Volume 27, No. 1 May 1997

ASTIN BULLETIN

A Journal of the International Actuarial Association

CONTENTS

Ceuterick

EDITORIAL POLICY

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

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

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

MEMBERSHIP

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

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

SUBSCRIPTION AND BACK ISSUES

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

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

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

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

INDEX TO VOLUMES 1-20

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

EDITORIAL

THE CHALLENGE FOR ASTIN IN THE 21st CENTURY

Perhaps I could start by mentioning two currently fashionable key phrases: "change management" and "teamwork". It is not my concern here to attribute precise meanings to these terms they are included as being indicative or symptomatic of underlying changes affecting the manner in which non-life insurance is being transacted at the end of the 20th century. Whilst it could be argued that the history of insurance is one of change, and that there is nothing new in the idea of teamwork, I think it is indisputable that, in Western Europe at least, change in the social and economic environments has forced a corresponding rate and depth of change in many aspects of insurance

To be specific, I need only refer to such developments as the burgeoning market in telesales insurance in the UK, with other countries variously following behind, the significant impact on the UK market of developments in mortgage related insurance; the problems which have beset Lloyds and, in a somewhat different vein, the stream of EC Directives not only having the effect of shaping internal markets, but introducing some degree of convergence between territories in aspects where diversity may have previously been the norm

Other developments include changes in solvency testing in the US, the securitisation of insurance risks and the increasing prominence given to linking risk arising from both insurance and its supporting assets.

Accompanying what might be regarded as market changes of this kind, the continuing evolution of computing power has brought undreamt-of capability to the desk of the most junior actuary. A consequence has been the continued tipping of the balance between, on the one hand, classical analysis and, on the other, numerical methods and simulation. Of course the old problems have not been force entirely off-stage—rather the onward march of processing capability has unveiled new problems which previously either did not arise in the conditions of the day, or could safely be put in the "too difficult" box with the expectation that competitors would do likewise—if indeed they recognised the problem. If solutions were needed in practice they could be provided by a non-actuarial management.

We now have a situation where what might be regarded as a surge of change is taking place across the insurance markets of the world. In turn, new problems in managing and controlling insurance and reinsurance operations are arising. In company with these developments, the force of competition, which decades ago might have been regarded as a gentlemanly, if not gentle, breeze, has suddenly become a gale

What does this mean for Astin?

To attempt to answer this, we have to look at the scope of Astin, which, as we all know, is concerned with actuarial studies in non-life insurance But what do "actuarial studies" embrace, either in terms of subject material or nature? Have we stretched the boundaries of the objects of our studies in line with the changing market scenario and the changing capabilities of modern technology? Have we got the right balance between "in-depth" academic studies of very specific topics and more superficial, less "respectable" examinations of a broader subject matter which does not lend itself so conveniently to a "nice" treatment?

Every member of Astin will have his own answers to these questions: perhaps I could try to stimulate discussion by looking again at familiar areas of activity.

For many years—since the formation of Astin—we have been concerned with a traditional subject matter embracing the areas of risk and ruin, moving more recently into such areas as claim reserving and risk costing (as distinct from rating).

If we look at what happens in an actual insurance operation, in arriving at a rate for a risk, it is difficult to deny that each of these areas should be represented. However, in practice, other considerations come into play whose significance may dwarf those mentioned (with the possible exception of claim reserves)

These areas—assuming we are concerned with setting rates in a competitive marketplace—would embrace (to select a few items at random):

— how to relate rates to risk in the presence of classificatory factors: for some of which only limited information, but for others extensive experience, may be available should we use explicit, purpose built models, neural networks, etc.,
— how to estimate outstanding claims for the purposes of rating, and to reflect risk and other factors in the basis used for claim development, given the existence in some cases of possibly vast historic stores of relevant detailed past experience;
— how to take into account competitors' activities,
— how to take into account more or less well-defined cycles of insurance-related experience;
— how and to what extent to take into account risk and return on assets supporting the insurance activity;
— how to define meaningful objectives, to which rates can be attuned, which reflect the rating cycle, uncertainty of experience, the need to relate risk and return to the performance of other capital markets, etc., etc

To take another example — after decades of papers on claims reserving, the methodology employed in practice is in most cases, I would guess, extremely basic and subjective. This most fundamental of actuarial activities I suspect suffers from the lack of a generally agreed basic approach which effectively utilises the extent of information available in a systematic way

Is something going wrong? If Astin was intended and is intended as no more than a group whose objectives either do not include practical usefulness of output, or include it only incidentally, then we could claim all is well. If, on the other hand, as a sub-group of IAA, its objective is to support the progress of actuarial science—and not least actuaries—then I suspect at the very least some of these issues deserve an airing

Let me make two suggestions:

— authors of papers to the Workshop Section of AB should be encouraged to write papers which describe problem areas they have encountered, without necessarily offering a solution;

— the Astin Committee itself should take stock of the extent to which

(a) actuaries are moving into less traditional areas of non-life insurance, and the extent to which they have the support of a range of actuarial methodologies.

(b) areas of insurance operation in which actuaries have only peripherally, if at all, been involved, now offer serious actuarial challenges.

The turn of the millennium represents a series of challenging opportunities for the profession — but only if it reaches out an grasps them before others develop the necessary skills

HARRY REID

ON THE DUALITY OF ASSUMPTIONS UNDERPINNING THE CONSTRUCTION OF LIFE TABLES

A. E. RENSHAW, PH D , S. HABERMAN, PH.D., F.I.A., AND P. HATZOPOULOS, M.SC.
of City University of London

ABSTRACT

We investigate the implications of a dual approach to the graduation of the force of mortality based on the modelling of the exposures as gamma random variables, as opposed to the modelling of the numbers of deaths as Poisson random variables.

KEYWORDS

Graduation, Life Tables; Exposure Response Models; Generalised Linear Models

1 INTRODUCTION

In this paper, we describe as the 'conventional' approach to graduation the method whereby the force of mortality is graduated by fitting a parameterised formula to the crude mortality rates under the assumption that the actual numbers of deaths are Poisson random variables conditional on the matching central exposures to the risk of death, e.g. Forfar, McCutcheon & Wilkie (1988) Under this approach, the Poisson assumption gives rise to a characteristic likelihood which is optimised to provide estimates for the parameters in the graduation formula. It has been noted, e.g. page 113 of Gerber (1995), that the same formal expression for the likelihood arises under the different assumption that the central exposures to the risk of death are gamma random variables conditional on the matching numbers of deaths The implications of adopting this dual approach for the parametric graduation process are investigated in this paper. Following Renshaw (1991), both approaches are formulated within the generalised linear modelling (GLM) framework , while the conclusions extend to include non-linear parameterised graduation formulae.

A brief description of the salient features of GLMs is presented in Section 2 for completeness The consequences of switching from the 'conventional' approach to the dual modelling approach when the data are based on head counts, or equivalently, on policy counts in the absence of duplicate policies, are discussed in Section 3. The implications for both approaches when duplicate policies are present in the data counts are then discussed in Section 4 and Section 5 respectively Finally an illustration of the implications of the switch from the 'conventional' approach to the dual approach, which reside largely in the reporting of the graduation, is presented in Section 6.

2 GENERALISED LINEAR MODELS

The purpose of this section is to provide a brief introduction to GLMs. A complete treatment of the theory and application can be found in McCullagh & Nelder (1989) and Francis, Green & Payne (1993).

The basis of a GLM is motivated, in the first instance, by the assumption that the data are sampled from a one parameter exponential family of distributions with log-likelihood

$$l = \frac{y\theta - b(\theta)}{\phi} + c(y,\phi)$$

for a single observation y, where θ is the canonical parameter and ϕ is the dispersion parameter, assumed known. It is then straightforward to demonstrate that

$$m = E(Y) = \frac{d}{d\theta} b(\theta) \text{ and } \mathrm{Var}(Y) = \phi \frac{d^2}{d\theta^2} b(\theta) = \phi\, b''(\theta).$$

We note that Var(Y) is the product of two quantities The quantity $b''(\theta)$ is called the variance function and depends on the canonical parameter and hence on the mean We can write this as V(m).

The log-likelihoods for some common distributions of interest and which conform to these properties are

$$l = y \log m - m - \log y!$$
$$\theta = \log m,\ b(\theta) = \exp\theta,\ V(m) = m,\ \phi = 1$$

for the Poisson distribution with mean m, and

$$l = \frac{-\frac{y}{m} + \log\frac{1}{m}}{\frac{1}{v}} + v\log y + v\log v - \log\Gamma(v)$$

$$\theta = -\frac{1}{m},\ b(\theta) = -\log(-\theta),\ V(m) = m^2,\ \phi = v^{-1}$$

for the gamma distribution mean m and variance m^2/v.

More generally a GLM is characterised by independent response variables $\{Y_u;\ u = 1, 2, \ldots, n\}$ for which

$$E(Y_u) = m_u,\ Var(Y_u) = \frac{\phi V(m_u)}{\omega_u} \tag{2.1}$$

comprising a variance function V, a scale parameter ($\phi > 0$) and prior weights ω_u.

Covariates enter via a linear predictor

$$\eta_u = \sum_{j=1}^{p} x_{uj}\beta_j$$

with specified structure (x_{uj}) and unknown parameters β_j linked to the mean response through a known differentiable monotonic link function g with

$$g(m_u) = \eta_u .$$

The special link function $g = \theta$, so that $\theta(m) = \eta$, is called the canonical link function. Examples are the log link in the case of the Poisson distribution and the reciprocal link in the case of the gamma distribution

The suffices or units u have structure, either intrinsic or imposed. The data comprise realisations $\{y_u\}$ of the independent response variables, matched to the structure of the units. Generally in any one study, the detail of the distribution and link are fixed, while the predictor structure may be varied

Model fitting is by maximising the quasi log-likelihood

$$q = q(\underline{y},\underline{m}) = \sum_{u=1}^{n} q_u = \sum_{u=1}^{n} \omega_u \int_{y_u}^{m_u} \frac{y_u - s}{\phi V(s)} ds \tag{2.2}$$

leading to the system of linear equations

$$\sum_{u=1}^{n} \omega_u \frac{y_u - m_u}{\phi V(m_u)} \frac{\partial m_u}{\partial \beta_j} = 0 \quad \forall j$$

in the unknown β_js These are solved numerically, e.g Francis, Green & Payne (1993), McCullagh & Nelder (1989). Detail of the construction of standard errors for the parameter estimators, based on standard statistical theory, is also to be found in these references Denote the resulting values of the parameter estimators, linear predictor and fitted values, for the current model c, $\hat{\beta}_j, \hat{\eta}_u$ and $\hat{m}_u$ respectively, where

$$\hat{m}_u = g^{-1}(\hat{\eta}_u), \hat{\eta}_u = \sum_{j=1}^{p} x_{uj} \hat{\beta}_j$$

For members of the exponential family of distributions, the quasi log-likelihood is synonymous with log-likelihood The maximal structure possible has the property that the fitted values are equal to the observed responses, that is $\hat{m}_u = y_u$ for all u, and is called the full or saturated model f.

The (unscaled) deviance of the current model c is

$$D(c,f) = d(\underline{y};\hat{\underline{m}}) = \sum_{u=1}^{n} d_u = \sum_{u=1}^{n} 2\omega_u \int_{\hat{m}_u}^{y_u} \frac{y_u - s}{V(s)} ds = -2\,\phi\, q(\underline{y};\hat{\underline{m}}),$$

in which the fitted values under the current and saturated models impact on the formula through the lower and upper limits of the integral respectively. The corresponding scaled deviance is

$$S(c,f) = d^*(\underline{y};\hat{\underline{m}}) = \frac{d(\underline{y},\hat{\underline{m}})}{\phi} = \sum_{u=1}^{n} 2\omega_u \int_{\hat{m}_u}^{y_u} \frac{y_u - s}{\phi V(s)} ds = -2\, q(\underline{y};\hat{\underline{m}}) \tag{2 3}$$

For fixed distribution, fixed link and hierarchical model structures c_1 and c_2, with c_2 nested in c_1, the difference in scaled deviance

$$S(c_2, f) - S(c_1, f)$$

may be referred, generally as an approximation, to the chi-square distribution with $v_2 - v_1$ degrees-of-freedom, where v_1 and v_2 denote the respective degrees-of-freedom.

Two types of residuals (which are identical only in the case of the Gaussian distribution, for which $V(s) = 1$) are of interest, the Pearson residuals

$$\frac{y_u - \hat{m}_u}{\sqrt{\frac{V(\hat{m}_u)}{\omega_u}}} \tag{2.4}$$

or the deviance residuals

$$\text{sign}(y_u - \hat{m}_u)\sqrt{d_u}$$

where d_u is the uth. component of the (unscaled) deviance above

3. HEAD OR POLICY COUNTS WITH NO DUPLICATES

3.1 Distribution Assumptions

In keeping with common practice, let

μ_x = the force of mortality at age x
${}_w p_x$ = the probability that a life aged x survives tot age $x + w$

and recall the basic identity

$${}_w p_x = \exp - \int_0^w \mu_{x+s} ds \tag{3 1}$$

with the implied assumption that μ_x is a function of age alone and is therefore assumed to be constant with respect to variations in calendar time within a fixed observation window.

Focus on a set of individual lives or policyholders. If the latter, and the data are based on policy counts, then it is assumed throughout this Section that all policyholders possess a single policy Individual members of the set are assumed to be observed between ages x and $x + 1$ in the *fixed* calendar period or observation window t to $t + t_0$, with pre-specified policy duration where relevant, and their survival experience is assumed throughout to be independent Typically $t_0 = 4$ years in many United Kingdom (UK) actuarial mortality studies. There is also interest in the case $t_0 = 1$ year when modelling trends in mortality, e g. Renshaw, Haberman & Hatzopoulos (1996). Within such a cell, identified in this instance by the *suffix* x, suppose an individual i enters observation at age v_{xi} and leaves it either by death ($I_{xi} = 1$) or by censorship

($I_{xi} = 0$) at age $v_{xi} + w_{xi}$ where $x \le v_{xi} < v_{xi} + w_{xi} \le x + 1$. Then it is well known see, e g Section 3.2 of Cox & Oakes (1984), that each such datum contributes an amount

$$L_{xi} = {}_{w_{xi}}p_{v_{xi}} \mu_{v_{xi}+w_{xi}}^{I_{xi}}$$

to the likelihood, or, on resorting to the use of expression (3 1), an amount

$$l_{xi} = \log L_{xi} = -\int_0^{w_{xi}} \mu_{v_{xi}+s} ds + I_{xi} \log \mu_{v_{xi}+w_{xi}}$$

to the log-likelihood. Thus the total contribution to the log-likelihood from such a cell is

$$l_x = \sum_{i=1}^{n_x} l_{xi} = \sum_{i=1}^{n_x} \left\{ -\int_0^{w_{xi}} \mu_{v_{xi}+s} ds + I_{xi} \log \mu_{v_{xi}+w_{xi}} \right\} \tag{3 2}$$

where the summation extends to all n_x individuals contributing to the experience in the cell If in addition μ_x is assumed to be piecewise constant with respect to age within each cell and accorded the central value $\mu_{x+1/2}$, expression (3.2) can be written as

$$l_x = -r_x \mu_{x+1/2} + a_x \log \mu_{x+1/2}$$

where

$$r_x = \sum_{i=1}^{n_x} w_{xi},\ a_x = \sum_{i=1}^{n_x} I_{xi}$$

denote the respective central exposure and actual number of deaths associated with cell x. The expression for the full log-likelihood

$$l = \sum_x l_x = \sum_x \{ -r_x \mu_{x+1/2} + a_x \log \mu_{x+1/2} \} \tag{3.3}$$

then follows by summation over all such cells. It is of specific interest to note that this expression may be interpreted in one of two ways

Firstly, and somewhat exclusively in the context of an actuarial graduation, expression (2.3) is identifiable as the kernel of the log-likelihood under the assumption that the actual numbers of deaths, a_x, are modelled as independent realisations of Poisson random variables A_x *conditional* on r_x, such that

$$A_x \sim \text{Poi}(r_x \mu_{x+1/2}).$$

For this case, the detail of the distributional requirements to set up the appropriate GLM (equation (2 1) with $u \equiv x$) is either

$$\text{responses } \{A_x\}, \text{with } m_x = r_x \mu_{x+1/2}, V(m_x) = m_x, \phi = 1, \omega_x = 1, \tag{3 4a}$$

or equivalently

$$\text{responses } \{A_x / r_x\}, \text{with } m_x = \mu_{x+1/2}, V(m_x) = m_x, \phi = 1, \omega_x = r_x \tag{3 4b}$$

Secondly, e.g. Section 11 5 of Gerber (1995), expression (3 3) is also identifiable as the kernel of the log-likelihood under the assumption that the exposures to risk, r_x, are modelled as independent realisations of gamma random variables R_x *conditional* on a_x, such that

$$R_x \sim \text{gam}(a_x, \mu_{x+1/2})$$

Superficially this result is perhaps a little unusual in-so-far as the gamma distribution is generally associated with two unknown parameters, whereas here, as with the Poisson distribution above, there is only a single parameter to estimate For this case, the detail of the distributional requirements to set up the appropriate GLM (equation (2 1) with $u \equiv x$) is either

$$\text{responses } \{R_x\}, \text{with } m_x = a_x \frac{1}{\mu_{x+1/2}}, V(m_x) = m_x^2, \phi = 1, \omega_x = a_x, \qquad (3\ 5a)$$

or equivalently

$$\text{responses } \{R_x / a_x\}, \text{with } m_x = \frac{1}{\mu_{x+1/2}}, V(m_x) = m_x^2, \phi = 1, \omega_x = a_x \qquad (3.5b)$$

The data comprise the ordered pairs of numbers of deaths and central exposures (a_x, r_x) over a range of ages x All of the r_xs are non-zero by implication, but it is conceivable that certain of the a_xs are zero. This is most likely to occur at the extremities of the age range were the data are sometimes sparse Note that while such data cells are retained in any analysis of the data based on distributional assumptions (3.4a & b), they are weighted out of any analysis based on distributional assumptions (3.5a & b)

3.2 Discussion

The optimisation of expression (3 3) under the former interpretation (based on the Poisson distribution) is central to the current graduation practice of the Continuous Mortality Investigation (CMI) Bureau in the UK, e g. Forfar *et al* (1988), while the optimisation of expression (3.3) under the alternative interpretation (based on the gamma distribution) would appear not to have been investigated previously in an actuarial graduation setting.

It is possible to derive the first set of assumptions, in which the number of actual deaths A_x form the response variables, by taking expectations and variances under the identity

$$A_x = \sum_{i=1}^{n_x} I_{xi}$$

where I_{xi} is the zero-one indicator random variable, introduced previously, in Section 3.1. It has the property

$$E(I_{xi}) = E(I_{xi}^2) = P(I_{xi} = 1) = 1 - \exp - \int_0^{u_{xi}} \mu_{x+s} ds$$

and is assumed to be independent for all individuals i. The results then follow under the assumption that μ_x is piecewise constant within cells, so that

$$E(I_{xi}) = (E(I_{xi}^2) = 1 - \exp(-\mu_{x+1/2} w_{xi}), \tag{3 6}$$

and on neglecting second and higher order therms in the power series expansion of $\exp(-\mu_{x+1/2} w_{xi})$, so that

$$\text{Var}(I_{xi}) = E(I_{xi}) \approx \mu_{x+1/2} w_{xi}.$$

Under the second set of assumptions, for which the responses satisfy

$$R_x = \sum_{i=1}^{n_x} W_{xi},$$

the individual exposures W_{xi} are modelled as random variables. Under the additional assumption that the individual exposures are independent and identically distributed, it follows trivially from the reproductive property of the gamma distribution that they have the gamma distribution

$$W_{xi} \sim \text{gam}(\frac{a_x}{n_x}, \mu_{x+1/2})$$

Again based on the reproductive property of the gamma distribution, note that it is also possible to construct the identical GLM by defining

$$R_x = \sum_{i=1}^{n_x} W_{xi} = \sum_{j=1}^{a_x} T_{xj}$$

in which the T_{xj}s are assumed to be independent and identically distributed gamma random variables, such that

$$T_{xj} \sim \text{gam}(1, \mu_{x+1/2}),$$

and where at least one death is recorded in every cell Here it is possible to interpret T_{xj} as the sum of randomly selected censored exposures W_{xi} the last of which is associated with a death

The target of the graduation process is the *force of mortality* μ_x under distribution assumptions (3.4a & b) and the *force of vitality* $1/\mu_x$ under distribution assumptions (3.4a & b). In using the latter description, we follow the terminology of Lambert (1772) see, e g Daw (1980)

The value of the scaled deviance, (expression 2.3, with $u \equiv x$) is identical under both sets of modelling assumptions (3 4a & b) and (3.5a & b) and is equal to

$$S(c,f) = \sum_x 2\left\{a_x \log\frac{a_x}{r_x \hat{\mu}_{x+1/2}} - (a_x - r_x \hat{\mu}_{x+1/2})\right\} \tag{3 7}$$

where $\hat{\mu}_x$ denotes the graduated values of μ_x provided deaths are recorded for all ages (i e. $a_x > 0 \; \forall \; x$) so that none of the terms are weighted out of the expression on the right hand side (RHS) of equation (3 7) under the dual modelling assumptions (3.5a &

b) This is perhaps a surprising result on the surface It reflects the fact that the same objective function, expression (3 3), which is embedded in the construction of the scaled deviance as the quasi log-likelihood function, (expression 2 2, with $u \equiv x$) is optimised when fitting the model structure (or graduation formula).

Subject to the weighting out of any data cells containing zero a_xs in the one case, the two sets of distribution assumptions lead to identical graduations for μ_x Thus, assumption (3 4a) with responses $\{a_x\}$ in combination with log-link based graduation formulae of the type

$$\log \mu_{x+1/2} = \sum_{j=0}^{p} h_{xj}\beta_j \tag{3 8}$$

so that

$$\log m_x = \eta_x = \log r_x + \log \mu_{x+1/2} = \log r_x + \sum_{j=0}^{p} h_{xj}\beta_j,$$

gives identical graduations to those obtained under assumption (3 5b) with responses $\{r_x\}$ so that

$$\log m_x = \eta_x = \log a_x - \log \mu_{x+1/2} = \log a_x + \sum_{j=0}^{p} h_{xj}\beta_j.$$

Typically the parameterised structure of the RHS of the graduation equation (3.8) is a polynomial in x with either the log r_x or log a_x terms declared as offsets, as the case may be The estimated values of the parameters β_j are identical in magnitude but opposite in sign in the two cases Similarly assumption (3.4b) with responses $\{a_x/r_x\}$ in combination with the power link graduation formulae of the type

$$\mu^{\gamma}_{x+1/2} = \sum_{j=0}^{p} h_{xj}\beta_j$$

gives identical graduations to those obtained under assumption (3 5b) with responses $\{r_x/a_x\}$ so that

$$\mu^{-\gamma}_{x+1/2} = \sum_{j=0}^{p} h_{xj}\beta_j$$

This time the estimated values of the parameters β_j are identical in both magnitude and sign in the two cases. Thus the general conclusions of this paper extend to non-linear parameterised graduation formulae via the identity link under the 'conventional' approach and the reciprocal link under the dual approach.

Let $e_x = r_x\mu_{x+1/2}$ denote the expected number of deaths predicted at age x, under the conventional graduation methodology encapsulated by equations (3.4a &b), and define the statistics

$$dev_x = a_x - e_x, \sqrt{V_x} = \sqrt{e_x}, zx = \frac{dev_x}{\sqrt{V_x}}, \ 100\frac{a_x}{e_x}. \tag{3.9}$$

It is common practice for these to be tabulated (subject to possible cell grouping in the tails of the age range) as part of the diagnostic checking procedure of a graduation. Note in particular that the statistic z_x is the Pearson residual of the corresponding GLM, (expression 2.3, with $u \equiv x$). Thus typically the value of the approximate chi-square statistic $\sum_x z_x^2$ is quoted as one of the many test statistics of a graduation. The equivalent statistics under the dual graduation methodology encapsulated by equations (3.5a or b) involving definition $\tilde{e}_x = a_x / \hat{\mu}_{x+1/2}$ or expected exposure predicted at age x, are

$$d\tilde{e}v_x = r_x - \tilde{e}_x, \sqrt{\tilde{V}_x} = \sqrt{\frac{\tilde{e}_x^2}{a_x}}, \tilde{z}_x = \frac{d\tilde{e}v_x}{\sqrt{\tilde{V}_x}}, \ 100\frac{r_x}{\tilde{e}_x}. \tag{3 10}$$

Again note that these statistics are defined in such a way that $\tilde{z}_x$ denotes the Pearson residual of the associated GLM (3.5a or b). The relationship between the values of the deviation under the dual and 'conventional' graduation methodologies, namely

$$d\tilde{e}v_x = \frac{-d\tilde{e}v_x}{\hat{\mu}_{x+1/2}}$$

implies that the residuals under the two methodologies have opposite signs. Although only strictly exact provided all the a_xs are positive, this relationship provides a very close approximation when the a_xs take zero values at the extremities of the age range concerned. Detailed examination of the respective formulae defining the Pearson residuals z_x and $\tilde{z}_x$ reveals that they differ in magnitude (and have opposite signs) On the other hand, because of the equality of the deviance components under the two methodologies established above, the deviance residuals defined by either

$$\text{sign}(dev_x)\sqrt{d_x} \text{ or } \text{sign}(d\tilde{e}v_x)\sqrt{d_x}$$

as the case may be, where d_x is the general term in the summation on the RHS of expression (3 7), are identical in magnitude (and opposite in sign) under the dual methodologies. It is also of interest to note that the final statistics quoted in expressions (3 9) and (3 10), corresponding to the respective dual modelling scenarios, are the reciprocals of one another prior to scaling by 100 Again both of these features are exact when all the a_xs are positive and represent a very close approximation when any of the a_xs are zero at the extremities of the age range

4 POLICY COUNTS WITH DUPLICATES: CLAIM NUMBER RESPONSE MODELS

4.1 Preliminaries

The data used in the construction of actuarial life tables are generally based on policy rather than head counts Consequently, the death of a policyholder with more than one policy will appear as more than one death in the raw data The resulting graduation needs to account for this overdispersion. for a review of the issues involved, readers should consult Forfar *et al.* (1988) and Renshaw (1992).

Let

D_{xi} = the number of policies held by policyholder i, age x
C_{xi} = the number of policies held by policyholder i, age x, resulting in a claim.

Assume that the random variables D_{xi} are i.i.d $\forall\, i$ and let D_x denote the generic type. For each i, the events $(C_{xi} = k \mid I_{xi} = 1)$ and $(D_{xi} = k)$ are such that

$$(C_{xi} = k \mid I_{xi} = 1) \Leftrightarrow (D_{xi} = k),\ k = 1, 2, 3, ..$$

and thus have identical probabilities. Define

$$P(D_x = k) = P(C_{xi} = k \mid I_{xi} = 1) = \begin{cases} \pi_x^{(k)} & k = 1,2,3,... \\ 0 & \text{otherwise} \end{cases}$$

where

$$\pi_x^{(k)} \geq 0,\ \sum_{k=1}^{\infty} \pi_x^{(k)} = 1$$

Denote

$$E(D_x) = E(C_{xi} \mid I_{xi} = 1) = \sum_{k=1}^{\infty} k\pi_x^{(k)} = {}_1\pi_x$$

and

$$E(D_x^2) = E(C_{xi}^2 \mid I_{xi} = 1) = \sum_{k=1}^{\infty} k^2\pi_x^{(k)} = {}_2\pi_x .$$

It also follows by definition that

$$P(C_{xi} = 0 \mid I_{xi} = 0) = 1$$

so that

$$E(C_{xi} \mid I_{xi} = 0) = E(C_{xi}^2 \mid I_{xi} = 0) = 0$$

Hence the *unconditional* distribution of C_{xi} is given by

$$P(C_{xi} = k) = \begin{cases} 1 - E(I_{xi}), & k = 0 \\ E(I_{xi})\pi_x^{(k)}, & k = 1,2,3, \end{cases}$$

for which

$$E(C_{xi}) = {}_1\pi_x E(I_{xi}),\ E(C_{xi}^2) = {}_2\pi_x E(I_{xi})$$

These equations, in combination with expression (3 6) for $E(I_{xi})$, on neglecting second and higher order terms in the power series expansion of $\exp(-\mu_{x+1/2}\, w_i)$, imply that

$$E(C_{xi}) \approx {}_1\pi_x \mu_{x+1/2} w_{xi} \text{ and } \mathrm{Var}(C_{xi}) \approx {}_2\pi_x \mu_{x+1/2} w_{xi}. \tag{4.1}$$

We also have an interest in the first two moments of the product random variable $D_{xi}\, I_{xi}$ Under the mild assumption that the number of policies, D_{xi}, held by policyholder i, aged x, is statistically independent of the mode of censorship, I_{xi}, it follows that

$$E(D_{xi} I_{xi}) = E(D_{xi})E(I_{xi}),\ \mathrm{Var}(D_{xi} I_{xi}) = E(D_{xi}^2)E(I_{xi}^2) - \{E(D_{xi})E(I_{xi})\}^2$$

These equations in combination with expressions (3 6), on neglecting second and higher order terms in the power series expansion of $\exp(-\mu_{x+1/2}\, w_i)$, then imply that

$$E(D_{xi} I_{xi}) \approx {}_1\pi_x \mu_{x+1/2} w_{xi} \text{ and } \mathrm{Var}(D_{xi} I_{xi}) \approx {}_2\pi_x \mu_{x+1/2} w_{xi} \tag{4.2}$$

4.2 Distribution Assumptions

Let

A'_x = the number of policies giving rise to a claim through deaths
r'_x = the cental exposure to the risk of death based on policies.

Note that
$$r'_x = \sum_{i=1}^{n_x} d_{xi} w_{xi}$$

where d_{xi} (≥ 1) denotes the number of policies held by policyholder i, reducing to r_x if and only if $d_{xi} = 1\ \forall\ i$. Throughout this Section the A'_xs are modelled as random variables *conditional* on r'_x. It follows on taking expectations and variances under any one of the following identities

$$A'_x = \sum_{i=1}^{A_x} D_{xi} \text{ (with } A_x > 0),\ A'_x = \sum_{i=1}^{n_x} C_{xi},\ A'_x = \sum_{i=1}^{n_x} D_{xi}\, I_{xi} \tag{4 3}$$

that the detail of the distributional requirements to set up the appropriate GLM (equation (2.1), with $u \equiv x$) is either

$$\text{responses } \{A'_x\}, \text{ with } m_x = r'_x \mu_{x+1/2},\ V(m_x) = m_x,\ \phi = 1,\ \omega_x = \phi_x^{-1}, \tag{4 4a}$$

or equivalently

$$\text{responses } \{A'_x / r'_x\}, \text{ with } m_x = \mu_{x+1/2},\ V(m_x) = m_x,\ \phi = 1,\ \omega_x = r'_x \phi_x^{-1}, \tag{4 4b}$$

where
$$\phi_x = \frac{{}_2\pi_x}{{}_1\pi_x}$$

4.3 Discussion

The result (4.4a) follows from the first of the identities (4.3) which, under the assumption that A_x is independent of the $\{D_{xi}\}$ implies, in combination with equations (3.4a)

$$E(A'_x) = E(D_x)E(A_x) = {}_1\pi_x r_x \mu_{x+1/2}$$

and

$$\text{Var}(A'_x) = \text{Var}(D_x)E(A_x) + \{E(D_x)\}^2 \text{Var}(A_x) = \frac{E(D_x^2)}{E(D_x)} E(A'_x) = \frac{{}_2\pi_x}{{}_1\pi_x} E(A'_x)$$

Under the independence of the terms in the respective summations, the same result follows trivially from either the second of the identities (4.3) in combination with equations (4 1), or the third of the identities (4.3) in combination with equations (4.2). In all three cases, the product term ${}_1\pi_x r_x$ in the expression for $E(A'_x)$ involving the unobserved central exposure based on lives has been replaced by r'_x, the observed central exposure based on policies The result (4.4b) follows trivially from result (4 4a)

The justification for (4.4a) based on the second of the identities (4.3) and equations (4.1) is a generalisation of the method described in Renshaw (1992) for initial exposures and the binomial response model. This work establishes a link with much earlier work on the modelling of duplicate policies using an empirical approach, e.g. Beard & Perks (1949).

A knowledge of the reciprocals of the overdispersion parameters ϕ_x is needed to form the weights, if the distributional assumptions (4 4) are to be fully implemented Insight into the potential variation of ϕ_x with x is provided by studies of the properties of so-called variance ratios, the empirical equivalent of ϕ_x, e.g Forfar *et al.* (1988). These are defined as

$$vr_x = \frac{\sum_i i^2 f_x^{(i)}}{\sum_i i f_x^{(i)}}$$

where $f_x^{(i)}$ denotes the proportion, at age x, of policyholders who have i policies and where

$$f_x^{(i)} \geq 0 \ \forall \ i = 1,2,3,\ldots, \ \sum_i i f_x^{(i)} = 1 \Rightarrow vr_x \geq 1$$

There are a number of alternative practical possibilities When available, variance ratios can be used as estimates for the dispersion parameters ϕ_x and graduation can proceed in accordance with assumptions (4 4) On the other hand, Forfar *et al.* (1988) acting for the CMI Bureau in the UK, elect to transform the data by dividing both the policy counts a'_x and exposures r'_x by the matching variance ratios prior to graduation with assumptions (3 4) displacing assumptions (4 4) When a detailed knowledge of the relevant variance ratios is not available for analysis a possible method of generating estimates for the dispersion parameters is described in Renshaw (1992). Alternatively, under the assumption that the underlying modelling distribution of the number of duplicate policies is identical across all ages x in the absence of any further detailed knowledge about this distribution, the dispersion parameters ϕ_x may be replaced by a constant scale (or dispersion) parameter ϕ in assumptions (4 4), e.g. Renshaw (1992) It is estimated as

$$\hat{\phi} = \frac{\text{unscaled deviance}}{\text{degrees - of - freedom}}$$

and is root $\sqrt{\hat{\phi}}$ used to scale the Pearson residuals z_x of expressions (3 9) or $\tilde{z}_x$ of expressions (3 10), by multiplying either V_x or $\tilde{V}_x$ by $\hat{\phi}$, as the case may be. Here the unscaled deviance is calculated using the expression on the RHS of equation (3.7). (Recall that ϕ was set to one when deriving this expression, so that the scaled deviance $S(c,f)$ is also the unscaled deviance in this instance.) This latter approach is closest in spirit to that adopted by Forfar *et al.* (1988) involving the transformation of the data prior to graduation in-so-far as it produces identical graduations, while allowing the presence of duplicate policies to impact solely on the second moment properties of the graduation process

5. POLICY COUNTS WITH DUPLICATES: EXPOSURE RESPONSE MODELS

5.1 Preliminaries

As before, let

D_{xi} = the number of policies held by policyholder i, age x
W_{xi} = the contribution to the exposure by policyholder i, age x

Recall that D_x, D_{xi} are assumed to be i.i d. $\forall\ i$ with

$$E(D_x) = {}_1\pi_x,\ E(D_x^2) = {}_2\pi_x.$$

Recall also the duality property of Section 3 2, namely that the central exposure to risk of death based on head counts, at age x

$$R_x = \sum_{i=1}^{n_x} W_{xi} \sim \text{gam}(a_x, \mu_{x+1/2}),$$

so that

$$E(R_x) = \frac{a_x}{\mu_{x+1/2}},\ E(R_x^2) = \frac{a_x(1+a_x)}{\mu_{x+1/2}^2}.$$

Consider the identity

$$R'_x = \sum_{i=1}^{n_x} D_{xi} W_{xi} \tag{5.1}$$

which defines the central exposure to risk of death based on policy counts, at age x Assuming that the number of policies held by an individual policyholder is independent of the corresponding contribution to the exposure to risk from that individual and that the individual exposures are independent, it follows from the identity (5 1) that

$$E(R'_x) = E(D_x)\sum_{i=1}^{n_x} E(W_{xi}) = E(D_x)E(R_x) = \frac{{}_1\pi_x a_x}{\mu_{x+1/2}} \tag{5.2}$$

and

$$E(R'^2_x) = E(D_x^2)E(\sum_{i=1}^{n_x} W_{xi})^2 = E(D_x^2)E(R_x^2) = \frac{{}_2\pi_x a_x(1+a_x)}{\mu^2_{x+1/2}} \tag{5 3}$$

after simplification.

