem and continuity property of the distortion pricing functional), it is possible to restrict the attention to risks, which belong to the following large set

S consists of all non-negative random variables with
finite means, such that the distribution functions of any
two of them cross finitely many times (*finite crossing condition*) (1.3)

For completeness, we show also that (1 1) satisfies the other properties ($P1$)-($P4$), where our exposé is intended to be essentially accessible from an elementary perspective.

The paper is organized as follows In Section 2, a number of more or less well-known, but quite complex, characterizations of stop-loss order are reviewed and proved in an elementary way. Since no such proofs have been found in the original and other papers (and books) consulted by the author, the present supplement to the existing literature will hopefully be helpful for future workers in this area (as it has been to the author). Section 3 is devoted to a derivation of the main properties of the distortion pricing principle. In Section 3.1 two recent proofs by Wang of the stop-loss order preserving property for the distortion pricing principle are invalidated through a simple counterexample A new proof is presented in Section 3.2 It is based on the important Hardy-Littlewood transform, which is known to characterize the stop-loss order by reduction to the usual stochastic order (Theorem 2.3), and the dangerousness characterization of stop-loss order under the finite crossing condition (1.3) (Theorem 2.2) Finally, we complete and summarize the main properties of the distortion pricing principle in Section 3.3.

2. Some equivalent characterizations of stop-loss order

Capital letters X, Y, ... denote random variables with distribution functions $F_X(x)$, $F_Y(x)$, ... and finite means μ_X, μ_Y, . . . The survival functions are denoted by $\bar{F}_X(x) = 1 - F_X(x)$, ... The stop-loss transform of a random variable X is defined by

$$\pi_X(x) := E\left[(X-x)_+\right] = \int_x^\infty \bar{F}_X(t)dt, \quad x \text{ in the support of } X \qquad (2.1)$$

The random variable X is said to precede Y in *stochastic order* or *stochastic dominance of first order*, a relation written as $X \le_{st} Y$, if $\bar{F}_X(x) \le \bar{F}_Y(x)$ for all x in the common support of X and Y. The random variables X and Y satisfy the *stop-loss order*, or equivalently the *increasing convex order*, written as $X \le_{sl} Y$ (or $X \le_{icx} Y$), if $\pi_X(x) \le \pi_Y(x)$ for all x. A sufficient condition for a stop-loss order relation is the *dangerousness order* relation, written as $X \le_D Y$, defined by the once-crossing condition

$$\begin{aligned} F_X(x) &\le F_Y(x) \text{ for all } x < c, \\ F_X(x) &\ge F_Y(x) \text{ for all } x \ge c, \end{aligned} \qquad (2.2)$$

where c is some real number, and the requirement $\mu_X \le \mu_Y$ (Lemma 2.1). By equal means $\mu_X = \mu_Y$, the ordering relations $\le_{sl}$ and $\le_D$ are precised by writing $\le_{sl,=}$ and $\le_{D,=}$. The partial stop-loss order by equal means is also called *convex order* and denoted by $\le_{cx}$. The probabilistic attractiveness of the partial order relations $\le_{st}$ and $\le_{sl}$ is corroborated by several invariance properties (e.g. Kaas et al. (1994), chap II.2 and III.2, or Shaked and Shanthikumar (1994)). For example, both of $\le_{st}$ and $\le_{sl}$ are closed under convolution and compounding, and $\le_{sl}$ is additionally closed under mixing and conditional compound Poisson summing

In applications, to establish stop-loss order comparison properties, one requires some fundamental facts and equivalent characterizations. First of all, the following well-known elementary equivalent statements hold:

(SL1) $X \le_{sl} Y$

(SL2) $E[\varphi(X)] \le E[\varphi(Y)]$ for all increasing convex functions $\varphi(x)$

(SL3) $E[\max(x, X)] \le E[\max(x, Y)]$ uniformly for all $x \in R$

A famous and widely known sufficient condition for stop-loss order is summarized in the following property.

Lemma 2.1. *(Karlin-Novikoff (1963) once-crossing condition, Lemma of Ohlin (1969)).* Let X and Y be random variables with distributions $F_X(x)$, $F_Y(x)$ and suppose that $X \leq_D Y$, as defined in (2.2). Then the stop-loss order relation $X \leq_{sl} Y$ is satisfied

Proof. By assumption, one has the inequalities

$$\max(x, X) \leq \max(x, Y), \quad x > c,$$
$$\min(x, X) \geq \min(x, Y), \quad x \leq c.$$

In particular, one obtains $E[\max(x, X)] \leq E[\max(x, Y)]$, $x > c$ By $(SL3)$ above, it remains to show the last inequality for $x \leq c$. This follows immediately from the identity

$$\max(x, X) = X + x - \min(x, X)$$

using the assumptions. □

A generalized version of the Karlin-Novikoff once-crossing conditions yields the following sign-change characterization of the stop-loss order. Without proof, one finds the relevant conditions in Taylor (1983), which attributes them to Stoyan (1977) However, the previous result by Taylor has not been formulated as a full characterization of stop-loss order

Theorem 2.1. *(Karlin-Novikoff-Stoyan-Taylor crossing conditions for stop-loss order).* Let $X, Y \in S$ be random variables with means μ_X, μ_Y, distributions $F_X(x)$, $F_Y(x)$ and stop-loss transforms $\pi_X(x)$, $\pi_Y(x)$. Suppose the distributions cross $n \geq 1$ times in the crossing points $t_1 < t_2 < \ldots < t_n$. Then one has $X \leq_{sl} Y$ if, and only if, one of the following is fulfilled:

Case 1 The first sign change of the difference $F_Y(x) - F_X(x)$ occurs from $-$ to $+$, there is an even number of crossing points $n = 2m$, and one has the inequalities

$$\pi_X(t_{2j-1}) \leq \pi_Y(t_{2j-1}), \quad j = 1, \ldots, m \tag{2.3}$$

Case 2 The first sign change of the difference $F_Y(x) - F_X(x)$ occurs from $+$ to $-$, there is an odd number of crossing points $n = 2m + 1$, and one has the inequalities

$$\mu_X \leq \mu_Y, \quad \pi_X(t_{2j}) \leq \pi_Y(t_{2j}), \quad j = 1, \ldots, m \tag{2.4}$$

Proof. Two cases must be distinguished.

Case 1 The first sign change occurs from $-$ to $+$

If $X \leq_{sl} Y$, the last sign change occurs from $+$ to $-$ (otherwise $\pi_X(x) > \pi_Y(x)$ for some $x \geq t_n$), hence $n = 2m$ is even Consider random variables $Z_0 = Y$, $Z_{m+1} = X$, and $Z_j, j = 1, .., m$ with distribution functions

$$F_j(x) = \begin{cases} F_X(x), & x \leq t_{2j-1}, \\ F_Y(x), & x \geq t_{2j-1} \end{cases} \tag{2 5}$$

For $j = 1, \ldots, m$, the Karlin-Novikoff once-crossing condition between Z_{j+1} and Z_j is fulfilled with crossing point t_{2j}. A partial integration shows the following mean formulas:

$$\mu_j := E[Z_j] = \mu_X - \pi_X(t_{2j-1}) + \pi_Y(t_{2j-1}), \quad j = 1, \ldots, m \tag{2.6}$$

Now, by Karlin-Novikoff, one has $Z_{j+1} \leq_D Z_j$, $j = 1, \ldots, m$, if, and only if, the inequalities $\mu_{j+1} \leq \mu_j$ are fulfilled, that is

$$\begin{aligned} &\pi_X(t_{2j-1}) - \pi_Y(t_{2j-1}) \leq \pi_X(t_{2j+1}) - \pi_Y(t_{2j+1}), \ j = 1, \ldots, m-1, \text{ and} \\ &\pi_X(t_{2m-1}) - \pi_Y(t_{2m-1}) \leq 0, \end{aligned} \tag{2 7}$$

which is equivalent to (2.3). Since obviously $Z_1 \leq_{st} Y$, one obtains the ordered sequence

$$X = Z_{m+1} \leq_D Z_m \leq_D \ldots \leq_D Z_1 \leq_{st} Z_0 = Y, \tag{2.8}$$

which is valid under (2.3) and implies the result.

Case 2 The first sign change occurs from $+$ to $-$

If $X \leq_{sl} Y$, then the last sign change occurs from $+$ to $-$, hence $n = 2m + 1$ is odd. Similarly to Case 1, consider random variables $Z_0 = Y$, $Z_{m+1} = X$, and $Z_j, j = 1, \ldots, m$, with distribution functions

$$F_j(x) = \begin{cases} F_X(x), & x \leq t_{2j}, \\ F_Y(x), & x \geq t_{2j}. \end{cases} \tag{2.9}$$

For $j = 0, 1, \ldots, m$, the once-crossing condition between Z_{j+1} and Z_j is fulfilled with crossing point t_{2j+1}. Using the mean formulas

$$\mu_j = E[Z_j] = \mu_X - \pi_X(t_{2j}) + \pi_Y(t_{2j}), \quad j = 1, \ldots, m, \tag{2.10}$$

the conditions for $Z_{j+1} \leq_D Z_j$, that is $\mu_{j+1} \leq \mu_j, j = 0, 1, ..., m$, are therefore

$$\begin{aligned} &\mu_X - \mu_Y \leq \pi_X(t_2) - \pi_Y(t_2), \\ &\pi_X(t_{2j}) - \pi_Y(t_{2j}) \leq \pi_X(t_{2j+2}) - \pi_Y(t_{2j+2}), \quad j = 1, .., m-1, \text{ and} \\ &\pi_X(t_{2m}) - \pi_Y(t_{2m}) \leq 0, \end{aligned} \tag{2,11}$$

which is equivalent to (2 4). One obtains the ordered sequence

$$X = Z_{m+1} \leq_D Z_m \leq_D \cdot \quad \leq_D Z_1 \leq_D Z_0 = Y, \tag{2.12}$$

which is valid under (2.4) and implies the result. □

It is instructive to relate this result with another (apparently simpler) known crossing characterization. Instead of crossing points, which describe the sign change properties of the distribution functions, consider slightly more general crossover points, which are defined as follows A pair $\{\xi, u\}$ of real numbers is a *crossover point* of the pair $\{F_1(x), F_2(x)\}$ of distribution functions if for $i \neq j \in \{1, 2\}$ one has

$$F_i(\xi^-) \leq F_j(\xi^-) \leq F_j(\xi) \leq F_i(\xi) \text{ and } u = F_j(\xi),$$

or equivalently

$$F_i^{-1}(u) \leq F_j^{-1}(u) \leq F_j^{-1}(u^+) \leq F_i^{-1}(u^+) \text{ and } \xi = F_j^{-1}(u).$$

How are the crossing points related to the crossover points? Clearly, every crossing point is a crossover point. Additionally, there are two crossover points, associated to the end points of the supports of $F_1(x)$, $F_2(x)$, where no actual sign change between the distributions occurs Let (a_i, b_i), $-\infty \leq a_i < b_i \leq \infty$, be the open support of $F_i(x)$, $i = 1, 2$, and set $\underline{a} = \min\{a_1, a_2\}$, $\bar{b} = \max\{b_1, b_2\}$. Then $(\underline{a}, \bar{b})$ is the open support of the pair $\{F_1(x), F_2(x)\}$, and $\{\underline{a}, 0\}$, $\{\bar{b}, 1\}$ are the remaining crossover points. The following characterization has been used by Kertz and Rosler (1992), again without proof.

Corollary 2.1 *(Crossover point characterization of the stop-loss order)* For $i = 1, 2$, let $X_i \in S$ be random variables with finite means μ_i, distributions $F_i(x)$, and stop-loss transforms $\pi_i(x)$. Then one has $X_1 \leq_{sl} X_2$ if, and only if, for all crossover points $\{\xi, u\}$ of the pair $\{F_1(x), F_2(x)\}$, the inequality $\pi_1(\xi) \leq \pi_2(\xi)$ is fulfilled.

Proof. It suffices to show that the conditions are *sufficient*. One needs the following additional criteria·

$$\begin{aligned} &\pi_1(\bar{b}) \leq \pi_2(\bar{b}) \Leftrightarrow b_1 \leq b_2, \\ &\pi_1(\underline{a}) \leq \pi_2(\underline{a}) \Leftrightarrow \mu_1 \leq \mu_2. \end{aligned} \tag{2.13}$$

The first one follows immediately from the integral representation $\pi_i(x) = \int_x^{\infty} \bar{F}_i(t)dt$. For the second one, we distinguish between two cases. If $\underline{a} > -\infty$, then the equivalence follows from the fact that $\pi_i(\underline{a}) = \mu_i - \underline{a}$, $i = 1, 2$. If $\underline{a} = -\infty$, the inequality

$$\mu_1 = \int_0^{\infty} \bar{F}_1(x)dx - \int_{-\infty}^{0} F_1(x)dx \leq \int_0^{\infty} \bar{F}_2(x)dx - \int_{-\infty}^{0} F_2(x)dx = \mu_2$$

can be rearranged to the inequality

$$\pi_1(\underline{a}) = \int_{-\infty}^{\infty} \bar{F}_1(x)dx \leq \pi_2(\underline{a}) = \int_{-\infty}^{\infty} \bar{F}_2(x)dx,$$

and vice versa. Since the set C of crossover points equals

$$C = \{\text{crossing points}\} \cup \{\underline{a}, 0\} \cup \{\bar{b}, 1\},$$

the inequalities $\pi_1(\xi) \leq \pi_2(\xi)$ for all $\{\xi, u\} \in C$ imply by the above criteria that the inequalities (2.3) and (2.4) required in Case 1 and Case 2 of the Theorem 2.1 are fulfilled. □

The simpler but less precise characterization by crossover points is often sufficient from the theoretical point of view (an example is Theorem 2 3 below) From a practical point of view, Theorem 2.1, together with the ordered sequences (2.8) and (2.12), yields the maximum amount of available information for a stop-loss order relation. In this respect, a detailed application of this result shows that $X_1 \leq_{sl} X_2$ if, and only if, the set C of crossover points is given as follows:

Case 1· $n = 2m$

$$C = \{\{a_1, 0\}, \{t_1, F_1(t_1)\}, \{t_2, F_2(t_2)\}, \{t_3, F_1(t_3)\}, \ldots, \{t_{2m}, F_2(t_{2m})\}, \{b_2, 1\}\},$$

Case 2. $n = 2m + 1$

$$C = \{\{a_2, 0\}, \{t_1, F_2(t_1)\}, \{t_2, F_1(t_2)\}, \{t_3, F_2(t_3)\}, \ldots, \{t_{2m+1}, F_2(t_{2m+1})\}, \{b_2, 1\}\}.$$

Some applications, which use the explicit characterization Theorem 2.1, are given in Hürlimann (1998a).

The once-crossing condition of dangerousness order formulated in Lemma 2.1 is not a transitive relation Though not a proper partial order, it is an important and main tool used to establish stop-loss order between two random variables In fact, the *transitive (stop-loss)-closure* of the order $\leq_D$, denoted by $\leq_{D^*}$, which is defined as the smallest partial order containing all pairs (X, Y) with $X \leq_D Y$ as a subset, identifies with the stop-loss order. To be precise, X precedes Y in the transitive (stop-loss-)closure of dangerousness, written as $X \leq_{D^*} Y$, if there is a sequence of random variables $Z_1, Z_2, Z_3,$, such that $X = Z_1$, $Z_i \leq_D Z_{i+1}$, and $Z_i \to Y$ in stop-loss convergence (equivalent to convergence in distribution plus convergence of the mean). The equivalence of $\leq_{D^*}$ and $\leq_{sl}$ is described in detail by Müller (1996) (see also Kaas and Heerwaarden (1992)). In case there are finitely many sign changes between the distributions, the stated result simplifies as follows.

Theorem 2.2. *(Dangerousness characterization of stop-loss order)* Let $X, Y \in S$ be random variables with finite means such that $X \leq_{sl} Y$. Then there exists a finite sequence of random variables $Z_1, Z_2, ..., Z_n$ such that $X = Z_1$, $Y = Z_n$ and $Z_i \leq_D Z_{i+1}$ for all $i = 1, ..., n-1$

Proof. This is Kaas et al. (1994), Theorem III.1.3 Alternatively, the ordered sequences (2.8) and (2 12) yield a more detailed constructive proof of this result. □

Other characterizations of the stop-loss order can be obtained by transforming the random variables, which must be compared A simple such result reduces a (degree one) stop-loss order comparison to a degree zero stop-loss order or usual stochastic order comparison by means of the Hardy-Littlewood maximal distribution. For any random variable X with finite mean and quantile function $F_X^{-1}(u)$, the *Hardy-Littlewood transform* X^H of X is defined by its quantile function on [0,1] through the formula

$$(F_X^H)^{-1}(u) = \begin{cases} \frac{1}{1-u}\int_u^1 F_X^{-1}(v)dv, & u < 1, \\ F_X^{-1}(1), & u = 1 \end{cases} \qquad (2\ 14)$$

Its name stems from the Hardy-Littlewood (1930) maximal function The random variable X^H is the least majorant with respect to $\leq_{st}$ among all random variables $Y \leq_{sl} X$ (e g. Meilijson and Nàdas (1979)). Its great importance in applied probability and related fields has been noticed by several further authors, among others Blackwell and Dubins (1963), Dubins and Gilat (1978), Rüschendorf (1991), and Kertz and Rosler (1990/92). A recent actuarial use has been proposed by the author (1998b)

Theorem 2.3. *(Reduction of stop-loss order to stochastic order)* For $i = 1, 2$, let $X_i \in S$ be random variables with finite means μ_i, distributions $F_i(X)$, and stop-loss transforms $\pi_i(x)$. Then one has $X_1 \leq_{sl} X_2$ if, and only if, one has $X_1^H \leq_{st} X_2^H$.

Proof. (Kertz and Rösler (1992), Lemma 1 8) The basic idea relies on the following geometric property. For each crossover point $\{\xi, u\}$, the identity

$$\int_\xi^\infty \{F_1(t) - F_2(t)\}dt = \int_u^1 \{F_2^{-1}(v) - F_1^{-1}(v)\}dv$$

expresses the fact that the area between F_1 and F_2 to the right of ξ equals the area between F_1^{-1} and F_2^{-1} to the right of u. From this and the Corollary 2.1 one obtains the result by means of the following equivalences:

$$X_1 \leq_{sl} X_2$$
$$\Leftrightarrow \pi_1(\xi) = \int_\xi^\infty \bar{F}_1(t)dt \leq \int_\xi^\infty \bar{F}_2(t)dt = \pi_2(\xi) \text{ for all crossover points } \{\xi, u\}$$
$$\Leftrightarrow \int_\xi^\infty \{F_1(t) - F_2(t)\}dt \geq 0 \text{ for all crossover points } \{\xi, u\}$$
$$\Leftrightarrow \int_u^1 \{F_2^{-1}(v) - F_1^{-1}(v)\}dv \geq 0 \text{ for all crossover points } \{\xi, u\}$$
$$\Leftrightarrow (F_1^H)^{-1}(u) \leq (F_2^H)^{-1}(u) \text{ for all } u \in [0,1]$$
$$\Leftrightarrow X_1^H \leq_{st} X_2^H. \qquad \square$$

By existence of a common mean $\mu_1 = \mu_2$, the resulting characterization of the convex order $X_1 \leq_{cx} X_2 \Leftrightarrow X_1^H \leq_{st} X_2^H$ is found in equivalent form in van der Vecht (1986), p. 69, which attributes the result to D. Gilat In this situation, there exists also the well-known higher degree stop-loss order reduction property of the integrated tail transform considered by van Heerwaarden (1991), p. 69, whose importance lies in actuarial ruin models (see e.g. Embrechts et al (1997)). For completeness, one may mention a further characterization of the convex order by means of Markov kernels, which goes back to Blackwell (1953), and still another one by means of fusions for probability measures as studied by Elton and Hill (1992). For this, the interested reader is referred to Szekli (1995).

3. Properties of the distortion pricing principle

First, we invalidate S. Wang's proofs of the stop-loss order preserving property ($P5$) for the distortion pricing principle through a simple counterexample Then we focus on a new proof of this important property. For completeness and convenience of the reader, elementary proofs of the other properties ($P1$)-($P4$) are also provided, where reference is made to related results in the literature.

3.1. A diatomic counterexample

For real numbers $0 < a_2 < a_1 < b_1 < b_2$ and for $i = 1, 2$ let X_i be a diatomic random variable with support $\{a_i, b_i\}$ and probabilities $\{p_i, 1 - p_i\}$, $0 < p_i < 1$, and mean $\mu_i = a_i + (1 - p_i)(b_i - a_i)$. Assume $\mu_1 \leq \mu_2$ and $p_2 < p_1$ Then the dangerousness order relation $X_1 \leq_D X_2$ (a sufficient condition for $\leq_{sl}$) holds because $\mu_1 \leq \mu_2$ and the survival functions satisfy the Karlin-Novikoff once-crossing condition (known as Ohlin's Lemma in actuarial science).

$$\begin{aligned} \bar{F}_1(x) \geq \bar{F}_2(x), \quad x < c, \\ \bar{F}_1(x) \leq \bar{F}_2(x), \quad x \geq c, \end{aligned} \tag{3.1}$$

with $c = a_1$. Set $g(x) = x^{\frac{1}{\rho}}$, $\rho \geq 1$, in (1) to get the PH-transform principle $\Pi_\rho[X] = \int_0^\infty \bar{F}(x)^{\frac{1}{\rho}} dx$. In the notation of Wang, one has

$$RHS(\rho) = \int_c^\infty \left\{ \bar{F}_2(x)^{\frac{1}{\rho}} - \bar{F}_1(x)^{\frac{1}{\rho}} \right\} dx = (1-p_2)^{\frac{1}{\rho}}(b_2 - a_1) - (1-p_1)^{\frac{1}{\rho}}(b_1 - a_1)$$

Wang (1996b), proof of Theorem 1, states that $RHS(\rho) \geq \bar{F}_2(a_1)^{\frac{1}{\rho}-1} RHS(1)$, or equivalently $(1-p_2)^{\frac{1}{\rho}-1} \geq (1-p_1)^{\frac{1}{\rho}-1}$ This is not correct because $x^{\frac{1}{\rho}-1}$ is decreasing over $(0, \infty)$ for $\rho > 1$ and $(1-p_2) > (1-p_1)$ by assumption. Similarly, Wang (1996a), proof of Theorem 1, states that $RHS(\rho) \geq \bar{F}_2(a_1)^{\frac{1}{\rho}-1} RHS(1)$, or equivalently $(1-p_2)^{\frac{1}{\rho}-1} \geq (1-p_1)^{\frac{1}{\rho}-1}$, which is false for the same reason Despite this, one has

$$\Pi_\rho[X_1] = a_1 + (1-p_1)^{\frac{1}{\rho}}(b_1 - a_1) \leq a_2 + (1-p_2)^{\frac{1}{\rho}}(b_2 - a_1) = \Pi_\rho[X_2],$$

and therefore a correct proof of $(P5)$ must be given.

3.2. An elementary proof of the stop-loss order preserving property

In a first step we suppose that $X, Y \in S$. The idea of the proof is simple. For each $X \geq 0$, let X^g be the *distortion transform* with survival function $\bar{F}_X^g(x) = g(\bar{F}_X(x))$ By Theorem 2.2 it suffices to show that $X \leq_D Y$ implies $X^g \leq_{sl} Y^g$, which in turns implies that $P[X] = E[X^g] \leq E[Y^g] = P[Y]$, hence $(P5)$. Furthermore, by Theorem 2.3 it suffices to show that $X \leq_D Y$ implies $(X^g)^H \leq_{st} (Y^g)^H$. (Note that the distributions of $(X^g)^H$ and $(X^H)^g$ differ in general.)

Suppose that $X \leq_D Y$, that is $E[X] \leq E[Y]$ and there exists $q \in (0,1)$ such that

$$\begin{aligned} F_X^{-1}(u) \geq F_Y^{-1}(u), \quad 0 \leq u < q, \\ F_X^{-1}(u) \leq F_Y^{-1}(u), \quad q \leq u \leq 1. \end{aligned} \tag{3.2}$$

For simplicity, assume that $g(x)$ (resp. $\gamma(x)$) is differentiable and has an inverse $g^{-1}(x)$ (resp. $\gamma^{-1}(x)$). Then the distortion transform X^g has quantile function $(F_X^g)^{-1} = (\gamma \circ F_X)^{-1}$, and using (2.14) one obtains for the Hardy-Littlewood distortion transform $(X^g)^H$ the relationships

$$(F_X^{g,H})^{-1}(u) = \frac{1}{1-u} \int_u^1 (\gamma \circ F_X)^{-1}(v) dv = \frac{1}{1-u} \int_{\gamma^{-1}(u)}^1 F_X^{-1}(v) d\gamma(v), 0 \leq u < 1 \tag{3.3}$$

Similar expressions hold with X replaced by Y One must show that $(F_X^{g,H})^{-1}(u) \leq (F_Y^{g,H})^{-1}(u)$ for all $u \in [0, 1]$, or equivalently

$$\int_w^1 \{F_Y^{-1}(v) - F_X^{-1}(v)\} d\gamma(v) \geq 0 \text{ for all } w \in [0, 1]. \tag{3 4}$$

If $w \geq q$ this is trivial by the second inequality in (3.2) Let now $0 \leq w < q < 1$ Since $\gamma(x)$ is convex, the derivative $\gamma'(x)$ is increasing, in particular $\gamma'(w) \leq \gamma'(q) \leq \gamma'(1)$. The affirmation follows from the following chain of equalities and inequalities

$$\begin{aligned}
&\int_w^1 \{F_Y^{-1}(v) - F_X^{-1}(v)\} d\gamma(v) \\
&= -\int_w^q \{F_X^{-1}(v) - F_Y^{-1}(v)\}\gamma'(v)dv + \int_q^1 \{F_Y^{-1}(v) - F_X^{-1}(v)\}\gamma'(v)dv \\
&\geq \gamma'(q) \cdot \int_w^1 \{F_Y^{-1}(v) - F_X^{-1}(v)\}dv \geq \gamma'(q) \cdot \int_0^1 \{F_Y^{-1}(v) - F_Y^{-1}(v)\}dv \\
&= \gamma'(q) \ \{E[Y] - E[X]\} \geq 0
\end{aligned} \tag{3 5}$$

This achieves the proof of the stop-loss order preserving property for the distortion pricing principle in case the finite crossing condition (1.3) holds.

In case $X \leq_{sl} Y$ and there are infinitely many crossing points, the equivalence of $\leq_{sl}$ and $\leq_{D^*}$ shows that there is a sequence of random variables $Z_1, Z_2, Z_3, \ldots$, such that $X = Z_1$, $Z_i \leq_D Z_{i+1}$, and $Z_i \rightarrow Y$ in stop-loss convergence For each $n \geq 1$ one has $X \leq_{sl} Z_n$ by Theorem 2.2 From the preceding first step, one obtains that $P[X] \leq P[Z_n]$. On the other side, the relation $Z_1 \leq_D Z_{i+1}$ implies $\min(Z_i, d) \leq_D \min(Z_{i+1}, d)$ for all d, from which one deduces by the first step that $P[\min(Z_m, d)] \leq P[\min(Z_m, d)]$ for all d, all $m \geq n$. Using this, the result follows from the inequality

$$P[Z_n] = \lim_{d\to\infty} P[\min(Z_n, d] \leq \lim_{d\to\infty} \left\{ \lim_{m\to\infty} P[\min(Z_m, d)] \right\} = \lim_{d\to\infty} P[\min(Y, d)] = P[Y]$$

The first and third equality is a continuity property satisfied by the Choquet integral, and a fortiori by the distortion pricing principle, which is a special case of it (see Denneberg (1994), or Axiom 4, Theorem 1 to 3 in Wang et al. (1997)). The second equality is an application of the dominated convergence theorem, which is allowed for risks with finite support

3.3. Other properties of the distortion pricing principle

It is now possible to complete and summarize the main properties of the distortion pricing principle Up to $(P5)$ an advanced proof of this is in Denneberg (1994), pp. 64 and 71.

Theorem 3.1. *(Main properties of the distortion pricing principle)* Let X be a non-negative random variable with survival function $\bar{F}_X(x)$, and quantile function $F_X^{-1}(u)$. Let $g(x)$ be a differentiable increasing concave function on [0,1] such that $g(0) = 0$, $g(1) = 1$. Then the functional $P[X] = \int_0^\infty g[\bar{F}(x)]dx = \int_0^1 F_X^{-1}(u)d\gamma(u)$ with $\gamma(x) = 1 - g(1-x)$, satisfies the properties $(P1) - (P5)$

Proof. $(P1)$-$(P3)$ are easily shown as follows (see also Denneberg (1990)).
$(P1)$ Since $g(x)$ is increasing concave on [0,1] and $g(0) = 0, g(1) = 1$, one has $g(x) \geq x$ and therefore $P[X] \geq \int_0^\infty \bar{F}(x)dx = E[X]$
$(P2)$ One first shows that $P[X]$ preserves $\leq_{st}$, which is obvious because $X \leq_{st} Y$ is equivalent with $F_X^{-1}(u) \leq F_Y^{-1}(u)$ for all $u \in (0,1)$ Since $X \leq_{st} Y := \sup[X]$, the property follows.
$(P3)$ This property follows from the facts $F_{X+b}^{-1}(u) = F_X^{-1}(u) + b$ and $F_{a\,X}^{-1}(u) = a\ F_X^{-1}(u)$ for $a \geq 0$.
$(P4)$ That this holds when $\gamma(x)$ has a bounded density is mentioned by Denneberg (1990). Using Wang (1995a), Appendix, one relaxes this condition as follows, where differentiability of $g(x)$ is here not assumed (The idea of proof is attributed to O. Hesselager). A simple property of concave functions is required.

Lemma 3.1. Let $0 < a < b$ and suppose $g(x)$ is concave for $x \geq 0$. Then for any $x \geq 0$ one has the inequality $g(x+b) - g(x+a) \leq g(b) - g(a)$.