5.2 Distribution Assumptions

Let

R'_x = the central exposure to the risk of death based on policies
a'_x = the number of policies giving rise to a claim through deaths

Throughout this section the R'_xs are modelled as random variables *conditional* on a'_x. It follows from equations (5.1), (5.2) and (5.3) that the detail of the distributional requirements to set up the appropriate GLM (equation (2 1), with $u \equiv x$) is either

$$\text{responses } \{R'_x\}, \text{ with } m_x = a'_x \frac{1}{\mu_{x+1/2}}, \; V(m_x) = m_x^2, \phi = 1, \omega_x = \psi_x^{-1}, \tag{5 4a}$$

or equivalently

$$\text{responses } \{R'_x / a'_x\}, \text{ with } m_x = \frac{1}{\mu_{x+1/2}}, \; V(m_x) = m_x^2, \phi = 1, \omega_x = \psi_x^{-1}, \tag{5.4b}$$

where this time

$$\psi_x = \left(\frac{{}_2\pi_x}{{}_1\pi_x^2} - 1\right) + \frac{{}_2\pi_x}{{}_1\pi_x}\frac{1}{a'_x}. \tag{5 5}$$

5.3 Discussion

In parallel with the previous case, this time the product term ${}_1\pi_x a_x$ in the expression for $E(R'_x)$ involving the unobserved number of deaths a_x based on head counts has been replaced by a'_x, the observed number of deaths base on policy counts. Again result (5 4b) follows trivially from result (5 4a)

A knowledge of the reciprocals of the dispersion parameters ψ_x is required to form the weights if the distribution assumptions (5.4a or b) are to be fully implemented In the event that the results of a study into the variance ratios for the policies in question are available, this will furnish estimates for the first two moments ${}_1\pi_x$ and ${}_2\pi_x$ of the number of duplicate policies so that modelling can proceed. Alternatively if it is assumed that the square of the coefficient of variation of the number of duplicate policies held by an individual is sufficiently small so as to make the first term on the RHS of expression (5.5) for ψ_x is negligible in comparison with the second term,

$$\psi_x = \phi_x \frac{1}{a'_x}$$

and the situation is analogous to that discussed in Section 4.3.

6. ILLUSTRATION

The dual methodologies are illustrated using the Pensioners' widows 1979-1982 experience reported in Table 15.5 of Forfar *et al* (1988). The data (a_x, r_x), comprising the numbers of deaths a_x and matching central exposures r_x, are reported in the age range 17 to 108 years inclusive. There are 2 + 5 = 7 completely empty cells in the extremities of the age range and 28 + 12 = 40 cells contain no reported deaths The detail of the graduation contained in the above Table is based on Gompertz's formula fitted by the 'conventional' approach, in which the numbers of deaths are modelled as Poisson random variables The data have been regraduated using both the 'conventional' approach based on assumptions (3.4a) with predictor-link formulation

$$\log m_x = \log r_x + \log \mu_{x+1/2} = \log r_x + \beta_0 + \beta_1\left(\frac{x-70}{50}\right),$$

and the dual approach based on assumptions (3 5a) with equivalent predictor-link formulation

$$\log m_x = \log a_x - \log \mu_{x+1/2} = \log a_x + \beta_0 + \beta_1\left(\frac{x-70}{50}\right),$$

where m_x denotes the respective mean responses The associated graduation formula, implied by these formulae, is taken from Forfar *et al* (1988). Some details of the respective fits including the parameter estimates are recorded in Table 6.1 The corresponding parameter estimates have opposite signs as expected, but differ slightly in absolute value because the data entries involving zero deaths feature only in the 'conventional' analysis Similarly the corresponding values of both the deviances and the degrees-of-freedom differ for the same reason. These differences are found to disappear when the 'conventional' analysis is applied to the reduced data set and identical graduations result as a consequence (subject to very minor differences induced by the numerical fitting algorithm operating under the two different approaches) An extract of both graduations based on the detail of Table 6 1 is reproduced in Table 6 2(a&b), along with detail of the associated statistics of expressions (3 9) and (3.10), as the case may be The detail of Table 6.2a is in complete agreement with that to be found in Table 15.5 of Forfar *et al* (1988), while the relatively minor effects of the excluded data under the dual modelling approach are demonstrated. The basic differences in the accompanying statistics used to monitor the effectiveness of a graduation under the two different approaches, as described in Section 3.2, can be verified

7 CONCLUSIONS

The 'conventional' actuarial approach to the construction of μ_x-graduations based on the fitting of a wide class of parameterised mathematical formulae by optimising the likelihood, in which the death counts are modelled as Poisson random variables *conditional* on the central exposures, is effectively equivalent to a dual approach in which the central exposures are modelled as gamma random variables *conditional* on the death counts The dual approaches lead to identical graduations provided deaths are recorded in all data cells, otherwise small differences occur in practice as a consequence of the loss of information from any data cells in which no deaths are recorded under the one approach Key differences occur in the diagnostic statistics of a graduation, with residuals being accorded opposite signs under the two different approaches In practice, a detailed knowledge of the specific nature of the empirical distributions on duplicate policies has only a minimal effect on the first moment of a graduation under the two formulations described here In the absence of this knowledge, these first moment properties may be neglected and a free standing constant scale (or dispersion) parameter introduced, under either formulation, to represent the second moment properties of a graduation in the presence of duplicate policies.

The dual approach to μ_x-graduation would appear to have distinct advantages over the 'conventional' approach to graduation, when it is adapted and applied to the construction of select mortality tables. This is discussed further in Renshaw & Haberman (1996), who successfully use the dual approach to model the log crude mortality ratios for individual select durations relative to the ultimate experience

REFERENCES

BEARD, R E & PERKS, W (1949) The Relation between the Distribution of Sickness and the Effect of Duplicates on the Distribution of Deaths *J I A*, 75, 75

COX, D R & OAKES, D (1984) Analysis of Survival Data *Chapman and Hall*

DAW, R H (1980) Johann Heinrich Lambert (1728-1777) *J I A*, 107, 345-363

FRANCIS, B, GREEN, M & PAYNE, C (1993) Eds The GLIM System Release 4 Manual *Clarendon Press Oxford*

FORFAR, D O , MCCUTCHEON, J J & WILKIE, A D (1988) On Graduation by Mathematical Formula *J I A* , 115 , 1-135 and *T F A* , 41, 97

GERBER, H U (1995) Life Insurance Mathematics (2nd edition) *Springer*

LAMBERT, J H (1772) Anmerkungen uber die Sterbluhkeit, Todtenlisten, Gebusthen un Ehen Vol III, 475-599

MCCULLAGH, P & NELDER, J A (1989) Generalized Linear Models (2nd edition) *Chapman & Hall*

RENSHAW, A E (1991) Actuarial Graduation Practice and Generalized Linear & Non-Linear Models *J I A* , 118, 295-312

RENSHAW, A E (1992) Joint Modelling for Actuarial Graduation and Duplicate Policies *J I A* , 119, 69-85

RENSHAW, A E & HABERMAN, S (1996) Dual Modelling and Select Mortality *I M E To appear*

RENSHAW, A E , HABERMAN, S & HATZOPOULOS, P (1996) Recent Mortality Trends in U K Male Assured Lives *B A J* , 2, 449-477

TABLE 6 1

PARAMETERS ESTIMATES WITH (STANDARD ERRORS)

	'conventional' approach	*dual approach*
	deviance is 60 98 with 83 d f	deviance is 45 99 with 50 d f
	scale parameter ϕ=1	scale parameter ϕ=1
	$\hat{\beta}_0$ = -3 553 (0 03923)	$\hat{\beta}_0$ =3 543 (0 03925)
	$\hat{\beta}_0$ = 4 317 (0 1966)	$\hat{\beta}_0$ = -4 332 (0 1979)

TABLE 6 2(a)

GRADUATION EXTRACT, 'CONVENTIONAL' METHOD

x	r_x	$\mu_{x+1/2}$	a_x	e_x	dev_x	$\sqrt{V_x}$	z_x	$100a_x/e_x$
17	0 5	0 00029	0	0 00	0 00	-	-	-
30	36 0	0 00091	0	0 03	-0 03	-	-	-
40	115 5	0 00215	0	0 25	-0 25	-	-	-
50	378 5	0 00509	3	1 93	1 07	-	-	-
60	1029 0	0 01208	14	12 43	1 57	3 53	0 45	112 6
65	1029 0	0 01860	21	19 14	1 86	4 37	0 43	109 7
70	941 0	0 02864	21	26 95	-5 95	5 19	-1 14	77 9
75	607 0	0 04410	33	26 77	6 23	5 17	1 20	123 3
80	323 5	0 06790	25	21 97	3 03	4 69	0 65	113 8
85	132 5	0 10455	11	13 85	11 60	3 72	-0 77	79 4
95	4 0	0 24790	2	0 99	1 01	-	-	-
108	2 0	0 76154	0	1 52	-1 52	-	-	-

TABLE 6.2(b)

GRADUATION EXTRACT, DUAL METHOD

x	a_x	$\mu_{x+1/2}$	r_x	$\tilde{e}_x$	$d\tilde{e}v_x$	$\sqrt{\tilde{V}_x}$	$\tilde{z}_x$	$100r_x/\tilde{e}_x$
17	0	0 00029	0 5					
30	0	0 00090	36 0					
40	0	0 00215	115 0					
50	3	0 00511	378 5	586 7	-208 2	338 7	-0 61	64 5
60	14	0 01216	1029 0	1151 1	-122 1	307 6	-0 40	89 4
65	21	0 01876	1029 0	1119 6	-90 6	244 3	-0 37	91 9
70	21	0 02893	941 0	725 9	215 1	158 4	1 36	129 6
75	33	0 04461	607 0	739 7	-132 7	128 8	-1 03	82 1
80	25	0 06880	323 5	363 4	-39 9	72 7	-0 55	89 0
85	11	0 10611	132 5	103 7	28 8	31 3	0 92	127 8
95	2	0 25237	4 0	7 9	-3 9	5 6	-0 70	50 5
108	0	0 77841	2 0					

ON THE BIVARIATE GENERALIZED POISSON DISTRIBUTION

RALUCA VERNIC

University "Ovidius" Constanta, Romania

ABSTRACT

This paper deals with the bivariate generalized Poisson distribution. The distribution is fitted 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. A comparative study with the bivariate Poisson distribution and with two bivariate mixed Poisson distributions has been carried out, based on data concerning natural events insurance in the USA and third party liability automobile insurance in France

KEYWORDS

Bivariate generalized Poisson distribution, generalized Poisson distribution, bivariate mixed Poisson distributions

1. INTRODUCTION

Whereas numerous bivariate discrete distributions are used in the statistic field (KOCHERLAKOTA and KOCHERLAKOTA, 1992), only a few of them, apart from the bivariate Poisson distribution, have been applied in the insurance field. It is worth noting the studys by PICARD (1976), LEMAIRE (1985) and PARTRAT (1993)

In this paper, we discuss the bivariate generalized Poisson distribution (BGPD) in detail. The distribution is derived from the generalized Poisson distribution (CONSUL, 1989; AMBAGASPITIYA and BALAKRISHNAN, 1994) using the trivariate reduction method. In section 2 we present some properties of the BGPD The method of moments is used in section 3 for estimation of the parameters We illustrate the usage of this method through two examples in section 4

2. BIVARIATE GENERALIZED POISSON DISTRIBUTION (BGPD)

2.1 Development of the distribution

We use the trivariate reduction method to construct the distribution (KOCHERLAKOTA and KOCHERLAKOTA, 1992). Let N_1, N_2 and N_3 be independent generalized Poisson random variables (GPD), $N_i \sim GPD\ (\lambda_i, \theta_i)$, $i = 1, 2, 3$. Let $X = N_1 + N_3$ and $Y = N_2 + N_3$ We get the joint probability function (p.f.) of (X, Y) as

$$P(X = r, Y = s) = \sum_{k=0}^{\min(r,s)} f_1(r-k) f_2(s-k) f_3(k), \tag{2.1}$$

where $f_i(n)$ is the p.f. of the random variable N_i

Since $N \sim GPD(\lambda, \theta)$, if its p f is given by (CONSUL and SHOUKRI, 1985)

$$f(n) = P(N = n) = \begin{cases} \dfrac{\lambda(\lambda + n\theta)^{n-1} \exp(-\lambda - n\theta)}{n!} \text{ for } n = 0,1,2,\ldots \\ 0 \quad , \quad \text{otherwise} \end{cases}, \tag{2.2}$$

where $\lambda > 0$, $\max(-1, -\lambda/m) \le \theta < 1$ and $m \ge 4$ is the largest positive integer for which $\lambda + \theta m > 0$ when $\theta < 0$, from (2.1) we have

$$P(X = r, Y = s) = p(r,s) = \lambda_1 \lambda_2 \lambda_3 \exp\{-(\lambda_1 + \lambda_2 + \lambda_3) - r\theta_1 - s\theta_2\}$$

$$\sum_{k=0}^{\min(r,s)} \frac{1}{(r-k)!(s-k)!k!} (\lambda_1 + (r-k)\theta_1)^{r-k-1} (\lambda_2 + (s-k)\theta_2)^{s-k-1} (\lambda_3 + k\theta_3)^{k-1} \tag{2.3}$$

$$\exp\{k(\theta_1 + \theta_2 - \theta_3)\},\ r, s \in N.$$

2.2 Properties of the distribution

Remark All the formulas that follows for the GPD are taken from AMBAGASPITIYA and BALAKRISHNAN (1994) and the general equations for a bidimensional distribution are from KOCHERLAKOTA and KOCHERLAKOTA (1992)

Probability generating function (pgf)

The pgf of a random variable N is defined by $\prod_N(t) = E(t^N)$ and the pgf of the pair of random variables (X, Y) is $\prod(t_1, t_2) = E(t_1^X t_2^Y)$

Let the pgf's of the random variables under consideration be $\prod_i(t)$, $i = 1, 2, 3$ Then the joint pgf of (X, Y) is

$$\prod(t_1, t_2) = \prod_1(t_1)\prod_2(t_2)\prod_3(t_1 t_2). \tag{2.4}$$

For simplicity, we assume the parameters $\theta_i > 0$, $i = 1, 2, 3$ AMBAGASPITIYA and BALAKRISHNAN (1994) has expressed the pgf of the GPD in terms of Lambert's W function when $\theta > 0$, as follows

$$\prod_N(t) = \exp\left\{-\frac{\lambda}{\theta}\left[W(-\theta z \exp(-\theta)) + \theta\right]\right\}, \tag{2.5}$$

where the Lambert's W function is defined as $W(x)\exp(W(x)) = x$. For more details about this function see CORLESS et al. (1994)

From (2 4) and (2 5), the pgf of (X, Y) is

$$\prod(t_1,t_2) = \exp\left\{-\frac{\lambda_1}{\theta_1}W(-\theta_1 t_1\exp(-\theta_1)) - \frac{\lambda_2}{\theta_2}W(-\theta_2 t_2\exp(-\theta_2)) - \right. \\ \left. -\frac{\lambda_3}{\theta_3}W(-\theta_3 t_1 t_2\exp(-\theta_3)) - \lambda\right\}, \tag{2.6}$$

with $\lambda = \lambda_1 + \lambda_2 + \lambda_3$.

Moment generating function (mgf)

If the mgf of N_i is $M_i(t)$, $i = 1, 2, 3$ then the mgf of (X, Y) is

$$M(t_1,t_2) = M_1(t_1)M_2(t_2)M_3(t_1+t_2) \tag{2.7}$$

The mgf of the GPD, when $\theta > 0$, is given by

$$M_N(t) = \exp\left\{-\frac{\lambda}{\theta}\left[W(-\theta\exp(-\theta+t)) + \theta\right]\right\} \tag{2.8}$$

Using (2.8) in (2.7) we get

$$M(t_1,t_2) = \exp\left\{-\frac{\lambda_1}{\theta_1}W(-\theta_1\exp(-\theta_1+t_1)) - \frac{\lambda_2}{\theta_2}W(-\theta_2\exp(-\theta_2+t_2)) - \right. \\ \left. -\frac{\lambda_3}{\theta_3}W(-\theta_3\exp(-\theta_3+t_1+t_2)) - \lambda\right\} \tag{2 9}$$

Moments

The expressions for the first four central moments of the GPD are as follows

$$\begin{aligned} E(N) &= \mu_1 = \lambda M \\ V(N) &= \mu_2 = \lambda M^3 \\ \mu_3 &= \lambda(3M-2)M^4 \\ \mu_4 &= 3\lambda^2 M^6 + \lambda(15M^2 - 20M + 6)M^5, \quad \text{where } M = (1-\theta)^{-1}. \end{aligned} \tag{2 10}$$

Since $X = N_1 + N_3$ and N_1, N_3 independent, we have $E(X) = E(N_1) + E(N_3)$ and $V(X) = V(N_1) + V(N_3)$, so that

$$\left\{\begin{aligned} E(X) &= \lambda_1 M_1 + \lambda_3 M_3 \\ V(X) &= \lambda_1 M_1^3 + \lambda_3 M_3^3 \\ E(Y) &= \lambda_2 M_2 + \lambda_3 M_3 \\ V(Y) &= \lambda_2 M_2^3 + \lambda_3 M_3^3 \end{aligned}\right\} \tag{2.11}$$

Let $\mu_{r,s} = E\left[(X-\mu_X)^r (Y-\mu_Y)^s\right]$ be the (r, s)[th] central moment of (X, Y). The equation for $\mu_{r,s}$ given $\mu_k^{(i)}$ the k[th] central moment of N_i, i = 1, 2, 3, is

$$\mu_{r,s} = \sum_{i=0}^{r}\sum_{j=0}^{s}\binom{r}{i}\binom{s}{j}\mu_i^{(1)}\mu_j^{(2)}\mu_{r+s-i-j}^{(3)}$$

Hence

$$\begin{Bmatrix} \mu_{11} = \lambda_3 M_3^3 \\ \mu_{21} = \mu_{12} = \lambda_3(3M_3 - 2)M_3^4 \end{Bmatrix} \tag{2.12}$$

This is enough to apply the method of moments.

Recurrence relations

The terms in the first row and column can be computed using the univariate generalized Poisson distribution, as is seen from

$$p(0,0) = \exp\{-\lambda\}$$

$$p(0,s) = \frac{\lambda_2(\lambda_2 + s\theta_2)^{s-1}}{s!}\exp\{-\lambda - s\theta_2\} = f(s,\lambda_2,\theta_2)\exp\{-(\lambda_1+\lambda_3)\}, \qquad s>0$$

$$p(r,0) = \frac{\lambda_1(\lambda_1 + r\theta_1)^{r-1}}{r!}\exp\{-\lambda - r\theta_1\} = f(r;\lambda_1,\theta_1)\exp\{-(\lambda_2+\lambda_3)\}, \qquad r>0$$

Given the probabilities in the first row and column, the probabilities for r ≥ 1, s ≥ 1 can be computed recursively as

$$p(r,s) = \lambda_3\exp\{\lambda\}\sum_{k=0}^{\min\{r,s\}}\frac{1}{k!}p(r-k,0)p(0,s-k)(\lambda_3 + k\theta_3)^{k-1}\exp\{-k\theta_3\}$$

Independence

Using (2 12) we have $cov(X,Y) = \lambda_3 M_3^3$, hence

$$\rho_{X,Y} = \frac{\lambda_3 M_3^3}{\left[(\lambda_1 M_1^3 + \lambda_3 M_3^3)(\lambda_2 M_2^3 + \lambda_3 M_3^3)\right]^{1/2}}.$$

Since $\lambda_3 \geq 0$ and $M_3 > 0$, it follows that for this model $\rho_{X,Y} \geq 0$. This shows that the condition of zero correlation is a necessary and sufficient condition for the independence of the random variables X and Y

Marginal distributions

The marginal distributions are.

$$P(X=r)=\lambda_1\lambda_3\exp\{-(\lambda_1+\lambda_3)-r\theta_3\}\sum_{i=0}^{r}\frac{(\lambda_1+i\theta_1)^{i-1}(\lambda_3+(r-i)\theta_3)^{r-i-1}}{i!(r-i)!}.$$
$$\exp\{-i(\theta_1-\theta_3)\}$$

$$P(Y=s)=\lambda_2\lambda_3\exp\{-(\lambda_2+\lambda_3)-s\theta_3\}\sum_{i=0}^{s}\frac{(\lambda_2+i\theta_2)^{i-1}(\lambda_3+(s-i)\theta_3)^{s-i-1}}{i!(s-i)!}.$$
$$\exp\{-i(\theta_2-\theta_3)\}.$$

In particular, if $\theta_1 = \theta_2 = \theta_3 = \theta$, this reduces to $X \sim GP(\lambda_1+\lambda_3, \theta)$ and $Y \sim GP(\lambda_2+\lambda_3, \theta)$.

3 ESTIMATION OF THE PARAMETERS: METHOD OF MOMENTS

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be a random sample of size n from the population. We will assume that the frequency of the pair (r, s) is n_{rs} for r = 0, 1, 2, . . , s = 0, 1, 2, . . We recall that $\sum_{r,s} n_{rs} = n$. Also

$$\begin{cases}\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{n}\sum_{r=0}r\ n_{r+}\ ,\quad \hat{\sigma}_X^2=\frac{1}{n}\sum_{r=0}(r-\bar{x})^2n_{r+}\\ \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}\sum_{s=0}s\ n_{+s}\ ;\quad \hat{\sigma}_Y^2=\frac{1}{n}\sum_{s=0}(s-\bar{y})^2n_{+s}\\ \hat{\mu}_{11}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{n}\sum_{r,s=0}r\ s\ n_{rs}-\bar{x}\ \bar{y}\\ \hat{\mu}_{21}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})=\frac{1}{n}\sum_{r,s=0}(r-\bar{x})^2(s-\bar{y})n_{rs}\end{cases}\quad (3.1)$$

The classical method of moments consists of equating the sample moments to their populations equivalents, expressed in terms of the parameters The number of moments required is six, equal to the number of parameters. Using (3 1), (2.11) and (2 12) we have

$$\begin{cases} \bar{x} = \lambda_1 M_1 + \lambda_3 M_3 \\ \bar{y} = \lambda_2 M_2 + \lambda_3 M_3 \\ \hat{\sigma}_X^2 = \lambda_1 M_1^3 + \lambda_3 M_3^3 \\ \hat{\sigma}_Y^2 = \lambda_2 M_2^3 + \lambda_3 M_3^3 \\ \hat{\mu}_{11} = \lambda_3 M_3^3 \\ \hat{\mu}_{21} = \lambda_3 (3M_3 - 2) M_3^4 \end{cases} \Leftrightarrow \begin{cases} M_3 = \dfrac{1 + \sqrt{1 + 3a}}{3} \\ \lambda_3 = \dfrac{\hat{\mu}_{11}}{M_3^3} \\ M_1 = \sqrt{\dfrac{\hat{\sigma}_X^2 - \hat{\mu}_{11}}{\bar{x} - \lambda_3 M_3}} \\ \lambda_1 = \dfrac{\bar{x} - \lambda_3 M_3}{M_1} \\ M_2 = \sqrt{\dfrac{\hat{\sigma}_Y^2 - \hat{\mu}_{11}}{\bar{y} - \lambda_3 M_3}} \\ \lambda_2 = \dfrac{\bar{y} - \lambda_3 M_3}{M_2} \end{cases}, \qquad (3.2)$$

where $a = \dfrac{\hat{\mu}_{21}}{\hat{\mu}_{11}}$

We use the fact that $\theta < 1$, so $M = \dfrac{1}{1-\theta} > 0$, when chosen the solution for M_i, $i = 1, 2, 3$.

4. NUMERICAL EXAMPLES

Example 1: The North atlantic coastal states in the USA (from Texas to Maine) can be affected by tropical cyclones. We divided these states into three geographical zones:

Zone 1. Texas, Louisiane, The Mississipi, Alabama;
Zone 2: Florida;
Zone 3: Other states

We were interested in studying the joint distribution of the pair (X, Y), where X and Y are the yearly frequency of hurricanes affecting respectively zone 1 and zone 3. To do that we used the data in table 1, first row in each cell, giving the realizations of (X, Y) observed during the 93 years from 1899 to 1991 (PARTRAT, 1993)

For these data we compute

$$\bar{x} = 0.74194, \quad \hat{\sigma}_X^2 = 0.62158, \quad \hat{\mu}_{11} = 0.02532,$$
$$\bar{y} = 0.47312, \quad \hat{\sigma}_Y^2 = 0.52885, \quad \hat{\mu}_{21} = 0.128341.$$

Under the hypothesis (X, Y) bivariate Poisson distributed $P_2(\lambda_1, \lambda_2, \mu)$, we have from PARTRAT (1993), method of maximum likelihood, the m l e $\hat{\lambda}_1 = 0.71876$,

$\hat{\lambda}_2 = 0.44994$, $\hat{\mu} = 0.02317$. The theoretical frequencies for $P_2\left(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\mu}\right)$ are given in table 1, middle row in each cell

TABLE 1

COMPARISON OF OBSERVED AND THEORETICAL YEARLY FREQUENCIES OF HURRICANES (1899-1991) HAVING AFFECTED ZONE 1 AND ZONE 3

Zone 3 / Zone 1	0	1	2	3	Σ
	27	9	3	2	41
0	28 24	12 71	2 86	0 48	44 29
	26 29	11 26	2 84	0 65	41 04
	24	13	1	0	38
1	20 30	9 79	2 35	0 42	32 86
	23 81	10 29	2 62	0 61	37 33
	8	2	1	0	11
2	7 29	3 75	0 96	0 19	12 19
	7 90	3 47	0 92	0 20	12 49
	1	0	2	0	3
3	2 12	1 16	0 32	0 06	3 66
	1 24	0 56	0 28	0 06	2 14
	60	24	7	2	
Σ	57 95	27 41	6 49	1 15	93
	59 24	25 58	6 66	1 52	

first row . observed frequency
middle row : theoretical frequency for P_2
last row . theoretical frequency for BGPD

The χ^2 goodness-of-fit test, after grouping in 7 categories (0, 0), (0, 1), (0, 2 and above), (1, 0), (1, 1), (2, 0), (other cases) to fulfill the Cochran criterium, lead us to $\chi^2_{obs} = \sum (obs - th)^2 / th = 5\ 96$ and a significance value $\hat{\alpha}$ verifying $0\ 20 \leq \hat{\alpha} \leq 0\ 54$.

We consider now the case of (X, Y) BGPD-distributed Then from the method of moments we have

$$\begin{cases} \lambda_1 = 0\ 81257, & \theta_1 = -0.10868 \\ \lambda_2 = 0.44555, & \theta_2 = 0.03995 \\ \lambda_3 = 0\ 00538, & \theta_3 = 0.40306 \end{cases}$$

The theoretical frequencies in this case are given in table 1, last row in each cell, and $\chi^2_{obs} = 2.66$ for the same categories; $0 \leq \hat{\alpha} \leq 0.85$.

Example 2: Automobile third party liability insurance.
The claims experience of a large automobile portfolio in France including 181038 liability policies was observed during the year 1989. The corresponding yearly claim frequencies, collected in table 2 (first row in each cell), have been divided into material damage only (type 1) and bodily injury (type 2) claims We obtain

$$\bar{x} = 0.05100,\quad \hat{\sigma}_X^2 = 0.05388,\quad \hat{\mu}_{11} = 0.00019,$$
$$\bar{y} = 0\ 00553,\quad \hat{\sigma}_Y^2 = 0.00552,\quad \hat{\mu}_{21} = 0.00023.$$

TABLE 2

COMPARISON OF OBSERVED AND THEORETICAL YEARLY FREQUENCIES

Type 2 / Type 1	0	1	2 and above	Σ
0	171345	918	2	172265 00
	171348 7	897 1	4 7	172250 50
	171348 7	897 5	4 6	172250 80
	171351 30	923 08	0 02	172274 40
1	8273	73	0	8346 00
	8275 5	86 3	0 7	8362 50
	8279 5	84 9	0 8	8365 20
	8248 39	71 01	0 14	8319 54
2	389	5	0	394 00
	398 2	6 2	0	404 40
	391 5	7 0	0 1	398 60
	415 41	3 52	1 37	420 30
3	31	1	0	32 00
	19 1	0 4	0	19 50
	21 3	0 6	0	21 90
	22 18	0 19	0 06	22 43
4 and above	1	0	0	1 00
	1 0	0 1	0	1 10
	1 4	0 1	0	1 50
	1 32	0 01	0	1 33
Σ	180039	997	2	
	180042 5	990 1	5 4	
	180042 4	990 1	5 5	181038.00
	180038 60	997 81	1 59	

first row observed frequency
second row · theoretical frequency for $P\text{–}G_2$
third row : theoretical frequency for $P\text{–}IG_2$
last row . theoretical frequency for BGPD

For the comparative study we have, from PARTRAT (1993)

- Bivariate Poisson Gamma $P\text{–}G_2\ (a;\ r,\ \beta)$ the m.l.e. $\begin{Bmatrix} \hat{a} = 0.10840 \\ \hat{r} = 1.00772 \\ \hat{\beta} = 19\ 75693 \end{Bmatrix}$.

The theoretical frequencies are provided in table 2, second row in each cell.

- Bivariate Poisson Inverse Gaussian $P\text{–}IG_2\ (a,\ \mu,\ \gamma)$ the m l.e. $\begin{Bmatrix} \hat{a} = 0\ 10840 \\ \hat{\mu} = 0\ 05101 \\ \hat{\gamma} = 0\ 05155 \end{Bmatrix}$.

The theoretical frequencies are provided in table 2, third row

Under the hypothesis (X, Y) BGPD, we have, using (3 1)
$\begin{Bmatrix} \hat{\lambda}_1 = 0\ 04945, & \hat{\theta}_1 = 0\ 02701 \\ \hat{\lambda}_2 = 0.00537, & \hat{\theta}_2 = -0\ 00266 \\ \hat{\lambda}_3 = 0.00016, & \hat{\theta}_3 = 0.04976 \end{Bmatrix}$, the theoretical frequencies are given in table 2, last row

The χ^2 goodness-of-fit test is applied on the 9 following categories: (0, 0), (0, 1), (0, 2 and above); (1, 0), (1, 1 and above); (2, 0); (3, 0); (4 and above, 0); (other cases) For this grouping we obtain

- In the $P\text{–}G_2$ case $\chi^2_{obs} = 11.94$ and a significance value $0.03 \le \hat{\alpha} \le 0\ 15$;
- In the $P\text{–}I\ G_2$ case. $\chi^2_{obs} = 8.8$ and a significance value $0.12 \le \hat{\alpha} \le 0.36$

In the BGPD case we used 7 categories (0, 0), (0, 1), (1, 0); (1, 1), (2, 0), (3, 0); (other cases), and we have $\chi^2_{obs} = 6\ 36$ with a significance value $0\ 00 \le \hat{\alpha} \le 0.4$.

REFERENCES

AMBAGASPITIYA, R S & BALAKRISHNAN, N (1994) On the compound generalized Poisson distributions *ASTIN Bulletin* **24**, 255-263

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

CONSUL, P C & SHOUKRI, M M (1985) The generalized Poisson distribution when the sample mean is larger than the sample variance *Communications in Statistics-Simulation and Computation* **14**, 1533-1547

CORLESS, R M , GONNET, G H , HARE, D E G & JEFFREY, D J (1994) The Lambert W function To appear in *Advances in Computational Mathematics*

KOCHERLAKOTA, S & KOCHERLAKOTA, K (1992) *Bivariate discrete distributions*, Marcel Dekker Inc

LEMAIRE, J (1985) *Automobile insurance Actuarial models*, Kluwer Publ

PARTRAT, C (1993) Compound model for two dependent kinds of claim, XXIVe *ASTIN Colloquium*, Cambridge

PICARD, P (1976) Generalisation de l'étude sur la survenance des sinistres en assurance automobile, *Bulletin de l'Institut des Actuaires Français*, **Vol. 297**, 204-267

RALUCA VERNIC
Department of Mathematics and Informatics
University "Ovidius" of Constanta
Bd Mamaia 124
8700 Constanta
Romania

ALLOWANCE FOR COST OF CLAIMS IN BONUS-MALUS SYSTEMS

JEAN PINQUET[1]

THEMA, University Paris X, 92001 Nanterre, France

ABSTRACT

The objective of this paper is to make allowance for cost of claims in experience rating. We design here a bonus-malus system for the pure premium of insurance contracts, from a rating based on their individual characteristics Empirical results are presented, that are drawn from a French data base of automobile insurance contracts.

KEYWORDS

Bayesian and heterogeneous models Number and cost residuals. Bonus-malus for frequency of claims, average cost per claim, and pure premium.