Proof. It is well-known that $g(x)$ is concave if, and only if, one has

$$\frac{g(y) - g(x)}{y - x} \geq \frac{g(z) - g(y)}{z - y} \text{ for all } 0 \leq x < y < z.$$

Two successive applications of this criterion to $a < b \leq x + a < x + b$, respectively $a < x + a < b < x + b$, yields the desired inequality. □

It suffices to show $(P4)$ for arbitrary Y and a discrete X taking values in $\{0, \ldots, n\}$. Indeed, applying $(P3)$, the result holds then for $X \in \{k, \ldots, n+k\}$ and $X \in \{kh, \ldots, (n+k)h\}$, $k \in N_+$, $h > 0$ arbitrary. Since any random variable can be approximated closely by a discrete random variable with small enough h, the property will hold for arbitrary X. One uses mathematical induction. For $n = 0$ the affirmation is obvious. To show the induction step $n \to n+1$ for (X, Y) with $X \in \{0, .., n+1\}$, let (X', Y') be distributed as $(X, Y|X > 0)$ Since $X' \in \{1, ..., n+1\}$ the induction hypothesis states that $P[X' + Y'] \le P[X'] + P[Y']$. With $\varepsilon = \Pr(X = 0)$ and $\bar{F}_{Y|0}(x) = \Pr(Y > x|U = 0)$ one has for $x > 0$:

$$\bar{F}_X(x) = (1 - \varepsilon)\bar{F}_{X'}(x),$$
$$\bar{F}_Y(x) = \varepsilon \bar{F}_{Y|0}(x) + (1 - \varepsilon)\bar{F}_{Y'}(x),$$
$$\bar{F}_{X+Y}(x) = \varepsilon \bar{F}_{Y|0}(x) + (1 - \varepsilon)\bar{F}_{X'+Y'}(x).$$

According to Lemma 3 1, one obtains for $x > 0$ that

$$g(\bar{F}_{X+Y}(x)) - g(\bar{F}_X(x))g(\bar{F}_Y(x)) \le g((1 - \varepsilon)\bar{F}_{X'+Y'}(x)) - g((1 - \varepsilon)\bar{F}_{X'}(x)) - g((1 - \varepsilon)\bar{F}_{Y'}(x)).$$

Observe now that $k(x) := \dfrac{g((1-\varepsilon)x)}{g(1-\varepsilon)}$ is increasing concave on [0,1] such that $k(0) = 0$, $k(1) = 1$. Integrate on both sides of the last inequality and use the induction assumption for the function $k(x)$ to see that

$$P[X+Y] - P[X] - P[Y]$$
$$\le g(1-\varepsilon) \left\{ \int_0^\infty k(\bar{F}_{X'+Y'}(x))dx - \int_0^\infty k(\bar{F}_{X'}(x))dx - \int_0^\infty k(\bar{F}_{Y'}(x))dx \right\} \le 0$$

This shows $(P4)$

Since the property $(P5)$ has been shown in Section 3.2, the proof is complete

□

Note added in proof. At the time this paper has been accepted for publication, the author has received a related paper by Dhaene et al. (1997). These authors present in particular an alternative proof of the stop-loss order preserving property of the distortion functional, whose idea is due to A. Müller Moreover, their Theorem 3 characterizes stop-loss order using the distortion functional in a way dual to the classical characterization $(SL1)$-$(SL3)$ based on the expected value functional. Finally, the author is grateful to A. Müller for pointing out an error in the elementary proof of Section 3.2.

REFERENCES

BLACKWELL, D (1953) Equivalent comparisons of experiments Annals of Mathematical Statistics 24, 265-272

BLACKWELL, D and L E DUBINS (1963) A converse to the dominated convergence theorem Illinois Journal of Mathematics 7, 508-514

BUHLMANN, H (1970) Mathematical Methods in Risk Theory Springer-Verlag

BUHLMANN, H (1980) An economic premium principle ASTIN Bulletin 11, 52-60

BUHLMANN, H (1984) The general economic premium principle ASTIN Bulletin 14, 13-21

BUHLMANN, H, GAGLIARDI, B, GERBER, H U and E STRAUB (1977) Some inequalities for stop-loss premiums ASTIN Bulletin 9, 75-83

CHATEAUNEUF, A, KAST, R and A LAPIED (1996) Choquet pricing for financial markets with frictions Mathematical Finance 6(3), 323-30

DENNEBERG, D (1985) Valuation of first moment risk for decision purposes in Finance and Insurance In Goppel, H, Henn, R (ed) 3 Tagung Geld, Banken und Versicherungen, Karlsruhe, 855-869 Verlag Versicherungswirtschaft, Karlsruhe

DENNEBERG, D (1990) Premium calculation why standard deviation should be replaced by absolute deviation ASTIN Bulletin 20, 181-190

DENNEBERG, D (1994) Non-Additive Measure and Integral Theory and Decision Library, Series B, vol 27 Kluwer Academic Publishers

DHAENE, J, WANG, S, YOUNG, V and M J GOOVAERTS (1997) Comonotonicity and maximal stop-loss premiums Submitted

DUBINS, L E and D GILAT (1978) On the distribution of the maxima of martingales Transactions of the American Mathematical Society 68, 337-38

ELTON, J and T P HILL (1992) Fusions of a probability distribution The Annals of Probability 20(1), 421-54

EMBRECHTS, P, KLUPPELBERG, C and Th MIKOSCH (1997) Modelling Extremal Events for Insurance and Finance Applications of Mathematics – Stochastic Modelling and Applied Probability, vol 33

HARDY, G H and J E LITTLEWOOD (1930) A maximal theorem with function-theoretic applications Acta Mathematica 54, 81-116

HEERWAARDEN, VAN A E (1991) Ordering of risks theory and actuarial applications Ph D Thesis, Tinbergen Research Series no 20, Amsterdam

HEERWAARDEN, VAN A E and R KAAS (1992) The Dutch premium principle Insurance Mathematics and Economics 11, 129-133

HEILMANN, W-R (1987) Grundbegriffe der Risikotheorie Verlag Versicherungswirtschaft, Karlsruhe (English translation (1988) Fundamentals of Risk Theory)

HURLIMANN, W (1993) Optimal stop-loss limits under non-expected utility preferences In Operations Research '92, 550-552 Physica-Verlag

HÜRLIMANN, W (1994) A note on experience rating, reinsurance and premium principles Insurance Mathematics and Economics 14, 197-204

HURLIMANN, W (1995a) Links between premium principles and reinsurance International Congress of Actuaries, Brussels, vol 2, 141-167

HURLIMANN, W (1995b) Transforming, ordering and rating risks Bulletin of the Swiss Association of Actuaries, 213-236

HURLIMANN, W (1997) On quasi-mean value principles Blatter der Deutschen Gesellschaft fur Versicherungsmathematik XXIII, Heft 1, 1-16

HURLIMANN, W (1998a) Truncation transforms, stochastic orders and layer pricing 26-th International Congress of Actuaries, June 1998, Birmingham

HURLIMANN, W (1998b) On distribution-free safe layer-additive pricing Appears in Insurance Mathematics and Economics

KAAS, R and A E VAN HEERWAARDEN (1992) Stop-loss order, unequal means, and more dangerous distributions Insurance Mathematics and Economics 11, 71-77

KAAS, R, HEERWARDEN, VAN A E and M J GOOVAERTS (1994) Ordering of Actuarial Risks CAIRE Education Series 1, Brussels

KARLIN, S and A NOVIKOFF (1963) Generalized convex inequalities Pacific Journal of Mathematics 13, 1251-1279

KERTZ, R P and U ROSLER (1990) Martingales with given maxima and terminal distributions Israel Journal of Mathematics 69, 173-192

KERTZ, R P and U ROSLER (1992) Stochastic and convex orders and lattices of probability measures, with a martingale interpretation Israel Journal of Mathematics 77, 129-164

LEMAIRE, J (1988) Actuarial challenges of reinsurance Proceedings of the 23-th International Congress of Actuaries, Helsinki

MEILIJSON, I and A NADAS (1979) Convex majorization with an application to the length of critical paths Journal of Applied Probability 16, 671-77

MULLER, A (1996) Ordering of risks a comparative study via loss-stop transforms Insurance Mathematics and Economics 17, 215-222
OHLIN, J (1969) On a class of measures of dispersion with application to optimal reinsurance ASTIN Bulletin 5, 249-66
RUSCHENDORF, L (1991) On conditional stochastic ordering of distributions Advances in Applied Probability 23, 46-63
SHAKED, M, SHANTHIKUMAR, J G (1994) Stochastic orders and their applications Academic Press, New York
STOYAN, D (1977) Qualitative Eigenschaften und Abschatzungen Stochastischer Modelle Akademie-Verlag, Berlin (English version (1983) Comparison Methods for Queues and Other Stochastic Models J Wiley, New York)
SZEKLI, R (1995) Stochastic Ordering and Dependence in Applied Probability Lecture Notes in Statistics 97 Springer-Verlag
TAYLOR, J M (1983) Comparisons of certain distribution functions Math Operationsforschung und Statistik, Ser Stat 14(3), 397-408
VAN DER VECHT, D P (1986) Inequalities for stopped Brownian motion C W I Tracts 21, Mathematisch Centrum, Amsterdam
WANG, S (1995a) Insurance pricing and increased limits ratemaking by proportional hazards transforms Insurance Mathematics and Economics 17, 43-54
WANG, S (1995b) Insurance pricing and increased limits ratemaking by proportional hazards transforms Proc XXV Int Congress of Actuaries Brussels, vol 2, 293-323
WANG, S (1995c) Riks loads on life/non-life insurance a unified approach XXVI ASTIN Colloquium, Leuven
WANG, S (1996a) Premium calculation by transforming the layer premium density ASTIN Bulletin 26, 71-92
WANG, S (1996b) Ordering of risks under PH-transforms Insurance Mathematics and Economics 18, 109-114
WANG, S, V R YOUNG and H H PANJER (1997) Axiomatic characterization of insurance prices Insurance Mathematics and Economics 21, 173-183

WERNER HURLIMANN
Mathematik KB L
"Winterthur"
Paulstr. 9
CH-8401 Winterthur
Switzerland

ON THE ANALYSIS OF THE TRUNCATED GENERALIZED POISSON DISTRIBUTION USING A BAYESIAN METHOD

BY

DAVID P M SCOLLNIK

Department of Mathematics and Statistics
University of Calgary

ABSTRACT

The generalized Poisson distribution with parameters θ and λ was introduced by Consul and Jain (1973) and has recently found several instances of application in the actuarial literature. The most frequently used version of the distribution assumes that $\theta > 0$ and $0 \leq \lambda < 1$, in which case the mean and variance are $\theta/(1-\lambda)$ and $\theta/(1-\lambda)^3$, respectively. These simple moment expressions, along with nearly all of the other theoretical results available for this distribution, fail when $\lambda < 0$ or $\lambda > 1$ (e.g., Johnson, Kotz, and Kemp, 1992, page 397). In these cases, even the definition of the probability mass function usually given in the literature is not properly normalized so that its values do not sum to unity. For this reason, it is common to truncate the support of the distribution and explicitly normalize the probability mass function. In this paper we discuss the estimation of the parameters of this truncated generalized Poisson distribution using a Bayesian method

KEYWORDS

Bayesian; bivariate; generalized Poisson; Langrangian Poisson; truncated; Markov chain Monte Carlo.

1. INTRODUCTION

A great many distributions are available for modelling discrete data arising in the insurance field. A large number of these discrete distributions are described in Chapter 3 of Klugman, Panjer, and Willmot (1997). Recently, some authors have also explored the use of Consul's Generalized Poisson Distribution (GPD) in actuarial settings. Consul (1990) demonstrated that the GPD, sometimes also known as the Lagrangian Poisson distribution, is a plausible model for claim frequency data, Goovaerts and Kaas (1991) and Ambagaspitiya and Balakrishnan (1994) presented recursive methods to compute the total claims distribution for certain compound GPD models, as did Hesselager (1997) for a class of compound Lagrangian distributions including the compound GPD; Scollnik (1995a) used the GPD, and its extension to a regression context, in order to model various sorts of claim frequency data and showed how Markov chain Monte Carlo (MCMC) methods could be used to implement Bayesian posterior and predictive analyses of these models (see also Scollnik, 1995b and 1995c); Famoye and Consul (1995) introduced a version of bivariate GPD (BGPD), discussed parameter estimation by the method of moments and double zero frequency and by the method of maximum likelihood, and fit the BGPD to a data set on accidents sustained by a group of shunters, Vernic (1997) considered the same BGPD as did Famoye and Consul, and used method of moments estimation to fit this BGPD to the aggregate amount of claims for a compound class of policies submitted to claims of two kinds whose yearly frequencies are *a priori* dependent.

The purpose of this paper is to clarify some points relating to the GPD which are frequently misrepresented in the literature and to discuss how Bayesian posterior and predictive analysis of the truncated GPD and of a truncated BGPD can proceed using MCMC methods. We begin with a discussion of GPD models.

2. Generalized Poisson Distribution Models

The probability mass function of the basic untruncated GPD is commonly given by

$$Pr(N=n)=p_n(\theta,\lambda,m)=\begin{cases}\theta(\theta+n\lambda)^{n-1}\frac{\exp(-\theta-n\lambda)}{n} & for\, n=0,1,2,\ldots,m\\ 0 & for\, n>m\ when\ \lambda<0,\end{cases} \quad (1)$$

and zero otherwise, where $\theta>0$, $\max(-1, -\theta/m)\leq\lambda<1$, and m is usually taken equal to the largest possible positive integer such that $\theta+m\lambda>0$ when λ is negative. Often it is explicitly further required that $m\geq 4$ (e.g., as in Vernic, 1997) in order to ensure that there are at least five classes with non-zero probability when λ is negative (see Consul, 1989, page 4), but this obviously need not be the case. At this time, we will review a few of the properties associated with (1) Most of these properties are documented in Consul's (1989) treatment of the GPD. Additional references will be introduced as required. The reader is forewarned that some authors switch the roles of the parameters θ and λ. We have adopted the parametrization found in Consul (1989) and Johnson, Kotz and Kemp (1992, page 396).

To begin with, suppose that $0 \leq \lambda < 1$ and the value of m is taken equal to ∞. For this case it is known that

$$E(N) = \frac{\theta}{1-\lambda} \quad and \quad Var(N) = \frac{\theta}{(1-\lambda)^3}, \tag{2}$$

so the variance of the GPD is always larger than or equal to the mean. It is apparent that this instance of the GPD reduces to the standard Poisson with parameter θ when $\lambda = 0$. The two moment expressions in (2), along with simple formulae for skewness and kurtosis and virtually all of the other theoretical results obtained relating to the GPD (e.g., Consul and Jain, 1973; Ambagaspitiya and Balakrishnan, 1995; Vernic, 1997), are only valid for the case of the GPD presently under consideration, i.e. when $\theta > 0$, $0 \leq \lambda < 1$, and $m = \infty$.

Henze and Klar (1995, page 1877) make the claim that this fact has not been emphasized enough in the literature, and point to a paper by Alzaid and Al-Osh (1993) in which it is tacitly assumed that (2) also holds for negative values of λ. Famoye and Consul (1995, page 128) recently made the same errant assumption, without alerting the reader as to its nature. It is also very common for authors to estimate the GPD parameters by equating empirical moments to the theoretical moments obtained in the special case described above, even when the sample variance is strictly less than the sample mean so that negative estimates of λ result (e.g., Consul, 1989, see also Vernic, 1997).

Actually, in order to permit cases where the variance is smaller than the mean, Consul and Jain (1973) had proposed to admit negative values of λ. However, when the value of λ is negative the probability mass function (1) is no longer normalized. To see this, suppose that $\theta = 1.6$, $\lambda = -0.75$ and $m = 2$. Then $Pr(N = 0) = 0.2019$, $Pr(N = 1) = 0.6839$, $Pr(N = 2) = 0.0724$, $Pr(N > 2) = 0.0$, and the sum of these supposedly exhaustive 'probabilities' is only 0.9582. This problem was not recognized in the early literature concerning the GPD (e.g., Consul and Jain, 1973) until Nelson (1975) indicated that a cautious approach was warranted in the use of the GPD model with negative values of λ. One solution to this problem is to simply normalize the function in (1) when $\lambda < 0$. In fact, (1) will generally need to be normalized except in the special case that $\theta > 0$, $0 \leq \lambda < 1$, and $m = \infty$ Accordingly, Consul and Famoye (1989) defined the probability mass function of the truncated GPD to be

$$Pr(N = n) = q_n(\theta, \lambda, m) = \frac{p_n(\theta, \lambda, m)}{K(\theta, \lambda, m)} \quad for\ n = 0, 1, 2, ..., m \tag{3}$$

and zero otherwise, where $\theta > 0$, $-\infty < \lambda < \infty$,

$$K(\theta, \lambda, m) = \sum_{n=0}^{m} p_n(\theta, \lambda, m)$$

and m is any positive integer such that $\theta + m\lambda > 0$. Usually, m is taken equal to the largest such value. Note that the definition of the truncated GPD extends the permitted range of the parameter λ to the entire real line.

When class frequencies are inappropriately calculated using (1) instead of (3), an error of truncation is said to occur Consul and Shoukri (1985) and Consul (1989, Section 9.1.1) have made an analysis of the error of truncation when $-1 < \lambda < 0$. The simulation study they conduct is not exhaustive, but it does appear to indicate that the error of truncation may be serious when the number of non-zero probability classes is 3 or 4 and the value of θ is approximately between 0 7 and 4.5. The reader can easily verify that the error of truncation may also be serious when $\lambda < -1$ or $\lambda > 1$.

Consul and Famoye (1989) studied the truncated GPD in some detail and discussed parameter inference using maximum likelihood (ML) estimation and estimation based upon the empirical mean and the ratio of the first two empirical class frequencies. Their main conclusion was that the ML estimates determined using (3) as the basis of the likelihood function are generally closer to the true values of the population parameters than are the ML estimates determined on the basis of (1) Hence, even though the error of truncation associated with using (1) may be small in some cases, they suggested that one should estimate the values of the parameters θ and λ using the truncated GPD model (3). It should be noted that the estimation methods persued by Consul and Famoye (1989) are implemented in such a way so as to determine estimates of θ and λ conditional upon a presumed known value of m. Since m is not known, Consul and Famoye (1989) simply set it equal to the value of the largest observation.

Bayesian estimation is a likelihood based style of inference that incorporates prior information on the unknown variables ML estimates are equivalent to the nodes of the Bayesian posterior distribution, when the prior distribution for the unknown variables is flat. However, the goal of a Bayesian analysis is generally not just a point estimate like the posterior mode (or mean or median), but a representation of the entire distribution for the unknown parameter(s) (Gelman, Carlin, Stern, Rubin, 1995, page 301). In the next Section, we discuss how a Bayesian analysis of the truncated GPD with an informative prior distribution can be accomplished using a MCMC approach. We emphasize that the Bayesian estimation method yields a posterior distribution for all of the unknown parameters, including m (cf. Consul and Famoye, 1989).

3. A Bayesian Analysis of the Truncated GPD Model

Consul and Famoye (1989) argue that any discrete probability model for a random variable N defined on the set of non-negative integers is automatically truncated in real life situations because the sample size is always finite and the probabilities for large values of N become so small so as to be unobservable. This is particularly true in an insurance setting when the number of claims per policy is small Assuming this context, we suppose that the sampling model is taken to be approximately truncated GPD as in (3) with parameters θ, λ, and m, so that

$$Pr(N = j|\theta, \lambda, m) = q_j(\theta, \lambda, m) = \frac{p_j(\theta, \lambda, m)}{K(\theta, \lambda, m)} \quad \text{for } j = 0, 1, 2, \ldots, m, \tag{4}$$

and zero otherwise, with $\theta > 0$ and $-\infty < \lambda < \infty$, with m equal to some positive integer such that $\theta + m\lambda > 0$, and with $1 \leq m \leq M$ so that there is at least one non-zero class with non-zero probability. Setting M equal to a value between 5 & 15, say, will generally suffice when the number of claims per policy or accidents per individual is small. We recognize that the value selected for the parameter M is formally an expression of *a priori* knowledge This is further discussed in the next paragraph. If the data consists of observed class frequencies $n_j, j = 0, \ldots, M$, with $n = n_0 + \ldots n_M$, then the likelihood function is of the form

$$l(\theta, \lambda, m) \propto \prod_{j=0}^{m} q_j(\theta, \lambda, m)^{n_j} = \frac{\prod_{j=0}^{m} p_j(\theta, \lambda, m)^{n_j}}{K(\theta, \lambda, m)^n} . \tag{5}$$

If the data includes some grouped class frequencies, then the likelihood function is modified in the obvious way. For example, if we observe the first two class frequencies n_0 and n_1 along with the grouped class frequency $g_2 = n_2 + \ldots + n_M$, then the likelihood function is of form

$$l(\theta, \lambda, m) \propto q_0(\theta, \lambda, m)^{n_0} q_1(\theta, \lambda, m)^{n_1} \{1 - q_0(\theta, \lambda, m) - q_1(\theta, \lambda, m)\}^{g_2}.$$

In order to complete the definition of a full probability model, it is now necessary to specify a prior distribution for the unknown parameters θ, λ, and m The reader is free to use any reasonable prior specification as befits the expert opinion that is available to him or her. For our presentation, we will consider 3 different forms of prior density specification (PDS). For the first PDS, we will assume that the parameters are distributed *a priori* in the following way:

$$p(\theta, \lambda, m) \propto p(\theta)p(\lambda)p(m) \quad \text{when } \theta + m\lambda > 0 , \tag{6}$$

and zero otherwise, with

$$p(\theta) \sim Gamma\ (1, 2) , \tag{7}$$

$$p(\lambda) \sim Normal\ (0, 0.1) , \tag{8}$$

$$p(m) \sim Uniform\ \{1, \ldots, M\} . \tag{9}$$

The *Gamma* distribution in (7) is parametrized so as to have mean and standard deviation both equal to 0.5, and the *Normal* distribution in (8) has standard deviation equal to 0.1 With respect to the *Uniform* distribution in (9), we are free to attach a hyper-prior distribution to the parameter M. We have not pursued this particular avenue, although in Section 5 we will compare the use of several different values of M in the context of a particular data analysis

Another approach is to forgo the introduction of M entirely, and rather specify a distribution $p(m)$ on the entirety of the non-negative integers (in effect, $M = \infty$). In this case, equations (6), (7) and (8) would be unchanged, and (9) might be replaced with

$$p(m) \sim Poisson\ (\mu)\ , \tag{10}$$

for some specified value $\mu > 0$ The parameter restrictions in effect would be $\theta > 0$, $-\infty < \lambda < \infty$, and $\theta + m\lambda > 0$. An analysis of the truncated GPD model incorporating this second form of PDS will also follow in Section 5.

Our third PDS will be similar to the two above, with the added restriction that $m = M$, for some specified value $M < \infty$. That is, our third analysis will be conditional on a fixed value of $m < \infty$.

By multiplying the likelihood and prior density functions together, we obtain the form of the posterior distribution up to a normalizing constant, that is

$$p(\theta, \lambda, m | n_0, \ \ldots, n_m) \propto p(\theta)p(\lambda)p(m)l(\theta, \lambda, m) \quad \text{when } \theta + m\lambda > 0\ , \tag{11}$$

and zero otherwise, with $\theta > 0$ and $-\infty < \lambda < \infty$. If we let n^* denote the value of the largest observation, then we also require that $m \in \{n^*, ..., M\}$. Here, either the value of $M < \infty$ is known as in the case of our first PDS, or else $M = \infty$ as in the second. In the case of our third PDS, M is assumed to be known and we further condition upon the assumption that $m = M < \infty$. At this stage, the complete probability model can be analysed using a numerical method. We propose the use of a MCMC method in order to complete the analysis of the posterior and predictive distributions.

4. Completing the Bayesian Analysis Using a MCMC Method

In order to complete the Bayesian analysis of the truncated GPD model, we adopt a MCMC method. In particular, we implement a 'single-component Metropolis-Hastings' (Gilks, Richardson, and Spiegelhalter, 1996, page 10), or 'variable-at-a-time Metropolis-Hastings' (cf. Chan and Geyer's discussion of Tierney's 1994 paper, page 1748; also, Haastrup and Arjas, 1996, page 156), algorithm. This algorithm simulates a realization of a Markov chain which has the posterior distribution of the unknown parameters θ, λ, and m as its equilibrium distribution The algorithm generates a sequence of simulated parameter values, $\theta^{(0)}$, $\lambda^{(0)}$, $m^{(0)}$, $\theta^{(1)}$, $\lambda^{(1)}$, $m^{(1)}$, . . , whose empirical distribution converges towards the posterior distribution of the unknown parameters. The posterior distribution can thus be approximated on the basis of these values, and the approximation can be made as exact as we desire by simply increasing the length of the simulation. Note that predictions can also be obtained by simply averaging the truncated GPD probability mass function over the sampled parameter values. That is, the probability mass function for a future observation N_f, given the observed class frequencies $n_0, \ldots, n_m$, can be estimated using the result that

$$
\begin{aligned}
Pr(N_f = j \mid n_0, \ldots, n_m) \qquad (12)\\
&= \sum_m \int \int Pr(N_f = j \mid \theta, \lambda, m) p(\theta, \lambda, m \mid n_0, \ldots, n_m)\, d\theta\, d\lambda \\
&\approx \sum_{i=B+1}^{B+L} \frac{Pr(N_f = j \mid \theta^{(i)}, \lambda^{(i)}, m^{(i)})}{L}.
\end{aligned}
$$

Here, B represents the number of iterations for which the Markov chain is allowed to 'burn-in' and L represents the number of iterations the Markov chain is run thereafter. A method for checking the convergence of the Markov chain by comparing several different and independently simulated sequences is given in Gelman, Carlin, Stern, and Rubin (1995, pages 330-333) If several different and independently simulated sequences are available, then the sample average in (12) should be taken over all of the available sample paths.

There are many ways of implementing the Markov chain described above. We proceed in the following manner. Let $\theta^{(0)}$, $\lambda^{(0)}$, and $m^{(0)}$ denote arbitrary starting values for the 3 random variables under examination In this context, the ith iteration of the single-component Metropolis-Hastings algorithm consists of 3 updating steps.

Step 1

We enter the first step of the ith iteration with values $\theta^{(i-1)}$, $\lambda^{(i-1)}$, and $m^{(i-1)}$. In this step, we update the value of θ by generating a candidate value θ^* from a proposal distribution indexed by $\theta^{(i-1)}$ with density $q_\theta(\theta|\theta^{(i-1)})$. The candidate value is accepted with probability

$$
min\left(1, \frac{p(\theta^*, \lambda^{(i-1)}, m^{(i-1)} \mid n_0, \ldots, n_m)\, q_\theta(\theta^{(i-1)} \mid \theta^*)}{p(\theta^{(i-1)}, \lambda^{(i-1)}, m^{(i-1)} \mid n_0, \ldots, n_m)\, q_\theta(\theta^* \mid \theta^{(i-1)})}\right), \qquad (13)
$$

where the density $p(\theta, \lambda, m \mid n_0, \ldots, n_m)$ is as given in equation (11). If the candidate value is accepted, we assign $\theta^{(i)}$ equal to θ^* Otherwise, $\theta^{(i)}$ is set equal to $\theta^{(i-1)}$;

Step 2

We enter the second step of the ith iteration with values $\theta^{(i)}$, $\lambda^{(i-1)}$, and $m^{(i-1)}$. In this step, we update the value of λ by generating a candidate value λ^* from a proposal distribution indexed by $\lambda^{(i-1)}$ with density $q_\lambda(\lambda|\lambda^{(i-1)})$. The candidate value is accepted with probability

$$
min\left(1, \frac{p(\theta^{(i)}, \lambda^*, m^{(i-1)} \mid n_0, \ldots, n_m)\, q_\lambda(\lambda^{(i-1)} \mid \lambda^*)}{p(\theta^{(i)}, \lambda^{(i-1)}, m^{(i-1)} \mid n_0, \ldots, n_m)\, q_\lambda(\lambda^* \mid \lambda^{(i-1)})}\right), \qquad (14)
$$

If the candidate value is accepted, we assign $\lambda^{(i)}$ equal to λ^*. Otherwise, $\lambda^{(i)}$ is set equal to $\lambda^{(i-1)}$,

Step 3
We enter the third and last step of the ith iteration with values $\theta^{(i)}$, $\lambda^{(i)}$, and $m^{(i-1)}$. In this step, we update the value of m by generating a candidate value m^* from a proposal distribution with density $q_m(m|m^{(i-1)})$. The candidate value is accepted with probability

$$min\left(1, \frac{p(\theta^{(i)}, \lambda^{(i)}, m^* \mid n_0, \ldots, n_m)\, q_m(m^{(i-1)} \mid m^*)}{p(\theta^{(i)}, \lambda^{(i)}, m^{(i-1)} \mid n_0, \ldots, n_m)\, q_m(m^* \mid m^{(i-1)})}\right), \tag{15}$$

If the candidate value is accepted, we assign $m^{(i)}$ equal to m^*. Otherwise, $m^{(i)}$ is set equal to $m^{(i-1)}$. This concludes the third step of the ith iteration, and we exit from it with the updated values $\theta^{(i)}$, $\lambda^{(i)}$, and $m^{(i)}$.

The specification of the proposal distributions $q_\theta(.|.)$, $q_\lambda(.|.)$, and $q_m(\,|\,)$ appearing in the steps above still remains. This is discussed in Section 5. It should be emphasized that the algorithm given above describes only one possible implementation of the single-component Metropolis-Hastings algorithm. A fuller discussion of this algorithm and other MCMC methods will not be presented at this time, since several such discussions are readily available in the texts by Carlin and Louis (1996, Section 5.4), Tanner (1996, Chapter 6), and Gelman, Carlin, Stern, and Rubin (1995, Chapter 11). Within the actuarial literature, the recent articles by Haastrup and Arjas (1996) and Scollnik (1995d) may prove instructive to a reader unfamiliar with these methods. Also, Pai (1997) discusses the use of MCMC to perform a Bayesian analysis to scrutinize the compound loss distribution.

5 Numerical Illustration

The data we analyse is taken from Adelstein (1949, p. 379) and gives the observed number of accidents in the age-group 26-30 years during the first year of service for a group of railyard shunters. The data appears in Table 1, and is underdispersed with a sample mean of 0 5815 and a sample variance of 0 5719. Consul and Famoye (1989) previously fit a truncated GPD model to this data and obtained the ML estimates $\hat{\theta} = 0.6115$ and $\hat{\lambda} = -0.0676$. However, Consul and Famoye (1989) proceeded by grouping the last three of the class frequencies appearing in Table 1 into a single class of frequencies greater than or equal to 4 and also appear to have set $m = 4$ for the purposes of estimation even though one worker experienced 6 accidents. Consequently, their ML estimates are adversely affected. Our own analysis will use the original form of the data presented by Adelstein.