INTRODUCTION

Bayesian models lead to a posteriori ratemaking of insurance contracts (Buhlmann (1967)) Suppose that the number of claims follows a Poisson distribution. A bonus-malus system for the frequency of claims is obtained if we consider that the parameter follows a gamma distribution (see Lemaire (1985, 1995)) This model may include a ratemaking of policyholders on an individual basis, the parameter of the Poisson distribution depending then on rating factors (see Dionne et al (1989, 1992)).

The allowance for severity of claims in experience rating can be achieved by considering the dichotomy between claims with material damage only, and claims including bodily injury (see Lemaire (1995)) In this model, the number of claims that caused bodily injury follows a binomial distribution, the parameter of which follows a beta distribution.

In this paper, the severity of claims will be taken into account by using their cost. The analysis of cost of claims makes clearly appear a positive correlation between the average cost per claim and the frequency risk (see Renshaw (1994), Pinquet et al (1992)) An a priori ratemaking will therefore be influenced by the allowance for costs Concerning the third party liability guaranty, it can be noted that.

- The settlement of claims with material damage is performed partly through fixed amount compensations from an insurance company to the third party

[1] Thanks to Georges Dionne for motivating this work, as well as Christian Gouriéroux, Eric Renshaw and two anonymous referees for comments This research received financial support from the Fédération Française des Sociétés d'Assurance

- The amount of compensations related to claims including bodily injury depends on the social position of the victim

Hence, it is difficult to explain the cost of these claims by the rating factors, and we shall investigate the damage guaranty in the empirical part of the paper

Allowing for cost of claims in bonus-malus systems can be achieved in the following way. starting from a rating model based on the analysis of number and cost of claims, two heterogeneity components are added They represent unobserved factors, that are relevant for the explanation of the severity variables Later on, we shall refer to any variable explained by a rating model (number, cost of claim, total cost of claims, and so on) as a "severity variable". These unobserved factors are, for instance, annual mileage for number distributions, and speed (and the driver's behaviour in general) for number and cost distributions. A bonus-malus coefficient can be related to the credibility estimation of a heterogeneity component

In this paper, costs of claims are supposed to follow gamma or log-normal distributions The rating factors, as well as the heterogeneity component, are included in the scale parameter of the distribution Considering that the heterogeneity component also follows a gamma or log-normal distribution, a credibility expression is obtained, which provides a predictor of the average cost per claim for the following period. For instance, a cost-bonus will appear after the first claim if its cost is inferior to the estimation made by the rating model

Experience rating with a bayesian model is possible only if there is enough heterogeneity in the data For instance, in the negative binomial model without covariates, the estimated variance of the heterogeneity component is equal to zero if the variance of the number of claims is inferior to their mean (see Pinquet et al (1992)) In that case, a priori and a posteriori tariff structures are the same, and the bayesian model fails.

A sufficient condition for the existence of a bonus-malus system derived from a bayesian model is provided in section 2 3 The existence is equivalent to an overdispersion of residuals related to the severity variable. This approach allows one to test for the presence of a hidden information, that is relevant for the explanation of the severity variables.

The heterogeneity on distributions for severity variables, that is not explained by the rating factors, is revealed through experience on policyholders The paper investigates the rate of this revelation, which is found to be lower for average cost per claim than for the frequency

For the sample considered here, the unexplained heterogeneity related to costs is stronger for gamma than for log-normal distributions Besides, the latter family gives a better fit to the data.

If the heterogeneity components on number and cost distributions are independent, the bonus-malus coefficient for pure premium is the product of the coefficients related to frequency and expected cost per claim. But one may think that the behavior of the policyholder influences the two heterogeneity components in a similar way, and so that they are positively correlated

Lastly, this paper proposes a bonus-malus system for the pure premium of insurance contracts, that admits a correlation between the two components Although the likelihood of a model based on number and costs of claims is not analytically tractable in the presence of such a correlation, consistent estimators for the parameters exist. The correlation between the number and cost heterogeneity components appears to be very low for the sample investigated here

1 A PRIORI RATEMAKING

Let us suppose a sample of policyholders indexed by i, the policyholder i being observed during T_i periods The analysis of the correlation between the number and cost heterogeneity components shows the necessity of considering a non constant number of periods for each policyholder. The working sample is presented in 1 3

1.1 Frequency of claims

We write

$$N_{it} \sim P(\lambda_{it})_{t=1,\dots,T_i}, \lambda_{it} = \exp(w_{it}\,\alpha)$$

to represent the Poisson model where n_{it}, the outcome of N_{it}, is the number of claims reported by the policyholder i in period t The parameter λ_{it} is a multiplicative function of the explanatory variables, the line-vector w_{it} represents their values, and α is the column-vector of the related parameters.

The frequency-premium (estimation of the expectation of N_{it}) is denoted as $\hat{\lambda}_{it} = \exp(w_{it}\,\hat{\alpha})$, and $nres_{it} = n_{it} - \hat{\lambda}_{it}$ is the number-residual for the policyholder i and period t. The maximum likelihood estimator of α is the solution to the equation:

$$\sum_{i,t} nres_{it}\, w_{it} = 0,$$

which is an orthogonality relation between the explanatory variables and the residuals The rating factors have in general a finite number of levels, and the explanatory variables are then indicators of these levels The preceding equation means that, for every sub-sample associated to a given level, the sum of the frequency premiums is equal to the total number of claims This property means that the preceding model provides the multiplicative tariff structure that does not mutualize the frequency-risk.

One may think of replacing n_{it} by tc_{it}, the total cost of claims (pure premium ratemaking) in the likelihood equation. When applied to the working sample, this non probabilistic model shows that the elasticity of the pure premium risk with respect to the frequency risk is greater than one (see section 1.4.1).

1.2 Models for average cost per claim and pure premium

1.2.1 Gamma distributions

Let c_{itj} be the cost of the j^{th} claim reported by the policyholder i in period t ($1 \leq j \leq n_{it}$, *if* $n_{it} \geq 1$). We shall suppose in the paper that the costs are strictly positive. This assumption gives another reason to discard the third party liability guaranty· owing to fixed amount compensations, a policyholder involved in a claim caused by the third party can make his insurance company earn money.

Considering gamma distributions, we write

$$C_{itj} \sim \gamma(d, b_{it}), b_{it} = \exp(z_{it}\beta),$$

or $b_{it} C_{itj} \sim \gamma(d)$. The coefficient b_{it} is a scale parameter, a multiplicative function of the covariates, that are represented by the line-vector z_{it}.

Let $\hat{c}_{it} = \hat{d} / \hat{b}_{it} = \hat{d} / \exp(z_{it}\hat{\beta})$ be the estimation of the average cost for each claim reported by the policyholder i in period t. If we suppose that the costs are independent, the maximum likelihood estimator of β is the solution of the following equation.

$$\sum_{i,t} (n_{it} - (tc_{it} / \hat{c}_{it})) z_{it} = \sum_{i,t} cres_{it} \, z_{it} = 0$$

The term $n_{it} - (tc_{it} / \hat{c}_{it})$ is the sum, for the claims reported by the policyholder i in period t, of their cost residual $1 - (c_{itj} / \hat{c}_{it})$. it is written $cres_{it}$ The likelihood equation in β can hence be interpreted as an orthogonality relation between the explanatory variables and cost-residuals.

The average cost per claim increases with the frequency risk (see 1 4.2), which confirms the previous conclusions about the risks related to frequency and pure premium

1.2.2 Log-normal distributions

The other distribution family considered in this paper is the normal distribution family for the logarithms of costs

$$\log C_{itj} \sim N(z_{it}\beta, \sigma^2) \Leftrightarrow \log C_{itj} = z_{it}\beta + \varepsilon_{itj}, \ \varepsilon_{itj} \sim N(0, \sigma^2).$$

The likelihood equation giving $\hat{\beta}$ is

$$\sum_{i,t} \left(\sum_{j} (\log c_{itj} - z_{it}\hat{\beta}) \right) z_{it} = \sum_{i,t} lcres_{it} \, z_{it} = 0.$$

This equation is also an orthogonality relation between explanatory variables and residuals.

1.2.3 Pure premium model

The total cost of claims reported by the policyholder i in period t may be written as·

$$TC_{it} = \sum_{j=1}^{N_{it}} C_{itj}$$

It is a sum of N_{it} i.i.d outcomes from a variable that we denote as C_{it}. The pure premium is· $E(TC_{it}) = E(N_{it}) \, E(C_{it})$.

1.3 Presentation of the working sample

The sample investigated in the paper is part of the automobile policyholders portfolio of a French insurance company It is composed of more than a hundred thousand policyholders The damage guaranty being considered here, only the contracts with that kind of guaranty were kept Policyholders can be observed over two years, and each anniversary date, changing of vehicle or coverage level entails a new period. Only claims concerning the damage guaranty and closed at the date of obtention of the data base were kept Reserved costs were thus avoided The rating factors retained for the estimation of number and cost distributions are

- The characteristics of the vehicle. group, class, age
- The characteristics of the insurance contract· type of use, level of the deductible, geographic zone

Other rating factors are the policyholder's occupation, as well as the year when the period began (in order to allow for a generation effect) These eight rating factors have a finite number of levels, the total number of which is 44 The explanatory variables are binary, and indicate the levels for the policyholders· in order to avoid collinearity, one level is suppressed for each rating factor, the intercept being kept anyway. Therefore, we shall consider (44-8)+1=37 covariates. With the notations of the paper, we obtain: $\alpha, \beta \in \mathbb{R}^{37}; w_{it}, z_{it} \in \{0,1\}^{37}$.

The estimated coefficients derived from the rating model depend on the level suppressed for each rating factor. Results that are independent from the suppressions are obtained by dividing the coefficients by their mean in the multiplicative model. These standardized coefficients can be compared with the relative severity of the levels

The periods having not the same duration, the parameter of the Poisson distribution must be proportional to the duration. The results given on the frequencies remain unchanged if, d_{it} being the duration of period t for the policyholder i, we write· $\lambda_{it} = d_{it} \exp(w_{it}\,\alpha)$, and $\hat{\lambda}_{it} = d_{it} \exp(w_{it}\,\hat{\alpha})$

The working sample includes 38772 policyholders and 71126 policyholders-periods These policyholders reported 3493 claims The average duration of the periods is nine months, and the annual frequency of the claims is 6 7%.

1.4 Empirical results

1.4.1 A priori rating for frequency and pure premium

When applied to the number of claims or their total cost, the Poisson models provide standardized coefficients, that can be compared with the relative severity of the levels For almost each rating factor, the variance of the coefficients related to the levels is inferior to the variance of the relative severity For instance, for the "type of use" rating factor, one gets

frequency	relative severity	standardized coefficient
professional use	1.623	1 278
standard use	0 982	0 992

pure premium	relative severity	standardized coefficient
professional use	1 747	1.177
standard use	0.979	0 995

The distributions of the policyholders among the levels of the different rating factors are not independent from one another Policyholders with a professional use have, for the other rating factors, more risky levels than the other policyholders The Poisson model does not mutualize the risk: hence these policyholders have, with respect to other rating factors, a level of relative severity equal to (1.747/1 177) – 1 = 48 4% more than the average, in term of pure premium.

The elasticity of the pure premium with respect to the frequency risk is equal to 1 52 on the sample, and the difference from 1 is significant (the related Student statistic is equal to 5.93) Hence, if the frequency risk is multiplied by two, the average cost per claim increases by $2^{0\,52} - 1 = 43.5\%$, and the pure premium increases by 187%.

This positive correlation between the risks on frequency and average cost per claim is observed on each rating factor, except for the geographical zone

1.4.2 A priori rating for average cost per claim

On the sample of claims, the gamma model leads to the following results (rating factor: type of use)

average cost	relative severity	standardized coefficient
professional use	1.076	0 933
standard use	0 996	1 003

The estimated elasticity of the average cost per claim with respect to the frequency is equal to 0 51, which confirms the results obtained in the preceding section.

2 EXPERIENCE RATING FOR FREQUENCY AND AVERAGE COST PER CLAIM

2.1 Heterogeneous models

In a bayesian framework, the allowance for a hidden information, relevant for the rating of risks, can be performed in the following way

- the starting point is an a priori rating model If y represents the severity variable(s), the likelihood of y will be written $f_0(y/\theta_1, x)$, where x is the vector of explanatory variables, and θ_1 the vector of parameters related to them
- A heterogeneity component (scalar, or vector) is added to the model, which measures the influence that unobserved variables have on the severity distribution. If u is this component, a distribution of y conditional on u and the explanatory variables is defined, and we denote its likelihood as $f_c(y/\theta_1, x, u)$ In practice, the a priori distribution is equal to the distribution defined conditionally on u, for some value u^0 of u $f_c(y/\theta_1, x, u^0) = f_0(y/\theta_1, x) \forall \theta_1, x, y$ If u is a scalar, $u^0 = 0$ or 1, according to the fact that u is included additively or multiplicatively in the conditional distribution

- The credibility estimation of u_i, the heterogeneity component for the policyholder i, leads to a bonus-malus system. It rests on a heterogeneous model, in which u_i is the outcome of a random variable U_i, the $(U_i)_{i=1,\ldots,p}$ being i i d. and their distribution being parameterized by θ_2. The likelihood of y_i in the model with heterogeneity is obtained by integrating the conditional likelihood over U_i, that is to say

$$f(y_i / \theta, x_i) = E_{\theta_2}[f_*(y_i / \theta_1, x_i, U_i)],$$

with $\theta = (\theta_1, \theta_2)$. The heterogeneity component vector on number and cost distributions will be denoted, for the policyholder i

$$U_i = \begin{pmatrix} U_{ni} \\ U_{ci} \end{pmatrix}.$$

where n stands for the numbers and c for the costs The link between heterogeneous and bayesian models is made clear in the example that follows

2.2 Examples of heterogeneous models

2.2.1 Number of claims

With the notations of 1 1, the distributions defined conditionally on u_{ni} are

$$N_{it} \sim P(\lambda_{it} u_{ni}), \text{ with } U_{ni} \sim \gamma(a, a)$$

in the heterogeneous model The expectation of U_{ni} is equal to one, and its variance is $1/a$ On a period, the number of claims distribution is negative binomial in the heterogeneous model

The negative binomial model can be considered as a Poisson model with a random component, if we write $\lambda_{it} U_{ni} = \tilde{\lambda}_{it}$ If the intercept is the first of k explanatory variables, and if e_1 is the first vector of the canonical base of $\mathbb{R}^k$, we have

$$\tilde{\lambda}_{it} = \exp(w_{it}\,\alpha + \log(U_{ni})) = \exp(w_{it}\,(\alpha + \log(U_{ni})e_1)) = \exp(w_{it}\,\tilde{\alpha}_i)$$

In the last expression of $\tilde{\lambda}_{it}$, the parameter $\tilde{\alpha}_i = \alpha + \log(U_{ni})e_1$ is random, and the formulation is bayesian But it is less tractable than that of the heterogeneous model, as well for bonus-malus computations as for statistical inference.

2.2.2 Gamma distributions for costs of claims

The heterogeneous models that follow, which allow us to design bonus-malus systems for average cost per claim, suppose the independence of heterogeneity components on the number and costs distributions The empirical results presented later will make this assumption plausible.

For the gamma model and with the notations of 1.2 1, the distributions conditional on u_{ci} are

$$C_{itj} \sim \gamma(d, b_{it} u_{ci}), \text{ with } U_{ci} \sim \gamma(\delta, \delta)$$

in the heterogeneous model The heterogeneity component is included, as the rating factors, in the scale parameter of the distribution

In the heterogeneous model, one can write: $C_{itj} = D_{itj}/(b_{it}U_{ci})$, with $D_{itj} \sim \gamma(d)$, $U_{ci} \sim \gamma(\delta,\delta)$, D_{itj} and U_{ci} being independent The variable C_{itj} follows a GB2 distribution (see Cummins et al (1990)), and D_{itj} represents the relative severity of the claim.

2.2.3 Log-normal distributions for costs of claims

With the notations of 1.2 2, the heterogeneous model is

$$\log C_{itj} = z_{it}\beta + \varepsilon_{itj} + U_{ci},\ U_{ci} \sim N(0,\sigma_U^2),$$

where the ε_{itj} and U_{ci} are independent. The variable ε_{itj} represents the relative severity of the claim

The heterogeneous model used to design a bonus-malus system for pure premium will be presented after the empirical results related to the preceding models.

2.3 A sufficient condition for the existence of a bonus-malus system derived from a bayesian model

Experience rating with a bayesian model is possible only if there exists enough heterogeneity on the data Considering for instance the negative binomial model without covariates, the estimated variance of the heterogeneity component is equal to zero if the variance of the number of claims is lower than their mean (see Pinquet et al. (1992)). In that case, a priori and a posteriori tariff structures do not differ, and the bayesian model fails.

A sufficient condition for the existence of a bonus-malus system derived from a bayesian model is provided here: it will be applied later on to the models for number and cost of claims

Let us start from a heterogeneous model, as defined in 2 1 The heterogeneity component is supposed to be scalar, and its distribution is parameterized by the variance σ^2 The parameters of the model are $\theta = (\theta_1, \sigma^2)$ and we shall write $\hat{\theta}^0 = (\hat{\theta}_1^0, 0)$, $\hat{\theta}_1^0$ being the maximum likelihood estimator of θ_1 in the a priori rating model.

If the right-derivative, with respect to σ^2, of the log-likelihood is positive in $\hat{\theta}^0$, $\hat{\sigma}^2$ will be positive in the heterogeneous model. The existence of a bonus-malus system is hence related to the sign of a lagrangian, which is part of the score test for nullity of σ^2 (see Rao (1948), Silvey (1959)). With the notations of 2 1, and denoting the lagrangian as $\mathcal{L}$, one can prove:

$$\sum_i \log f(y_i/\hat{\theta}_1^0, \sigma^2, x_i) - \sum_i \log f_0(y_i/\hat{\theta}_1^0, x_i) = \mathcal{L}\sigma^2 + o(\sigma^2), \text{ with}$$

$$\mathcal{L} = \frac{1}{2}\sum_i (res_i^2 - s_i);$$

$$res_i = \left(\frac{\partial}{\partial u}\log f_*(y_i/\hat{\theta}_1^0, x_i, u)\right)_{u=u^0};\ s_i = -\left(\frac{\partial^2}{\partial u^2}\log f_*(y_i/\hat{\theta}_1^0, x_i, u)\right)_{u=u^0}$$

See Pinquet (1996b) for a proof, and references to a recent literature. The term res_t is a residual, which is related to those encountered in the likelihood equations for numbers and costs. The condition for existence of a bonus-malus system is

$$L > 0 \Leftrightarrow \sum_t res_t^2 > \sum_t s_t$$

It can be interpreted as an overdispersion condition on residuals.

2.4 Prediction with heterogeneous models and bonus-malus systems

Let us suppose a policyholder observed on T periods: $Y_T = (y_1, \ldots, y_T)$ is the sequence of severity variables, and $X_T = (x_1, \ldots, x_T)$ that of the covariates. The sequences X_T and Y_T take the place of x_t and y_t in the preceding sections. The date of forecast T must be explicited here, and the individual index can be suppressed, since the policyholder can be considered separately. Besides, belonging to the working sample is not mandatory for this policyholder.

We want to predict a risk for the period $T+1$, by means of a heterogeneous model. For the period t, this risk R_t is the expectation of a function of Y_t (y_t is the outcome of Y_t). For instance, Y_t is the sequence of both number and costs of claims in period t, and R_t, the pure premium, is the expectation of the total cost.

We now include a heterogeneity component u, as defined in 2.1. The distribution of Y_t conditional on u depends on θ_1, x_t and u. This applies to R_t, and we can write $R_t = h_{\theta_1}(x_t)\, g(u)$, for the three types of risk dealt with later (frequency of claims, average cost per claim, pure premium), g being a real-valued function.

A predictor for the risk in period $T+1$ can be written as $h_{\theta_1}(x_{T+1})\, \hat{g}^{T+1}(u)$, with $\hat{g}^{T+1}(u)$ a credibility estimator of $g(u)$, defined from:

$$\hat{g}^{T+1}(u) = \arg\min_a E_{\theta_2}\left[(g(U)-a)^2\, f_*(Y_T/\theta_1, X_T, U)\right],$$

$$f_*(Y_T/\theta_1, X_T, U) = \prod_{t=1}^{T} f_*(y_t/\theta_1, x_t, U).$$

The expectation is taken with respect to U, and one obtains

$$\hat{g}^{T+1}(u) = E_\theta[g(U)/X_T, Y_T] = \frac{E_{\theta_2}[g(U)\, f_*(Y_T/\theta_1, X_T, U)]}{E_{\theta_2}[\, f_*(Y_T/\theta_1, X_T, U)]},$$

the expectation of $g(U)$ for the posterior distribution of U. Replacing θ_1 and θ_2 by their estimations in the heterogeneous model, we obtain the a posteriori premium

$$\hat{R}_{T+1}^{T+1} = h_{\hat{\theta}_1}(x_{T+1}) E_{\hat{\theta}}[g(U)/X_T, Y_T],$$

computed for period $T+1$. It can be written as

$$\left(h_{\hat{\theta}_1}(x_{T+1}) E_{\hat{\theta}_2}[g(U)]\right) \times \frac{E_{\hat{\theta}}[g(U)/x_1, \ldots, x_T; y_1, \ldots, y_T]}{E_{\hat{\theta}_2}[g(U)]}$$

The first term is an a priori premium, based on the rating factors of the current period. The second one is a bonus-malus coefficient it appears as the ratio of two expectations of the same variable, computed for prior and posterior distributions Owing to the equality $E_\theta[E_\theta(g(U)/X_T,Y_T)] = E_\theta[g(U)] = E_{\theta_2}[g(U)]$, the rating is balanced.

2.5 Bonus-malus for frequency of claims

2.5.1 Theoretical results

With the notations of 2 2.1 and 2.4, we write: $y_t = n_t$, $x_t = w_t$, $\theta_1 = \alpha$; $R_t = E(N_t) = \lambda_t u$, $h_{\theta_1}(x_t) = \lambda_t$, $g(u) = u$; $X_T = (w_1, \ldots, w_T), Y_T = (n_1, \ldots, n_T)$. The posterior distribution of U is a $\gamma(a + \sum_t n_t, a + \sum_t \lambda_t)$ (see Dionne et al (1989, 1992)) Hence:

$$E_\theta[U/w_1, \ldots, w_T, n_1, \ldots, n_T] = \hat{u}^{T+1} = \frac{a + \sum_{t=1}^{T} n_t}{a + \sum_{t=1}^{T} \lambda_t} \tag{1}$$

Replacing λ_t by $\hat{\lambda}_t = \exp(w_t \hat{\alpha})$ and a by $\hat{a}$ in equation (1) leads to the bonus-malus coefficient. There will be a frequency-bonus if the estimator of $\hat{u}^{T+1} - 1$ is negative, or if the number-residual $\sum_t (n_t - \hat{\lambda}_t)$ is negative

Considering in equation (1) that N_t follows a Poisson distribution, with a parameter $\lambda_t u$, $\hat{u}^{T+1}$ converges towards u when T goes to $+\infty$ The heterogeneity on number distributions, which is not explained by the rating factors, is hence revealed completely with time. It may be interesting to investigate the distribution of bonus-malus coefficients on a portfolio of policyholders, as well as its time evolution (see section 2 5.2 for empirical results)

We explicit now the condition for existence of a bonus-malus system for frequencies On the working sample, and with the notations in 2 2.1, one can write

$$\log f_\kappa(y_i/\hat{\theta}_1^0, x_i, u) = \sum_t \left[n_{it}(\log \hat{\lambda}_{it} + \log u) - \hat{\lambda}_{it} u - \log(n_{it}!)\right],$$

with $\hat{\lambda}_{it} = \exp(w_{it}\hat{\alpha}^0)$, $\hat{\alpha}^0$ being the estimator of α in the a priori rating model With the notations of 2 3, and with $u^0 = 1$, we obtain

$$res_i = \sum_t \left(n_{it} - \hat{\lambda}_{it}\right), s_i = \sum_t n_{it}, L > 0 \Leftrightarrow \sum_i nres_i^2 > \sum_i n_i,$$

where $nres_i = \sum_t (n_{it} - \hat{\lambda}_{it})$ is the number-residual for policyholder i, and $n_i = \sum_t n_{it}$ is the number of claims reported by this policyholder on all periods This condition means that, considering the total number of claims, its variance is superior to its mean, the variance being calculated conditionally on the explanatory variables. This empirical overdispersion condition can be related to the theoretical overdispersion of the negative binomial model: if $N_i \sim P(\lambda_i U_i), U_i \sim \gamma(a,a)$(with $a = 1/\sigma^2$), one gets. $V(N_i) = \lambda_i + \lambda_i^2\sigma^2 > \lambda_i = E(N_i)$

A score test for nullity of σ^2 can be performed from the Lagrange multiplier $\mathcal{L} = (1/2)\sum_i (nres_i^2 - n_i)$ The previous remarks allow us to reject the nullity of σ^2 if $\mathcal{L}$ is large enough If the number of policyholders goes to infinity, $\xi^L = \mathcal{L}/\sqrt{\hat{V}(\mathcal{L})}$ converges towards a $N(0,1)$ distribution. One can prove that $\hat{V}(\mathcal{L}) = 1/2\sum_i \hat{\lambda}_i^2$, with $\hat{\lambda}_i = \sum_t \hat{\lambda}_{it}$ If $u_{1-\varepsilon}$ is the quantile at the level $1-\varepsilon$ of a $N(0,1)$ distribution, the null hypothesis $\sigma^2 = 0$ will be rejected at the level ε if $\xi^L \geq u_{1-\varepsilon}$.

Besides, the lagrangian provides an estimator of the parameters. Starting from $\hat{\alpha}^0$ and $\widehat{\sigma^2}^0 = 0$ in the algorithm of the likelihood maximisation, one gets at the following step

$$\hat{\alpha}^1 = \hat{\alpha}^0; \widehat{\sigma^2}^1 = \frac{\mathcal{L}}{\hat{V}(\mathcal{L})} = \frac{\sum_i nres_i^2 - n_i}{\sum_i \hat{\lambda}_i^2} = \frac{\sum_i \left[(n_i - \hat{\lambda}_i)^2 - n_i\right]}{\sum_i \hat{\lambda}_i^2} \tag{2}$$

The estimators $\hat{\alpha}^1$ and $\widehat{\sigma^2}^1$ can be shown to be consistent for the negative binomial model (see Pinquet (1996b) for demonstrations)

2.5.2 Empirical results

From the sample described in 1.3, we obtain

$$\sum_i nres_i^2 = \sum_i (n_i - \hat{\lambda}_i)^2 = 3709.24; \sum_i n_i = n = 3493,$$

and experience rating is possible for frequencies Without explanatory variables (apart from total duration of observation for each policyholder), one obtains: $\sum_i nres_i^2 = 3746\ 25$ The sum of square of residuals decreases when explanatory variables are added, and the condition for existence of a bonus-malus system is more restrictive when they are present. This is logical because they are a cause of heterogeneity on a priori distributions

Besides, $\sum_i \hat{\lambda}_i^2 = 389\ 48$, and the estimator of σ^2 given in (2) is

$$\hat{\sigma}^2 = \frac{\mathcal{L}}{\hat{V}(\mathcal{L})} = \frac{\sum_i nres_i^2 - \sum_i n_i}{\sum_i \hat{\lambda}_i^2} = \frac{216.24}{389\ 48} = 0.555.$$

As a comparison, the maximum likelihood estimation for the negative binomial model is $\hat{\sigma}^2 = 0\ 576$. The score test for nullity of σ^2 is based on the statistic

$$\xi^L = \frac{L}{\sqrt{\hat{V}(L)}} = \frac{\sum_i nres_i^2 - \sum_i n_i}{\sqrt{2\sum_i \hat{\lambda}_i^2}} = \frac{216.24}{\sqrt{778.96}} = 7.75,$$

and the null hypothesis is rejected. Examples of bonus-malus coefficients derived from the credibility formula are developped in actuarial and econometric literature (see Lemaire (1985), Dionne et al. (1989,1992))

Evolution throughout time of bonus-malus coefficients, as well as a posteriori premiums related to them, will be investigated for the risks related to frequency and average cost per claim. We consider here a simulated portfolio, derived from the working sample. In this portfolio, the characteristics of each policyholder in the sample are those of the first period, and we suppose that they remain unchanged. If this assumption does not hold individually, it is however plausible on the whole population. Investigating the distribution of bonus-malus coefficients in the heterogeneous model, one can measure their dispersion on the portfolio by estimating their coefficient of variation after T years (see Pinquet (1996a)). Considering the frequencies, with the tariff structure obtained in 1.4.1 and $\hat{\sigma}^2 = 0.576$, we obtain:

TABLE 1

REVELATION THROUGHOUT TIME OF HETEROGENEITY RELATED TO NUMBER DISTRIBUTIONS

	Coefficients of variation (frequency of claims) a priori premium 0.372				
	T=1	T=5	T=10	T=20	T=+∞
bonus-malus coefficient	0.144	0.300	0.392	0.494	0.759
a posteriori premium	0.411	0.515	0.590	0.673	0.891

The coefficient of variation is a measure of the relative dispersion of bonus-malus coefficients and premiums. Apart from the a priori premium, the elements of the preceding table are an estimation of the expectation in the heterogeneous model. After nine years, the relative dispersion of the bonus-malus coefficients exceeds that of the a priori premium. This means that, after nine years, the heterogeneity revealed by the observation of policyholders becomes more important than that explained by the rating factors.

2.6 Bonus-malus for average cost per claim (gamma distributions)

2.6.1 Theoretical results

With the notations in 2.2.2 and 2.4, we can write: $y_i = (c_{ij})_{j=1,\ldots,n_i}, x_i = z_i$; $R_i = E(C_{ij}) = d/(b_i u)$; $\theta_1 = (\beta, d)$; $h_{\theta_1}(x_i) = d/b_i$; $g(u) = 1/u$. The bonus-malus coefficient on average cost per claim for period $T+1$ is derived from the credibility estimator of $1/u$ Since the a priori distribution of U is a $\gamma(\delta,\delta)$, with a density proportional to $f_\delta(u)=\exp(-\delta u)u^{\delta-1}$, one gets:

$$f_\delta(u)\times f_*(Y_T/\theta_1,X_T,u)=\exp((\delta+\sum_{t,j}b_t c_{tj})u)u^{d(\sum_t n_t)+\delta-1},$$

times a coefficient independent of u The posterior distribution of U is therefore a $\gamma(\delta+d(\sum_t n_t),\ \delta+\sum_{t,j}b_t c_{tj})$, and:

$$\widehat{1/u}^{T+1}=E_\theta\left[\frac{1}{U}\Big/X_T,Y_T\right]=\frac{\delta+\sum_{t,j}b_t c_{tj}}{\delta-1+d(\sum_t n_t)}$$

We have $E_{\theta_2}(1/U)=\delta/(\delta-1)$ (we suppose $\delta>1$, a necessary condition for $1/U$ to have a finite expectation) Omitting the period index, and writing S_T for the set of claims reported by the policyholder during the first T periods, the bonus-malus coefficient is

$$\frac{E_{\hat\theta}\left[\frac{1}{U}\Big/X_t,Y_t\right]}{E_{\hat\theta_2}\left[\frac{1}{U}\right]}=\frac{\hat\eta+\sum_{j\in S_t}(c_j/E_{\hat\theta}(C_j))}{\hat\eta+|S_T|},\tag{3}$$

where we wrote: $\eta=(\delta-1)/d$, $E_\theta(C_j)=E_{\theta_2}(d/(b_jU))=(d/b_j)(\delta/(\delta-1))$. The rating structure derived from (3) is obviously balanced. Writing $E_{\hat\theta}(C_j)=\hat c_j$, and $cres_T=\sum_{j\in S_T}(1-(c_j/\hat c_j))$ the cost-residual for the policyholder, there will be a cost-bonus if the cost-residual is positive The bonus is then equal to

$$1-\frac{\hat\eta+\sum_{j\in S_T}c_j/\hat c_j}{\hat\eta+|S_T|}=\frac{cres_T}{\hat\eta+|S_T|}$$

The time evolution of the distribution of bonus-malus coefficients is investigated in 2 6 2 Considering the simulated portfolio defined in 2 5.2, the heterogeneity unexplained by the rating factors is revealed more slowly for cost than for number distributions This is not surprising, as far as no claim means no information on the cost distribution — if there is no correlation between the two heterogeneity components — whereas no claim generates frequency-bonus.