TABLE I

ADELSTEIN'S (1949) SHUNTERS ACCIDENTS DATA
FIRST YEAR OF SHUNTING AGE 26-30 YEARS

Number of Accidents	*Number of Men*
0	121
1	85
2	19
3	1
4	0
5	0
6	1

We proceed to analyse Adelstein's data using the truncated GPD model along with each PDS introduced in Section 3. We utilise the MCMC method described in Section 4. A few specifics concerning the implementation of the Markov chain are worthy of note. For the univariate proposal distributions associated with the parameters θ and λ, we found that normal distributions centered at the current value of the parameter in question and with standard deviation of 0.05, that is

$$q_\theta(\theta|s) \sim Normal\ (s, 0.05) \quad \text{and} \quad q_\lambda(\lambda|s) \sim Normal\ (s, 0.05),$$

yielded acceptance rates in the 50 to 75 per cent range. The proposal distribution for the parameter m was taken to be *Poisson* with mean μ in the case of the analysis incorporating the second PDS, that is

$$q_m(m|s) \equiv q_m(m) \sim Poisson\ (\mu)\ .$$

This makes Step 3 of the algorithm an independence sampler (Gilks, Richardson, and Spiegelhalter, 1996, page 9; also, Tierney, 1994, page 1706) since $q_m(m|m^{(t-1)})$ no longer depends on the value of $m^{(t-1)}$. For the analysis incorporating the first PDS, exact draws of m from its full conditional posterior distribution were used. In this case, the acceptance probability (15) is always equal to 1. For the analysis incorporating the third PDS, no draws of m were required since this analysis assumed that the value of m was fixed and known.

For each analysis, four realizations of a Markov chain were simulated Each chain was permitted to run for 10,000 iterations. The results of the first 5,000 iterations were discarded as 'burn-in', and convergence of the Markov chains for each analysis was formally monitored by applying the diagnostic of Gelman, Carlin, Stern, and Rubin (1995, page 330-333) to the output of iterations 5001 through 10,000. The behaviour of the realised Markov chain sample paths associated with one of the simulations (corresponding to the second PDS with $\mu = 10$) is illustrated in Figures 2, 3 and 4. In these plots, it is apparent that the simulated Markov chains are well on their way towards convergence by the 100th iteration in each case. Estimated posterior distributions for the parameters θ, λ, and m are presented in Figures 5, 6 and 7. These posterior distributions are estimated on the basis of the 20,000 (4 times 5,000) simulated draws for each parameter from its posterior distribution.

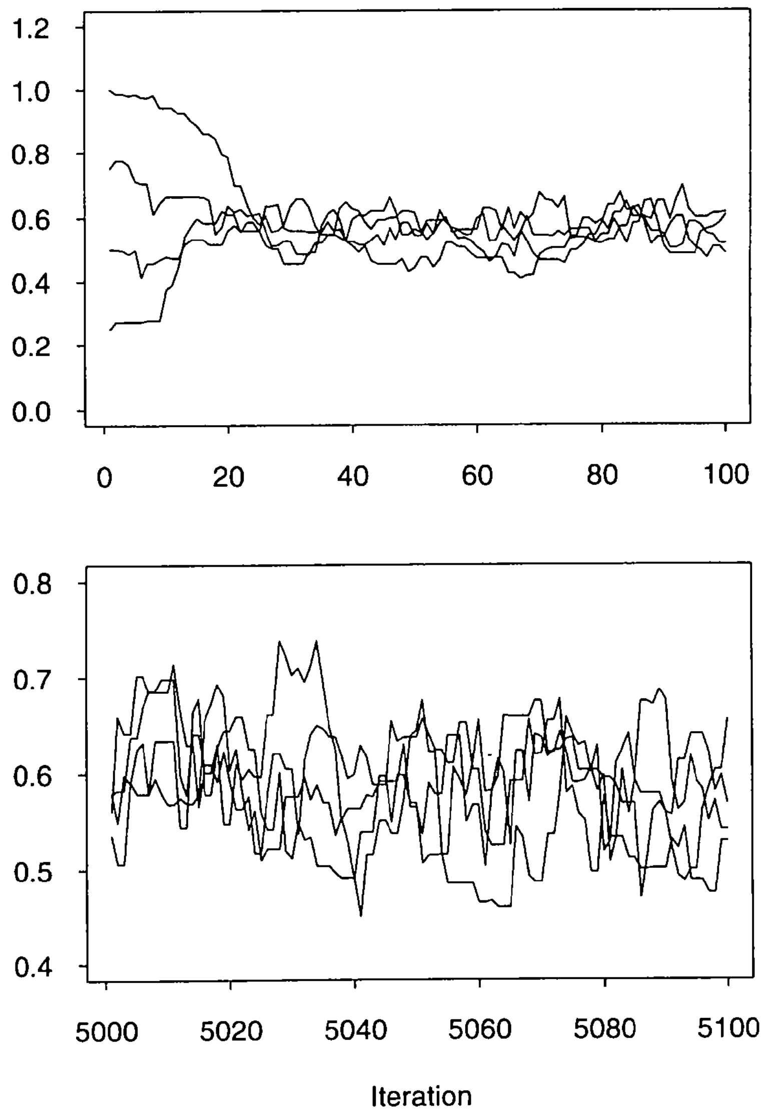

FIGURE 1 Sample Paths for the Parameter θ
Iterations 1 to 100 and 5000 to 5100
(Second PDS with μ = 10)

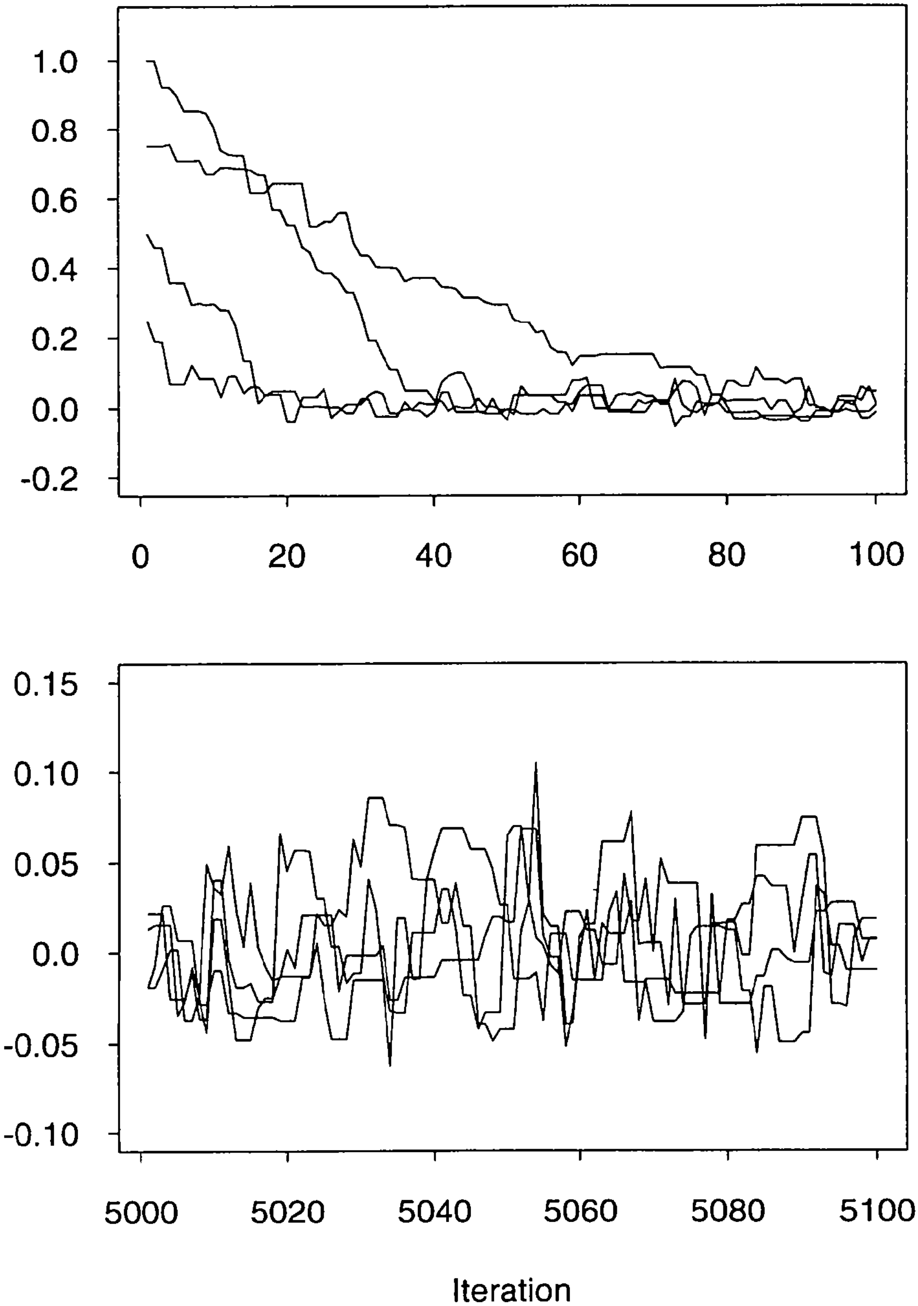

FIGURE 2 Sample Paths for the Parameter λ
Iterations 1 to 100 and 5000 to 5100
(Second PDS with μ = 10)

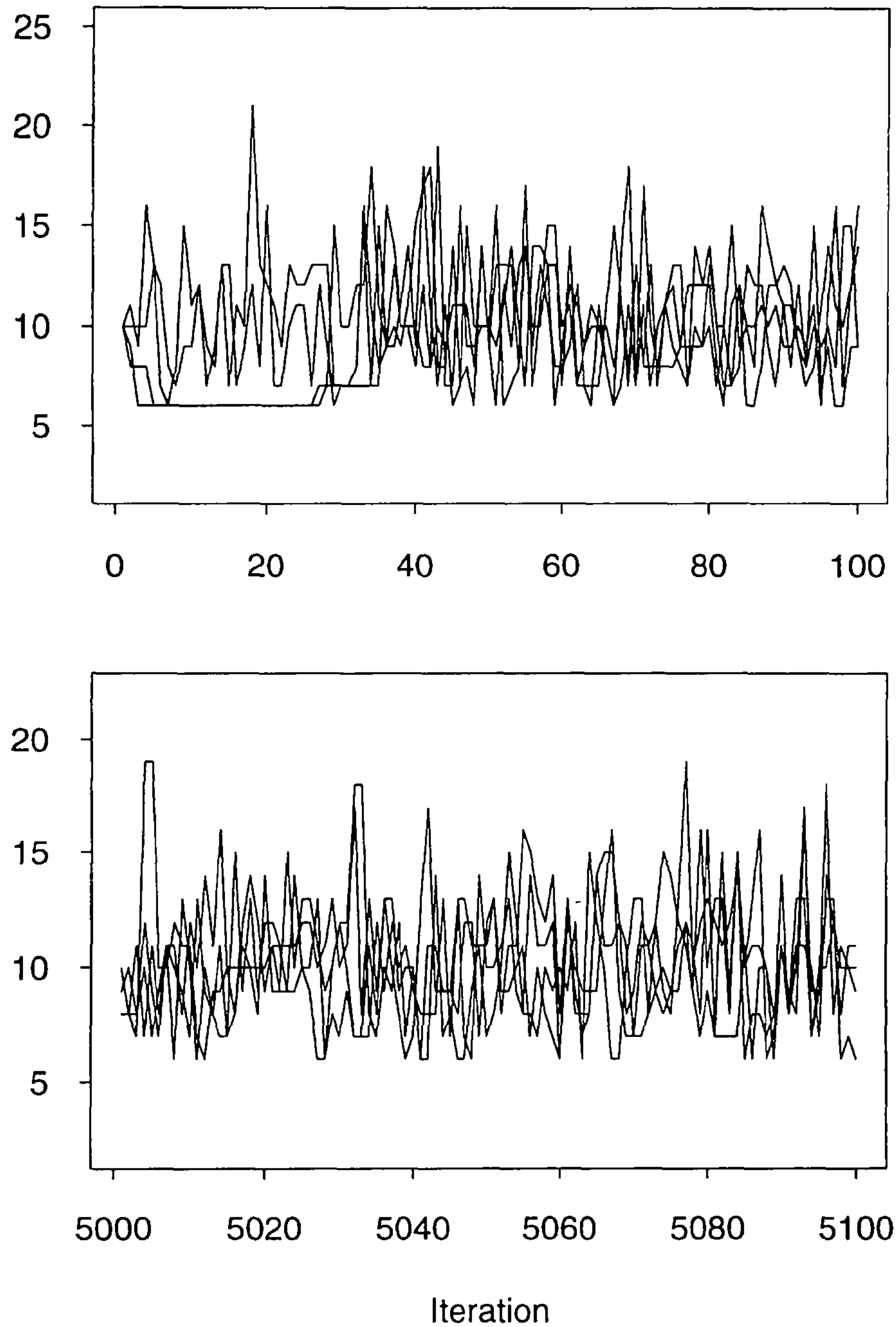

FIGURE 3 Sample Paths for the Parameter m
Iterations 1 to 100 and 5000 to 5100
(Second PDS with $\mu = 10$)

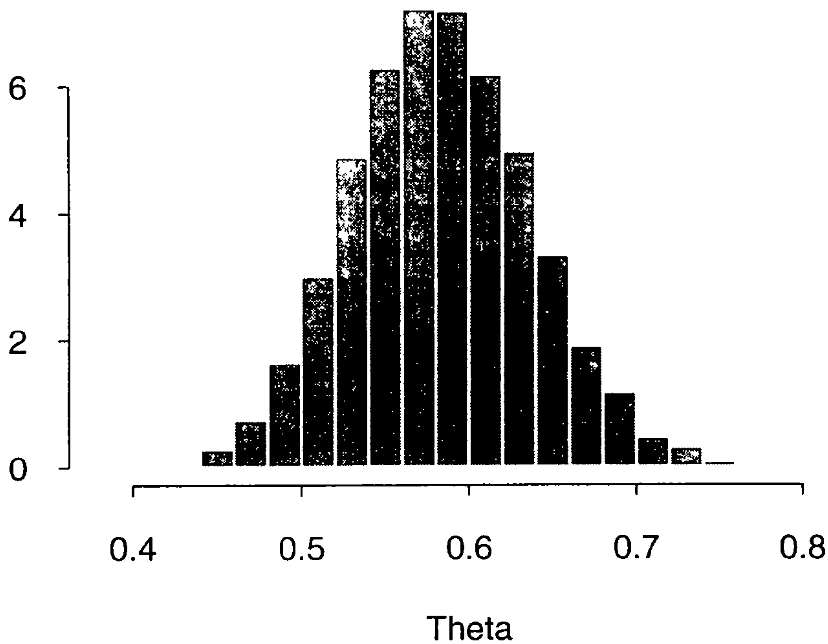

FIGURE 4 Estimated Posterior Density Functions for the Parameter θ (Second PDS with $\mu = 10$)

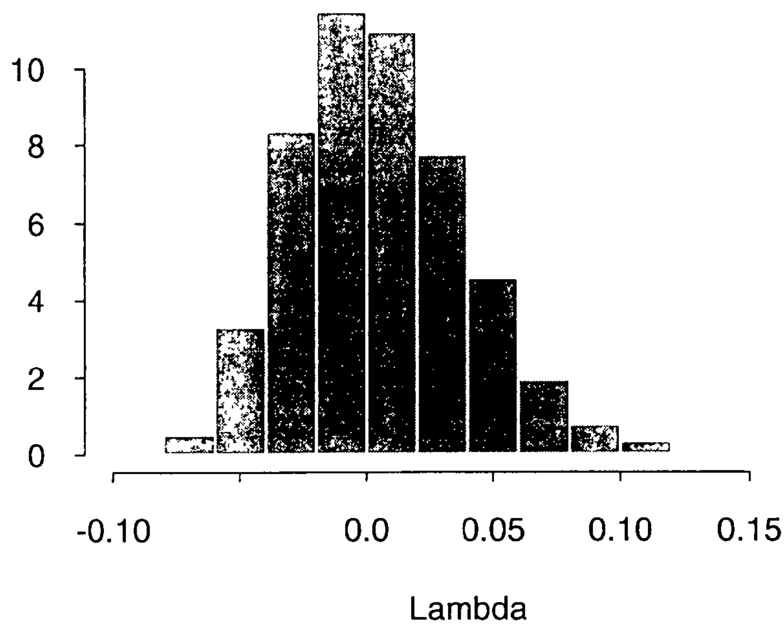

FIGURE 5 Estimated Posterior Density Functions for the Parameter λ (Second PDS with $\mu = 10$)

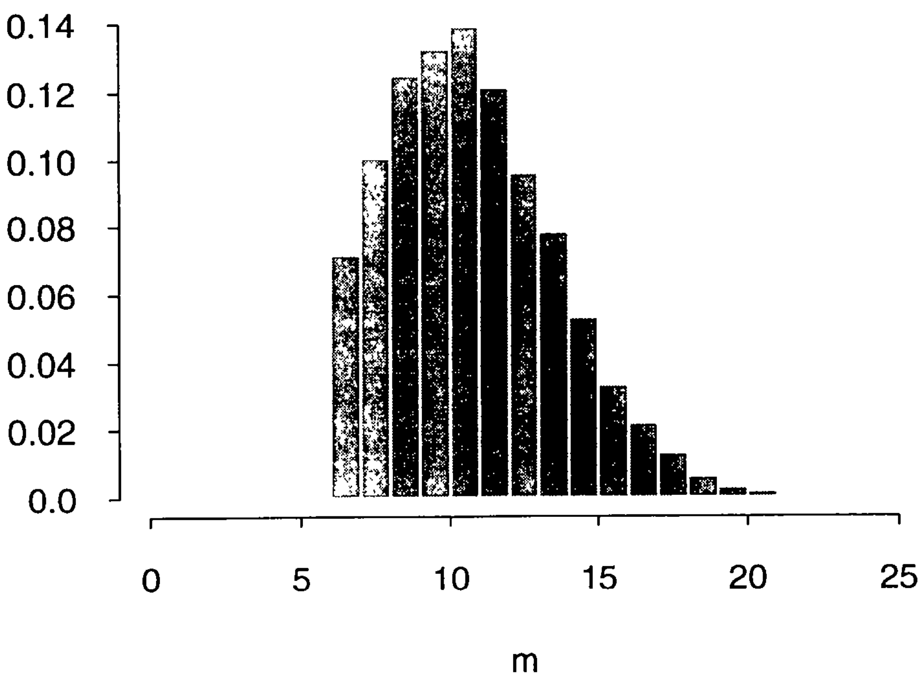

FIGURE 6 Estimated Posterior Density Functions for the Parameter m (Second PDS with μ = 10)

Summary results for all of our analyses appear in Tables 2 through 7. From Tables 3, 5, and 7, one can observe that predictive inferences are largely unaffected by the particular choice of PDS.

TABLE 2

ESTIMATED POSTERIOR MEANS AND SDs FOR THE PARAMETERS θ, λ, AND m RESULTING UNDER THE FIRST PDS FOR 3 VALUES OF M ($m \leq M$)

Parameter	$M = 6$	$M = 10$	$M = 25$
θ	0 5837 (0 0556)	0 5861 (0 0541)	0 5807 (0 0536)
λ	0 0034 (0 0353)	0 0009 (0 0340)	0 0085 (0 0316)
m	6 (0)	8 0016 (1 4196)	14 9864 (5 7347)

TABLE 3

THE ESTIMATED PREDICTIVE DISTRIBUTION $Pr(N_f = n_f | n_0, \ldots, n_m)$
RESULTING UNDER THE FIRST PDS FOR 3 VALUES OF M ($m \leq M$)

n_f	$M = 6$	$M = 10$	$M = 25$
0	0 5587	0 5573	0 5603
1	0 3237	0 3251	0 3213
2	0 0951	0 0953	0 0952
3	0 0190	0 0189	0 0194
4	0 0030	0 0029	0 0032
5	0 0004	0 0004	0 0005
≥ 6	0 0001	0 0001	0 0001
mean	0 5857	0 5867	0 5858
(*sd*)	(0 7711)	(0 7697)	(0 7749)

TABLE 4

ESTIMATED POSTERIOR MEANS AND SDs FOR THE PARAMETERS θ, λ, AND m
RESULTING UNDER THE SECOND PDS FOR 4 DIFFERENT VALUES OF μ

Parameter	$\mu = 2$	$\mu = 5$	$\mu = 10$	$\mu = 25$
θ	0 5810	0 5831	0 5828	0 5774
	(0 0529)	(0 0544)	(0 0531)	(0 0529)
λ	0 0034	0 0032	0 0051	0 0150
	(0 0346)	(0 0350)	(0 0335)	(0 0291)
m	6 32473	7 2713	10 3290	24 6669
	(0 6319)	(1 4242)	(2 7903)	(5 0021)

TABLE 5

THE ESTIMATED PREDICTIVE DISTRIBUTION $Pr(N_f = n_f | n_0, \ldots, n_m)$
RESULTING UNDER THE SECOND PDS FOR 4 DIFFERENT VALUES OF μ

n_f	$\mu = 2$	$\mu = 5$	$\mu = 10$	$\mu = 25$
0	0 5602	0 5590	0 5591	0 5621
1	0 3231	0 3236	0 3230	0 3184
2	0 0945	0 0949	0 0951	0 0953
3	0 0188	0 0189	0 0192	0 0201
4	0 0030	0 0030	0 0031	0 0034
5	0 0004	0 0004	0 0004	0 0005
≥ 6	0 0001	0 0001	0 0001	0 0001
mean	0 5830	0 5849	0 5858	0 5863
(*sd*)	(0 7692)	(0 7704)	(0 7724)	(0 7803)

TABLE 6

ESTIMATED POSTERIOR MEANS AND SDs FOR THE PARAMETERS θ, λ, AND m
RESULTING UNDER THE THIRD PDS FOR 3 VALUES OF $M(m = M)$

Parameter	$M = 6$	$M = 10$	$M = 25$
θ	0 5844	0 5838	0 5748
	(0 0532)	(0 0545)	(0 0511)
λ	0 0025	0 0029	0 0160
	(0 0343)	(0 0333)	(0 0279)
m	6	10	25
	(0)	(0)	(0)

TABLE 7

THE ESTIMATED PREDICTIVE DISTRIBUTION $Pr(N_f = n_f | n_0, \ldots, n_m)$
RESULTING UNDER THE THIRD PDS FOR 3 VALUES OF $M(m = M)$

n_f	$M = 6$	$M = 10$	$M = 25$
0	0 5583	0 5586	0 5636
1	0 3243	0 3238	0 3175
2	0 0951	0 0951	0 0948
3	0 0189	0 0189	0 0200
4	0 0030	0 0030	0 0034
5	0 0004	0 0004	0 0005
≥ 6	0 0001	0 0001	0 0001
mean	0 5857	0 5854	0 5842
(*sd*)	(0 7700)	(0 7703)	(0 7796)

6. FUTURE RESEARCH THE CORRELATED TRUNCATED BGPD MODEL

Famoye and Consul (1995) and Vernic (1997) have both considered a BGPD (bivariate GPD) formed by applying the method of trivariate reduction. This method proceeds as follows: let N_1, N_2 and N_3 be independent GPD random variables with respective parameters (θ_1, λ_1), (θ_2, λ_2), and (θ_3, λ_3). Then the random vector (X, Y) is said to have a correlated BGPD if $X = N_1 + N_2$ and $Y = N_2 + N_3$. Unfortunately, both Famoye and Consul (1995) and Vernic (1997) implicitly permit the parameters λ_i, $i = 1, 2, 3$, to take on negative values but fail to correct the definitions of the affected GPD and BGPD distributions by appropriately truncating and normalizing them.

In order to correct this problem, we define a correlated truncated BGPD by the method of trivariate reduction. Let N_1, N_2 and N_3 be independent truncated GPD random variables with respective parameters $(\theta_1, \lambda_1, m_1)$, $(\theta_2, \lambda_2, m_2)$, and $(\theta_3, \lambda_3, m_3)$. Then the random vector (X, Y) will be said to have a correlated truncated BGPD if $X = N_1 + N_2$ and $Y = N_2 + N_3$ as before. It should be possible to implement Bayesian posterior and predictive inferences for this distribution by using an extension of the MCMC method described in Sections 3 and 4 along with a data augmentation method to simulate the unobserved values of N_1, N_2 and N_3, given the observations X and Y along with the current simulated values of the parameters $(\theta_i, \lambda_i, m_i)$, $i = 1, 2, 3$. This procedure will be further explained, and also applied to a numerical example, in a paper to follow.

ACKNOWLEDGEMENTS

The author wishes to thank the referees for their comments and suggestions. This research was partially supported by a grant from the Natural Sciences and Research Council of Canada.

REFERENCES

ADELSTEIN, A M (1949) Accident Proneness A Criticism of the Concept Based Upon an Analysis of Shunters' Accidents *Journal of the Royal Statistical Society*, A **115**, 354-410

ALZAID, A A , and AL-OSH, M.A (1993) Some Autoregressive Moving Average Processes with Generalized Poisson Marginal Distributions *Ann Inst Statist Math* **45**, 223-232

AMBAGASPITIYA, R S and BALAKRISHNAN, N (1994) On the Compound Generalized Poisson Distributions *ASTIN Bulletin* **24** (2), 255-263

CARLIN, B P , and LOUIS, T A (1996) *Bayes and Empirical Bayes Methods for Data Analysis* Chapman and Hall, New York

CONSUL, P C (1994) Some Bivariate Families of Lagrangian Probability Distributions *Communications in Statistics Theory and Method* **23**, 2895-2906

CONSUL, P C (1990) A Model for Distribution of Injuries in Auto-Accidents *Mitteilungen der Schweiz Vereinigung der Versicherungsmathematiker* **1**, 161-168

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

CONSUL, P C and FAMOYE, F (1989) The Truncated Generalized Poisson Distribution and its Estimation *Communications in Statistics Theory and Method* **18** (10), 3635-3648

CONSUL, P C , and SHOUKRI, M M (1985) The Generalized Poisson Distribution when the Sample Mean is Larger than the Sample Variance *Communications in Statistics Simulation and Computation* **14** (3), 667-681

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

FAMOYE, F and CONSUL, P C (1995) Bivariate Generalized Poisson Distribution with Some Applications *Metrika* **42**, 127-138

GELMAN, A , CARLIN, J B , STERN, H S , and RUBIN, D B (1995) *Bayesian Data Analysis* Chapman and Hall, New York

GILKS, W R , RICHARDSON, S and SPIEGELHALTER, D J (1996) Introducing Markov Chain Monte Carlo In *Markov Chain Monte Carlo in Practice* Edited by Gilks, W R , Richardson, S , and Spiegelhalter, D J Chapman and Hall, New York

GOOVAERTS, M J , and KAAS, R (1991) Evaluating Compound Generalized Poisson Distributions Recursively *ASTIN Bulletin* **21**, 193-197

HAASTRUP, S , and ARJAS, E (1996) Claims Reserving in Continuous Time a Nonparametric Bayesian Approach *ASTIN Bulletin* **26** (2), 139-164

HENZE, N and KLAR, B (1995) Bootstrap Based Goodness of Fit Tests for the Generalized Poisson Model *Communications in Statistics Theory and Method* **24** (7), 1875-1896

HESSELAGER, O (1997) Recursions for a class of compound Lagrangian distributions *Swiss Association of Actuaries Bulletin* **1**, 95-101

JOHNSON, N L , KOTZ, S , and KEMP, A W (1992) *Univariate Discrete Distributions* Second Edition John Wiley & Sons Inc, New York

KLUGMAN, S A , PANJER, H H , and WILMOT, G E (1997) *Loss Models From Data to Decisions* In press

NELSON, D L (1975) Some Remarks on Generalizations of the Negative Binomial and Poisson Distributions *Technometrics* **17**, 135-136

PAI, J S (1997) Bayesian Analysis of Compound Loss Distributions *Journal of Econometrics* **79**, 129-146

SCOLLNIK, D P M (1995a) The Bayesian Analysis of Generalized Poisson Models for Claim Frequency Data Utilising Markov Chain Monte Carlo Methods *Actuarial Research Clearing House* **1995.1**, 339-356

SCOLLNIK, D P M (1995b) Bayesian Analysis of Two Overdispersed Poisson Models *Biometrics* **51**, 1117-1126

SCOLLNIK, D P M (1995c) Bayesian Analysis of Two Overdispersed Poisson Regression Models *Communications in Statistics Theory and Method* **24** (11), 2901-2918

SCOLLNIK, D P M (1995d) Simulating Random Variates from Makeham's Distribution and from Others with Exact or Nearly Log-concave Densities (with discussion) *Transactions of the Society of Actuaries* **1995** XLVII, 409-454

TANNER, M A (1996) *Tools for Statistical Inference Methods for the Exploration of Posterior Distributions and Likelihood Functions* Third Edition Springer-Verlag, New York

TIERNEY, L (1994) Markov Chains for Exploring Posterior Distributions (with discussion). *Annals of Statistics* **22**, 1701-1762

VERNIC, R (1997) On the Bivariate Generalized Poisson Distribution *Astin Bulletin* **27** (1), 23-31

DAVID P.M. SCOLLNIK
Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta, Canada
scollnik@acs.ucalgary.ca

A NOTE ON THE NET PREMIUM FOR A GENERALIZED LARGEST CLAIMS REINSURANCE COVER

By RAOUL M BERGLUND

Department of Mathematics
Åbo Akademi University

ABSTRACT

In the present paper the author gives net premium formulae for a generalized largest claims reinsurance cover If the claim sizes are mutually independent and identically 3-parametric Pareto distributed and the number of claims has a Poisson, binomial or negative binomial distribution, formulae are given from which numerical values can easily be obtained The results are based on identities for compounded order statistics.