Let us apply to this model the condition allowing experience rating. For the working sample, we denote S_i as the set of claims reported by the policyholder over the T_i periods. One can write

$$\log f_*(y_i/\hat\theta_1^0,x_i,u)=\sum_{j\in S_i}(\hat d^0\log u-\hat b_{ij}^0 c_{ij}u)+z_i,$$

where z_i does not depend on u With the notations of 2 3 and with $u^0=1$, we obtain:

$$res_i = \sum_{j \in S_i} (\hat{d}^0 - \hat{b}_{ij}^0 c_{ij});\ s_i = n_i \hat{d}^0;\ L > 0 \Leftrightarrow \frac{1}{n} \sum_i cres_i^2 > \frac{1}{\hat{d}^0}$$

The total number of claims over the sample is n, and $cres_i$ is the cost-residual for the policyholder i This residual is equal to 0 without claims, and otherwise. $cres_i = \sum_{j \in S_i} (1 - (c_{ij} / \hat{c}_{ij}^0)) = \sum_{j \in S_i} cres_{ij}$, where $\hat{c}_{ij}^0 = \hat{d}^0 / \hat{b}_{ij}^0$ is the estimator for the expectation of C_{ij} Now, we have: $E(1 - (C_{ij} / E(C_{ij})))^2 = V(C_{ij}) / E^2(C_{ij}) = CV^2(C_{ij}) = 1/d$, if $C_{ij} \sim \gamma(d, b_{ij})$ The condition for existence of a bonus-malus system is hence related to the square of coefficients of variation

2.6.2 Empirical results

Considering the working sample, one obtains:

$$\frac{1}{n} \sum_i cres_i^2 = 1.092; \frac{1}{\hat{d}^0} = 0.821,$$

and experience rating for average cost of claims is possible For the sample of policyholders that reported claims, the maximum likelihood estimators for the GB2 model are.

$$\hat{\delta} = 3.620, \hat{d} = 1.807, \hat{\eta} = (\hat{\delta} - 1)/\hat{d} = 1.45.$$

The bonus (negative in case of malus) related to average cost per claim is equal to $cres_T / (\hat{\eta} + |S_T|)$ It remains equal to zero as long as there are no claims. After the first claim, if we consider the cases where the ratio actual cost-predicted cost is equal, either to 0.5 or to 2, the related cost-residuals are equal to 0.5 and -1 respectively The multiplicative coefficient $1/(1 + \hat{\eta})$ being equal to 0.408, we obtain a cost-bonus of 20.4% in the first case, and a cost-malus of 40.8% in the second case This coefficient is independent of the period during which the claim occurs

The distributions of bonus-malus coefficients and a posteriori premiums can be investigated on the simulated portfolio defined in 2 5 2 With the tariff structures obtained in 1 4 1 and 1 4.2 and $\hat{\delta} = 3.62$, we obtain (see Pinquet (1996a))

TABLE 2

REVELATION THROUGHOUT TIME OF HETEROGENEITY RELATED TO COST DISTRIBUTIONS

	Coefficients of variation (expected cost per claim) a priori premium 0 401				
	T=1	T=5	T=10	T=20	T=+∞
bonus-malus coefficient	0 128	0 268	0 356	0 453	0 786
a posteriori premium	0 427	0 504	0 568	0 648	0 937

The relative dispersion of the bonus-malus coefficients exceeds the dispersion of the a priori premium after fourteen years Unexplained heterogeneity on cost distributions is revealed more slowly than it was for numbers

2.7 Bonus-malus for average cost per claim (log-normal distributions)

2.7.1 Theoretical results

With the notations in 2.2.2 and 2.4, we write $y_t = (\log c_{tj})_{j=1,\dots,n_t}$; $x_t = z_t$, $\log C_{tj} \sim N(z_t\beta + u, \sigma^2) \Rightarrow R_t = E(C_{tj}) = \exp(z_t\beta + u + (\sigma^2/2))$, $\theta_1 = (\beta, \sigma^2)$, $h_{\theta_1}(x_t) = \exp(z_t\beta + (\sigma^2/2))$; $g(u) = \exp(u)$. The bonus-malus coefficient is derived from the credibility estimator of $\exp(u)$. Now

$$f_{\sigma_U^2}(u) \times f_*(Y_T/\theta_1, X_t, u) = \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_U^2} + \frac{tn_T}{\sigma^2}\right)\left(u - \frac{tlc_T - E_{\theta_1}(TLC_T)}{tn_T + (\sigma^2/\sigma_U^2)}\right)^2\right]$$

times a coefficient independent from u We wrote $tn_T = \sum_{t=1}^{T} n_t$, $tlc_T = \sum_{j\in S_T} \log c_j$, $E_{\theta_1}(TLC_T) = \sum_{j\in S_T} E_{\theta_1}(\log C_j)$; S_T is the set of claims reported by the policyholder during the T periods $(|S_T| = tn_T)$, and the period index is omitted Hence, the posterior distribution of U is

$$U/(X_T, Y_T) \sim N\left(\frac{tlc_T - E_{\theta_1}(TLC_T)}{tn_T + (\sigma^2/\sigma_U^2)}, \frac{1}{(1/\sigma_U^2) + (tn_T/\sigma^2)}\right)$$

The bonus-malus coefficient for period T+1 is equal to

$$\frac{E_{\hat{\theta}}[\exp(U)/X_T, Y_T]}{E_{\hat{\theta}_2}[\exp(U)]} = \exp\left[\frac{lcres_T - (tn_T\hat{\sigma}_U^2/2)}{\left(\hat{\sigma}^2/\hat{\sigma}_U^2\right) + tn_T}\right],$$

writing $lcres_T = \sum_{j\in S_T} lcres_j$, $lcres_j = \log c_j - E_{\hat{\theta}_1}(\log C_j)$.

The condition for existence of a bonus-malus sytem is easily interpretable with the log-normal model We have

$$\log f_*(y_t/\hat{\theta}_1^0, x_t, u) = -\sum_{j\in S_t} \frac{(lcres_{tj} - u)^2}{2\,\widehat{\sigma^2}^0}$$

plus terms that do not depend on u, with $lcres_{tj} = \log(c_{tj}) - z_{tj}\hat{\beta}^0$. With $u^0 = 0$ (see 2 3), the existence condition is·

$$\sum_t \frac{(\sum_{j\in S_t} lcres_{tj})^2}{(\widehat{\sigma^2}^0)^2} - \frac{n}{\widehat{\sigma^2}^0} = \frac{1}{(\widehat{\sigma^2}^0)^2}\left[\sum_t\left(\sum_{j\in S_t} lcres_{tj}\right)^2 - n\,\widehat{\sigma^2}^0\right] > 0$$

Now, in the a priori rating model. $n\widehat{\sigma^2}^0 = \sum_{t,j} lcres_{tj}^2$, with $\widehat{\sigma^2}^0$ the maximum likelihood estimator of σ^2. Experience rating is possible if

$\sum_i \left(\sum_{j\in S_i} lcres_{ij}\right)^2 - \sum_{i,j} lcres_{ij}^2$ is positive, that is to say if

$$\sum_{i/n_i\geq 2}\ \sum_{j,k\in S_i, j\neq k} lcres_{ij}\, lcres_{ik} > 0$$

This condition means that, for claims related to policyholders having reported several of them, cost-residuals have rather the same sign. If the first claim has a cost greater than its prediction, it will be the same on average for the following ones.

One can prove that, if L is the lagrangian with respect to σ_U^2, we have

$$\hat{V}(L) = \frac{\sum_i n_i(n_i - 1)}{2(\widehat{\sigma^2}^0)^2} \Rightarrow \widehat{\sigma_U^2}^1 = \frac{L}{\hat{V}(L)} = \frac{\sum_{i/n_i\geq 2}\ \sum_{j,k\in S_i, j\neq k} lcres_{ij}\, lcres_{ik}}{\sum_i n_i(n_i - 1)},$$

and that $\widehat{\sigma_U^2}^1$ is an consistent estimator of σ_U^2 (see Pinquet (1996a)). It appears to be the average, for the policyholders having reported several claims, of the product of residuals associated to couples of different claims

2.7.2 Empirical results

From the working sample, we obtain $\sum_{i/n_i\geq 2}\sum_{j,k\in S_i, j\neq k} lcres_{ij}\, lcres_{ik} = 100\ 80$, and experience rating is possible Hence

$$\widehat{\sigma_U^2}^1 = \frac{\sum_{i/n_i\geq 2}\ \sum_{j,k\in S_i, j\neq k} lcres_{ij}\, lcres_{ik}}{\sum_i n_i(n_i - 1)} = \frac{100\ 80}{590} = 0.171.$$

The nullity of σ_U^2 is tested for with $\xi^L = L/\sqrt{\hat{V}(L)} = 2.86$ The critical value for a one-sided test at a level of 5% is 1.645, and the null hypothesis is rejected The maximum likelihood estimators of σ_U^2 and σ^2 in the heterogeneous model are: $\hat{\sigma}_U^2 = 0\ 172$, $\hat{\sigma}^2 = 0.855$.

Bonus-malus coefficients can be computed from the examples considered with the gamma distributions (one claim, and a ratio actual cost-expected cost equal to 0 5 or 2) The residual associated to a claim is the logarithm of the latter ratio In the first case, the bonus-malus coefficient is equal to

$$\exp\left[\frac{lcres_T - (m_T\hat{\sigma}_U^2/2)}{\left(\hat{\sigma}^2/\hat{\sigma}_U^2\right) + m_T}\right] = \exp\left[\frac{-\log 2 - 0.086}{(0\ 855/0.172) + 1}\right] = 0\ 878,$$

and is associated to a cost-bonus of 12 2% In the second case, the bonus-malus coefficient is equal to 1 107, and implies a cost-malus of 10 7% These results can be compared with 20 4% and 40.8%, the boni and mali derived from the gamma distributions, although the ratios actual cost-expected cost are different in the two models. They must be different, since the cost-residuals in the gamma and log-normal models are equal to $1-(c_{ij}/\hat{c}_{ij}^{gamma})$ and $\log(c_{ij}/\hat{c}_{ij}^{\log-normal})$ respectively, whereas they fulfill the same orthogonality relations with respect to the covariates.

Considering the simulated portfolio defined in 2.5.2, the heterogeneity on cost distributions that is unexplained by the a priori rating model is more important for gamma than for log-normal distributions This can be seen by comparing the limits of the coefficients of variation for the bonus-malus coefficients, as we did in sections 2 5 2 and 2 6 2 For the GB2 model, this limit is the coefficient of variation of $1/U, U \sim \gamma(\hat{\delta},\hat{\delta})$ (see Pinquet (1996a)) With $\hat{\delta}=3.62$, it is equal to $1/\sqrt{\hat{\delta}-2}=0\,786$ Considering the log-normal model, the limit is the coefficient of variation of exp (U), $U \sim N(0,\hat{\sigma}_U^2)$

With $\hat{\sigma}_U^2=0.172$, it is equal to $\sqrt{\exp(\hat{\sigma}_U^2)-1}=0\,433$.

This result can be related to a comparison between the two a priori rating models If F_{θ_1,x_j} is the continuous distribution function of Y_j (here equal to the cost of the claim j, or its logarithm) $\varepsilon_j=F_{\theta_1,x_j}(Y_j)$ is uniformly distributed on [0,1] Computing the residuals $e_j, e_j=F_{\hat{\theta}_1^0,x_j}(Y_j)$, and rearranging e_j in the increasing order, by $e_{(1)}\leq \ldots \leq e_{(n)}$, we derive the Komolgorov-Smirnov statistic $KS=\sqrt{n}\max_{1\leq j\leq n}|(j/n)-e_{(j)}|$ We obtain KS=2 83 (resp KS=1.04) for the gamma (resp log-normal) distribution family. The latter family seems to fit the data better than the gamma family, and will be retained for the bonus-malus system on pure premium

The two last results can be related to each other. there is more unexplained heterogeneity for gamma than for log-normal distributions, and the latter provide a better fit to the data This fact raises a question: is apparent heterogeneity only explained by hidden information, or can it be also explained by the fact that the model does not make the best use of observable information?

3 BONUS-MALUS FOR PURE PREMIUM

3.1 The heterogeneous model

From the preceding results, we shall retain log-normal rather than gamma distributions for costs Besides, they are better integrated in a heterogeneous model with a joint distribution for the two heterogeneity components related to the number and cost distributions We retain here a bivariate normal distribution The parameters of the related heterogeneous model can be estimated consistently, although the likelihood is not analytically tractable

A way to derive consistent estimators for heterogeneous models is proposed in Pinquet (1996b) It is based on the properties of extremal estimators, the maximum likelihood estimator being of this type. The estimators of the parameters of the a priori rating model have a limit if the actual distributions include heterogeneity, and this limit is tractable in the model investigated here Consistent estimators are then obtained from a method of moments using the scores with respect to the variances and the covariances of the heterogeneity components

The heterogeneous model is hence composed of Poisson distributions on numbers, log-normal distributions on costs, and of bivariate normal distributions for the two heterogeneity components. The notations are the following.

- The distributions conditional on u_{ni} and u_{ci}, the heterogeneity components for number and cost distributions of the policyholder i, are

$$N_{it} \sim P(\lambda_{it}\exp(u_{ni})), \log C_{itj} = z_{it}\beta + \varepsilon_{itj} + u_{ci}, \text{ with}$$

$$\lambda_{it} = \exp(w_{it}\alpha), \varepsilon_{itj} \sim N(0,\sigma^2), t = 1,\ldots,T_i; j = 1, \ldots, n_{it}$$

- In the heterogeneous model, U_{ni} and U_{ci} follow a bivariate normal distribution with a null expectation and a variance equal to

$$V = \begin{pmatrix} V_{nn} & V_{nc} \\ V_{cn} & V_{cc} \end{pmatrix}.$$

The parameters of the model are

$$\theta_1 = \begin{pmatrix} \alpha \\ \beta \\ \sigma^2 \end{pmatrix}, \theta_2 = \begin{pmatrix} V_{nn} \\ V_{cn} \\ V_{cc} \end{pmatrix}$$

Bonus-malus coefficients are computed in the heterogeneous model from the expression given in section 2.4

$$\frac{E_{\hat{\theta}}[g(U)/X_I, Y_T]}{E_{\hat{\theta}_2}[g(U)]} = \frac{E_{\hat{\theta}_2}\left[g(U)f(Y_I/\hat{\theta}_1, X_T, U)\right]}{E_{\hat{\theta}_2}[g(U)]\,E_{\hat{\theta}_2}\left[g(U)f(Y_I/\hat{\theta}_1, X_T, U)\right]} \tag{4}$$

We can write.

- $g(u_n, u_c) = \exp(u_n)$ for frequency
- $g(u_n, u_c) = \exp(u_c)$ for average cost per claim
- $g(u_n, u_c) = \exp(u_n + u_c)$ for pure premium,

because the expectations of N_t, C_{tj} and TC_t are respectively proportional to $\exp(u_n)$, $\exp(u_c)$ and $\exp(u_n + u_c)$, if computed conditionally on u_n and u_c The mathematical expectations that lead to the bonus-malus coefficients (see equation (4)) can be estimated if we can write $U = f_{\theta_2}(S)$, where the distribution of S is independent from θ_2 it is enough to simulate outcomes of S Such an expression can be obtained by writing the Choleski decomposition of the variances-covariances matrix, i e.

$$V = \begin{pmatrix} V_{nn} & V_{nc} \\ V_{cn} & V_{cc} \end{pmatrix} = T_\varphi T_\varphi'; \; T_\varphi = \begin{pmatrix} \varphi_{nn} & 0 \\ \varphi_{cn} & \varphi_{cc} \end{pmatrix} \Rightarrow V = \begin{pmatrix} \varphi_{nn}^2 & \varphi_{nn}\varphi_{cn} \\ \varphi_{nn}\varphi_{cn} & \varphi_{nn}^2 + \varphi_{cc}^2 \end{pmatrix}$$

One can write for the policyholder i

$$U_i = \begin{pmatrix} U_{ni} \\ U_{ci} \end{pmatrix} = T_\varphi S_i ; S_i = \begin{pmatrix} S_{ni} \\ S_{ci} \end{pmatrix}, S_i \sim N(0, I_2),$$

and we have $U_i = f_{\theta_2}(S_i)$, φ being related to V, hence to θ_2. The likelihood used in the bonus-malus expression (see equation (4)) is obtained as the product of the likelihoods related to numbers and costs With the notations of 2 4, we have

$$\log f_*(Y_T / \theta_1, X_T, U) =$$

$$-\left(\sum_t \lambda_t\right)\exp(U_n) + \left(\sum_t n_t\right) U_n - \sum_{t,j} \frac{(\log c_{tj} - z_t\beta - U_c)^2}{2\sigma^2}, \text{ with}$$

$$X_T = (x_1, \ldots, x_T); x_t = (w_t, z_t), Y_T = (y_1, \ldots, y_T), y_t = (n_t, (c_{tj})_{j=1,\ldots n_t}),$$

plus terms that do not depend on the heterogeneity components Replacing θ_1 by $\hat{\theta}_1$, we obtain

$$f_*(Y_T/\hat{\theta}_1, X_T, U) = \exp(V_T) \times \text{terms independent from } U, \text{ with}$$

$$V_T = -\left(\sum_t \hat{\lambda}_t\right)\exp(U_n) + m_T U_n - \frac{m_T U_c^2 - 2U_c lcres_T}{2\hat{\sigma}^2} \tag{5}$$

A bonus-malus coefficient for a policyholder and for the period T+1 depends then on:

- $\sum_t \hat{\lambda}_t$, which is proportional to the frequency premium of the policyholder on all periods This premium is equal to

$$\hat{E}(TN_T) = \sum_t \hat{\lambda}_t \hat{E}\left[\exp(U_n)\right] = \left(\sum_t \hat{\lambda}_t\right)\exp\frac{\hat{\varphi}_{nn}^2}{2} = \left(\sum_t \hat{\lambda}_t\right)\exp\frac{\hat{V}_{nn}}{2}.$$

- m_T, the number of claims reported by the policyholder during the T periods
- $lcres_T$, the sum of residuals on the logarithm of costs of claims reported by the policyholder it represents their relative severity.

From equation (4), bonus-malus coefficients on frequency, expected cost per claim, and pure premium are respectively equal to

$$\frac{\hat{E}[\exp(U_n + V_T)]}{\hat{E}[\exp(U_n)]\,\hat{E}[\exp(V_T)]}, \frac{\hat{E}[\exp(U_c + V_T)]}{\hat{E}[\exp(U_c)]\,\hat{E}[\exp(V_T)]}, \frac{\hat{E}[\exp(U_n + U_c + V_T)]}{\hat{E}[\exp(U_n + U_c)]\,\hat{E}[\exp(V_T)]}.$$

The coefficients are estimated by simulations of outcomes of S_n and S_c For instance, we infer that the estimated covariance

$$\widehat{Cov}\left(\frac{\exp(U_n)}{E[\exp(U_n)]}, \frac{\exp(V_T)}{E[\exp(V_T)]}\right)$$

is a frequency-malus The existence of boni and mali for the different risks can be interpreted through the sign of estimated covariances

The a posteriori premium is obtained by the expression given in section 2 4

$$\hat{R}_{t+1}^{T+1} = \left(h_{\hat{\theta}_1}(x_{T+1}) E_{\hat{\theta}_2}[g(U)]\right) \frac{E_{\hat{\theta}}[g(U)/X_T, Y_T]}{E_{\hat{\theta}_2}[g(U)]}$$

The first term is the a priori premium It is an estimation of

$$\lambda_{T+1} \exp(z_{T+1}\beta) E[\exp(U_n + U_c)] = \exp\left(w_{T+1}\alpha + z_{T+1}\beta + \frac{(\varphi_{nn} + \varphi_{cn})^2 + \varphi_{cc}^2}{2}\right),$$

because $U_n + U_c = (\varphi_{nn} + \varphi_{cn}) S_n + \varphi_{cc} S_c$.

Besides, $(\varphi_{nn} + \varphi_{cn})^2 + \varphi_{cc}^2 = V_{nn} + 2V_{cn} + V_{cc}$.

We should have consistent estimators for the parameters, in order to derive bonus-malus coefficients. A method to obtain such estimators was quoted in the introduction. When applied to the preceding model, it leads to the following results

We write $\hat{\alpha}^0$, $\hat{\beta}^0$, $\widehat{\sigma^2}^0$ the estimators of the parameters in the a priori rating model, and $\hat{\lambda}_i = \sum_t \exp(w_{it}\hat{\alpha}^0)$, $tlc_i = \sum_{t,j} \log(c_{itj})$, $E_{\theta_1}(TLC_i) = \sum_t n_{it} z_{it}\beta$, $t\hat{l}c_i = E_{\hat{\theta}_1^0}(TLC_i) = \sum_t n_{it} z_{it}\hat{\beta}^0$

The variances and covariances of the two heterogeneity components are consistently estimated by:

$$\hat{V}_{nn} = \log(1 + \hat{V}_{nn}^1),\ \hat{V}_{nn}^1 = \frac{\sum_i (n_i - \hat{\lambda}_i)^2 - n_i}{\sum_i \hat{\lambda}_i^2};\ \hat{V}_{cn} = \frac{\sum_i (n_i - \hat{\lambda}_i)(tlc_i - t\hat{l}c_i)}{\left(\sum_i \hat{\lambda}_i^2\right)(1 + \hat{V}_{nn}^1)},$$

$$\hat{V}_{cc} = \frac{\sum_i \left[(tlc_i - t\hat{l}c_i)^2 - n_i \widehat{\sigma^2}^0\right]}{\left(\sum_i \hat{\lambda}_i^2\right)(1 + \hat{V}_{nn}^1)} - \hat{V}_{cn}^2 \tag{6}$$

Consistent estimators of φ_{nn}, φ_{cn} and φ_{cc} are given by the solutions of the equation

$$T_{\hat{\varphi}} T'_{\hat{\varphi}} = \hat{V}$$

The estimators of φ are used in the computation of bonus-malus coefficients, remember that $U_i = T_\varphi S_i$ ($S_i \sim N(0, I_2)$), and that the coefficients are estimated through simulations of outcomes of S_i. As for the parameters of the a priori rating model, they are consistently estimated by

$$\hat{\alpha} = \hat{\alpha}^0 - \frac{\hat{V}_{nn}}{2} e_{n,1},\ \hat{\beta} = \hat{\beta}^0 - \hat{V}_{cn} e_{c,1},\ \hat{\sigma}^2 = \widehat{\sigma^2}^0 - \hat{V}_{cc} \tag{7}$$

The intercepts are supposed to be the first of the k_n and k_c explanatory variables for the number and cost distributions, and $e_{n,1}$ (resp $e_{c,1}$) are the first vectors of the canonical base of $\mathbb{R}^{k_n}$ (resp $\mathbb{R}^{k_c}$)

3. 2 Empirical results

The numerical results $\sum_i (n_i - \hat{\lambda}_i)^2 - n_i = 216.24$; $\sum_i \hat{\lambda}_i^2 = 389\,48$, already used for bonus-malus on frequencies, lead to.

$$\hat{V}_{nn}^1 = \frac{\sum_i (n_i - \hat{\lambda}_i)^2 - n_i}{\sum_i \hat{\lambda}_i^2} = 0.555, \hat{V}_{nn} = \log(1 + \hat{V}_{nn}^1) = 0\,442 \Rightarrow \hat{\varphi}_{nn} = \sqrt{\hat{V}_{nn}} = 0\,665$$

In this paper, two distribution families are considered for the heterogeneity component related to numbers We first took into account the gamma, and now the log-normal family (writing the heterogeneity component in a multiplicative way)

Considering an insurance contract without claims, we can compare the boni derived from the two models The sum $\sum_t \hat{\lambda}_t$ being the cumulated frequency premium in the negative binomial model, the bonus for the policyholder is equal to

$$1 - \frac{\hat{a}}{\hat{a} + \sum_t \hat{\lambda}_t} = \frac{\sum_t \hat{\lambda}_t}{\hat{a} + \sum_t \hat{\lambda}_t} = \frac{\hat{V}_{nn}^1 \sum_t \hat{\lambda}_t}{1 + \left(\hat{V}_{nn}^1 \sum_t \hat{\lambda}_t\right)}, (\hat{a} = 1 / \hat{V}_{nn}^1).$$

For the log-normal family, the bonus can be written as

$$-\widehat{Cov}\left(\frac{\exp(U_n)}{E[\exp(U_n)]}, \frac{\exp(V_T)}{E[\exp(V_T)]}\right), U_n = \varphi_{nn} S_n, V_T = -\sum_t \hat{\lambda}_t \exp(U_n),$$

with $S_n \sim N(0,1)$ With the values of $\hat{V}_{nn}^1$ and $\hat{\varphi}_{nn}$ computed precendently, one obtains for example

TABLE 3

COMPARISON OF FREQUENCY-BONUS COEFFICIENTS FOR TWO DISTRIBUTIONS ON THE HETEROGENEITY COMPONENT (CONTRACTS WITHOUT CLAIMS REPORTED)

frequency premium	**0.05**	**0.1**	**0.2**	**0.5**	**1**	**2**
bonus (%, gamma distributions)	2 7	5 3	10	21 7	35 7	52 6
bonus (%, log-normal distributions)	2 6	5 1	9 4	19 3	30 3	43 6

The boni derived from log-normal distributions on the heterogeneity component are lower than those derived from the gamma distributions. The difference is all the more important since the frequency premium is high

Let us estimate the covariance between the two heterogeneity components:

$$\sum_i (n_i - \hat{\lambda}_i)(tlc_i - \hat{tlc}_i) = 7.96 \Rightarrow \hat{V}_{cn} = \frac{\sum_i (n_i - \hat{\lambda}_i)(tlc_i - \hat{tlc}_i)}{\left(\sum_i \hat{\lambda}_i^2\right)(1 + \hat{V}_{nn}^1)} = 0\ 013.$$

One can think of relating a positive or negative sign of the covariance to the fact that the average cost per claim increases or decreases with the number of claims reported by the policyholder To see this, suppose that the duration of observation is the same for all the policyholders, and that the intercept is the only explanatory variable for number and cost distributions We would then have

$$\hat{\lambda}_i = \bar{n}, \hat{tlc}_i = n_i \overline{\log c} \Rightarrow \sum_i (n_i - \hat{\lambda}_i)(tlc_i - \hat{tlc}_i) = \sum_i (n_i - \bar{n}) n_i (\overline{\log c}^i - \overline{\log c}) =$$

$$\sum_{i / n_i \geq 2} (n_i - 1) n_i (\overline{\log c}^i - \overline{\log c}), \text{ because } \sum_i n_i (\overline{\log c}^i - \overline{\log c}) = 0.$$

We wrote $\overline{\log c}^i$ for the logarithms of costs of claims reported by the policyholder i, computed on average. The estimator of the covariance would be positive if the average of the logarithms of costs of claims related to the policyholders that reported several of them was superior to the global mean

On the working sample, the number of claims reported by the policyholder had little influence on the average cost

The preceding results justify the allowance for a non constant number of periods related to the observation of policyholders To see this, we remark that the more severe is a claim, the greater is the probability to change the vehicule afterwards. Hence, there is less severity on average for several claims reported on the same car If policyholders were not kept in the sample after changing cars, a negative bias would appear in the estimation of the correlation coefficient between the heterogeneity components. Now, keeping the policyholder in the sample as long as possible leads us to consider a non constant number of periods.

When computing bonus-malus coefficients for average cost per claim, we used (see 2 7 2)

$$\sum_i \left[(tlc_i - \hat{tlc}_i)^2 - n_i \widehat{\sigma^2}^{(0)} \right] = \sum_{i / n_i \geq 2} \sum_j \sum_{k \in S_i, j \neq k} lcres_{ij} lcres_{ik} = 100\ 80$$

A bonus-malus system for average cost per claim can be considered if the observation of the ratio actual cost-expected cost for a claim brings information for the following claims. If the last expression is positive, the cost residuals of claims related to policyholders having reported several of them have rather the same sign The relative severity of a claim is associated to the sign of the residual, and it may be interesting to compare the sign of residuals for claims related to policyholders having reported two of them.

Considering the working sample, we obtain

number of policyholders having reported two claims	negative residual (second claim)	positive residual (second claim)
negative residual (first claim)	74	46
positive residual (first claim)	36	70

The sign of the residual does not change for 64% of policyholders having reported two claims

From equation (6), we infer

$$\hat{V}_{cc} = \frac{\sum_i (tlc_i - t\hat{l}c_i)^2 - n_i \widehat{\sigma^2}^0}{\left(\sum_i \hat{\lambda}_i^2\right)(1+\hat{V}_{nn}^1)} - \hat{V}_{cn}^2 = 0\ 166, \text{ and } \hat{r}_{cn} = \frac{\hat{V}_{cn}}{\sqrt{\hat{V}_{cc}\hat{V}_{nn}}} = 0\ 048$$

The correlation coefficient between the heterogeneity components is positive, but close to zero Hence

$$\hat{V}_{cn} = \hat{\varphi}_{nn}\hat{\varphi}_{cn} \Rightarrow \hat{\varphi}_{cn} = 0.020,\ \hat{V}_{cc} = \hat{\varphi}_{cn}^2 + \hat{\varphi}_{cc}^2 \Rightarrow \hat{\varphi}_{cc} = 0\ 407$$

The boni for average cost per claim and pure premium for the contracts without claims can be computed, and results can be compared to those obtained for frequency. From the expressions

$$-\widehat{Cov}\left(\frac{\exp(U_c)}{E[\exp(U_c)]}, \frac{\exp(V_T)}{E[\exp(V_T)]}\right),\ -\widehat{Cov}\left(\frac{\exp(U_n+U_c)}{E[\exp(U_n+U_c)]}, \frac{\exp(V_T)}{E[\exp(V_T)]}\right)$$

we obtain

TABLE 4

BONI FOR AVERAGE COST PER CLAIM AND PURE PREMIUM (CONTRACTS WITHOUT CLAIM REPORTED)

frequency premium	**0.05**	**0.1**	**0.2**	**0.5**	**1**	**2**
average cost per claim bonus (%)	0 1	0 1	0 2	0 5	0 9	1 5
pure premium bonus (%)	2 7	5 3	9 7	19 9	31 2	44 7

Because of the positive correlation between the two heterogeneity components, a cost-bonus appears in the absence of claims, but it is very low.

We now compute bonus-malus coefficients for policyholders that reported one claim They are a function of the cost-residual $lcres_T = \log(c_1) - z_1\hat{\beta}$ (c_1 is the cost of the claim, and z_1 represents the policyholder's characteristics when the claim occured), and of the frequency premium From equations (5) and (7), we have

$$V_T = -\sum_t \hat{\lambda}_t \exp(U_n) + U_n - \frac{U_c^2 - 2U_c lcres_T}{2\hat{\sigma}^2},$$

$$\hat{\sigma}^2 = \widehat{\sigma^2}^0 - \hat{V}_{cc} = \frac{\sum_{i,j} lcres_{ij}^2}{n} - \hat{V}_{cc} = \frac{3588}{3493} - 0.166 = 0.861$$

We recall that the bonus-malus coefficients on frequency, expected cost per claim and pure premium are respectively equal to

$$\frac{\hat{E}[\exp(U_n + V_T)]}{\hat{E}[\exp(U_n)]\,\hat{E}[\exp(V_T)]};\ \frac{\hat{E}[\exp(U_c + V_T)]}{\hat{E}[\exp(U_c)]\,\hat{E}[\exp(V_T)]};\ \frac{\hat{E}[\exp(U_n + U_c + V_T)]}{\hat{E}[\exp(U_n + U_c)]\,\hat{E}[\exp(V_T)]}.$$

We obtain for example (the bonus-malus coefficients are given in percentage)

TABLE 5
BONUS-MALUS COEFFICIENTS (POLICYHOLDERS HAVING REPORTED ONE CLAIM)

frequency coefficient			**frequency premium**			
$lcres_T$	**0.05**	**0.1**	**0.2**	**0.5**	**1**	**2**
-1	147 4	142 1	133 1	113 9	94 5	73 4
-0 5	148 4	143	133 8	114 5	95	73 7
0	149 3	143 7	134 6	115	95 3	74
0 5	150 1	144 6	135 3	115 6	95 7	74 3
1	151	145 6	136	116 1	96 2	74 6

average cost per claim coefficient			**frequency premium**			
$lcres_T$	**0.05**	**0.1**	**0.2**	**0.5**	**1**	**2**
-1	84 8	84 7	84 6	84 3	84	83 5
-0 5	92	91 9	91 7	91 4	91	90 5
0	99 7	99 6	99 5	99 1	98 7	98 1
0 5	108 1	108	107 8	107 5	107	106 4
1	117 1	117	116 9	116 5	116	115 4

pure premium coefficient			**frequency premium**			
$lcres_T$	**0.05**	**0.1**	**0.2**	**0.5**	**1**	**2**
-1	124 6	120	112 2	95 6	78 9	60 9
-0 5	136 1	131	122 3	104 2	86	66 3
0	148 4	142 7	133 3	113 5	93 5	72 2
0 5	161 8	155 7	145 4	123 7	101 9	78 5
1	176 6	170	158 4	134 7	111	85 4

Because of the positive correlation between the two heterogeneity components, the frequency coefficients increase with the cost-residual, which is related to the severity of the claim In the same way, the coefficients related to average cost per claim decrease with the frequency premium, but these variations are very low Because of the correlation, the coefficients related to pure premium are not equal to the product of the coefficients for frequency and expected cost per claim. Here also, differences are very low

4. CONCLUDING REMARKS

We recall the main results obtained in this paper

- The unexplained heterogeneity with respect to the cost distributions depends strongly on the choice of the distribution family.
- Besides, it is revealed more slowly throughout time than for number distributions
- On the working sample, the correlation between the heterogeneity components on the number and cost distributions is very low.

In the long run, it would be desirable to relax the assumption of invariance of the heterogeneity components with respect to time Because of this invariance, the age of claims has no influence on the bonus-malus coefficients Now, the fact that an ancient claim has the same influence on the coefficients that a recent one is questionable. The allowance for an innovation at each period for the heterogeneity components would raise new problems, and would make it necessary to observe policyholders on many periods.

REFERENCES

BUHLMANN, H (1967) Experience Rating and Credibility *ASTIN Bulletin* **4**, 199-207

CUMMINS, J D, DIONNE, G, MC DONALD, J B and PRITCHETT, B M (1990) Application of the GB2 Distribution in Modelling Insurance Loss Processes *Insurance Mathematics and Economics* **9**, 257-272

DIONNE, G and VANASSE, C (1989) A Generalization of Automobile Insurance Rating Models The Negative Binomial Distribution with a Regression Component *ASTIN Bulletin* **19**, 199-212

DIONNE, G and VANASSE, C (1992) Automobile Insurance Ratemaking in the Presence of Asymmetrical Information *Journal of Applied Econometrics* **7**, 149-165

LEMAIRE, J (1985) *Automobile Insurance Actuarial Models* Huebner International Series on Risk, Insurance and Economic Security

LEMAIRE, J (1995) *Bonus-Malus Systems in Automobile Insurance* Huebner International Series on Risk, Insurance and Economic Security

PINQUET, J, ROBERT, J C, PESTRE, G and MONTOCCHIO, L (1992) Tarification a Priori et a Posteriori des Risques en Assurance Automobile *Mémoire au Centre d'Etudes Actuarielles*

PINQUET, J (1996a) Allowance for Costs of Claims in Bonus-Malus Systems *Proceedings of the ASTIN colloquium, Copenhague 1996*

PINQUET, J (1996b) Hétérogénéité Inexpliquée *Document de travail THEMA n° 11*

RAO, C R (1948) Large Sample Tests of Statistical Hypothesis Concerning Several Parameters with Applications to Problems of Estimation *Proceedings of the Cambridge Philosophical Society* **44**, 50-57

RENSHAW, E A (1994) Modelling the Claims Process in the Presence of Covariates *ASTIN Bulletin* **24**, 265-285

SILVEY, S D, (1959) The Lagrange Multiplier Test *Annals of Mathematical Statistics* **30**, 389-407

JEAN PINQUET
Université de Paris X, U F R de Sciences Economiques
200, Avenue de la République 92001 NANTERRE CEDEX
Phone 33 1 49 81 72 45, Fax: 33 1 40 97 71 42,
E-mail. pinquet@u-paris10 fr

EXCESS OF LOSS REINSURANCE AND THE PROBABILITY OF RUIN IN FINITE HORIZON

MARIA DE LOURDES CENTENO[1]
ISEG, Technical University of Lisbon

July, 1995

ABSTRACT

The upper bound provided by Lundberg's inequality can be improved for the probability of ruin in finite horizon, as Gerber (1979) has shown This paper studies this upper bound as a function of the retention limit, for an excess of loss arrangement, and compares it with the probability of ruin

KEYWORDS

Excess of loss, reinsurance; finite time ruin probability

1 INTRODUCTION

Several studies about the effect of reinsurance on the ultimate probability of ruin (for example Gerber (1979), Waters (1979), Bowers, Gerber, Hickman, Jones and Nesbitt (1987), Centeno (1986) and Hesselager (1990)) have concentrated their attention on the effect of reinsurance on the adjustment coefficient.