KEYWORDS

Net premium; Reinsurance, LCR; ECOMOR, Compounded order statistics

1 INTRODUCTION

An expression for the pure premium for the largest claim reinsurance cover was already introduced by AMMETER (1964a) and for the p largest claims reinsurance cover by AMMETER (1964b) Simple formulae were presented under the assumptions that the claim sizes obeyed a one parametric Pareto distribution and the number of claims was Poisson distributed For the same claim size distribution KUPPER (1971) gave a formula for the largest claim reinsurance when the number of claims was geometrically distributed and CIMINELLI (1976) considered a negative binomial distribution BERLINGER (1972) extended the results by AMMETER and deduced the variance for the p largest claims reinsurance cover Net premium for a general claim size and claim number distribution was given by KREMER (1985) and for some generalized claim number distributions and a general claim size distribution by KREMER (1988a). The results in the latter were, however, not so practical for a specific claim size distribution. The author of this paper gives net premium formulae for a generalized largest claims reinsurance cover, assuming that the claim sizes are mutually independent and identically 3-parametric Pareto distributed and when the number of claims has a Poisson, binomial or negative binomial distribution The formulae presented in this paper are simple and easily calculated.

2 PRELIMINARIES

From now on, let $X_1, X_2, \ldots, X_N$ denote non-negative, mutually independent and identically distributed claim sizes, which are independent of the number of claims N that occur in a given time period. Denote by

$$X_{N\,1} \geq X_{N\,2} \geq \ldots \geq X_{N\,N}$$

the claims ordered in a decreasing size. The i-th largest claim is called the i-th ordered claim or more generally the i-th compounded order statistic Let

$$f_i\ [0,\infty) \to [0,\infty)$$

$(i \geq 1)$ be measurable functions, that satisfy

$$f_i(0) = 0 \text{ and } \sum_{i=1}^{n} f_i(y_i) \in \left[0, \sum_{i=1}^{n} y_i\right]$$

for all $0 \leq y_n \leq \ldots \leq y_2 \leq y_1$ This representation was first made by KREMER (1982) and the following main definition by KREMER (1984):

Definition. The reinsurance treaty defined by

$$R_N(X_{N\,1}, X_{N\,2}, \ldots, X_{N\,N}) = R_N = \sum_{i=1}^{N} f_i(X_{N\,i}),$$

which determins the reinsurers share of the total loss $\sum_{i=1}^{N} X_i$, is called a reinsurance treaty based on ordered claims

We are especially interested in the case

$$f_i(x) = a_i\, x,$$

where a_i, $i \geq 1$, are real constants. This reinsurance treaty is defined as the generalized largest claims cover (KREMER 1988b) We get for

$$a_1 = a_2 = \ldots = a_p = 1, a_i = 0\ \forall\ i > p$$

the so called LCR(p) treaty covering the p largest claims and for

$$a_1 = a_2 = \ldots = a_{p-1} = 1, a_p = 1 - p, a_i = 0\ \forall\ i > p$$

the so called ECOMOR(p) treaty covering all claims in excess of the p-th largest claim

We will subsequently use some special functions. The incomplete gamma function is defined as

$$\Gamma(a, x) = \int_0^x e^{-u} u^{a-1} du \qquad , a > 0,\ x \geq 0$$

and the complete gamma function as $\lim_{x\to\infty} \Gamma(a,x) = \Gamma(a)$ The incomplete beta function is defined as

$$B(a, b, x) = \int_0^x u^{a-1}(1-u)^{b-1}du = \int_0^{\frac{x}{1-x}} u^{a-1}(1+u)^{-(a+b)}\,du \quad ,a,b > 0, 0 \le x < 1$$

and the complete beta function as $\lim_{x\to 1} B(a,b,x) = B(a,b)$. The complete beta function and the complete gamma function are related by

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

3 FORMULAE FOR THE NET PREMIUM

The two most common risk loaded premium principles, the variance principle and the standard deviation principle, are based on the expectation and the variance of a certain risk For a generalized largest claims reinsurance cover the expectation is given by

$$E[R_N] = \sum_{i=1}^{\infty} a_i\, E[X_{N,i}]$$

and the variance by

$$Var[R_N] = \sum_{i=1}^{\infty} a_i^2 E[X_{N,i}^2] + 2\sum_{j=2}^{\infty}\sum_{i=1}^{j-1} a_i a_j E[X_{N,i} X_{N,j}] - \left(\sum_{i=1}^{\infty} a_i E[X_{N,i}]\right)^2$$

The following theorem is due to CIMINELLI (1976) and KREMER (1985), where

$$\phi(s) = \sum_{n=0}^{\infty} P(N=n)s^n$$

denotes the probability generating function of N, which is assumed to have derivatives $\phi^{(i)}$on (0,1) of each order $i \ge 1$

Theorem 1 If the claim sizes $X_1, X_2, \ldots, X_N$ have a continuous distribution function F the density function of the i-th ordered claim is given by

$$P(X_{N,i} = x) = \frac{1}{\Gamma(i)} f(x)[1-F(x)]^{i-1}\phi^{(i)}(F(x))$$

and the joint density function of the i-th and j-th ordered claims $(0 < i < j)$ is given by

$$P(X_{N,i} = x_i, X_{N,j} = x_j) = \frac{1}{\Gamma(i)\Gamma(j-i)}[1-F(x_i)]^{i-1}[F(x_i)-F(x_j)]^{j-i-1}\phi^{(j)}(F(x_j))f(x_i)f(x_j)$$

Theorem 2 If the claim sizes $X_1, X_2, \ldots, X_N$ have a continuous distribution function F the k-th moment around the origin of the i-th ordered claim is given by

$$E[X_{N,i}^k] = \frac{1}{\Gamma(i)} \int_0^1 F^{-1}(u)^k [1-u]^{i-1} \phi^{(i)}(u) du$$

and the expectation of the cross product of the i-th and j-th ordered claims $(0 < i < j)$ is given by

$$E[X_{N,i} X_{N,j}] =$$
$$\frac{1}{\Gamma(i)\Gamma(j-i)} \int_0^1 F^{-1}(v)(1-v)^{j-1} \phi^{(j)}(v) \int_0^1 F^{-1}(1-u(1-v)) u^{i-1} (1-u)^{j-i-1} du dv.$$

Proof The first part of the statement follows from theorem 1 after the substitution $u = F(x)$ For the second part we have for $0 < i < j$ and $0 \leq X_{N,j} \leq X_{N,i}$ that

$$E[X_{N,i} X_{N,j}] =$$
$$C \int_0^\infty \int_{x_j}^\infty x_i x_j [1 - F(x_i)]^{i-1} [F(x_i) - F(x_j)]^{j-i-1} \phi^{(j)}(F(x_j)) f(x_i) f(x_j)\, dx_i\, dx_j,$$

where

$$C = \frac{1}{\Gamma(i)\Gamma(j-i)}$$

After substituting $u = \frac{1-F(x_i)}{1-F(x_j)}$ and $v = F(x_j)$ we obtain

$$E[X_{N,i} X_{N,j}] =$$
$$C \int_0^1 \int_0^1 F^{-1}(1-u(1-v)) F^{-1}(v) u^{i-1} [1-v]^{j-1} [1-u]^{j-i-1} \phi^{(j)}(v) du dv =$$
$$C \int_0^1 F^{-1}(v) [1-v]^{j-1} \phi^{(j)}(v) \int_0^1 F^{-1}(1-u(1-v)) u^{i-1} [1-u]^{j-i-1} du dv$$

□

From now on we will focus on the case where the claim sizes are distributed according to the 3-parametric Pareto distribution

$$F(x) = 1 - \left(\frac{d+\beta}{x+\beta}\right)^\alpha \qquad x \geq d > 0. \tag{3.1}$$

where the parameters α, β and d satisfy $\alpha > 0$ and $b > -d$ The distribution (3 1) is the most used claim size distribution, especially if there is a possibility of large claims In the literature the 3-parametric Pareto distribution is sometimes also called the "shifted" Pareto distribution (RYTGAARD 1990) or the complete Pareto distribution (DAYKIN et al 1994). Since

$$F^{-1}(x) = \frac{d+\beta}{(1-x)^{\frac{1}{\alpha}}} - \beta$$

the expectations of theorem 2 becomes after binomial expansion and simplifications

$$E[X_{N,i}^k] = \frac{1}{\Gamma(i)} \sum_{h=0}^{k} \binom{k}{h} (d+\beta)^{k-h} (-\beta)^h \int_0^1 (1-u)^{i-\frac{k-h}{\alpha}-1} \phi^{(i)}(u) du$$

and for $\alpha > \frac{1}{i}$

$$E[X_{N,i} X_{N,j}] = \frac{1}{\Gamma(i)}$$

$$\left[A_1 \int_0^1 (1-v)^{j-\frac{2}{\alpha}-1} \phi^{(j)}(v) dv - A_2 \int_0^1 (1-v)^{j-\frac{1}{\alpha}-1} \phi^{(j)}(v) dv + A_3 \int_0^1 (1-v)^{j-1} \phi^{(j)}(v) dv \right]$$

where

$$A_1 = (d+\beta)^2 \frac{\Gamma(i-\frac{1}{\alpha})}{\Gamma(j-\frac{1}{\alpha})}$$

$$A_2 = \beta(d+\beta) \left[\frac{\Gamma(i)}{\Gamma(j)} + \frac{\Gamma(i-\frac{1}{\alpha})}{\Gamma(j-\frac{1}{\alpha})} \right]$$

$$A_3 = \beta^2 \frac{\Gamma(i)}{\Gamma(j)}$$

The restriction on the parameter α is needed to get a finite expression Assuming further that the number of claims N is Poisson distributed

$$P(N=n) = \frac{\lambda^n}{n!} e^{-\lambda} \qquad \lambda > 0, n \geq 0, \tag{3 2}$$

negative binomially distributed

$$P(N=n) = \frac{\Gamma(r+n)}{\Gamma(r) n!} \left(\frac{1}{1+\lambda} \right)^r \left(\frac{\lambda}{1+\lambda} \right)^n \quad r, \lambda > 0, n \geq 0 \tag{3 3}$$

or binomially distributed

$$P(N=n) = \binom{m}{n} q^n (1-q)^{m-n} \quad 0 \le q \le 1,\ n = 0, 1, \ \ , m, \tag{3.4}$$

where m is a non-negative integer, we have the following corollaries

Corollary 3. Assume that the claim sizes $X_1, X_2, \ \ ., X_N$ are Pareto distributed (3 1) and that the claim number N is Poisson distributed (3 2) Then the k-th moment around the origin of the i-th ordered claim is, for $\alpha > \frac{k}{i}$ given by

$$E[X_{N\ i}^k] = \frac{1}{\Gamma(i)} \sum_{h=0}^{k} \binom{k}{h} (d+\beta)^{k-h} (-\beta)^h \lambda^{\frac{k-h}{\alpha}} \Gamma(i - \frac{k-h}{\alpha}, \lambda)$$

and the expectation of the cross product of the i-th and j-th ordered claims $(0 < i < j)$ is, for $\alpha > \max\{\frac{1}{i}, \frac{2}{j}\}$, given by

$$E[X_{N\ i} X_{N\ j}] = \frac{1}{\Gamma(i)} \left[A_1 \lambda^{\frac{2}{\alpha}} \Gamma(j - \tfrac{2}{\alpha}, \lambda) - A_2 \lambda^{\frac{1}{\alpha}} \Gamma(j - \tfrac{1}{\alpha}, \lambda) + A_3 \Gamma(j, \lambda) \right]$$

Proof Since the j-th derivative of the probability generating function ϕ for a Poisson distributed random variable (3 2) is given by

$$\phi^{(j)}(s) = \lambda^j e^{\lambda(s-1)}$$

we have, for $\gamma > 0$, that

$$\int_0^1 (1-u)^{\gamma-1} \phi^{(j)}(u) du = \lambda^j \int_0^1 (1-u)^{\gamma-1} e^{\lambda(u-1)} du$$

After the substitution $t = \lambda(1-u)$ we obtain

$$\int_0^1 (1-u)^{\gamma-1} \phi^{(j)}(u) du = \lambda^{j-\gamma} \int_0^\lambda t^{\gamma-1} e^{-t} dt$$

$$= \lambda^{j-\gamma} \Gamma(\gamma, \lambda).$$

which gives the result.

□

Corollary 4 Assume that the claim sizes X_1, X_2 $\ \ $, X_N are Pareto distributed (3.1) and that the claim number N is negative binomially distributed (3 3). Then the k-th moment around the origin of the i-th ordered claim is, for $\alpha > \frac{k}{i}$, given by

$$E[X_{N:i}^k] = \frac{1}{B(i,r)} \sum_{h=0}^{k} \binom{k}{h} (d+\beta)^{k-h} (-\beta)^h \lambda^{\frac{k-h}{\alpha}} B(i - \tfrac{k-h}{\alpha}, r + \tfrac{k-h}{\alpha}; \tfrac{\lambda}{1+\lambda})$$

and the expectation of the cross product of the i-th and j-th ordered claims $(0 < i < j)$ is, for $\alpha > \max\{\frac{1}{i}, \frac{2}{j}\}$, given by

$$E[X_{N:i} X_{N:j}] = \frac{1}{B(j,r)} \frac{\Gamma(j)}{\Gamma(i)}$$

$$\left[A_1 \lambda^{\frac{2}{\alpha}} B(j - \tfrac{2}{\alpha}, r + \tfrac{2}{\alpha}, \tfrac{\lambda}{1+\lambda}) - A_2 \lambda^{\frac{1}{\alpha}} B(j - \tfrac{1}{\alpha}, r + \tfrac{1}{\alpha}, \tfrac{\lambda}{1+\lambda}) + A_3 B(j; i, \tfrac{\lambda}{1+\lambda})\right]$$

Proof Since the j-th derivative of the probability generating function ϕ for a negative binomially distributed random variable (3.3) is given by

$$\phi^{(j)}(s) = \frac{\Gamma(r+j)}{\Gamma(r)} \lambda^j [1 - \lambda(s-1)]^{-(r+j)}$$

we have, for $\gamma > 0$, that

$$\int_0^1 (1-u)^{\gamma-1} \phi^{(j)}(u) du = \frac{\Gamma(r+j)}{\Gamma(r)} \lambda^j \int_0^1 (1-u)^{\gamma-1} [1 - \lambda(u-1)]^{-(i+j)} du$$

After the substitution $t = \lambda(1-u)$ we obtain

$$\int_0^1 (1-u)^{\gamma-1} \phi^{(j)}(u) du = \frac{\Gamma(r+j)}{\Gamma(r)} \lambda^{j-\gamma} \int_0^{\lambda} t^{\gamma-1} (1+t)^{-(r+j)} dt$$

$$= \frac{\Gamma(r+j)}{\Gamma(r)} \lambda^{j-\gamma} B(\gamma; r + j - \gamma, \tfrac{\lambda}{1+\lambda}),$$

from which the result follows after simplification.