Centeno (1986) has used an algorithm suggested by Panjer (1986) to calculate the probability of ultimate ruin, incorporating reinsurance, to show with some examples that the behaviour of this probability and Lundberg's inequality are very similar, both considered as functions of the retention level, provided that the initial reserve is not too small This is consistent with the figures obtainded more recently by Dickson and Waters (1994) for some other examples and using a different algorithm for the probability of ultimate ruin In this paper, Dickson and Waters have also calculated finite horizon ruin probabilities, after reinsurance, by adapting the algorithm of De Vylder and Goovaerts (1988) and by an approximation provided by the translated Gamma process Through an example they show that in continuous time for an excess of loss arrangement, the optimal retention limit in finite horizon can be quite far from the optimum value in infinite horizon. Of course, the sequence of optimal retention levels

[1] Research performed under contract n° SPES-CT91-0063

converges to the infinite horizon optimal level as the time increases But, for a finite horizon, Lundberg's inequality can be improved The purpose of this paper is to show how we can use this improvement to redefine the "optimal" retention limit for an excess of loss arrangement, and to compare this inequality with the ruin probability in finite horizon and continuous time for some examples Of course, the same methodology can be applied to proportional reinsurance provided that, the moment generating function of the individual claim amounts distribution exists

2 ASSUMPTIONS AND PRELIMINARIES

In the classical risk process, the insurer's surplus at time t is denoted $U(t)$, with

$$U(t) = u + ct - S(t),$$

where u is the initial surplus, c is the premium income per unit of time, assumed to be received continuously, and $S(t)$ is the aggregate claims occurred up to time t. $\{S(t)\}_{t \geq 0}$ is assumed to be a compound Poisson process and without loss of generality the Poisson parameter is assumed to be 1, which means that "*time t*" is the interval during which t claims are expected Let $G(x)$ denote the individual claim amount distribution function and again without loss of generality, let us assume that this distribution has mean 1, which means that the monetary unit chosen is the expected amount of a claim We further assume that $G(0) = 0$, with $0 < G(x) < 1$ for $x > 0$ and also that G is such that its moment generating function exists for $x < T$ for some $0 < T \leq \infty$, and that

$$\lim_{r \to T} E\left[e^{rX}\right] = \infty. \tag{1}$$

We assume that c is greater than 1, i.e. it is greater than the expected aggregate claims in each period. Let θ be such that $c = 1 + \theta$

The ruin probability before time t is

$$\psi(u,t) = \Pr\{U(s) < 0 \text{ for some } s, 0 < s \leq t\}.$$

Of course $\psi(u, t)$ is not greater than the ultimate probability of ruin, denoted as $\psi(u)$. Therefore the upper bound given by Lundberg's inequality is valid for finite horizon. Gerber (1979), pp 139, has shown that this bound can be improved in finite horizon He proved that for $u \geq 0$ and $t > 0$

$$\psi(u,t) \leq \min_{r \geq R}\left\{e^{-ru+t[M_X(r)-1-rc]}\right\}, \tag{2}$$

where $M_X(r)$ is the moment generating function of the individual claim amounts and R denotes the adjustment coefficient, defined as the unique positive root of

$$M_X(r) - 1 = cr \tag{3}$$

In the following we refer to expression (2) as Gerber's inequality After an integration by parts, inequality (2) can be written as

$$\psi(u,t) \leq \min_{r \geq R}\left\{e^{-ru+rt[\int_0^\infty e^{rx}(1-G(x))dx-c]}\right\}, \tag{4}$$

and the equation defining the adjustment coefficient as

$$\int_0^\infty e^{rx}(1-G(x))\,dx = c \tag{5}$$

Now suppose that the insurer has an excess of loss arrangement such that when a claim X occurs he is responsible for min $\{X, M\}$, paying in return per unit of time a reinsurance premium $c(M)$, which we assume to be calculated according to the expected value principle with loading coefficient ξ, i.e.

$$c(M) = (1+\xi)\int_M^\infty (1-G(x))\,dx \tag{6}$$

Assuming that the reinsurance premiums are paid continuously, the insurer's surplus at time t is

$$U(M;t) = u + (c - c(M))t - \sum_{k=1}^{N(t)} \min\{X_k, M\},$$

where $N(t)$ denotes the number of claims up to time t. The ruin probability before time t is

$$\psi(M,u,t) = \Pr\{U(M,s) < 0 \text{ for some } s, 0 < s \le t\}.$$

After this arragement Gerber's inequality becomes

$$\psi(M;u,t) \le \min_{r \ge R(M)} \left\{ e^{-ru+rt\left[\int_0^M e^{rx}(1-G(x))dx-(c-c(M))\right]} \right\}, \tag{7}$$

where $R(M)$ denotes the adjustment coefficient after reinsurance, i e the unique positive root of

$$\int_0^M e^{rx}(1-G(x))\,dx = c - c(M), \tag{8}$$

when it exists or zero otherwise. Such a root exists if and only if the expected profit after reinsurance is positive

We know that the value of M that maximises the adjustment coefficient, when the excess of loss reinsurance premium is calculated according to the expected value principle with $\xi > \theta$, is such that

$$M = \frac{1}{R}\ln(1+\xi), \tag{9}$$

(see for example Waters (1979)), minimising then the upper bound provided by Lundberg's inequality.

In the next section we will study the problem that consists in choosing M in such a way that the upper bound provided by (7) is minimised

3. THE PROBLEM AND ITS SOLUTION

We define as "optimal" retention the value of M that minimises the upper bound of the probability of ruin given by (7). We can write (7) as

$$\psi(M;u,t) \le \exp\left(\min_{r \ge R(M)} f(r,M,u,t) \right), \tag{10}$$

where

$$f(r,M,u,t) = -ru + rt\left[\int_0^M e^{rx}(1-G(x))dx - (c-c(M)) \right]. \tag{11}$$

In the next result we will study the condition under which (11), as a function of r, possesses a minimum

Result 1

(i) For each $M > 0, f(r, M; u, t)$, defined by (11), for $r > 0$, has a local minimum and it is unique if and only if the expected surplus at t is positive

(ii) Suppose that the expected surplus at time t is positive and let $\hat{r}(M)$ be the value of r at which the local minimum of $f(r, M; u, t)$ occurs. Then $\hat{r}(M) \ge R(M)$, where $R(M)$ is the unique positive root of (8) if it exists or zero otherwise, if and only if

$$\frac{u}{t} \ge R(M)\int_0^M x e^{R(M)x}(1-G(x))dx. \tag{12}$$

Proof:

(i) It is clear that for $M > 0$

$$\lim_{r\to 0} f(r,M,u,t) = 0$$

and, by assumption (1), that also for any $M > 0$

$$\lim_{r\to\infty} f(r,M;u,t) = +\infty.$$

On the other hand

$$\frac{\partial f}{\partial r} = -u + t\int_0^M e^{rx}(1-G(x))dx - t(c-c(M)) + rt\int_0^M x e^{rx}(1-G(x))dx \tag{13}$$

and

$$\frac{\partial^2 f}{\partial r^2} = 2t\int_0^M x e^{rx}(1-G(x))dx + rt\int_0^M x^2 e^{rx}(1-G(x))dx. \tag{14}$$

As (14) is strictly positive for any $M > 0$, then $f(r, M; u, t)$ will have a minimum if and only if the limit of (13) is negative as $r \to 0$ But

$$\lim_{r\to 0}\frac{\partial f}{\partial r} = -u + t\left[\int_0^M (1-G(x))dx - (c-c(M)) \right],$$

which is negative if and only if the expected surplus at time t is positive

(ii) $\hat{r}(M)$ is the solution of

$$\frac{\partial f}{\partial r} = 0, \tag{15}$$

with $\partial f/\partial r$ given by (13) It is clear that $\hat{r}(M)$ will be greater than or equal to $R(M)$ if and only if $\partial f/\partial r$ is non positive at the point $r = R(M)$, i e if and only if condition (12) holds

Let M_0 be the minimum of the values for which the expected surplus at time t is non negative, i e.

$$M_0 = \min\left\{M \quad M \geq 0 \text{ and } u + t\left[c - c(M) - \int_0^M (1 - G(x))dx\right] \geq 0\right\} \tag{16}$$

Note that M_0 will be zero if and only if $u/t \geq \xi - \theta$ Then the following corollary follows from the previous proof

Corolary 1.1 For each $M > M_0$,

$$\psi(u,t;M) \leq \begin{cases} e^{f(\hat{r}(M),u,t,M)} & \text{if } \frac{u}{t} > R(M)\int_0^M xe^{R(M)x}(1 - G(x))dx \\ e^{f(R(M),u,t,M)} & \text{if } \frac{u}{t} \leq R(M)\int_0^M xe^{R(M)x}(1 - G(x))dx \end{cases} \tag{17}$$

where $R(M)$ is the unique positive solution of (8) if it exists or zero otherwise and $\hat{r}(M)$ is the unique positive solution of

$$\int_0^M e^{rx}(1 - G(x))dx - (c - c(M)) + r\int_0^M xe^{rx}(1 - G(x))dx = \frac{u}{t}. \tag{18}$$

Hence we can conclude that for some values of M it will be possible to improve the upper bound given by Lundberg's inequality, which implies that in some cases the value of M that minimises the upper bound provided by Gerber's inequality is different from the value of M that maximises the adjustment coefficient As this maximum is attained at the unique solution of (8) satisfying (9) we can conclude that this value is different from the minimiser of Gerber's inequality if and only if

$$\frac{u}{t} > R^* \int_0^{\frac{1}{R^*}\ln(1+\xi)} xe^{R^* x}(1 - G(x))dx, \tag{19}$$

where R^* is the unique solution of

$$\int_0^{\frac{1}{r}\ln(1+\xi)} e^{rx}(1 - G(x))dx = c - c\left(\frac{1}{r}\ln(1+\xi)\right) \tag{20}$$

Let us study the behaviour of Gerber's bound as a function of the retention limit Notice that

$$\begin{aligned}\min_{M\geq M_0}\psi(u,t;M) &\leq \exp\left(\min_{M\geq M_0}\min_{r\geq R(M)} f(r,M;u,t)\right)\\ &= \exp\left(\min_{r\geq R(M)}\min_{M\geq M_0} f(r,M;u,t)\right).\end{aligned} \tag{21}$$

Differentiating $f(r, M, u, t)$ with respect to M and considering (6) we get

$$\frac{\partial f}{\partial M} = rt(1-G(M))\left(e^{rM}-(1+\xi)\right), \tag{22}$$

and differentiating twice

$$\frac{\partial^2 f}{\partial M^2} = rt\left[re^{rM}(1-G(M))+\left((1+\xi)-e^{RM}\right)g(M)\right]. \tag{23}$$

which implies that the first derivative is zero if and only if

$$M = \frac{1}{r}\ln(1+\xi), \tag{24}$$

and that the second derivative is positive whenever (24) holds This means that for fixed r, u and t, $f(r, M, u, t)$ has a local minimum, which is unique and attained at the point $M = \frac{1}{r}\ln(1+\xi)$.

Let $r_0 = \frac{1}{M_0}\ln(1+\xi)$ with M_0 given by (16). (Note that r_0 will be finite if and only if $u/t < \xi - \theta$.)

So, minimising $f(r, M; u, t)$ for $r \geq R(M)$ and $M \geq M_0$, is equivalent to minimising $f(r, \frac{1}{r}\ln(1+\xi), u, t)$ for $R^* \leq r \leq r_0$, where R^* is the unique solution to (20)

Differentiating $f(r, \frac{1}{r}\ln(1+\xi), u, t)$ with respect to r we get

$$\begin{aligned}\frac{\partial}{\partial r} f(r,\frac{1}{r}\ln(1+\xi);u,t) \quad = \quad & -u + t\int_0^{\frac{1}{r}\ln(1+\xi)} e^{rx}(1-G(x))dx\\ & -t\left(c - c\left(\frac{1}{r}\ln(1+\xi)\right)\right)\\ & +rt\int_0^{\frac{1}{r}\ln(1+\xi)} xe^{rx}(1-G(x))dx,\end{aligned} \tag{25}$$

and differentiating twice we get

$$\begin{aligned}\frac{\partial^2}{\partial r^2} f(r,\frac{1}{r}\ln(1+\xi),u,t) &= 2t\int_0^{\frac{1}{r}\ln(1+\xi)} xe^{rx}(1-G(x))dx \\ &+rt\int_0^{\frac{1}{r}\ln(1+\xi)} x^2e^{rx}(1-G(x))dx \\ &-\frac{t}{r^2}(\ln(1+\xi))^2(1+\xi)\left(1-G(\frac{1}{r}\ln(1+\xi))\right) \\ &= t\int_0^{\frac{1}{r}\ln(1+\xi)} x^2e^{rx}dG(x),\end{aligned} \tag{26}$$

which is positive, implying that $f(r,\frac{1}{r}\ln(1+\xi),u,t)$ is a convex function of r That the three terms sum to the right hand side of (26), can be easily checked, by integrating by parts this last expression Hence we can conclude that there is at most one solution to

$$\frac{\partial}{\partial r} f\left(r,\frac{1}{r}\ln(1+\xi);u,t\right)=0 \tag{27}$$

and that when it exists it is the global minimum of $f(r,\frac{1}{r}\ln(1+\xi),u,t)$.

But on one hand

$$\lim_{r\to 0} f\left(r,\frac{1}{r}\ln(1+\xi),u,t\right)=0$$

and

$$\lim_{r\to 0}\frac{\partial}{\partial r} f\left(r,\frac{1}{r}\ln(1+\xi),u,t\right)=-u-\theta t<0.$$

On the other hand, if $u/t < \xi-\theta$, then r_0 will be finite and

$$\lim_{r\to r_0} f\left(r,\frac{1}{r}\ln(1+\xi),u,t\right)=r_0t\int_0^{M_0}(e^{r_0x}-1)(1-G(x))dx\geq 0$$

and if $u/t \geq \xi-\theta$ then

$$\begin{aligned}\lim_{r\to r_0} f\left(r,\frac{1}{r}\ln(1+\xi),u,t\right) &= \lim_{r\to r_\infty} f\left(r,\frac{1}{r}\ln(1+\xi)\right) \\ &= \lim_{r\to r_\infty}\left(-r(u-t(\xi-\theta))\right)=-\infty,\end{aligned}$$

so we can state the following result

Result 2

If $u/t \geq \xi-\theta$ then the upper bound to the ruin probability before time t, given by (10), attains its minimum at $M = 0$

If $u/t < \xi-\theta$ then the upper bound, considered as a function of M has an absolute mi-nimum which is attained at the point $M=\frac{1}{r^*}\ln(1+\xi)$ with $r^* = \max(\hat{r}, R^*)$ where $\hat{r}$ is the solution to

$$\int_0^{\frac{1}{r}\ln(1+\xi)} e^{rx}(1-G(x))dx - \left(c - c\left(\frac{1}{r}\ln(1+\xi)\right)\right) + r\int_0^{\frac{1}{r}\ln(1+\xi)} xe^{rx}(1-G(x))dx = \frac{u}{t} \quad (28)$$

and R^* is the unique solution to

$$\int_0^{\frac{1}{r}\ln(1+\xi)} e^{rx}(1-G(x))dx = (1+\theta) - (1+\xi)\int_{\frac{1}{r}\ln(1+\xi)}^{\infty} (1-G(x))dx, \quad (29)$$

if such a root exists or zero otherwise.

4. EXAMPLES

In this section we give some examples for the problem studied in the previous section and compare the values obtained for the upper bound given by Gerber's inequality with the values of Lundberg's bound and the values of ruin probability in finite horizon.

Example 1: Let us consider first the case of exponential individual claim amounts, i e $G(x) = 1 - e^{-x}$ for $x > 0$. Then the excess of loss reinsurance premium is $c(M) =$ (1 +) e^{-M} and

$$M_0 = -\ln\left(\frac{u+t\theta}{t\xi}\right)$$

Equation (8) defining the adjustment coefficient $R(M)$ is, in this case, equivalent to

$$\left(1-e^{-(1-r)M}\right)/(1-r) = (1+\theta) - (1+\xi)e^{-M}, \quad (30)$$

and equation (18) defining $\hat{r}(M)$ is equivalent to

$$\left(\frac{1}{1-r} + \frac{r}{(1-r)^2}\right)\left(1-e^{-(1-r)M}\right) - \frac{r}{1-r}Me^{-(1-r)M} - \left[(1+\theta)-(1+\xi)e^{-M}\right] = \frac{u}{t} \quad (31)$$

$\hat{r}(M)$ will be greater than $R(M)$ if and only if

$$\frac{u}{t} > \frac{R(M)}{1-R(M)}\left[\frac{1}{1-R(M)}\left(1-e^{-(1-R(M))M}\right) - Me^{-(1-R(M))M}\right] \quad (32)$$

Equations (30) and (31) can be solved for each M by standard numerical techniques given values of θ and ξ.
If $u/t < \xi - \theta$ the upper bound to $\psi(M; u, t)$ given by (10) is attained at the point

$$M = \frac{1}{r^*}\ln(1+\xi) \quad (33)$$

with $r^* = \max(\hat{r}, R^*)$ where $\hat{r}$ is the solution to equation (31) with M substituted by the right-hand side of (33) and R^* is the solution to equation (30) again with M substituted by the right-hand side of (33)

Let $\theta = 0.2$ and $\xi = 0.4$. In this case the value of M that minimises the upper bound provided by Lundberg's inequality is $M = 1.486$, which gives a value for the adjustment coefficient of $R^* = 0.226466$. When we minimise the upper bound provided by Gerber's inequality we get a different solution for the excess of loss retention limit if $u / t > 0.12075$, the solution being $M = 0$ if $u / t \geq 0.2$. Table 1 gives the optimal M for different values of u / t

TABLE 1

'OPTIMAL' RETENTION AS A FUNCTION OF u/t, WITH CLAIM AMOUNTS EXPONENTIALLY DISTRIBUTED

u/t	0.125	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
M	1.427	1.357	1.219	1.078	0.932	0.779	0.611	0.412	0

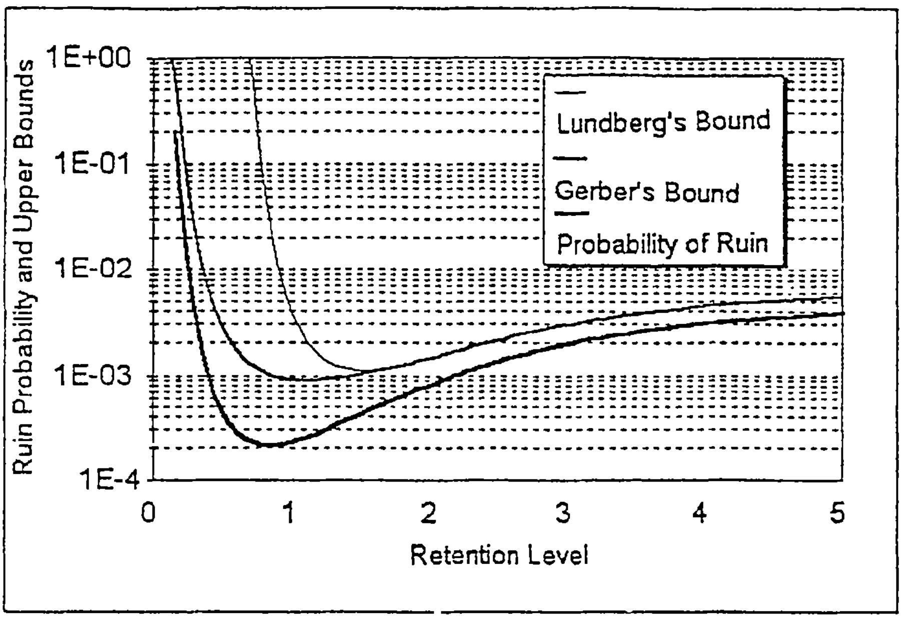


FIGURE 1 CLAIM AMOUNTS EXPONENTIALLY DISTRIBUTED

Figure 1 shows calculated values of $\psi(M, u, t)$, Gerber's upper bound and Lundberg's upper bound for $u = 30$ and $t = 200$

Table 2 gives the values attained by these functions at the minimum of each of them (rounded to two decimal places)

TABLE 2

'OPTIMAL' VALUES WITH CLAIM AMOUNTS EXPONENTIALLY DISTRIBUTED

M	$\psi(M;30,200)$	Gerber's bound	Lundberg's bound
0.83	0 218 x 10^{3}	0 101 x 10^{2}	0 252 x 10^{1}
1.08	0 257 x 10^{3}	0 896 x 10^{3}	0 219 x 10^{2}
1.49	0 442 x 10^{3}	0 104 x 10^{2}	0 112 x 10^{2}

The efficiency measure defined by Dickson and Waters (1994) goes from 49% (= ψ(0 83;30, 200)/ψ(1 49, 30, 200)) for the minimiser of Lundberg's bound to 85% (= ψ(0.83,30, 200)/ψ(1.08, 30, 200)) for the minimiser of Gerber's bound.

The probabilities, in all the examples, were calculated using the algorithm of De Vylder and Goovaerts (1988) as re-scaled by Dickson and Waters (1991) and adjusted to take into account reinsurance

We started by discretizing the individual claim amounts (before reinsurance) on $1/\beta$, $2/\beta$, ..., using the method suggested by De Vylder and Goovaerts (1988) Then, for each value of M we have calculated the net premium (after reinsurance) in the new monetary unit, after which we have calculated the distribution function F of the aggregate claim amounts after reinsurance in a period of time with the rescaled Poisson parameter (in this case - with $t = 1$ - the inverse of the net premium). In this way the rescaling parameter depends on the value of the retention.[2]

Then we have used the recursion formula

$$\hat{\psi}(w,1) = 1 - F(w+1),\ w \le \overline{w} + (\overline{n} - 1),$$

$$\hat{\psi}(w,n) = 1 - F(w+1) + \sum_{j=0}^{w+1} f_j \hat{\psi}(w+1-j, n-1),\ w \le \overline{w} + (\overline{n} - n), n = 2, \quad , \overline{n},$$

where $\overline{w} = u\beta$ and $\overline{n} = \{tP\}$ where P denotes the net premium in the new monetary unit and $\{x\}$ denotes the least integer greater than or equal to x

We have used the approximation

$$\overline{\psi}(w,n) \cong \frac{1}{2}\left(\hat{\psi}(w-1,n) + \hat{\psi}(w,n)\right)$$

with $\hat{\psi}(w-1,n)$ to be zero if w is zero, as suggested by De Vylder and Goovaerts (1988), for probabilities in continuous time

TABLE 3

'OPTIMAL' VALUES WITH CLAIM AMOUNTS PARETO DISTRIBUTED

M	$\psi(M;30,200)$	Gerber's bound	Lundberg's bound
0.83	0 102 x 10^{2}	0 549 x 10^{2}	1 000
1.03	0 109 x 10^{2}	0 523 x 10^{2}	0 644
2.33	0 356 x 10^{2}	0 977 x 10^{2}	0 013

[2] Note that with this rescaling we are restricted to evaluate the ruin probabilities for a positive net premium

As tP may be not an integer we have used the following interpolation to calculate the probabilities of the original process

$$\psi(M,u,t) \cong \overline{\psi}(u\beta,tP) \cong (\{tP\}-tP)\overline{\psi}(u\beta,\{tP\}-1)+(tP-(\{tP\}-1))\overline{\psi}(u\beta,\{tP\})$$

In the calculations of Table 2 we have taken $\beta = 100$ and the control parameter, ε, was set at 3 x 10^{-9} This parameter is used for the calculations in the De Vylder and Goovaerts algorithm (see De Vylder and Goovaerts (1988), p 7) For the calculations of the ruin probabilities necessary to perform Figure 1 we have used $\beta = 20$.

Example 2: Consider now the case where $G(x) = 1 - (1+x)^{-2}$, i e. individual claims follow a Pareto (2,1) distribution. Let $\theta = 0\ 2$ and $\xi = 0\ 4$ as in the previous example In this case the equations defining $R(M)$ and $\hat{r}(M)$ require numerical calculations of integrals of the kind

$$\int_0^M e^{rx}(1-G(x))dx$$

Instead of using standard numerical techniques to calculate them, we have calculated $R(M)$ and $\hat{r}(M)$ based on the discretized distribution. Figure 2 shows the ruin probability before time $t = 200$, for $u = 30$, and both Gerber's and Lundberg's bounds

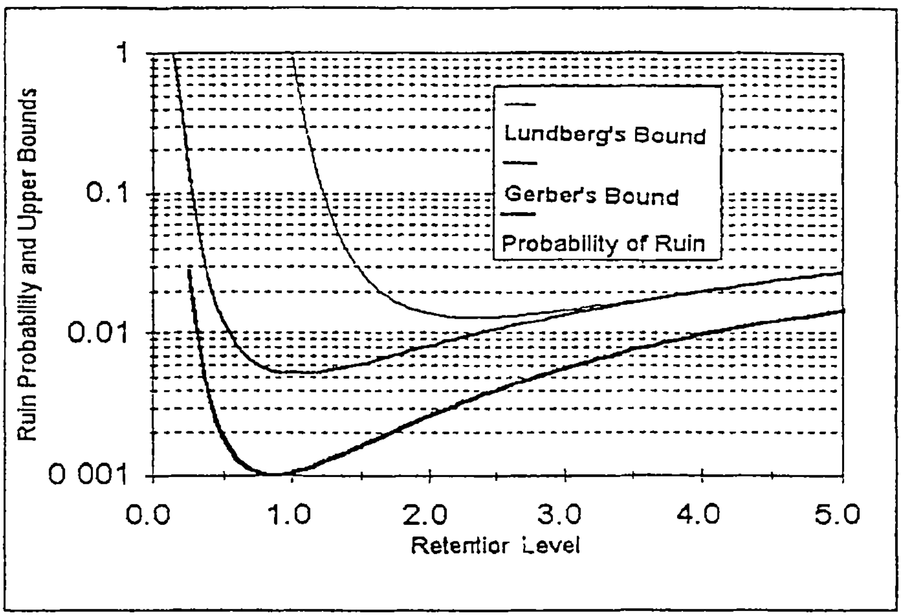


FIGURE 2 CLAIM AMOUNTS PARETO DISTRIBUTED

Table 3 equivalent to Table 2, but for the Pareto distribution The figures are even more indicative.

5. CONCLUDING REMARKS

As we have already mentioned, the optimal retention limit, when the probability of ruin in continuous time with a finite horizon is minimised, can be quite far from the optimal value when the probability of ruin in continuous time with an infinite horizon is considered However, the calculations of the ruin probabilities in finite horizon are very time consuming, making this criterion less appealing.

Gerber's bound is computationally much easier to deal with than the ruin probability and in the examples considered it provides a solution that is very close to the solution obtained when the probability of ruin is used. The disadvantage of using Gerber's bound is that this bound is not always an improvement on Lundberg's bound - it depends on the value of the ratio of u to t Our advice would be to use Gerber's bound, if it provides an improvement to Lundberg's bound, and use an approximation such as that provided by the translated Gamma process otherwise

We have shown that when the reinsurance premium calculation principle is the expected value principle, Gerber's bound has a unique minimum. However, this is not true in general. When this is not the case, in all the examples considered, the probability of ruin had a similar behaviour Some care should be taken in these cases.

REFERENCES

BOWERS, N L, GERBER, H U, HICKMAN, J C, JONES, D A and NESBITT, C J (1987) *Actuarial Mathematics*, Society of Actuaries, Chicago

CENTENO, L (1986) Measuring the effects of reinsurance by the adjustment coefficient, *Insurance Mathematics and Economics 5* 169-182

DE VYLDER, F AND GOOVAERTS, M J (1988) Recursive calculation of finite-time ruin probabilities, *Insurance Mathematics and Economics 7* 1-8

DICKSON, D AND WATERS, H (1991) Recursive calculation of survival probabilities, *Astin Bulletin 21* 199-221

DICKSON, D AND WATERS, H (1994) Reinsurance and ruin, *University of Melbourne, CAS, Research Paper Series*

GERBER, H U (1979) *An Introduction to Mathematical Risk Theory*, S S Huebner Foundation Monographs, University of Pensylvania

HESSELAGER, O (1990) Some results on optimal reinsurance in terms of the adjustment coefficient *Scandinavian Actuarial Journal* pp 80-95

PANJER, H H (1986) Direct calculation of ruin probabilities, *The Journal of Risk and Insurance 53* 521-529

WATERS, H (1979) Excess of loss reinsurance limits, *Scandinavian Actuarial Journal* pp 37-43

CREDIBILITY THEORY AND GENERALIZED LINEAR MODELS

J A NELDER [1] and R J VERRALL [2]

ABSTRACT

This paper shows how credibility theory can be encompassed within the theory of Hierarchical Generalized Linear Models. It is shown that credibility estimates are obtained by including random effects in the model. The framework of Hierarchical Generalized Linear Models allows a more extensive range of models to be used than straightforward credibility theory. The model fitting and testing procedures can be carried out using a standard statistical package Thus, the paper contributes a further range of models which may be useful in a wide range of actuarial applications, including premium rating and claims reserving

KEYWORDS

Credibility Theory, Hierarchical Generalized Linear Models; Generalized Linear Models; Premium Rating Random-Effect Models

1 INTRODUCTION

Credibility theory began with the papers by Mowbray (1914) and Whitney (1918). In those papers, the emphasis was on deriving a premium which was a balance between the experience of an individual risk and of a class of risks Buhlmann (1967) showed how a credibility formula can be derived in a distribution-free way, using a least-squares criterion. Since then, a number of papers have shown how this approach can be extended· see particularly Buhlmann and Straub (1970), Hachemeister (1975), de Vylder (1976, 1986). The survey by Goovaerts and Hoogstad (1987) provides an excellent introduction to these papers.

[1] Department of Mathematics
Imperial College
Huxley Building
180 Queen's Gate
LONDON
SW7 2BZ

[2] Department of Actuarial Science and Statistics
City University
Northampton Square
LONDON
EC1V 0HB

The underlying assumption of credibility theory which sets it apart from formulae based on the individual risk alone is that the risk parameter is regarded as a random variable This naturally leads to a Bayesian model, and there have been a large number of papers which adopt the Bayesian approach to credibility theory: for example Jewell (1974, 1975), Klugman (1987), Zehnwirth (1977) Klugman (1992) gives an introduction to the use of Bayesian methods, covering particularly aspects of credibility theory. A recent review of Bayesian methods in actuarial science and credibility theory is given by Makov *et al* (1996)

It can be shown that, under suitable assumptions, a credibility formula can be derived as the best linear approximation to the Bayesian estimate, using a quadratic loss function Jewell (1974) showed that for an exponential family of distributions, the credibility formula is the same as the exact formula, so long as the conjugate prior distribution and a natural parametrisation is used This result will be derived in a different way in section 3, in order to place the basic model of credibility within a wider framework. The choice of structure for the collective and the parameterisation will be discussed in more detail. Since exponential families form the basis of Generalized Linear Models (GLMs) - see McCullagh and Nelder (1989) - it is natural to seek an extension of credibility theory encompassing the full range of models which can be formulated as GLMs. This is particularly apposite as GLMs have many very natural applications in the actuarial field: see, for example Renshaw (1991), Renshaw and Verrall (1994) This will also make possible more applications of credibility theory.

The main purpose of this paper is to show how credibility theory can be incorporated into the general framework of GLMs and implemented in the statistical package Genstat Although the formulation of the credibility model is similar in many ways to the Bayesian approach, our approach is likelihood-based rather than Bayesian. The dispersion parameters will be estimated directly from the data without specifying prior distributions No prior estimates for the parameters need to be supplied. All assumptions used in the model can be checked using, for example, appropriate residual analyses Recent advances in the statistical literature on GLMs allow unobserved random effects to be estimated along with the parameter vector in the linear predictor A useful recent paper is Breslow and Clayton (1993) which covers the theory of generalized linear mixed models (GLMMs) GLMMs allow the inclusion of normally distributed random effects and have been applied to a wide variety of statistical problems We use the theory of Lee and Nelder (1996), which develops hierarchical generalized linear models (HGLMs). HGLMs also allow the inclusion of random effects, but these are not restricted to be normally distributed Pure random-effect models, in which no fixed effects are included in the linear predictor, are known in the actuarial literature as credibility models. They form one part of a much wider class of models which have many potential applications to actuarial data

Thus, the purpose of this paper is further to unify the actuarial theory; to show how modern statistical methods can be used to enable credibility theory to be applied in a standard statistical package, to allow extensions of basic credibility theory and to show how the assumptions made can be checked This last point is important, since we regard many aspects of actuarial work as exercises is statistical modeling, rather than a dogmatic application of risk theory models

It should be noted that the theory can be applied to models that specify only the mean and variance functions, using quasi-likelihood (Wedderburn, 1974, Nelder and Pregibon, 1987) - see section 5

The paper is set out as follows. Section 2 contains a brief introduction to GLMs and derives some results which will be used elsewhere. Section 3 shows how credibility theory can be treated within the context of HGLMs Section 4 outlines more general HGLMs, and how they are likely to be used for actuarial data Section 5 outlines some extensions to the models in sections 3 and 4

2 INTRODUCTION TO GLMS

This section contains a brief introduction to GLMs, and derives some of the key results which will be used later in the paper. A complete treatment of the theory and application of GLMs can be found in McCullagh and Nelder (1989).

The basis of GLMs is the assumption that the data are sampled from a one-parameter exponential family of distributions We first describe these and some of their fundamental properties

Consider a single observation y A one-parameter exponential family of distributions has a log-likelihood of the form

$$\frac{y\theta - b(\theta)}{\varphi} + c(y, \varphi) \tag{2 1}$$

where θ is the canonical parameter
and φ is the dispersion parameter, assumed known

Haberman and Renshaw (1996) review the application of Generalized Linear Models in actuarial science, and include a section on loss distributions. In actuarial applications, many distributions belonging to one-parameter exponential families are useful However, Haberman and Renshaw (1996) show how it is also possible to fit certain heavy-tailed distributions using Generalized Linear Models

Some examples of such families are given below It is straightforward to show that

$$\mu = E(Y) = \frac{db(\theta)}{d\theta} \tag{2 2}$$

$$\text{and } Var(Y) = \frac{d^2 b(\theta)}{d\theta^2}\varphi \tag{2.3}$$

Note that $Var(Y)$ is the product of two quantities $\frac{d^2 b(\theta)}{d\theta^2}$ is called the variance function and depends on the canonical parameter (and hence on the mean) We can write this as $V(\mu)$, since equation (2 2) shows that θ is a function of μ

$$\text{Thus } V(\mu) = \frac{d^2 b(\theta)}{d\theta^2} \tag{2 4}$$

In actuarial applications, it is possible to include deterministic volume measures in the definition of *Var(Y)*. A GLM may be defined by specifying a distribution, as above, together with a link function and a linear predictor. The link function defines the relationship between the linear predictor and the mean. The linear predictor takes the form

$$\eta = X\beta \tag{2 5}$$

where β is parameter vector
and X is defined by the design.