□

Corollary 5. Assume that the claim sizes $X_1, X_2, \ldots, X_N$ are Pareto distributed (3.1) and that the claim number N is binomially distributed (3.4). Then the k-th moment around the origin of the i-th ordered claim is, for $\alpha > \frac{k}{i}$, given by

$$E[X_{N:i}^k] = i \binom{m}{i} \sum_{h=0}^{k} \binom{k}{h} (d+\beta)^{k-h} (-\beta)^h q^{\frac{k-h}{\alpha}} B(i - \tfrac{k-h}{\alpha}, m - i + 1, q)$$

and the expectation of the cross product of the i-th and j-th ordered claims $(0 < i < j)$ is, for $\alpha > \max\{\frac{1}{i}, \frac{2}{j}\}$, given by

$$E[X_{N:i} X_{N:j}] = \binom{m}{j} \frac{\Gamma(j+1)}{\Gamma(i)} [A_1 q^{\frac{2}{\alpha}} B(j - \tfrac{2}{\alpha}, m - j + 1, q) - A_2 q^{\frac{1}{\alpha}} B(j - \tfrac{1}{\alpha}, m - j + 1; q)$$
$$+ A_3 B(j; m - j + 1, q)]$$

Proof Since the j-th derivative of the probability generating function ϕ for a binomially distributed random variable (3 4) is given by

$$\phi^{(j)}(s) = \frac{\Gamma(m+1)}{\Gamma(m-j+1)} q^j [qs + 1 - q]^{m-j} \quad j \leq m,$$

we have, for $\gamma > 0$, that

$$\int_0^1 (1-u)^{\gamma-1} \phi^{(j)}(u) du = \frac{\Gamma(m+1)}{\Gamma(m-j+1)} q^j \int_0^1 (1-u)^{\gamma-1} (qu + 1 - q)^{m-j} du$$

After the substitution $t = q(1-u)$ we obtain

$$\int_0^1 (1-u)^{\gamma-1} \phi^{(j)}(u) du = \frac{\Gamma(m+1)}{\Gamma(m-j+1)} q^{j-\gamma} \int_0^q t^{\gamma-1} (1-t)^{m-j} dt$$
$$= \frac{\Gamma(m+1)}{\Gamma(m-j+1)} q^{j-\gamma} B(\gamma, m-j+1, q),$$

from which the result follows after simplification

□

If $0 < \alpha < 1$, which indicates a very heavy tailed distribution, we have according to the results above that the first moment around origin of a certain number of the largest ordered claims does not exist We could therefore consider the number of ordered claims, for which the first moment around the origin does not exist, as a measure for how dangerous a Pareto distribution is Since many computer programs have built-in routines for computing the complete gamma, incomplete gamma and the incomplete beta function, the expectations in results above can be calculated easily

If the claim sizes obey an exponential distribution

$$F(x) = 1 - e^{-\beta(x-a)} \quad \beta > 0, \; x \geq a,$$

we cannot get useful expressions for the moments around the origin and the cross product by applying theorem 2 Using well known results from order statistics for a deterministic number of claims (DAVID 1970) and then the iterativity of the expectation operator, expression for the pure premium can be constructed. Exponentially distributed claim sizes have been studied by KUPPER (1971) and KREMER (1985 and 1986).

4 A NUMERICAL EXAMPLE

Let the distribution for the claim sizes be Pareto distributed (3.1) with $d = 0$ For the insurance line under consideration the method of moments gives the following parameter estimates $\hat{\alpha} = 2.3401$ and $\hat{\beta} = 13692$. Since the most import claim number distributions are the Poisson and the negative binomial, we will restrict the example to them. Using the same estimation method we have the following parameter estimates. Poisson $\hat{\lambda} = 79\,667$, negative binomial $\hat{\lambda} = 1\,0865$ and $\hat{r} = 73\,326$ We have the following numerical results

Expectation of LCR(p) and ECOMOR(p) treaties

p	Poisson	negative binomial	p	Poisson	negative binomial
1	124 597	124 368	1	0	0
2	190 099	189 738	2	59 095	58 997
3	238 679	238 215	3	92 937	92 783
4	278 390	277 837	4	119 548	119 350
5	312 395	311 763	5	142 369	142 133
LCR(p)-treaty			ECOMOR (p)-treaty		

Standard deviation of LCR(p) and ECOMOR(p) treaties

p	Poisson	negative binomial	p	Poisson	negative binomial
1	178 069	178 129	1	0	0
2	191 632	191 860	2	134 587	134 549
3	198 847	199 254	3	182 222	182 206
4	203 797	204 389	4	188 799	188 815
5	207 581	208 363	5	193 255	193 405
LCR(p)-treaty			ECOMOR (p)-treaty		

The difference between the numerical values for Poisson and the negative binomial cases is quite small. If we assume that in the incomplete beta function b is large and a is bounded we have the following asymptotic representation (ABRAMOWITZ and STEGUN 1972)

$$\frac{B(a;b;x)}{B(a;b)} = \frac{\Gamma\left(a, \frac{x(2b+a-1)}{2-x}\right)}{\Gamma(a)} + O(b^{-2}).$$

This explains the similarity in the numerical results above. This suggests, that the Poisson distribution might be the right claim number model if the parameter value r is large and λ is small in the negative binomial distribution

REFERENCES

ABRAMOWITZ, M and STEGUN, I (1972) *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables* Dover, New York

AMMETER, H (1964a) Note concerning the distribution function of the total loss excluding the largest individual claims *ASTIN Bulletin* **3**, 132-143

AMMETER, H (1964b) The rating of "largest claim" reinsurance covers *Quarterly letter from the Allgemeene Reinsurance Companies, Jubilee number* **9**, 5-17

BERLINGER, B (1972) Largest claims reinsurance (LCR) A quick method to calculate LCR-risk rates from excess of loss risk rates *ASTIN Bulletin* **10**, 54-58

CIMINELLI, E (1976) On the distribution of the highest claims and its application to the automobile insurance liability *Transactions of the 20th International Congress of Actuaries*, 501-517

DAVID, H A (1970) *Order Statistics* John Wiley & Sons, New York

DAYKIN, C D, PENTIKÄINEN, T and PESONEN, M (1994) *Practical Risk Theory for Actuaries* Chapman & Hall, London

KREMER, E (1982) Rating of largest claims and ECOMOR reinsurance treaties for large portfolios *ASTIN Bulletin* **13**, 47-56

KREMER, E (1984) An asymptotic formula for the net premium of some reinsurance treaties *Scandinavian Actuarial Journal* **1984**, 11-22

KREMER, E (1985) Finite formulae for the premium of the general reinsurance treaty based on ordered claims *Insurance Mathematics and Economics* **4**, 233-238

KREMER, E (1986) Simple formulas for the premiums of the LCR and ECOMOR treaties under exponential claim sizes *Blatter der Deutschen Gesellschaft fur Versicherungsmathematik* **17**, 457-469

KREMER, E (1988a) Ruckversicherungspramien unter Verallgemeinerten Schadenzahlverteilungen *Transactions of the 23rd International Congress of Actuaries* 73-79

KREMER, E (1988b) A general bound for the net premium of the largest claims reinsurance covers *ASTIN Bulletin* **18**, 69-78

KUPPER, J (1971) Contributions to the theory of the largest claim cover *ASTIN Bulletin* **6**, 134-146

RYTGAARD, M (1990) Estimation in the Pareto distribution *ASTIN Bulletin* **20**, 201-216

RAOUL M BERGLUND
Åbo Akademi University
Department of Mathematics
Fanriksgatan 3 B
20500 Åbo
Finland

BOOK REVIEWS

S.A KLUGMAN, H.H. PANJER and G E. WILLMOT (1998). *Loss Models. From Data to Decisions.* Wiley, New York.

Hogg and Klugmann (1984) gives an excellent introduction into the stochastic modeling of insurance losses. A key feature of that book is the attention given to the special character of insurance data. Rather than having embarked on a second (revised) edition, the present three authors have decided to write a new text, keeping the main ideas of Hogg and Klugman (1984), but adding numerous topics which every actuary, whether practicing or academic, ought to know

Loss Models "is organized around the principle that actuaries build models in order to analyze risks and make decisions about managing the risks based on conclusions drawn from the analysis". It is to be stressed that the text mainly looks at the liability side of insurance· the losses. These are put together in a global risk model where uncertainty may enter at the claim-size level (Chapter 2) and at the claim-frequency level (Chapter 3). Combining these two levels leads to an aggregate model (Chapter 4). The premium side of the coin is treated through credibility theory (Chapter 5). Long term stability questions are discussed via the classical limit theorems for ultimate ruin (Chapter 6).

So far, various existing texts present, at least from a chapter heading point of view, similar material Where are the novelties? First of all, this text is extensive in its 644 pages. That means that all of the above topics are treated in a fair amount of detail. Secondly, numerical examples together with accompanying exercises and case studies are abundant. On each topic introduced, the reader is asked to calculate actual numbers (i e take decisions) based on data Many of the exercises presented stem from actuarial examination papers. Answers to selected ones are given.

This brings me to the key question· "What is the intended readership?" As the book assumes no specific prerequisites beyond basic courses in linear algebra, analysis and elementary probability and statistics, the readership is broad. Anyone interested in acquiring the basic stochastic techniques which practicing actuaries use daily will find this text useful. The necessary statistical and probabilistic techniques are introduced if and when needed Computability is always a concern. no theory without numbers The style of writing is relaxed, yet also concise. A slight loss of conciseness is present towards the end of the text where basic results of Poisson processes and Brownian motion are derived· for instance the proof of the interarrival-time characterisation of the homogeneous Poisson process leaves the critical reader a bit in the cold when it comes to achieving independence (the usual step-by-step "proof"), also the reader could have benefited from a warning that the reflection principle for brownian motion (figuring on the cover!), though intuitively clear, needs a proof (strong Markov property). Similar warnings could have been made in the chapter on ruin theory. Also, I found the Index, and to some extend the References a bit wanting. These "flaws" however should not diminish my admiration for this book: it is a most useful addition to the actuarial literature Especially from the more applied, industrial side: If I were recruiting a new, young actuary of which I would know that he or she had a through knowledge of the material treated in **Loss Models**, I would be most glad As such, this book will no doubt become a classic reference.

REFERENCE

HOGG, R and KLUGMAN, S. (1984) *Loss Distributions*. Wiley, New York.

PAUL EMBRECHTS
ETH Zurich

THOMAS MACK (1997): *Schadenversicherungsmathematik*. Sonderauflage von Heft 28 der Schriftenreihe Angewandte Versicherungsmathematik der Deutschen Gesellschaft für Versicherungsmathematik e.V Verlag Versicherungswirtschaft e.V. Karlsruhe, 1997 IISN 0178-8116, ISBN 3-88487-582-5

The more I was reading in this book the more I got interested in it and at the same time I found it was a pity that it is not written in English because there is no doubt a great number of potential readers not mastering the German language sufficiently. So let's hope that it will soon be translated into English

Thomas Mack's present book on actuarial sciences in Non-Life insurance is subdivided into the four main parts

Part 1 Basics
Part 2 Pricing
Part 3 Reserving
Part 4 Risksharing

In the first part both the classical individual and collective model of risk theory are dealt with, complemented by a third approach where the portfolio is assumed to consist of a number of subportfolios in which each risk has the same claims degree distribution. In the same first chapter there is already a section on pricing where the author proposes the so-called covariance principle, i.e., the total security loading is distributed onto the individual risks proportionally to the covariance between the claims potential of that risk and the one of the entire portfolio. Furthermore there is an interesting part discussing the practically important fact that a company can still underwrite a certain share of a risk even if the total premium for it is less than what according to the company's standard would be required as a technical minimum.

In the second part on pricing there is at the beginning an extensive discussion on how to define-more or less homogenous-risk categories as a basis for the construction of a tariff Several statistical procedures are proposed for this like cluster analysis, maximum likelihood and minimum square procedures as well as some parametric approaches. Next comes credibility theory, experience rating and the construction of bonus malus systems followed by a small section on the truncation of large individual claims that distort the normal claims statistics.

Part three on claims reserves is visibly the chapter where the author could draw most from his vast practical experience. Among many other things also a credibility approach for assessing claims reserves is discussed here. But basically this chapter deals with three different statistical procedures, namely two non-parametric ones (one additive and the other multiplicative) and a parametric approach which is called "cross-classified". Although most of this chapter is very much practically oriented (last but not least, I think, because of the proposed separations "claims frequency/severity" on one hand and "IBNR/IBNER" on the other), there is this theorem on page 279 which is of remarkable theoretical interest and which would read in English:

"The maximum likelihood estimator for the claims reserve within the cross-classified Poisson model with positive increments is identical with the chain ladder reserve."

The last chapter is on risksharing, i e., on coinsurance, reinsurance and retrocession and right at the beginning the important distinction between proportional and non-proportional risksharing is made The chapter closes with some general observations on risk management and solvency.

I found reading in this book refreshing because of many original thoughts and approaches which are not commonly known and I would just like to express my hope again that it should be translated into English soon.

ERWIN STRAUB

GUIDELINES TO AUTHORS

1. Papers for publication should be sent in quadruplicate to one of the Editor

Paul Embrechts,
Department of Mathematics, ETH-Zentrum,
CH-8092 Zurich, Switzerland

or to one of the Co-Editors

René Schnieper,
Zurich Insurance Company,
P O Box, CH-8022 Zurich, Switzerland

Andrew Cairns
Department of Actuarial Mathematics
and Statistics
Heriot-Watt University
Edinburgh EH 14 4AS, United Kingdom

Submission of a paper is held to imply that it contains original unpublished work and is not being submitted for publication elsewhere

Receipt of the paper will be confirmed and followed by a refereeing process, which will take about three months

2. Manuscripts should be typewritten on one side of the paper, double-spaced with wide margins The basic elements of the journal's style have been agreed by the Editors and Publishers and should be clear from checking a recent issue of *ASTIN BULLETIN* If variations are felt necessary they should be clearly indicated on the manuscript

3. Papers should be written in English or in French Authors intending to submit longer papers (e g exceeding 30 pages) are advised to consider splitting their contribution into two or more shorter contributions

4. The first page of each paper should start with the title, the name(s) of the author(s), and an abstract of the paper as well as some major keywords An institutional affiliation can be placed between the name(s) of the author(s) and the abstract

5. Footnotes should be avoided as far as possible

6. Upon acceptance of a paper, any figures should be drawn in black ink on white paper in a form suitable for photographic reproduction with lettering of uniform size and sufficiently large to be legible when reduced to the final size

7. References should be arranged alphabetically, and for the same author chronologically Use a, b, c, etc to separate publications of the same author in the same year For journal references give author(s), year, title, journal (in italics, cf point 9), volume (in boldface, cf point 9), and pages For book references give author(s), year, title (in italics), publisher, and city

   Examples

   BARLOW, R E and PROSCHAN, F (1975) *Mathematical Theory of Reliability and Life Testing* Holt, Rinehart, and Winston, New York

   JEWELL, W S (1975a) Model variations in credibility theory In *Credibility Theory and Applications* (ed P M KAHN), pp 193–244, Academic Press, New York

   JEWELL, W S (1975b) Regularity conditions for exact credibility *ASTIN Bulletin* **8**, 336–341

   References in the text are given by the author's name followed by the year of publication (and possibly a letter) in parentheses

8. The address of at least one of the authors should be typed following the references

Continued overleaf

Guidelines to Authors *continued from inside back cover*

9. Italics (boldface) should be indicated by single (wavy) underlining Mathematical symbols will automatically be set in italics, and need not be underlined unless there is a possibility of misinterpretation
   Information helping to avoid misinterpretation may be listed on a separate sheet entitled 'special instructions to the printer' (Example of such an instruction Greek letters are indicated with green and script letters with brown underlining, using double underlining for capitals and single underlining for lower case)

10. Electronic Typesetting using Word Perfect 5 1 is available Authors who wish to use this possibility should ask one of the editors for detailed instructions

11. Authors will receive from the publisher two sets of page proofs together with the manuscript One corrected set of proofs plus the manuscript should be returned to the publisher within one week Authors may be charged for alterations to the original manuscript

12. Authors will receive 50 offprints free of charge Additional offprints may be ordered when returning corrected proofs A scale of charges will be enclosed when the proofs are sent out