For a single observation, X is a row vector, and for a set of observations, X is the design matrix

The linear predictor is related to the mean by $\eta = g(\mu)$ The function g is called the link function, and the special case $g(\mu) = \theta$ is called the canonical link function

By way of illustration, the log-likelihoods for some common distributions are given below

(i) Normal

The log-likelihood is $\dfrac{\mu y - \frac{1}{2}\mu^2}{\sigma^2} - \dfrac{y^2}{2\sigma^2} - \dfrac{1}{2}\log(2\pi\sigma^2)$

Thus, $\theta = \mu$ and the canonical link function is the identity function.

$b(\theta) = \dfrac{\theta^2}{2}$ and $c(y,\theta) = -\dfrac{y^2}{2\sigma^2} - \dfrac{1}{2}\log(2\pi\sigma^2)$

$V(\mu) = 1$ and $\varphi = \sigma^2$

(ii) Poisson

The log-likelihood is $y \log\mu - \mu - \log y'$

$\theta = \log \mu$ and the canonical link is the log function

$b(\theta) = e^\theta$ and $c(y, \varphi) = -\log y'$

$V(\mu) = \mu$ and $\varphi = 1$

(iii) Binomial

Suppose $R \sim$ *Binomial* (m, μ). Define $Y = \dfrac{R}{m}$. Then the log-likelihood is

$$\frac{y \log \frac{\mu}{1-\mu} - \log(1-\mu)}{\frac{1}{m}} + \log\binom{m}{my}$$

Hence $\theta = \log \dfrac{\mu}{1-\mu}$, and the canonical link function is the logit function

$b(\theta) = \log(1 + e^\theta)$ and $c(y,\varphi) = \log\dbinom{m}{my}$

$V(\mu) = \mu(1-\mu)$ and $\varphi = \frac{1}{m}$.

Note that this parameterisation may be unfamiliar because of the definition of Y However, it enables us to give a coherent theory in the following section

(iv) Gamma (with mean μ and variance $\frac{\mu^2}{\nu}$).

The log-likelihood is $\frac{-\frac{y}{\mu} + \log\frac{1}{\mu}}{\frac{1}{\nu}} + \nu\log y + \nu\log\nu - \log\Gamma(\nu)$

$\theta = -\frac{1}{\mu}$ and the canonical link is the reciprocal function.

$b(\theta) = -\log(-\theta)$ and $c(y,\varphi) = \nu\log y + \nu\log\nu - \log\Gamma(\nu)$.

$V(\mu) = \mu^2$ and $\varphi = \nu^{-1}$.

This section has given a brief introduction to GLMs The following section shows how standard credibility theory can be applied in this context Section 4 will show how more general models can be formulated

3. THE BUHLMAN MODEL FOR EXPONENTIAL FAMILIES

In this section, we derive the credibility formulae for exponential families of distributions, under the assumptions made by Buhlmann (1967) It is possible to extend this to other models for example the assumptions of Buhlmann and Straub (1970) can be incorporated using weight functions This section derives just the credibility formulae A brief discussion of the estimation of the dispersion parameters is given in section 4, where the appropriate references are cited.

Denote the data by y_{ij} for $i = 1, 2, \ldots, t$; $j = 1, 2, \ldots, n_i$ Assume for the moment, as is common in credibility applications, that $n_i = k$, $\forall\ i$, but note that this restriction is not necessary for the derivation of HGLMs

Thus, i indexes the risks within the collective In credibility theory, it is assumed that each risk has a risk parameter, which we denote by ξ_i for risk i

The assumptions of the model of Buhlmann (1967) are

(i) The risks, and hence ξ_i, are independently, identically distributed.

(ii) $y_{ij} \mid \xi_i$ are independently, identically distributed.

We assume that $y_{ij} \mid \xi_i$ is distributed according to an exponential family Define $m(\xi_i) = E\left[y_{ij} \mid \xi_i\right]$ Note that under the assumptions of the model, $E\left[y_{ij} \mid \xi_i\right]$ does not depend on J. Hence the canonical parameter for observation y_{ij} does not depend on j, and we assume that it can be written as follows

$$\theta_i' = \theta(m(\xi_i)) = \theta(u_i) \tag{3 1}$$

where θ is the canonical link function and u_i is a random effect for group i. Thus, for the standard credibility model, $m(\xi_i) = u_i$ Define $v_i = \theta(u_i)$; then, in this case,

$$\theta_i' = v_i. \tag{3 2}$$

Again, note that there is no j dependence here Note also that this also implies that $Var(y_{ij} \mid \xi_i)$ does not depend on j

This has defined the distribution of the random variable within each risk, conditional on the risk parameter. It is also necessary to define the structure of the collective - the distribution of $\{\xi_i \quad i = 1, \quad , t\}$. This is often done by defining a Bayesian prior distribution; here we use the same form of distribution for the random effects, but do not perform a Bayesian analysis Instead, we define a "hierarchical likelihood", h, which we maximize.

We define the conjugate hierarchical generalized linear model (HGLM) by defining the kernel of the log-likelihood for $\theta(u_i)$ as

$$a_1\theta_i' - a_2 b(\theta_i') \tag{3.3}$$

In the actuarial literature, this distribution (the distribution of the random effects) is known as the structure of the collective Note that we define the log-likelihood of ξ_i implicitly through that of $\theta(m(\xi_i))$. We have conditioned on ξ_i through $m(\xi_i) = u_i$, since it is the latter that we wish to estimate.

From (3.3) and the distribution of $y_{ij} \mid \xi_i$, we may define a hierarchical log-likelihood as

$$h = \sum_{i,j} l(\theta_i', y_{ij} \mid v_i) + \sum_i l(v_i) \tag{3 4}$$

$$= \sum_{ij} \left(\frac{y_{ij}\theta_i' - b(\theta_i')}{\varphi} \right) + c(y_{ij}, \varphi) + a_1\theta_i' - a_2 b(\theta_i') \tag{3 5}$$

When the distribution of both the data and the random effects is normal, this is Henderson's joint log-likelihood (Henderson 1975). In other cases, it is an obvious extension of the joint log-likelihood, called the hierarchical log-likelihood We have now defined a hierarchical generalized linear model (HGLM), in this case the conjugate HGLM In the particular case described in this section, the linear predictor for y_{ij} consists solely of a random effects term which is modelled in the second stage of the likelihood, (3 2) It is possible to incorporate more structure into the model by including fixed effects and generalizing the form of the random effects model However, in this section we are concerned solely with showing that the estimates obtained under the basic model described above are the usual credibility estimates Thus, we require an estimate of $m(\xi_i) = u_i$ The mean random effects $\{u_i : i = 1, \ldots, t\}$ are estimated by maximizing the hierarchical likelihood, (3 4), as follows.

Using (2 2)

$$\frac{\partial b(\theta(u_i))}{\partial v_i} = u_i .$$

Hence $$\frac{\partial h}{\partial v_i} = \sum_{j=1}^{k}\left(\frac{y_{ij} - u_i}{\varphi}\right) + a_1 - a_2 u_i$$

Equating $\frac{\partial h}{\partial v_i}$ to 0 gives

$$y_{i+} - k\hat{u}_i + \varphi a_1 - \varphi a_2 \hat{u}_i = 0 \tag{3.6}$$

where $y_{i+} = \sum_{j=1}^{k} y_{ij}$

Hence $$\hat{u}_i = \frac{y_{i+} + \varphi a_1}{k + \varphi a_2} = Z\bar{y}_i + (1 - Z)m$$

where $\bar{y}_i = \frac{1}{k} y_{i+}$, $Z = \frac{k}{k + \varphi a_2}$ and $m = \frac{a_1}{a_2}$.

Thus, we have shown that, with the choice of distribution for the random effects defined in (3.3), and using the canonical link function, the estimate of u_i is in the form of a credibility estimate provided $E(m(\xi_i)) = \frac{a_1}{a_2}$ This is straightforward to show, and was also proved by Jewell (1974). The density of u_i is proportional to

$$e^{a_1\theta_i' - a_2 b(\theta_i')}$$

Now $$\frac{\partial e^{a_1\theta_i' - a_2 b(\theta_i')}}{\partial \theta_i'} = \left(a_1 - a_2\frac{\partial b(\theta_i')}{\partial \theta_i'}\right)e^{a_1\theta_i' - a_2 b(\theta_i')} = (a_1 - a_2 m(\xi_i))e^{a_1\theta_i' - a_2 b(\theta_i')}$$

Integrating over the natural range of θ_i', and assuming $e^{a_1\theta_i' - a_2 b(\theta_i')}$ is zero at the end points, we have

$$a_1 - a_2 E[m(\xi_i)] = 0 .$$

Hence, using (2 2), $$E[m(\xi_i)] = E[u_i] = \frac{a_1}{a_2}$$

Thus, we have shown that the credibility estimate is the same as the estimate obtained using a conjugate HGLM with pure random effects. This shows that credibility theory is closely connected to the statistical theory of random-effect models Of course, it is possible to widen the scope of the models considerably. Fixed effects terms can also be included in the model, other link functions may be considered and the form of the random-effect models can be generalized

It is possible to formulate the pure random-effect model in another way, by including a fixed effect which is constant for all the data This means that the overall mean is estimated as a fixed effect and the randomeffects model departures from this overall mean There is no effect on the credibility estimates, but the above derivation is, in some ways, closer to the actuarial theory

The results in this section are closely related to those of Jewell (1974). The present approach differs in that it is not presented as a Bayesian procedure, and the emphasis is on the modelling aspects encapsulated within Generalized Linear Models

The estimation of the dispersion parameters is discussed in section 4. This includes the estimation of φ and of a_1 and a_2. It should be noted that if a constant fixed effect is included in the model, as outlined above, there is only one parameter to estimate in the distribution of u_i For this reason we adopt this approach henceforth

By way of illustration, we consider the four exponential families outlined in section 2 Note that we can derive the density of u_i from the density of $\theta(u_i)$, defined in (3.3) The density of u_i is proportional to

$$e^{a_1\theta_i' - a_2 b(\theta_i')} \frac{\partial \theta(u_i)}{\partial u_i}$$

$$\frac{e^{a_1\theta_i' - a_2 b(\theta_i')}}{V(u_i)} \tag{3.7}$$

(i) Normal

The random effects have log-likelihood whose kernel is

$$a_1 u_i - a_2 \frac{u_i^2}{2}$$

i e. $u_i \sim N(m, \sigma_0^2)$ $\quad a_1 = \frac{m}{\sigma_0^2},\ a_2 = \frac{1}{\sigma_0^2}$ and $m = E[u_i] = \frac{a_1}{a_2}$

(ii) Poisson

u_i has a likelihood proportional to

$$\frac{e^{a_1 \log u_i - a_2 u_i}}{u_i}$$

Hence $u_i \sim$ Gamma, parameters a_1 and a_2, and $m = E[u_i] = \frac{a_1}{a_2}$

(iii) Binomial

u_i has a likelihood proportional to

$$\frac{\exp\left[a_1 \log\left(\frac{u_i}{1-u_i}\right) - a_2 \log\left(\frac{1}{1-u_i}\right)\right]}{u_i(1-u_i)}$$

i e $u_i \sim$ Beta, parameters a_1 and $a_2 - a_1$, and $m = E[u_i] = \frac{a_1}{a_2}$

(iv) Gamma

u_i has a likelihood proportional to

$$\frac{\exp\left(\frac{-a_1}{u_i} + a_2 \log u_i\right)}{u_i^2}$$

i e $u_i \sim$ inverse gamma and $m = E[u_i] = \frac{a_1}{a_2}$

Having shown that the estimates obtained using conjugate HGLMs for a simple random-effect model are the usual credibility estimates, we now define a more general framework which encompasses credibility models

4 HIERARCHICAL GENERALIZED LINEAR MODELS

Standard GLMs model differences between groups, parametric variation and other effects as fixed effects in the linear predictor Random-effect models can be combined with standard GLMs in order to formulate models with both fixed effects and the random effects of credibility models. To do this, we define an extended linear predictor for a single observation as

$$\eta' = \eta + v \tag{4 1}$$

where $\eta = X\beta$, as in (2 5)
and v is a strictly monotonic function of $u, v = v(u)$

When $v = 0$, (4.1) reduces to the standard linear predictor for GLMs. When $\eta = 0$ and $v = \theta(u)$, we have the basis credibility model described in section 3.

The hierarchical log-likelihood, (3.4), becomes

$$h = \sum_{i,j} l(\beta, y_{ij} \mid v_i) + \sum_i l(v_i)$$

where $v_i = v(u_i)$

The maximum hierarchical likelihood extimates (MHLEs) of β and u are obtained from the pair of equations

$$\frac{\partial h}{\partial \beta} = 0 \quad \text{and} \quad \frac{\partial h}{\partial v} = 0$$

which may be solved iteratively using the procedures written by the second author for the statistical package Genstat

We consider here the case when the canonical link function is used for the fixed effects and $v = \theta(u)$ In this case, equation (3 1) for observation y_{ij} becomes

$$\theta'_{ij} = \theta_{ij} + \theta(u_i) \tag{4 2}$$

where $\theta_{ij} = X_{ij}\beta$
θ is the canonical link function

and X_{ij} is the row from the design matrix for the fixed effects which relates to y_{ij}

The same log-likelihood is used for $\theta(u_i)$, as in (3 3) Then the kernel of h is

$$\frac{\sum_{i,j}(y_{ij}\theta'_{ij} - b(\theta'_{ij}))}{\varphi} + \sum_i l(v_i)$$

Hence
$$\frac{\partial h}{\partial \beta_k} = \frac{\sum_{i,j}(y_{ij} - u'_{ij}))x_{kij}}{\varphi} \tag{4.3}$$

and
$$\frac{\partial h}{\partial v_i} = \frac{\sum_{i,j}(y_{ij} - u'_{ij})) + \varphi a_1}{\varphi} - a_2 u_i \tag{4 4}$$

where $u'_{ij} = E[y_{ij} | u_i] = E[y_{ij} | \xi_i]$,
β_k is the kth parameter in the fixed effects
and x_{kij} is the kth entry of the row vector X_{ij}

Note that in this case, unlike that in section 3, $E[y_{ij} | \xi_i] \neq u_i$ Instead,

$$\theta(u'_{ij}) = \eta_{ij} + \theta(u_i) \tag{4 5}$$

which implies that $\mu'_{ij} = u_i$ when $\eta_{ij} = 0$.

We include the overall mean as a fixed effect and require that the random effects then have the appropriate mean (eg 0 for the identity link function).

The dispersion parameters given the fixed and random effects are estimated by maximising the h-likelihood after a suitable adjustment. The adjustment, which results in an adjusted profile h-likelihood, is necessary because the marginal maximum likelihood estimates may be biased. Further justifications for this adjustment can be found in Cox and Reid (1987) and Lee and Nelder (1996) For the normal distribution, unbiased estimates are obtained More details on estimation theory for random-effect GLMs can be found in McGilchrist (1994) and Schall (1991).

The joint estimates of the mean effects (fixed and random) and the dispersion parameters are obtained by iterating between the two sets of estimating equations. These processes may be conveniently carried out in Genstat, for which a set of procedures is available from the second author.

For the distributions illustrated in section 1, the likelihoods of the random effects are again appropriate, but the estimate will be different because of the difference between (3 1) and (4 2)

5 DISCUSSION

It is possible to extend the class of models to which these methods may be applied by specifying just the mean and variance functions This is useful when greater flexibility is required in the modelling assumptions For example, Renshaw and Verrall (1994) show that the chain-ladder technique in claims reserving is essentially equivalent to GLM with a Poisson likelihood and an appropriate linear predictor. By specifying just the mean and variance function, this model may be applied to a much wider class of data than is implied by the Poisson assumption (which obviously requires the variance to equal the nean). This involves the use of extended quasi-likelihood (Wedderburn 1974, Nelder and Pregibon 1987). For HGLMs, the equivalent extension is the extended quasi-h-likelihood, in which the extended quasi-likelihood is used in the hierarchical likelihood This extension makes it possible, for example, to include random effects in the chain-ladder linear model to allow a connection between accident years.

HGLMs may also be of use when a particular factor is hard to model parametrically An example of this, which has been mentioned above, is claims reserving, when it is inappropriate to model the accident years as completely independent, but a parametric relationship is also inappropriate. The same comment applies in motor premium rating, when it is usual to group a factor such as the age of the policyholder. Such a grouping may be inappropriate, as it may be crude or doubtful because it has been decided before the analysis of the data (for example, according to the present rating structure). However, it is often inappropriate, because of computational and theroretical considerations, to treat the ages as completely separate or to apply a parametric model In this situation, HGLMs may be useful

Applications in life insurance include similar premium-rating situations as in general insurance, and also graduation theory The use of HGLMs for graduation would have some similarities to Whittaker graduation, which can be regarded as a GLM with a stochastic linear predictor (Verrall, 1993).

REFERENCES

BRESLOW, N E and CLAYTON D G (1993) Approximate inference in generalised linear mixed models *J Am Statist Ass*, **88**, 9-25

BUHLMANN, H (1967) Experience Rating and Credibility *ASTIN Bulletin*, **4**, 199-207

BUHLMANN, H and STRAUB, E (1970) Credibility for Loss Ratios *ARCH*, **1972.2.**

COX, D R and REID, N (1987) Parameter orthogonality and approximate conditional inference *J R Statist Soc B*, **49**, 1-39

GOOVAERTS, M and HOOGSTAD, W (1987) *Credibility Theory* Surveys of Actuarial Studies No 4, Rotterdam Nationale-Nederlanden

HABERMAN, S and RENSHAW, A E, (1996) *Generalized Linear Models and actuarial science* The Statistician, **45**, 407-436

HACHEMEISTER, C R (1975) Credibility for Regression Models with Applications to Trend, in *Credibility Theory and Applications*, P Kalm, ed, New York Academic Press

HENDERSON, C R (1975) Best linear unbiased estimation and prediction under a selection model *Biometrics* **31**, 423-447

JEWELL, W S (1974) Credible means are exact Bayesian for exponential families *ASTIN Bulletin*, **8**, 77-90

JEWELL, W S (1975) The use of collateral data in credibility theory a hierarchical model *Giornale dell'Instituto Italiano degli Attuari*, **38**, 1-16

KLUGMAN, S (1987) Credibility for Classification Ratemaking via the Hierarchical Normal Linear Model *Proc of the Casualty Act Soc*, **74**, 272-321

KLUGMAN, S (1992) *Bayesian Statistics in Actuarial Science with Emphasis on Credibility* Kluwer Academic Publishers

LEE Y and NELDER, J A (1996) Hierarchical Generalized Linear Models *J R Statist Soc* B, 58, 619-678

McCULLAGH, P and NELDER, J A (1989) *Generalized Linear Models* 2nd Edition London Chapman and Hall

McGILCHRIST, C A (1994) Estimation in generalized mixed models *J R Statist Soc B*, **56**, 61-69

MAKOV, U E, SMITH, A F M and LIU, Y -H *Bayesian methods in actuarial science* The Statistician, **45**, 503-515

MOWBRAY, A H (1914) How extensive a payroll exposure is necessary to give a dependable pure premium *Proc of the Casualty Act Soc*, **1**, 24-30

NELDER, J A and LEE, Y (1992) Likelihood, quasi-likelihood and pseudo-likelihood Some coomparisons *J R Statist Soc B*, **54**, 273-284

NELDER J A and PREGIBON, D (1987) An extended quasi-likelihood function *Biometrika*, **74**, 221-231

RENSHAW A E (1991) Actuarial Graduation Practice and Generalized Linear and Non-Linear Models *J Inst Acts*, **118**, 295-312

RENSHAW, A E and VERRALL, R J (1994) A stochastic model underlying the chain-ladder technique, *Proceedings, ASTIN Colloquium*, 1994

SCHALL, R (1991) Estimating in generalized linear models with random effects *Biometrika*, **78**, 719-727

VERRALL, R J (1993) A state space formulation of Whittaker graduation, with extensions *Insurance Mathematics and Economics*, **13**, 7-14

DE VYLDER, F (1976) Optimal semilinear credibility *Bull of the Assoc of Swiss Acts*, **78**, 27-40

DE VYLDER, F (1986) General regression in multidimensional credibility theory, in *Insurance and Risk Theory*, ed M GOVAERTS, J, HAEZENDONCK and F DE VYLDER, Reidel

WEDDERBURN, R W M (1974) Quasi-likelihood functions, generalized linear models and the Gauss-Newton method *Biometrika*, **61**, 439-447

WHITNEY, A W (1918) The theory of experience rating *Proc of the Casualty Act Soc*, **4**, 274-292

ZEHNWIRTH, B (1977) The mean credibility formula is a Bayes rule *Scandinavian Actuarial Journal*, **4**, 212-216

CREDIBILITY IN THE REGRESSION CASE REVISITED
(A LATE TRIBUTE TO CHARLES A HACHEMEISTER)

H BUHLMANN

ETH Zurich

A GISLER

Winterhur-Versicherungen

ABSTRACT

Many authors have observed that Hachemeisters Regression Model for Credibility – if applied to simple linear regression – leads to unsatisfactory credibility matrices they typically 'mix up' the regression parameters and in particular lead to regression lines that seem 'out of range' compared with both individual and collective regression lines We propose to amend these shortcomings by an appropriate definition of the regression parameters:

– intercept
– slope

Contrary to standard practice the intercept should however not be defined as the value at time zero but as the value of the regression line at the barycenter of time. With these definitions regression parameters which are uncorrelated in the collective can be estimated separately by standard one dimensional credibility techniques

A similar convenient reparametrization can also be achieved in the general regression case The good choice for the regression parameters is such as to turn the design matrix into an array with orthogonal columns

1. THE GENERAL MODEL

In his pioneering paper presented at the Berkeley Credibility Conference 1974, Charlie Hachemeister introduced the following General Regression Case Credibility Model

a) Description of individual risk r.

risk quality θ_r

observations (random vector)

$$\begin{pmatrix} X_{1r} \\ X_{2r} \\ \cdot \\ X_{nr} \end{pmatrix} = X_r$$

with distribution $dP(X_r/\theta_r)$ and where X_{ir} = observation of risk r at time i

b) Description of collective.
$\{\theta_r, (r = 1, 2, \ldots, N)\}$ are i i d with structure function $U(\theta)$

We are interested in the (unknown)
individually correct pure premiums $\mu_i(\theta_r) = E[X_{ir}/\theta_r]$ $(i = 1, 2, \ldots, n)$

$$\begin{pmatrix} \mu_1(\theta_r) \\ \mu_2(\theta_r) \\ \vdots \\ \mu_n(\theta_r) \end{pmatrix} = \mu(\theta_r) \quad \text{where } \mu_i(\theta_r) = \text{individual pure premium at time } i$$

and we suppose that these individual pure premiums follow a regression pattern

R) $$\mu(\theta_r) = Y_r \beta(\theta_r),$$

where $\mu(\theta_r) \sim n$-vector, $\beta(\theta_r) \sim p$-vector and $Y_r \sim n * p$-matrix (= design matrix)

Remark:

The model is usually applied for $p < n$ and maximal rank of Y_r, in practice p is much smaller than n (e g. $p = 2$).

The goal is to have credibility estimator $\hat{\beta}(\theta_r)$ for $\beta(\theta_r)$
which by linearity leads to the credibility estimator $\hat{\mu}(\theta_r)$ for $\mu(\theta_r)$.

2 THE ESTIMATION PROBLEM AND ITS RELEVANT PARAMETERS AND SOLUTION (GENERAL CASE)

We look for

$$\hat{\beta}(\theta_r) = \mathbf{a} + AX_r$$
$$\mathbf{a} \sim p - \text{vector}$$
$$A \sim p * n \text{ matrix}$$

The following quantities are the "relevant parameters" for finding this estimator

$$E[Cov[X_r, X'_r / \theta_r]] = \Phi_r \qquad \Phi_r \sim n * n \text{ matrix (regular)} \tag{1}$$

$$Cov[\beta(\theta_r), \beta'(\theta_r)] = \Lambda \qquad \Lambda \sim p * p \text{ matrix (regular)} \tag{2}$$

$$E[\beta(\theta_r)] = \mathbf{b} \qquad \mathbf{b} \sim p - \text{vector} \tag{3}$$

We find the credibility formula

$$\hat{\beta}(\theta_r) = (I - Z_r)\mathbf{b} + Z_r \mathbf{b}_r^X \tag{4}$$

where

$$Z_r = (I - W_r^{-1}\Lambda^{-1})^{-1} = (W_r + \Lambda^{-1})^{-1} W_r = \Lambda(\Lambda + W_r^{-1})^{-1} \quad \text{~ credibility matrix}(p * p) \tag{5}$$

$$W_r = Y_r' \Phi_r^{-1} Y_r \quad \text{~ auxiliary matrix} \quad (p * p) \tag{6}$$

$$\mathbf{b}_r^X = W_r^{-1} Y_r' \Phi_r^{-1} X_r \quad \text{~ individual estimate} \quad (p * 1) \tag{7}$$

Discussion:

The generality under which formula (4) can be proved is impressiv, but this generality is also its weakness Only by specialisation it is possible to understand how the formula can be used for practical applications Following the route of Hachemeisters original paper we hence use it now for the special case of simple linear regression.

3 SIMPLE LINEAR REGRESSION

Let

$$Y_r = Y = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ . & \\ 1 & n \end{pmatrix}$$

and

$$\beta(\theta_r) = \begin{pmatrix} \beta_0(\theta_r) \\ \beta_1(\theta_r) \end{pmatrix}$$

hence R) becomes

$$\mu_i(\theta_r) = \beta_0(\theta_r) + i \ \beta_1(\theta_r) \tag{8}$$

which is one of the most frequently applied regression cases Assume further that Φ_r is diagonal, i.e. that observations X_{ir}, X_{jr} given θ_r are uncorrelated for $i \neq j$
To simplify notation, we drop in the following the index r, i e we write Φ instead of Φ_r, W instead of W_r and Z instead of Z_r

Hence

$$\Phi = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & . & \\ 0 & & & \sigma_n^2 \end{pmatrix} \tag{9}$$

e g $\sigma_i^2 = \dfrac{\sigma_2}{V_i}$, V_i = "volume" of observation at time i

Let $$\Lambda = \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{pmatrix} \quad \tau_{01} = \tau_{10}$$

We find

$$W = Y' \Phi^{-1} Y = \begin{pmatrix} \sum_{k=1}^{n} \frac{1}{\sigma_k^2} & \sum_{k=1}^{n} \frac{k}{\sigma_k^2} \\ \sum_{k=1}^{n} \frac{k}{\sigma_k^2} & \sum_{k=1}^{n} \frac{k^2}{\sigma_k^2} \end{pmatrix} \tag{10}$$

It is convenient to write

$$\sigma_k^2 = \frac{\sigma^2}{V_k}, \quad V_\cdot = \sum_{k=1}^{n} V_k$$

(which is always possible for diagonal Φ) Hence we have

$$W = \frac{V_\cdot}{\sigma^2} \begin{pmatrix} \sum_k \frac{V_k}{V_\cdot} & \sum_k k \frac{V_k}{V_\cdot} \\ \sum_k k \frac{V_k}{V_\cdot} & \sum_k k^2 \frac{V_k}{V_\cdot} \end{pmatrix}$$

Think of $\frac{V_k}{V_\cdot}$ as <u>sampling weights</u>, then we have

$$W = \frac{V_\cdot}{\sigma^2} \begin{pmatrix} 1 & E^{(s)}[k] \\ E^{(s)}[k] & E^{(s)}[k^2] \end{pmatrix} \tag{11}$$

where $E^{(s)}$, $Var^{(s)}$ denote the <u>moments with respect to the sampling distribution</u>

One then also finds (see (7))

$$\begin{aligned} \mathbf{b}_r^X &= W^{-1} Y' \Phi^{-1} X_r \\ &= \frac{1}{Var^{(s)}[k]} \begin{pmatrix} E^{(s)}[k^2] \; E^{(s)}[X_{kr}] - E^{(s)}[k] \; E^{(s)}[kX_{kr}] \\ E^{(s)}[kX_{kr}] - E^{(s)}[k] \; E^{(s)}[X_{kr}] \end{pmatrix} \end{aligned} \tag{12}$$

$$\text{where} \quad E^{(s)}[kX_{kr}] = \sum_k \frac{V_k}{V_\cdot} k X_{kr}, \; E^{(s)}[X_{kr}] = \sum_k \frac{V_k}{V_\cdot} X_{kr}$$

Remark:

It is instructive to verify by direct calculation that the values given by (12) to b_{0r}^X, b_{1r}^X are identical with those obtained from

$$\sum_{k=1}^{n} V_k (X_{kr} - b_{0r}^X - k b_{1r}^X)^2 = \min!$$

The calculations to obtain the credibility matrix Z (see (5)) are as follows

$$\Lambda^{-1} = \frac{1}{\tau_0^2 \tau_1^2 - \tau_{01}^2} \begin{pmatrix} \tau_1^2 & -\tau_{01} \\ -\tau_{01} & \tau_0^2 \end{pmatrix} = \begin{pmatrix} \rho_0^2 & +\rho_{01} \\ +\rho_{01} & \rho_1^2 \end{pmatrix}$$

Abbreviate

$$\rho_0^2 \ \frac{\sigma^2}{V} = h_0$$
$$\rho_1^2 \ \frac{\sigma^2}{V} = h_1 \tag{13}$$
$$\rho_{01} \ \frac{\sigma^2}{V} = h_{01}$$

Hence

$$W + \Lambda^{-1} = \frac{V}{\sigma^2}\begin{pmatrix} 1+h_0 & E^{(s)}[k]+h_{01} \\ E^{(s)}[k]+h_{01} & E^{(s)}[k^2]+h_1 \end{pmatrix}$$

$$(W+\Lambda^{-1})^{-1} = \frac{\sigma^2}{V}\cdot\frac{1}{\underline{(1+h_0)(E^{(s)}[k^2]+h_1)-(E^{(s)}[k]+h_{01})^2}\atop N}\begin{pmatrix} E^{(s)}[k^2]+h_1 & -(E^{(s)}[k]+h_{01}) \\ -(E^{(s)}[k]+h_{01}) & 1+h_0 \end{pmatrix}$$

$$Z = (W+\Lambda^{-1})^{-1} \ W \tag{14}$$

$$Z = \frac{1}{N}\begin{pmatrix} Var^{(s)}[k]+h_1-h_{01}E^{(s)}[k] & E^{(s)}[k]h_1 - E^{(s)}[k^2]h_{01} \\ h_0E^{(s)}[k]-h_{01} & Var^{(s)}[k]+h_0E^{(s)}[k^2]-h_{01}E^{(s)}[k] \end{pmatrix}$$

Discussion:

The credibility matrix obtained is not satisfactory from a practical point of view

a) individual weights are not always between zero and one.

b) both intercept $\hat{\beta}_0(\theta_r)$ of the credibility line and slope $\hat{\beta}_1(\theta_r)$ of the credibility line may not lie between intercept and slope of individual line and collective line

Numerical examples:

$n = 5 \quad V_k \equiv 1$

collective regression line. $b_0 = 100 \quad b_1 = 10$

individual regression line· $b_0^X = 70 \quad b_1^X = 7$

Example 1 $\sigma = 20 \quad \tau_0 = 10 \quad \tau_1 = 5 \quad \tau_{10} = 0$

resulting credibility line. $\hat{\beta}_0(\theta_r) = 88.8 \quad \hat{\beta}_1(\theta_r) = 3.7$

Example 2 $\sigma = 20 \quad \tau_0 = 100'000 \quad \tau_1 = 5 \quad \tau_{10} = 0$

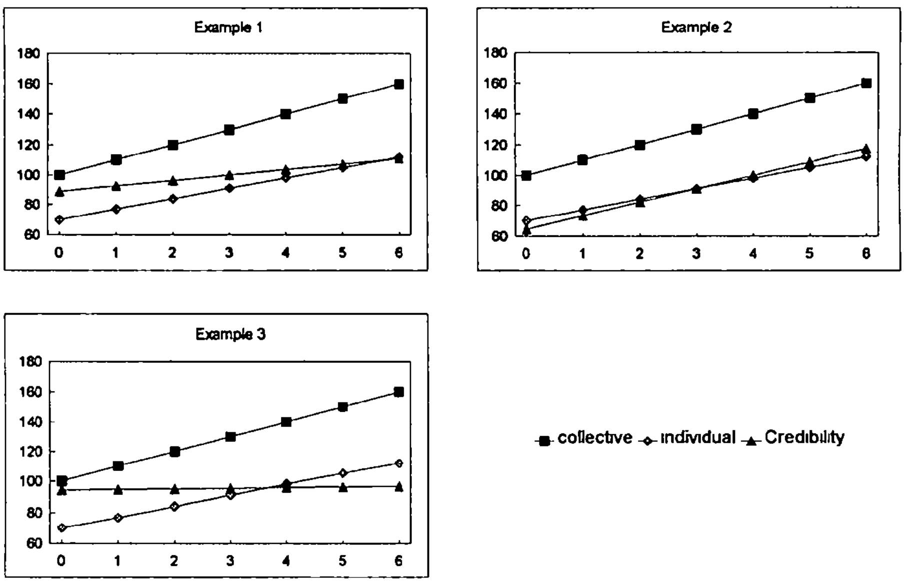
resulting credibility line: $\hat{\beta}_0(\theta_r) = 64\ 5 \quad \hat{\beta}_1(\theta_r) = 8.8$

Example 3 $\sigma = 20 \quad \tau_0 = 10 \quad \tau_1 = 100'000 \quad \tau_{10} = 0$

resulting credibility line $\hat{\beta}_0(\theta_r) = 94\ 7 \quad \hat{\beta}_1(\theta_r) = 0.3$

Comments:

In none of the 3 examples do both, intercept and slope of the credibility line, lie between the collective and the individual values In example 2 there is a great prior uncertainty about the intercept (τ_0 very big) One would expect that the credibility estimator gives full weight to the intercept of the individual regression line and that $\hat{\beta}_0(\theta_r)$ nearly coincides with b_0^X. But $\hat{\beta}_0(\theta_r)$ is even smaller than b_0 and b_0^X In example 3 there is a great prior uncertainty about the slope and one would expect, that $\hat{\beta}_1(\theta_r) \cong b_1^X$ But $\hat{\beta}_1(\theta_r)$ is much smaller than b_1 and b_1^X
For this reason many actuaries have either considered Hachemeisters regression model as not usable or have tried to impose artificially additional constraints (e g De Vylder (1981) or De Vylder (1985)) Dannenburg (1996) discusses the effects of such constraints and shows that they have serious drawbacks This paper shows that by an appropriate reparametrization the defects of the Hachemeister model can be made to disappear and that hence no additional constraints are needed.



4 SIMPLE LINEAR REGRESSION WITH BARYCENTRIC INTERCEPT

The idea, that choosing the time scale in such a way as to have the intercept at the barycenter of time, is already mentioned in Hachemeisters paper, although it is then not used to make the appropriate model assumptions. Choosing the intercept at the barycenter of the time scale means formally that our design matrix is chosen as

$$Y = \begin{pmatrix} 1 & 1 - E^{(s)}[k] \\ 1 & 2 - E^{(s)}[k] \\ . & \\ 1 & n - E^{(s)}[k] \end{pmatrix}$$

Remark:

It is well known, that any linear transformation of the time scale (or more generally of the covariates) does not change the credibility estimates. But what we do in the following changes the original model by assuming that the matrix Λ is now the covariance matrix of the 'new' vector $\beta(\theta_r), \beta_0(\theta_r)$ now being the intercept at the barycenter of time instead of the intercept at the time zero.
In our general formulae obtained in section 3 we have to replace

$$E^{(s)}[k] \leftarrow 0 \quad E^{(s)}[k^2] \leftarrow Var^{(s)}[k]$$

It is also important that sample variances and covariances are <u>not</u> changed by this shift of time scale.

We immediately obtain

$$\begin{aligned} b_{0r}^{s} &= E^{(s)}[X_{kr}] \\ b_{1r}^{s} &= \frac{Cov^{(s)}(k, X_{kr})}{Var(s)[k]} \end{aligned} \qquad (12_{bar})$$

and

$$Z = \frac{1}{(1+h_0)(Var^{(s)}[k]+h_1) - h_{01}^2} \begin{pmatrix} Var^{(s)}[k]+h_1 & -Var^{(s)}[k]h_{01} \\ -h_{01} & Var^{(s)}[k](1+h_0) \end{pmatrix} \qquad (14_{bar})$$

These formulae are now becoming very well understandable, in particular the crosseffect between the credibility formulae for intercept and slope is only due to their correlation in the collective (off diagonal elements in the matrix Λ) In case of no <u>correlation between regression parameters in the collective</u> we have

$$Z = \frac{1}{(1+h_0)(Var^{(s)}[k]+h_1)} \begin{pmatrix} Var^{(s)}[k]+h_1 & 0 \\ 0 & Var^{(s)}[k](1+h_0) \end{pmatrix} \qquad (14_{sep})$$

which separates our credibility matrix into two separate one-dimensional credibility formulae with credibility weights

$$Z_{11} = \frac{1}{1+h_0} = \frac{1}{1+\frac{\sigma^2}{\tau_0^2 V.}} = \frac{V.}{V. + \frac{\sigma^2}{\tau_0^2}} \qquad (15)$$

$$Z_{22} = \frac{Var^{(s)}[k]}{Var^{(s)}[k]+h_1} = \frac{Var^{(s)}[k]}{Var^{(s)}[k]+\frac{\sigma^2}{\tau_1^2 V.}} = \frac{V\,Var^{(s)}[k]}{V.Var^{(s)}[k]+\frac{\sigma^2}{\tau_1^2}}$$

Remark:

Observe the classical form of the credibility weights in (15) with volumes $V.$ for Z_{11} and $V\ Var^{(s)}[k]$ for Z_{22}.

Numerical examples

The model assumptions of the following three examples numbered 4 – 6 are exactly the same as in the examples numbered 1 – 3 of the previous section with the only difference that the first element of the vector $\beta(\theta_r)$ now represents the intercept at the barycenter Thus we have.

collective regression line. $b_0 = 130 \quad b_1 = 10$

individual regression line $b_0^X = 91 \quad b_1^X = 7$

The resulting credibility lines are:

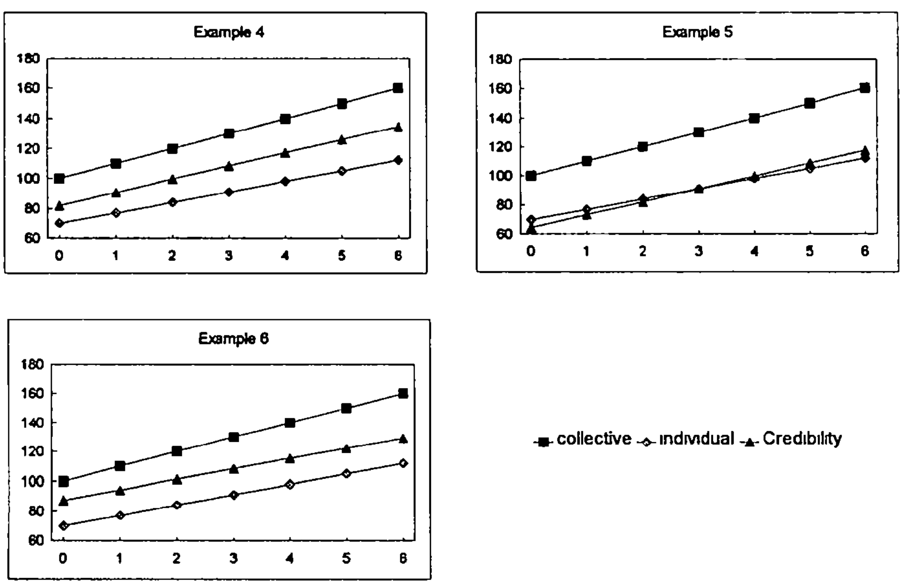
Example 4. $\hat{\beta}_0(\theta_r) = 108\ 3 \quad \hat{\beta}_1(\theta_r) = 8.8$

Example 5. $\hat{\beta}_0(\theta_r) = 91.0 \quad \hat{\beta}_1(\theta_r) = 8\ 8$

Example 6: $\hat{\beta}_0(\theta_r) = 108\ 3 \quad \hat{\beta}_1(\theta_r) = 7\ 0$

Comments:

Intercept and slope of the credibility lines are always lying between the values of the individual and of the collective regression line In example 5 (respectively in example 6) the intercept $\hat{\beta}_0(\theta_r)$ (respectively the slope $\hat{\beta}_1(\theta_r)$) coincides with b_0^X (resp. b_1^X). It is also interesting to note that the credibility line of example 5 is exactly the same as the one of example 2.



5 HOW TO CHOOSE THE BARYCENTER?

Unfortunately the barycenter for each risk is shifting depending on the individual sampling distribution. There is usually no way to bring – simultaneously for all risks – the matrices Y, W, Z into the convenient form as discussed in the last section. This discussion however suggests that the most reasonable parametrization is the one using the intercept at the barycenter of the collective This has two advantages: it is the point to which individual barycenters are (in the sum of least square sense) closest and the orthogonality property of parameters still holds for the collective.
In the following we work with this parametrization and assume that the regression parameters in this parametrization are uncorrelated.

Hence we work from now on with the regression line

$$\alpha_0(\theta_r) + (k-K)\alpha_1(\theta_r),$$

where K is the barycenter of the collective i.e $K = \sum_{i=1}^{n} \frac{V_i}{V} i$.
We assume also that the collective parameters are uncorrelated, i e

$$\Lambda^{(\alpha)} = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \tau_1^2 \end{pmatrix}$$

If we shift to the individual barycenter $E^{(s)}[k]$ we obtain the line·

$$\beta_0(\theta_r) + (k - E^{(s)}[k])\beta_1(\theta_r)$$

Hence

$$\beta_1(\theta_r) = \alpha_1(\theta_r)$$

$$\beta_0(\theta_r) = \alpha_0(\theta_r) + \alpha_1(\theta_r)\overbrace{(E^{(s)}[k] - K)}^{\Delta} \tag{16}$$

$$\Lambda(\beta) = \begin{pmatrix} \tau_0^2 + \tau_1^2(E^{(s)}[k]-K)^2 & \tau_1^2(E^{(s)}[k]-K) \\ \tau_1^2(E^{(s)}[k]-K) & \tau_1^2 \end{pmatrix} = \begin{pmatrix} \tau_0^2 + \Delta^2\tau_1^2 & \Delta\tau_1^2 \\ \Delta\tau_1^2 & \tau_1^2 \end{pmatrix}$$

For the β-line we have further

$$\rho_0^2 = \frac{1}{\tau_0^2}, \qquad h_0^{(\beta)} = \frac{\sigma^2}{\tau_0^2\ V} = h_0^{(\alpha)}$$

$$\rho_1^2 = \frac{1}{\tau_1^2} + \Delta^2\frac{1}{\tau_0^2}, \qquad h_1^{(\beta)} = \frac{\sigma^2}{\tau_1^2 \cdot V} + \Delta^2\frac{\sigma^2}{\tau_0^2\ V.} = h_1^{(\alpha)} + \Delta^2 h_0^{(\alpha)}$$

$$\rho_{01} = -\Delta\frac{1}{\tau_0^2}, \qquad h_{01}^{(\beta)} = \Delta\frac{\sigma^2}{\tau_0^2\ V} = -\Delta h_0^{(\alpha)}$$

has two orthogonal columns (using the weights of the sampling distribution) This is the clue for the general regression case. The good choice of the regression parameters is such as to render the design matrix into an array with orthogonal columns

6.2 The Barycentric Model

Let

$$Y = \begin{pmatrix} Y_{11} & Y_{12} & Y_{1p} \\ Y_{21} & \vdots & \vdots \\ \vdots & \vdots & . \\ Y_{n1} & Y_{n2} & Y_{np} \end{pmatrix}$$

and assume volumes $V_1, V_2, \quad, V_n$ and let be $V = \sum_{k=1}^{n} V_k$.

We think of column j in Y as a random variable Y_j which assumes Y_{jk} with sampling weight $\frac{V_k}{V}$ in short $P^{(s)}[Y_j = Y_{jk}] = \frac{V_k}{V}$ where $P^{(s)}$ stands for the sampling distribution

As in the case of simple linear regression it turns out that also in the general case this sampling distribution allows a concise and convenient notation.

We have from (9)

$$\Phi^{-1} = \frac{V}{\sigma_2}\begin{pmatrix} \frac{V_1}{V} & & \\ & \frac{V_2}{V.} & \\ & & \frac{V_n}{V} \end{pmatrix}$$

and from (10)

$$W = Y' \Phi^{-1} Y = (w_{ij})$$

where

$$w_{ij} = \frac{V}{\sigma_2} E^{(s)}[Y_i \quad Y_j]$$

Under the barycentric condition we find

$$W = \frac{V}{\sigma^2}\begin{pmatrix} E^{(s)}[Y_1^2] & & 0 \\ & E^{(s)}[Y_2^2] & \\ 0 & & E^{(s)}[Y_p^2] \end{pmatrix} \tag{18}$$

i.e. a matrix of diagonal form.

Assuming non-correlation for the corresponding parametrization we have

$$\Lambda = \begin{pmatrix} \tau_1^2 & & 0 \\ & \tau_2^2 & \\ 0 & & \tau_p^2 \end{pmatrix} \quad \Lambda^{-1} = \frac{V}{\sigma^2}\begin{pmatrix} h_1 & & 0 \\ & h_2 & \\ 0 & & h_p \end{pmatrix}$$

with
$$h_j = \frac{1}{\tau_j^2}\,\frac{\sigma^2}{V}$$

Hence

$$(W+\Lambda^{-1}) = \frac{V.}{\sigma^2}\begin{pmatrix} E^{(s)}[Y_1^2]+h_1 & & 0 \\ & E^{(s)}[Y_2^2]+h_2 & \\ 0 & & E^{(s)}[Y_p^2]+h_p \end{pmatrix}$$

and finally

$$Z = (W+\Lambda^{-1})W = \begin{pmatrix} \frac{E^{(s)}[Y_1^2]}{E^{(s)}[Y_1^2]+h_1} & 0 \\ 0 & \frac{E^{(s)}[Y_p^2]}{E^{(s)}[Y_p^2]+h_p} \end{pmatrix} \tag{19}$$

(19) shows that our credibility matrix is of diagonal form. Hence the multidimensional credibility formula breaks down into p one dimensional formulae with credibility weights.

$$Z_{jj} = \frac{V\;E^{(s)}[Y_j^2]}{V\;E^{(s)}[Y_j^2]+\frac{\sigma^2}{\tau_j^2}} \tag{20}$$

Observe the "volume" $V.E^{(s)}[Y_j^2]$ for the j-th component

6.3 The Summary Statistics for the Barycentric Model

From (7) we have

$$\mathbf{b}_r^{\text{`}} = W^{-1}Y^{-1}\Phi^{-1}\mathbf{X}_r = C\mathbf{X}_r$$

where the elements of C are

$$c_{ij} = \frac{1}{E^{(s)}[Y_i^2]}\;Y_{ij}\,\frac{V_j}{V} \tag{21}$$

hence
$$b_{ir}^{\text{`}} = \frac{1}{E^{(s)}[Y_i^2]}\sum_{j=1}^{n} X_{ij}Y_{jr}\,\frac{V_j}{V}$$

or

$$b_{ir}^{s} = \frac{E^{(s)}[Y_i \ X_r]}{E^{(s)}[Y_i^2]} \quad i = 1,2,\ldots p \tag{22}$$

6.4 How to find the Barycentric Reparametrization

We start with the design matrix
Y and its column vectors $Y_1, Y_2, \ldots, Y_p$
and want to find the new design matrix
Y^* with orthogonal column vectors $Y_1^*, Y_2^*, \ldots, Y_p^*$

The construction of the vectors Y_k^* is obtained recursively

i) Start with $Y_1^* = Y_1$

ii) If you have constructed $Y_1^*, Y_2^*, \ldots, Y_{k-1}^*$, you find Y_k^* as follows

a) Solve $E^{(s)}[(Y_k - a_1^* Y_1^* - a_2^* Y_2^* - \ldots - a_{k-1}^* Y_{k-1}^*)^2] = \min!$
over all values of $a_1, a_2, \ldots, a_{k-1}$

b) Define $Y_k^* := Y_k - a_1^* Y_1^* - a_2^* Y_2^* - \ldots - a_{k-1}^* Y_{k-1}^*$

Remarks:

i) obviously this leads to Y_k^* such that

$$E^{(s)}[Y_k^* \ Y_l^*] = 0 \quad \text{for all} \quad l < k \tag{23}$$

ii) The procedure of orthogonalisation is called weighted Gram-Schmitt in Numerical Analysis

iii) The result of this procedure depends on the order of the colums of the original matrix Hence there might be several feasible solutions.

With the new design matrix Y^* we can now also find the new parameters $\beta_j^*(\theta_r)$ $j = 1,2, \ldots p$ The regression equation becomes

R) $$\mu(\theta_r) = Y^* \beta^*(\theta_r)$$

which reads componentwise

$$\mu_j(\theta_r) = \sum_{j=1}^{p} Y_{ij}^* \beta_j^*(\theta_r).$$

Multiply both sides by $Y_{ik}^* \dfrac{V_i}{V}$ and sum over i

$$\sum_{i=1}^{n} Y_{ik}^* \mu_i(\theta_r) \frac{V_i}{V.} = \sum_{j=1}^{p} \sum_{i=1}^{n} Y_{ik}^* Y_{ij}^* \beta_j^*(\theta_r) \frac{V_i}{V}$$

leading to

$$E^{(s)}[Y_k^* \mu(\theta_r)] = E^{(s)}[(Y_k^*)^2] \ \beta_k^*(\theta_r) \tag{24}$$

where, on the right hand side, we have used the orthogonality of Y_k^* and Y_j^* for $j \neq k$
Hence

$$\beta_k^*(\theta_r) = \frac{E^{(s)}[Y_k^* \mu(\theta_r)]}{E^{(s)}[(Y_k^*)^2]} \quad k = 1,2, \ldots, p \tag{25}$$

which defines our new parameters in the barycentric model
You should observe that this transformation of the regression parameters $\beta_j(\theta_r)$ may lead to new parameters $\beta_j^*(\theta_r)$ which are sometimes difficult to interprete In each application one has therefore to decide whether the orthogonality property of the design matrix or the interpretability of the regression parameters is more important
Luckily – as we have seen – there is no problem with the interpretation in the case of simple linear regression and interpretability is also not decisive if we are interested in prediction only

6.5 An example

Suppose that we want to model $\mu_k(\theta_r)$ as depending on time in a quadratic manner, *i e*

$$\mu_k(\theta_r) = \beta_0(\theta_r) + k\beta_1(\theta_r) + k^2\beta_2(\theta_r)$$

Our design matrix is hence of the following form

$$Y = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \cdot & \vdots & \cdot \\ 1 & k & k^2 \\ \vdots & \vdots & \\ 1 & n & n^2 \end{pmatrix}$$

Let us construct the design matrix Y^* with orthogonal columns.
Following the procedure as outlined in 6.4 we obviously have for the first two columns those obtained in the case of simple linear regression (measuring time from its barycenter) and we only have to construct Y_3^*
Formally

$$Y^* = \begin{pmatrix} 1 & 1 - E^{(s)}[k] & Y_{13}^* \\ 1 & 2 - E^{(s)}[k] & Y_{23}^* \\ \cdot & \cdot & \vdots \\ \cdot & \cdot & \\ 1 & n - E^{(s)}[k] & Y_{n3}^* \end{pmatrix}$$

To find Y_3^* we must solve

$$\sum_{k=1}^{n}\left(k^2 - a_1^* - a_2^*(k - E^{(s)}[k])\right)^2 \frac{V_k}{V} = \min!$$

Using relation (23) we obtian

$$a_1^* = E^{(s)}[k^2]$$
$$a_2^* = \frac{E^{(s)}[k^2(k - E^{(s)}[k])]}{Var^{(s)}[k]}$$

Hence we get

$$Y_{k3}^* = k^2 - E^{(s)}[k^2] - \frac{E^{(s)}[k^2(k - E^{(s)}[k])]}{Var^{(s)}[k]}(k - E^{(s)}[k]) \quad k = 1, 2, \ldots, n \tag{26}$$

and from

$$\mu_i(\theta) = \sum\nolimits_{j=1}^{3} Y_{ij}^* \beta_j^*(\theta_r) \qquad \text{R)}$$

we get both

– the interpretation of $\beta_j^*(\theta_i)$ (use (25))

– the prediction $\hat{\mu}_i(\theta) = \sum\nolimits_{j=1}^{3} Y_{ij}^* \hat{\beta}_j^*(\theta_i)$

where $\hat{\beta}_i^*(\theta_r)$ is the credibility estimator. Due to orthogonality of Y^* it can be obtained componentwise

7. FINAL REMARKS

Our whole discussion of the general case is based on a particular fixed sampling distribution. As this distribution typically varies from risk to risk Y^*, β^* and Z^* depend on the risk r and we cannot achieve orthogonality of Y^* simultaneously for all risks r This is the problem which we have already discussed in section 5 The observations made there apply also to the general case and the basic lesson is the same You should construct the orthogonal Y^* for the sampling distribution of the whole collective which then will often lead to "nearly orthogonal" design matrices for the individual risks which again "nearly separates" the credibility formula into componentwise procedures

The question not addressed in this paper is the one of choice of the number of regression parameters In the case of simple linear regression this question would be. Should you use a linear regression function, a quadratic or a higher order polynominal? Generally the question is. How should one choose the design matrix to start with? We hope to address this question in a forthcoming paper

ACKNOWLEDGEMENT

The authors wish to thank Peter Buhlmann, University of California, Berkeley for very stimulating ideas and discussions which have led to the formulation of the general case barycentric credibility.

REFERENCES

HACHEMEISTER, C (1975) *Credibility for regression models with application to trend* Credibility Theory and Applications, (e d D M Kahn), 129-163 Academic Press, New York

DANNENBURG, D (1996) *Basic actuarial credibility models* PhD Thesis University of Amsterdam

DE VYLDER, F (1981) *Regression model with scalar credibility weights* Mitteilungen Vereinigung Schweizerischer Versicherungsmathematiker, Heft 1, 27-39

DE VYLDER, F (1985) *Non-linear regression in credibility theory* Insurance Mathematics and Economics **4**, 163-172

PROF. H BUHLMANN
Mathematics Department
ETH Zurich
CH – 8092 Zurich

DR. A. GISLER
"Winterthur", Swiss Insurance Co.
P O. Box 357
CH – 8401 Winterthur

THE SWISS RE EXPOSURE CURVES AND THE MBBEFD[1] DISTRIBUTION CLASS

STEFAN BERNEGGER

ABSTRACT

A new two-parameter family of analytical functions will be introduced for the modelling of loss distributions and exposure curves The curve family contains the Maxwell-Boltzmann, the Bose-Einstein and the Fermi-Dirac distributions, which are well known in statistical mechanics. The functions can be used for the modelling of loss distributions on the finite interval [0, 1] as well as on the interval [0, ∞]. The functions defined on the interval [0, 1] are discussed in detail and related to several Swiss Re exposure curves used in practice The curves can be fitted to the first two moments μ and σ of a loss distribution or to the first moment μ and the total loss probability p.

1 INTRODUCTION

Whenever possible, the rating of non proportional (NP) reinsurance treaties should not only rely on the loss experience of the past, but also on actual exposure. For the case of per risk covers, exposure rating is based on risk profiles All risks of similar size (SI, MPL or EML) belonging to the same risk category are summarized in a risk band For the purpose of rating, all the risks belonging to one specific band are assumed to be homogeneous They can thus be modelled with the help of one single loss distribution function.

The problem of exposure rating is how to divide the total premiums of one band between the ceding company and the reinsurer The problem is solved in two steps First, the overall risk premiums (per band) are estimated by applying an appropriate loss ratio to the gross premiums. In a second step, these risk premiums are divided into risk premiums for the retention and risk premiums for the cession Due to the nature of NP reinsurance, this is possible only with the help of the loss distribution function.

However, the correct loss distribution function for an individual band of a risk profile is hardly known in practice. This lack of information is overcome with the help of distribution functions derived from large portfolios of similar risks. Such distribution functions are available in the form of so-called exposure curves These curves directly permit the extraction of the risk premium ratio required by the reinsurer as a function of the deductible.

[1] Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac distribution

ASTIN BULLETIN Vol 27, No 1 1997, pp 99-111

Often, underwriters have only a finite number of discrete exposure curves at their disposal. These curves are available in graphical or tabulated form, and are also implemented in computerized underwriting tools One of the curves must be selected for each risk band, but it is not always clear which curve should be used In such cases, the underwriter might also want to use a virtual curve lying between two of the discrete curves available to him.

This can be achieved by replacing the discrete curves with analytical exposure curves Each set of parameters then defines another curve. If a continuous set of parameters is available, the exposure curves can be varied smoothly within the whole range of available curves However, the curves must fulfill certain conditions which restrict the range of the parameters In addition, practical problems can arise if a curve family with many (more than two) parameters is used. It might then become very difficult to find a set of parameters which can be associated with the information available for a class of risks. This problem can be overcome if a curve family is restricted to a one- or two-parameter subclass and if new parameters are introduced which can easily be interpreted by the underwriters

In the following, the MBBEFD class of analytical exposure curves will be introduced As will be seen, this class is very well suited for the modelling of exposure curves used in practice. Before analysing the MBBEFD curves in detail, some general relations between a distribution function and its related exposure curve will be discussed in section 2 These relations permit the derivation of the conditions to be fulfilled by exposure curves The new, two-parameter class of distribution functions will then be introduced in section 3 Finally, several practical aspects, and the link to the well known Swiss Re property exposure curves Y_i, will be discussed in section 4

Conventions

Following the notation used by Daykin et al in [1], we will denote stochastic variables by bold letters, e.g. **X** or **x**. Monetary variables are denoted by capital letters, for instance, X or M, while ratio variables are denoted by small letters, for instance, x = X/M.

2. DISTRIBUTION FUNCTION AND EXPOSURE CURVE

2.1. Definition of the exposure curve

In the following, the relation between the distribution function F(x) defined on the interval [0, 1] and its limited expected value function L(d) = E[min(d, **x**)] will be discussed. Here, d = D/M and **x** = **X**/M represent the normalized deductible and the normalized loss, respectively. M is the maximum possible loss (MPL) and **X** ≤ M the gross loss The deductible D is the cedent's maximum retention under a non proportional reinsurance treaty M L (d) is the expected value of the losses retained by the cedent while M · (L(1) – L(d)) is the expected value of the losses paid by the reinsurer Thus, the ratio of the pure risk premiums retained by the cedent is given by the relative limited expected value function G(d) = L(d)/L(1) [1] The curve representing this function is also called the **exposure curve**

$$G(d) = \frac{L(d)}{L(1)} = \frac{\int_0^d (1-F(y))dy}{\int_0^1 (1-F(y))dy} = \frac{\int_0^d (1-F(y))dy}{E[x]} \tag{2.1}$$

Because of 1 – F(x) ≥ 0 and F'(x) = f(x) ≥ 0, G(d) is an increasing and concave function on the interval [0, 1] In addition, G(0) = 0 and G(1) = 1 by definition

2.2. Deriving the distribution function from the exposure curve

If the exposure curve G(x) is given, the corresponding distribution function F(x) can be derived from:

$$G'(d) = \frac{1-F(d)}{E[x]} \tag{2.2}$$

With F(0) = 0 and G'(0) = 1/E[x] one obtains

$$F(x) = \begin{cases} 1 & x = 1 \\ 1 - \dfrac{G'(x)}{G'(0)} & 0 \le x < 1 \end{cases} \tag{2.3}$$

Thus, F(x) and G(x) are equivalent representations of the loss distribution

2.3. Total loss probability and expected value

The probability p for a total loss equals 1 – F(1⁻) and the expected (or average) loss μ equals E[**x**]. These two functionals of the distribution function F(x) can be derived directly from the derivatives of G(x) at x = 0 and x = 1:

$$\begin{aligned} \mu &= E[x] = \frac{1}{G'(0)} \\ p &= 1 - F(1^-) = \frac{G'(1)}{G'(0)} \end{aligned} \tag{2.4}$$

The fact that G(x) is a concave and increasing function on the interval [0 ,1] with G(0) = 0 and G(1) = 1 implies:

$$G'(0) \ge 1 \ge G'(1) \ge 0 \tag{2.5}$$

This is also reflected in the relation:

$$0 \le p \le \mu \le 1 \tag{2.6}$$

2.4. Unlimited distributions

If the distribution function F(X) is defined on the interval [0, ∞], the above relations have to be slightly modified. In this case there is no finite maximum loss M However, the deductible D and the losses **X** can be normalized with respect to an arbitrary reference loss X_0, i e **x** = **X**/X_0 and d = D/X_0 G(d) is still a concave and increasing function with G(0) = 0 and G(∞) = 1 The expected value μ = E[**x**] is also given by 1/G'(0), but there are no total losses, i.e. G'(∞) = 0

3 THE MBBEFD CLASS OF TWO-PARAMETER EXPOSURE CURVES

3.1. Definition of the curve

In this section we will investigate the exposure curves and the related distribution functions defined by.

$$G(x) = \frac{\ln(a+b^x) - \ln(a+1)}{\ln(a+b) - \ln(a+1)} \tag{3 1 a}$$

The distribution function belonging to this exposure curve is given by

$$F(x) = \begin{cases} 1 & x = 1 \\ 1 - \dfrac{(a+1)b^x}{a+b^x} & 0 \le x < 1 \end{cases} \tag{3.1 b}$$

The denominator and the term -ln(a + 1) in the nominator of (3.1 a) ensure that the boundary conditions G(0) = 0 and G(1) = 1 are fulfilled As will be seen below, the cases a = {-1, 0, ∞} or b = {0, 1, ∞} have to be treated separately.

*Distribution functions of the type (3 1), defined on the interval [0, ∞] or [-∞, ∞], are very well known in statistical mechanics (**Maxwell-Boltzmann**, **Bose-Einstein**, **Fermi-Dirac** and **Planck** distribution) The implementation of these functions in risk theory does not mean that the distribution of insured losses can be derived from the theory of statistical mechanics However, the MBBEFD distribution class defined in (3.1) shows itself to be very appropriate for the modelling of empirical loss distributions on the interval [0, 1].*

3.2. New parametrisation

The parameters {a, b} are restricted to those values, for which $G_{a,b}(x)$ is a real, increasing and concave function on the interval [0, 1] It is easier to fulfill this condition by using the inverse g = 1/p of the total loss probability p as a curve parameter and to replace the parameter a in (3 1).

$$g = \frac{a+b}{(a+1)b}, \quad a = \frac{(g-1)b}{1-gb} \tag{3 2}$$

On the one hand, the condition $0 \le p \le 1$ is fulfilled only for $g \ge 1$. On the other hand, G(x) is a real function only for $b \ge 0$. It can be shown that no other restrictions regarding the set of parameters are necessary

However, cases $b = 1$ (i.e. $a = -1$), $b = 0$ or $g = 1$ (i.e. $a = 0$) and $b \cdot g = 1$ (i.e. $a = \infty$) must be treated as special cases. The cases $b \cdot g = 1$ (i.e. $a = \infty$), $b \cdot g > 1$ (i.e. $a < 0$) and $b \cdot g < 1$ (i.e. $a > 0$) correspond to the MB, the BE and the FD distribution, respectively (cf. figure 4.1). By considering special cases $b = 1$, $g = 1$ and $b \cdot g = 1$ separately, all real, increasing and concave functions G(x) on the interval [0, 1] with G(0) = 0 and G(1) = 1 belonging to the MBBEFD class (3.1) can be represented as follows:

$$G_{b,g}(x) = \begin{cases} x & g = 1 \vee b = 0 \\ \dfrac{\ln(1+(g-1)x)}{\ln(g)} & b = 1 \wedge g > 1 \\ \dfrac{1-b^x}{1-b} & bg = 1 \wedge g > 1 \\ \dfrac{\ln\left(\dfrac{(g-1)b+(1-gb)b^x}{1-b}\right)}{\ln(gb)} & b > 0 \wedge b \ne 1 \wedge bg \ne 1 \wedge g > 1 \end{cases} \qquad (3.3)$$

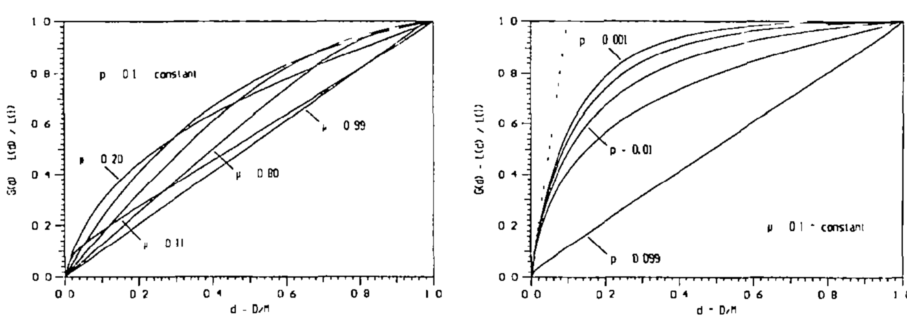


FIGURE 3.1 a) Set of MBBEFD exposure curves with constant parameter $g = 1/p = 10$ and $\mu = E[x] = 0.11$, 0.2, 0.4, 0.6, 0.8, 0.99

FIGURE 3.1 b) Set of MBBEFD exposure curves with constant $\mu = E[x] = 0.1$ and $p = 1/g = 0.099$, 0.031, 0.01, 0.0031, 0.001 The dashed line with slope $1/\mu$ represents the tangent at d = 0

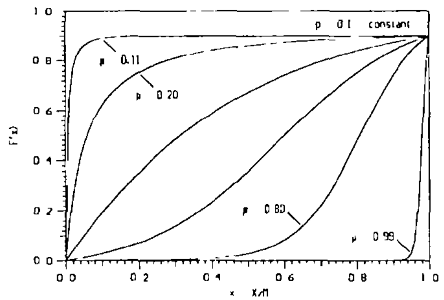


FIGURE 3 2 a) Distribution functions belonging to exposure curves of figure 3 1 a)

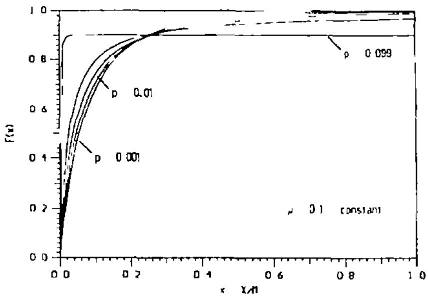


FIGURE 3 2 b) Distribution functions belonging to exposure curves of figure 3 1 b)

Examples of MBBEFD exposure curves are shown in figure 3 1 A set of curves with constant total loss probability p = 0 1 (i.e. g = 10) is represented in figure 3.1 a). Figure 3.1 b) contains a set of curves with constant expected value $\mu = 0.1$ The corresponding distribution functions are shown in figures 3.2 a) and b)

3.3. Derivatives

The derivatives of the exposure curves are given by

$$G'(x) = \begin{cases} 1 & g = 1 \vee b = 0 \\ \dfrac{g-1}{\ln(g)(1+(g-1)x)} & b = 1 \wedge g > 1 \\ \dfrac{\ln(b)b^x}{b-1} & bg = 1 \wedge g > 1 \\ \dfrac{\ln(b)(1-gb)}{\ln(gb)\left((g-1)b^{1-x}+(1-gb)\right)} & b > 0 \wedge b \neq 1 \wedge bg \neq 1 \wedge g > 1 \end{cases} \tag{3 4}$$

with

$$G'(0) = \begin{cases} 1 & g = 1 \vee b = 0 \\ \dfrac{g-1}{\ln(g)} & b = 1 \wedge g > 1 \\ \dfrac{\ln(b)}{b-1} = \dfrac{\ln(g)g}{g-1} & bg = 1 \wedge g > 1 \\ \dfrac{\ln(b)(1-gb)}{\ln(gb)(1-b)} & b > 0 \wedge b \neq 1 \wedge bg \neq 1 \wedge g > 1 \end{cases} \tag{3 4 a}$$

and

$$G'(1) = \begin{cases} 1 & g = 1 \vee b = 0 \\ \dfrac{g-1}{\ln(g)g} & b = 1 \wedge g > 1 \\ \dfrac{\ln(b)b}{b-1} = \dfrac{\ln(g)}{g-1} & bg = 1 \wedge g > 1 \\ \dfrac{\ln(b)(1-gb)}{\ln(gb)g(1-b)} & b > 0 \wedge b \neq 1 \wedge bg \neq 1 \wedge g > 1 \end{cases} \quad (3\,4\,\text{b})$$

The relation p = G'(1)/G'(0) = 1/g is obtained immediately from (3.4 a) and (3 4 b)

3.4. Expected value

According to (2 4) the expected value μ is given by·

$$\mu = E[x] = \frac{1}{G'(0)} = \begin{cases} 1 & g = 1 \vee b = 0 \\ \dfrac{\ln(g)}{g-1} & b = 1 \wedge g > 1 \\ \dfrac{b-1}{\ln(b)} = \dfrac{g-1}{\ln(g)g} & bg = 1 \wedge g > 1 \\ \dfrac{\ln(gb)(1-b)}{\ln(b)(1-gb)} & b > 0 \wedge b \neq 1 \wedge bg \neq 1 \wedge g > 1 \end{cases} \quad (3.5)$$

The expected value μ is represented as a function of the parameters b and g in figure 3 3 and discussed below in section 3.7.

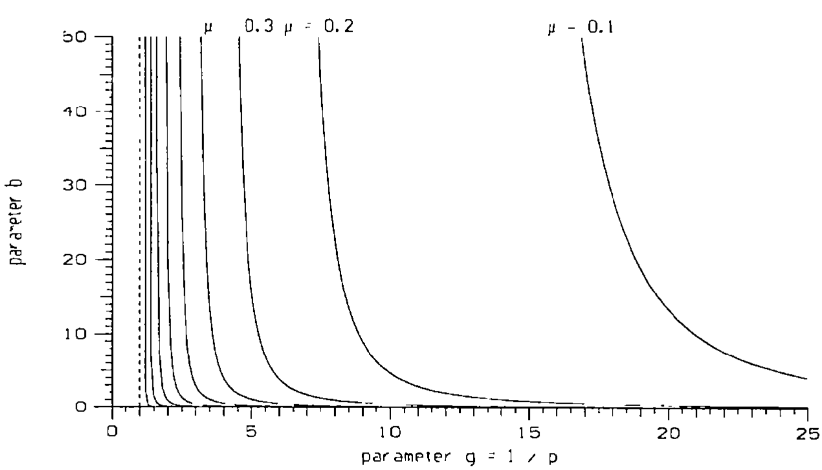


FIGURE 3 3 Parameter b as a function of g = 1/p for μ = E[x] = 0 1, 0 2, 0 9
The dashed line at g = 1 and the horizontal line at b = 0 represent the parameter sets {b, g} with μ = 1

3.5. Distribution function

According to (2 3), the distribution function belonging to the exposure curve $G_{b,g}(x)$ is given by.

$$F(x)=\begin{cases}1 & x=1\\ 0 & x<1\wedge(g=1\vee b=0)\\ 1-\dfrac{1}{1+(g-1)x} & x<1\wedge b=1\wedge g>1\\ 1-b^{x} & x<1\wedge bg=1\wedge g>1\\ 1-\dfrac{1-b}{(g-1)b^{1-x}+(1-gb)} & x<1\wedge b>0\wedge b\neq 1\wedge bg\neq 1\wedge g>1\end{cases}\qquad(3\ 6)$$

The distribution functions belonging to the exposure curves of figure 3.1 are represented in figure 3 2 The set of distribution functions with constant total loss probability p = 0.1 (g = 10) is shown in figure 3 2 a). Figure 3 2 b) contains the set of distribution functions with constant expected value $\mu = 0\ 1$

3.6. Density function

Because of the finite probability p = 1/g for a total loss, the density function f(x) = F'(x) is defined only on the interval [0, 1).

$$f(x)=\begin{cases}0 & g=1\vee b=0\\ \dfrac{g-1}{\left(1+(g-1)x\right)^{2}} & b=1\wedge g>1\\ -\ln(b)b^{x} & bg=1\wedge g>1\\ \dfrac{(b-1)(g-1)\ln(b)b^{1-x}}{\left((g-1)b^{1-x}+(1-gb)\right)^{2}} & b>0\wedge b\neq 1\wedge bg\neq 1\wedge g>1\end{cases}\qquad(3\ 7)$$

3.7. Discussion

It is instructive to analyse the expected value $\mu = \mu(b, g)$ as a function of the parameters b and g (3.5). Figure 3.3. shows the range of permitted parameters in the {b, g} plane and the curves with constant expected value μ. One can see in figure 3 3 that $\mu_g(b)$ is a decreasing function of b (for g > 1 constant) and that $\mu_b(g)$ is a decreasing function of g (for b > 0 constant)

$$\begin{aligned}&\frac{\partial}{\partial b}\mu_g(b)\le 0\\ &\frac{\partial}{\partial g}\mu_b(g)\le 0\end{aligned}\qquad g>1\wedge b>0\qquad(3\ 8)$$

The expected value μ is related as follows to the extreme values of the parameters b and g

$$\lim_{b\to 0}\mu_g(b)=1; \qquad \lim_{b\to\infty}\mu_g(b)=1/g=p$$
$$\lim_{g\to 1}\mu_b(g)=1; \qquad \lim_{g\to\infty}\mu_b(g)=0 \tag{3.9}$$

3.8. Unlimited distributions

So far, only distributions defined on the interval [0, 1] have been discussed However, as the MB, the BE and the FD distributions are defined on the interval [-∞, ∞] or [0, ∞], the MBBEFD distribution class can also be used for the modelling of loss distributions on the interval [0, ∞] If the losses **X** and the deductible D are normalized with respect to an arbitrary reference loss X_0, then **x** = **X**/X_0 and d = D/X_0 The above formula can now be modified as follows:

$$G_{b,g}(x)=\begin{cases}1-b^x & bg=1\wedge g>1\\[2ex] \dfrac{\ln\left(\dfrac{(g-1)b+(1-gb)b^x}{1-b}\right)}{\ln\left(\dfrac{(g-1)b}{1-b}\right)} & 0<b<1\wedge bg\neq 1\wedge g>1\end{cases} \tag{3.10}$$

$$G'(x)=\begin{cases}-\ln(b)b^x & bg=1\wedge g>1\\[2ex] \dfrac{\ln(b)(1-gb)}{\ln\left(\dfrac{(g-1)b}{1-b}\right)\left((g-1)b^{1-x}+(1-gb)\right)} & 0<b<1\wedge bg\neq 1\wedge g>1\end{cases} \tag{3.11}$$

$$G'(0)=\begin{cases}-\ln(b) & bg=1\wedge g>1\\[2ex] \dfrac{\ln(b)(1-gb)}{\ln\left(\dfrac{(g-1)b}{1-b}\right)(1-b)} & 0<b<1\wedge bg\neq 1\wedge g>1\end{cases} \tag{3.11 a}$$

$$G'(1)=\begin{cases}-\ln(b)b & bg=1\wedge g>1\\[2ex] \dfrac{\ln(b)(1-gb)}{\ln\left(\dfrac{(g-1)b}{1-b}\right)g(1-b)} & 0<b<1\wedge bg\neq 1\wedge g>1\end{cases} \tag{3.11 b}$$

$$G'(\infty)=0 \tag{3.11 c}$$

$$F(x)=\begin{cases}1-b^x & bg=1\wedge g>1\\[2ex] 1-\dfrac{1-b}{(g-1)b^{1-x}+(1-gb)} & 0<b<1\wedge bg\neq 1\wedge g>1\end{cases} \tag{3.12}$$

The restriction $0 < b < 1$ is obtained immediately from (3.12) and the condition $F(\infty) = 1$, while the restriction $g > 1$ is obtained from (3 10), where the argument of the logarithm in the denominator must be greater than 0 The same restriction is also obtained from the relation $p = G'(1)/G'(0) = 1/g$, which is still valid The parameter g is thus the inverse of the probability p of having a loss **X** exceeding the reference loss X_0

4 CURVE FITTING

4.1. Expected value μ and total loss probability p

Because of (3 8) and (3.9), there exists exactly one distribution function belonging to the MBBEFD class for each given pair of functionals p and μ (cf figure 3 3), provided that p and μ fulfill the conditions (2 6) The curve parameter $g = 1/p$ is obtained directly The second curve parameter b can be calculated with the help of (3 5) Here, the following cases must be distinguished:

$$\begin{array}{lll} a) & \mu = 1 & \Rightarrow b = 0 \\ b) & \mu = \dfrac{g-1}{\ln(g)g} & \Rightarrow b = 1/g \\ c) & \mu = \dfrac{\ln(g)}{g-1} & \Rightarrow b = 1 \\ d) & \mu = 1/g & \Rightarrow b = \infty \\ e) & else & \Rightarrow 0 < b < \infty \wedge b \neq 1/g \wedge b \neq 1 \end{array} \quad (4\ 1)$$

In the general case e), the parameter b has to be calculated iteratively by solving the equation·

$$\mu = \frac{\ln(gb)(1-b)}{\ln(b)(1-gb)} \quad (4\ 2)$$

Because $\mu_g(b)$ is a decreasing function of b (3 8), the iteration causes no problems. An upper and a lower limit for b can be derived directly from (4 1).

4.2. Expected value μ and standard deviation σ

It is also possible to find a MBBEFD distribution assuming the first two moments (e g. μ and σ) are known, provided the moments fulfill certain conditions The first two moments of a distribution function with total loss probability p are given by:

$$\begin{aligned} \mu &= E[x] = p + \int_0^1 xf(x)dx \\ \mu^2 + \sigma^2 &= E[x^2] = p + \int_0^1 x^2 f(x)dx \leq \mu \end{aligned} \quad (4\ 3)$$

According to (4.3) the first two moments of F(x) and p must fulfill the following conditions

$$\begin{aligned} \mu^2 &\leq E[x^2] \leq \mu \\ p &\leq E[x^2] \end{aligned} \tag{4.4}$$

Calculation of g and b

Basic idea:
1. Start with $p^* = E[x^2] \geq p$ as a first estimate (upper limit) for p, and calculate b^* and g^* for the given functionals μ and p^* with the method described in 4.1 above.
2. Compare the second moment $E^*[x^2]$ with the given moment $E[x^2]$ and find a new estimate for p^*.
3. Repeat until $E^*[x^2]$ is close enough to $E[x^2]$

If the first moment μ is kept constant, then the second moment $E^*[x^2]$ will be an increasing function of p^*. Thus the parameters g and b can be calculated without complications

Remark. The second moment of the MBBEFD distribution has to be calculated numerically This is best done by replacing F(x) with a discrete distribution function which has the same upper tail area $L(x_{i+1}) - L(x_i)$ as F(x) on each discretized interval $[x_i, x_{i+1}]$

4.3. The MBBEFD distribution class and the Swiss Re Y_i property exposure curves

The Swiss Re Y_i exposure curves ($i = 1 \ldots 4$) are very well known and widely used by non proportional property underwriters As will be shown in this section, all these curves can be approximated very well with the help of a subclass of the MBBEFD exposure curves. In a first step, the parameters b_i and g_i have been evaluated for each curve i By plotting the points belonging to these pairs of parameters in the {b, g} plane, we found that the points were lying on a smooth curve in the plane In a next step, this curve was modelled as a function of a single curve parameter c. Finally, the parameters c_i representing the curves Y_i were evaluated

The subclass of the one-parameter MBBEFD exposure curves is defined as follows:

$$G_c(x) = G_{b_c, g_c}(x) \tag{4.5}$$

with

$$\begin{aligned} b_c &= b(c) = e^{3.1-0.15(1+c)c} \\ g_c &= g(c) = e^{(0.78+0.12c)c} \end{aligned} \tag{4.6}$$

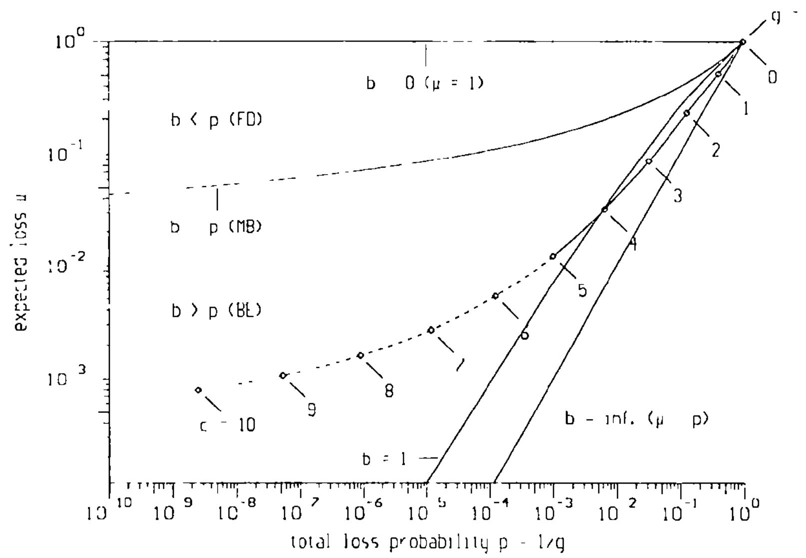


FIGURE 4.1 Range of parameters of the exposure curves $G_{b,p}(x)$. The expected value μ is shown as a function of $p = 1/g$ for special cases $b = 0$, $b = p$, $b = 1$ and $b = \infty$. In addition, p and μ are shown as a function of the curve parameter c for c = 0 … 10. The dashed part of this curve has no empirical counterparts.

The position of the curves c = 0 … 10 in the {p, μ} plane is shown in figure 4.1. Here, the special cases b = 0, p, 1, ∞ and g = 1 are also shown.

The curves defined by c = 0.0, ..., 5.0, which are shown in figure 4.2, are related as follows to several exposure curves used in practice:

- The curve c = 0 represents a distribution of total losses only because of g(0) = 1.
- The four curves defined by c = {1.5, 2.0, 3.0 and 4.0} coincide very well with the Swiss Re curves {Y_1, Y_2, Y_3, Y_4}.
- The curve defined by c = 5.0 coincides very well with a Lloyd's curve used for the rating of industrial risks.

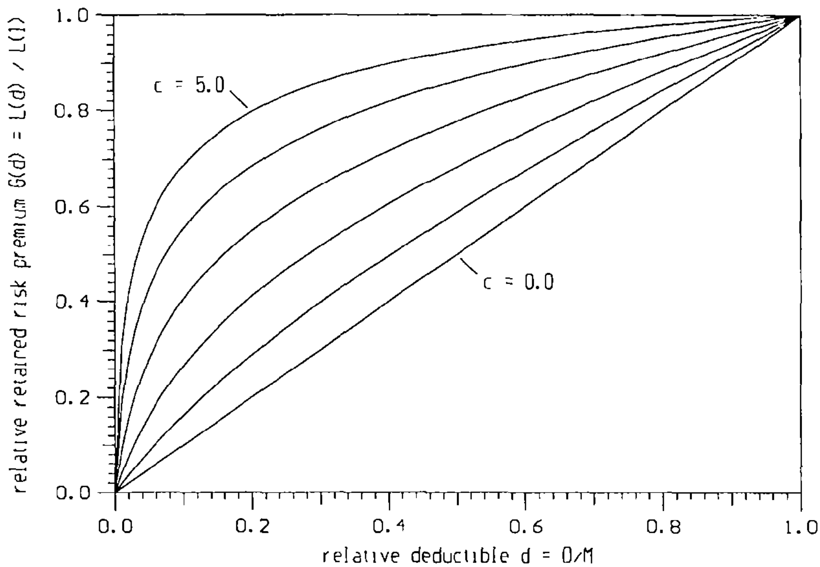


FIGURE 4 2 One-parameter subclass of the MBBEFD exposure curves, shown for c = 0 0, 1 0, 2 0, 3 0, 4 0 and 5 0

Thus, the exposure curves defined in (4.6) are very well suited for practical purposes The underwriter can use curve parameters which are very familiar to him In addition, the class of exposure curves defined by (4 6) is continuous and the underwriter has at his disposal all curves lying between the individual curves Y_i, too

REFERENCES

C D DAYKIN, T PENTIKAINEN AND M PESONEN (1994) "Practical Risk Theory for Actuaries" *Chapman & Hall*, London

A SEMI-PARAMETRIC PREDICTOR OF THE IBNR RESERVE

LOUIS G. DORAY[1]

Université de Montréal

ABSTRACT

We develop a semi-parametric predictor of the IBNR reserve in a macro-model when the claim amount for a certain accident and development year can be expressed in a loglinear form composed of a deterministic part and a random error. We need to make assumptions only on the first two moments of the error, without any specified parametric assumption on its distribution. We give its properties, present its advantages and compare the estimates obtained with various predictors of the IBNR reserve, parametric and non-parametric, using a data set.

KEYWORDS

Chain-ladder; regression, least-squares; smearing estimator.

1 INTRODUCTION

In a macro-model, claims are grouped by accident year (year in which the accident giving rise to a claim occurrs) and development year (number of years elapsed since the accident), and data are presented in a trapezoidal array. Taylor (1986) presents a comprehensive survey of various macro methods and models, both deterministic and stochastic, developed to predict incurred but not reported (IBNR) reserves; it is usually assumed that the pattern of cumulative claims incurred or paid is stable across the development years, for each accident year. The problem of setting IBNR reserves consists in predicting for each accident year, the ultimate amount of claims incurred and subtracting the amount already paid by the insurer.

To illustrate the predictor proposed in this paper, we will use the cumulative claims appearing in Doray (1996), which represent the liability claims in thousands of dollars incurred by a Canadian insurance company over the ten-year period 1978-1987. We will perform the analysis on the incremental claims (see Table 1), obtained by differencing successive cumulative amounts, and assume that they are independent. Section 2 presents the loglinear model used, and section 3 the semi-parametric predictor of the IBNR reserve, finally, we compare various predictors of the reserve with the claims of Table 1.

[1] The author gratefully acknowledges the financial support of the Natural Sciences and Engineering Research Council of Canada

TABLE 1

INCREMENTAL CLAIMS INCURRED

	Development year					
Accident year	1	2	3	4	5	6
1978	8489	1296	924	580	246	126
1979	12970	1796	1435	859	654	265
1980	17522	2783	1469	1023	423	652
1981	21754	2584	1163	783	887	355
1982	19208	2341	1220	619	841	703
1983	19604	2469	1223	1247	612	
1984	21922	2311	1141	1508		
1985	25038	3363	2144			
1986	32532	4474				
1987	39862					

2 A LOGLINEAR MODEL

We consider models of the form $Y_i = \exp(X_i\beta + \sigma\varepsilon_i)$, or expressed as a loglinear regression model,

$$Z_i = \ln Y_i = X_i\beta + \sigma\varepsilon_i, \quad Y_i > 0 \tag{2 1}$$

where Y_i is the ith element of the data vector Y, of dimension n, X is the regression matrix of dimension n x p, whose ith row is the vector X_i, element (i, j) is denoted X_{ij}, and where we assume that the unit vector is in the column space of X, β is the vector (of dimension p) of unknown parameters to be estimated, and ε_i are independent random errors with mean 0 and variance 1

For the regression parameters, various choices are possible, for example $\alpha_i + \beta_j$ for the stochastic chain ladder model, where i is the accident year and j, the development year, or $\alpha + \beta \ln j + \gamma j + \iota(i + j - 2)$, as in Zehnwirth (1990).

This paper does not study models which rely on parametric assumptions for the distribution of the error ε, instead, we present a semi-parametric regression model which does not assume any particular density for ε, but uses its first two moments only

3 A PREDICTOR IMPLIED BY THE SMEARING ESTIMATOR

Let us represent by Y_k a value to be predicted, corresponding to a cell in the lower right unobserved triangle of Table 1 ($i = 6,$, 10 and $j = 12 - i,$., 6) Doray (1996) analyzed the two types of errors involved in the prediction of the value Y_k by its expected value, the estimation error on the parameter β from past values and the process error ε_k for a future value, yielding $X_k\tilde{\beta} + \tilde{\sigma}\varepsilon_k$, where X_k is the vector of coefficients of the parameters corresponding to Z_k.

According to Gauss-Markov theory, the least-square estimator $\tilde{\beta} = (X'X)^{-1}X'Z$ is the minimum variance linear unbiased estimator of β, for any distribution of ε such that $E(\varepsilon) = 0$ and $Var(\varepsilon) = 1$. The variance σ^2 is estimated by the mean-square error

$\tilde{\sigma}^2 = (Z - X\tilde{\beta})'(Z - X\tilde{\beta})/(n-p)$ For a fixed vector X_k, $X_k\tilde{\beta}$ is an unbiased and consistent estimator of $X_k\beta$, but $\exp(X_k\tilde{\beta})$ is not in general an unbiased or consistent estimator of $E(Y_k)$ The assumption that ε is normal influences only the efficiency of the estimator $\tilde{\beta}$, if the true error is not normal, the estimator $\tilde{\beta}$ is still consistent and minimum variance linear unbiased. If ε is normal, $\exp(X_k\tilde{\beta} + \tilde{\sigma}^2/2)$ is a consistent estimator for $E(Y_k)$, however, the predictor for Y_k will not be consistent if the assumption that ε is $N(0, 1)$ is wrong.

Duan (1983) proposed the following smearing estimator for the expected value of Y_k, $\frac{1}{n}\sum_{i=1}^{n}\exp(X_k\tilde{\beta} + \tilde{\sigma}\tilde{\varepsilon}_i)$, where $\tilde{\varepsilon}_i = Z_i - X_i\tilde{\beta}$ denotes the least-squares residual. He shows that under certain regularity conditions, the smearing estimator of $E(Y_k)$ is weakly consistent and notes that for small σ^2, its relative efficiency compared to the simple estimator $\exp(X_k\tilde{\beta} + \tilde{\sigma}^2/2)$ is very high when the error distribution is normal (for $\sigma^2 \leq 1\ 00$ and rank $(X) \geq 3$, it is at least 94%) This efficiency increases as σ^2 decreases or rank (X) increases

Using the smearing estimator, we can define the following semi-parametric predictor of the IBNR reserve,

$$\hat{\theta}_{SP} = \sum_k \frac{1}{n}\sum_{i=1}^{n}\exp(X_k\tilde{\beta} + \tilde{\sigma}\tilde{\varepsilon}_i) = \left(\sum_k \exp(X_k\tilde{\beta})\right) \times \left(\frac{1}{n}\sum_{i=1}^{n}\exp(\tilde{\sigma}\tilde{\varepsilon}_i)\right),$$

where $\sum_k$ denotes a summation over all cells in the lower triangle to be predicted.

4 COMPARISON OF VARIOUS PREDICTORS

We can obtain a simple approximation for $\hat{\theta}_{SP}$ when σ^2 is small by using the first three terms of the Taylor's series expansion for $\exp(\tilde{\sigma}\tilde{\varepsilon}_i)$, and the facts that $\sum_{i=1}^{n}\tilde{\varepsilon}_i = 0$ and $\sum_{i=1}^{n}\tilde{\sigma}^2\tilde{\varepsilon}_i^2/2 = (n-p)\tilde{\sigma}^4/2$,

$$\hat{\theta}_{SP} \cong \hat{\theta}_A = \left(\sum_k \exp(X_k\tilde{\beta})\right) \times \left[1 + (n-p)\tilde{\sigma}^4/2n\right].$$

In Table 2, we compare the predicted values of the IBNR reserve obtained with the non-parametric predictors, $\hat{\theta}_{SP}$, $\hat{\theta}_A$, the chain-ladder ($\hat{\theta}_{CL}$), and predictors obtained when ε_i's in (2 1) are assumed to be i i.d $N(0, 1)$, the uniformly minimum variance unbiased predictor of Doray (1996)

$$\hat{\theta}_U = {}_0F_1\left(\frac{n-p}{2}; \frac{n-p}{4}\tilde{\sigma}^2\right)\sum_k \exp(X_k\tilde{\beta}),$$

where ${}_0F_1(\alpha, z)$ is the hypergeometric function defined as

$${}_0F_1(\alpha, z) = \sum_{j=0}^{\infty}\frac{z^j}{j!(\alpha)_j}, \text{ with } (\alpha)_j = \alpha(\alpha+1)\ .\ (\alpha+j-1),\ j \geq 1, \text{ and } (\alpha)_0 = 1,$$

the predictor of Kremer (1982), $\hat{\theta}_K = \sum_k \exp(X_k\tilde{\beta})$, and the simple estimator $\hat{\theta}_1 = \sum_k \exp(X_k\tilde{\beta} + \tilde{\sigma}^2/2)$ The model used was the stochastic chain ladder model $(\alpha_i + \beta_j)$, on the claims of Table 1. We notice that $\hat{\theta}_A$, $\hat{\theta}_U$ and $\hat{\theta}_1$ are of the form C x $\hat{\theta}_K$, where C is a factor depending only on $\tilde{\sigma}^2$

In conclusion, the smearing estimator possesses four important properties It is easily calculated, consistent, highly efficient if the error ε has a normal distribution and robust against departure from the assumed parametric distribution for ε It can also be used with transformations other than exponential The semi-parametric predictor of the IBNR reserve based on the smearing estimator will share those properties and present a worthwhile alternative to predictors based on full parametric assumptions.

TABLE 2

PREDICTION OF THE IBNR RESERVE

Predictor	Predicted value
$\hat{\theta}_{SP}$	23,552
$\hat{\theta}_A$	23,589
$\hat{\theta}_{CL}$	23,919
$\hat{\theta}_U$	24,403
$\hat{\theta}_K$	23,549
$\hat{\theta}_1$	24,404

REFERENCES

DORAY, L G (1996) UMVUE of the IBNR Reserve in a Lognormal Linear Regression Model, *Insurance Mathematics and Economics* **18,** 43-57

DUAN, N (1983) Smearing Estimate A Nonparametric Retransformation Method *Journal of the American Statistical Association* **78,** 605-610

KREMER, E (1982) IBNR claims and the two-way model of ANOVA *Scandinavian Actuarial Journal,* 47-55

TAYLOR G C (1986) Claims reserving in non-life insurance North-Holland, Amsterdam

ZEHNWIRTH, B (1990) Interactive Claims Reserving Forecasting System, Version 5 3, Manual Volume 1, Benhar Nominees Pty Ltd , Australia

DR LOUIS DORAY

Département de mathématiques et de statistique

Université de Montréal

C.P 6128, Succ Centre-ville Montréal, Qué , Canada, H3C 3J7

ESTIMATING THE TAILS OF LOSS SEVERITY DISTRIBUTIOINS USING EXTREME VALUE THEORY

ALEXANDER J. MCNEIL

Departement Mathematik
ETH Zentrum
CH-8092 Zurich

March 7, 1997

ABSTRACT

Good estimates for the tails of loss severity distributions are essential for pricing or positioning high-excess loss layers in reinsurance We describe parametric curve-fitting methods for modelling extreme historical losses These methods revolve around the generalized Pareto distribution and are supported by extreme value theory. We summarize relevant theoretical results and provide an extensive example of their application to Danish data on large fire insurance losses

KEYWORDS

Loss severity distributions, high excess layers; extreme value theory, excesses over high thresholds; generalized Pareto distribution

1 INTRODUCTION

Insurance products can be priced using our experience of losses in the past We can use data on historical loss severities to predict the size of future losses One approach is to fit parametric distributions to these data to obtain a model for the underlying loss severity distribution; a standard reference on this practice is Hogg & Klugman (1984).

In this paper we are specifically interested in modelling the tails of loss severity distributions This is of particular relevance in reinsurance if we are required to choose or price a high-excess layer In this situation it is essential to find a good statistical model for the largest observed historical losses It is less important that the model explains smaller losses, if smaller losses were also of interest we could in any case use a mixture distribution so that one model applied to the tail and another to the main body of the data However, a single model chosen for its overall fit to all historical losses may not provide a particularly good fit to the large losses and may not be suitable for pricing a high-excess layer

Our modelling is based on extreme value theory (EVT), a theory which until comparatively recently has found more application in hydrology and climatology (de Haan

1990, Smith 1989) than in insurance As its name suggests, this theory is concerned with the modelling of extreme events and in the last few years various authors (Beirlant & Teugels 1992, Embrechts & Kluppelberg 1993) have noted that the theory is as relevant to the modelling of extreme insurance losses as it is to the modelling of high river levels or temperatures.

For our purposes, the key result in EVT is the Pickands-Balkema-de Haan theorem (Balkema & de Haan 1974, Pickands 1975) which essentially says that, for a wide class of distributions, losses which exceed high enough thresholds follow the generalized Pareto distribution (GPD) In this paper we are concerned with fitting the GPD to data on exceedances of high thresholds This modelling approach was developed in Davison (1984), Davison & Smith (1990) and other papers by these authors

To illustrate the methods, we analyse Danish data on major fire insurance losses. We provide an extended worked example where we try to point out the pitfalls and limitations of the methods as well their considerable strengths

2 MODELLING LOSS SEVERITIES

2.1 The context

Suppose insurance losses are denoted by the independent, identically distributed random variables $X_1, X_2, \ldots$ whose common distribution function is $F_X(x) = P\{X \leq x\}$ where $x > 0$ We assume that we are dealing with losses of the same general type and that these loss amounts are adjusted for inflation so as to be comparable.

Now, suppose we are interested in a high-excess loss layer with lower and upper attachment points r and R respectively, where r is large and $R > r$ This means the payout Y_i on a loss X_i is given by

$$Y_i = \begin{cases} 0 & \text{if } 0 < X_i < r, \\ X_i - r & \text{if } r \leq X_i < R, \\ R - r & \text{if } R \leq X_i < \infty. \end{cases}$$

The process of losses becoming payouts is sketched in Figure 1. Of six losses, two pierce the layer and generate a non-zero payout One of these losses overshoots the layer entrirely and generates a capped payout.

Two related actuarial problems concerning this layer are.

1. The pricing problem Given r and R what should this insurance layer cost a customer?
2. The optimal attachment point problem. If we want payouts greater than a specified amount to occur with at most a specified frequency, how low can we set r?

To answer these questions we need to fix a period of insurance and know something about the frequency of losses incurred by a customer in such a time period. Denote the unknown number of losses in a period of insurance by N so that the losses are $X_1, \ldots, X_N$ Thus the aggregate payout would be $Z = \sum_{i=1}^{N} Y_i$

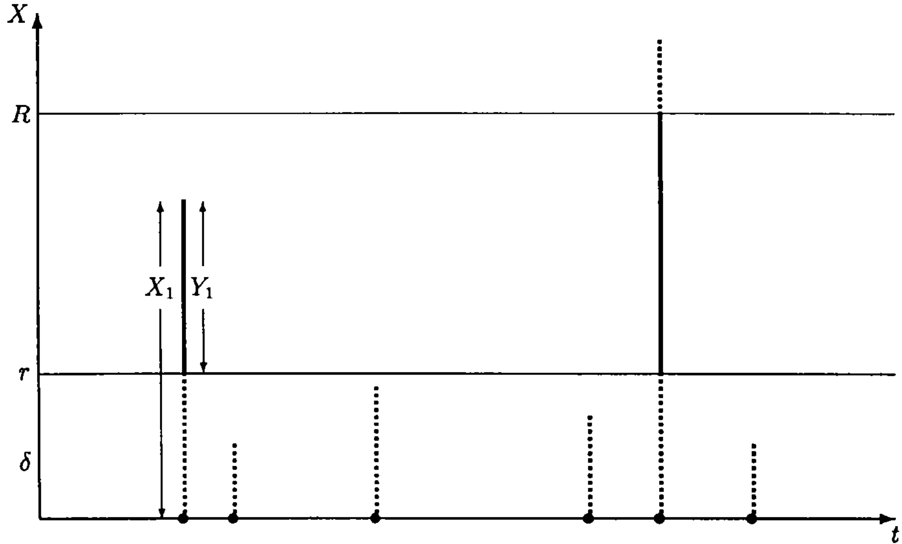


FIGURE 1 Possible realizations of losses in future time period

A common way of pricing is to use the formula Price = $E[Z] + k.var[Z]$, so that the price is the expected payout plus a risk loading which is k times the variance of the payout, for some k This is known as the variance pricing principle and it requires only that we can calculate the first two moments of Z

The expected payout $E[Z]$ is known as the pure premium and it can be shown to be $E[Y_i]E[N]$. It is clear that if we wish to price the cover provided by the layer (r,R) using the variance principle we must be able to calculate $E[Y_i]$, the pure premium for a single loss We will calculate $E[Y_i]$ as a simple price indication in later analyses in this paper However, we note that the variance pricing principle is unsophisticated and may have its drawbacks in heavy tailed situations, since moments may not exist or may be very large An insurance company generally wishes payouts to be rare events so that one possible way of formulating the attachment point problem might be choose r such that $P\{Z > 0\} < p$ for some stipulated small probability p That is to say, r is determined so that in the period of insurance a non-zero aggregate payout occurs with probability at most p

The attachment point problem essentially boils down to the estimation of a high quantile of the loss severity distribution $F_X(x)$ In both of these problems we need a good estimate of the loss severity distribution for x large, that is to say, in the tail area We must also have a good estimate of the loss frequency distribution of N, but this will not be a topic of this paper

2.2 Data Analysis

Typically we will have historical data on losses which exceed a certain amount known as a displacement It is practically impossible to collect data on all losses and data on small losses are of less importance anyway Insurance is generally provided against significant losses and insured parties deal with small losses themselves and may not report them

Thus the data should be thought of as being realizations of random variables truncated at a displacement δ, where $\delta \ll r$ This displacement is shown in Figure 1; we only observe realizations of the losses which exceed δ.

The distribution function (d f) of the truncated losses can be defined as in Hogg & Klugman (1984) by

$$F_{X^\delta}(x) = P\{X \le x \mid X > \delta\} = \begin{cases} 0 & \text{if } x \le \delta, \\ \dfrac{F_X(x) - F_X(\delta)}{1 - F_X(\delta)} & \text{if } x > \delta, \end{cases}$$

and it is, in fact, this d f that we shall attempt to estimate

With adjusted historical loss data, which we assume to be realizations of independent, identically distributed, truncated random variables, we attempt to find an estimate of the truncated severity distribution $F_X^{\delta}(x)$ One way of doing this is by fitting parametric models to data and obtaining parameter estimates which optimize some fitting criterion – such as maximum likelihood But problems arise when we have data as in Figure 2 and we are interested in a very high-excess layer

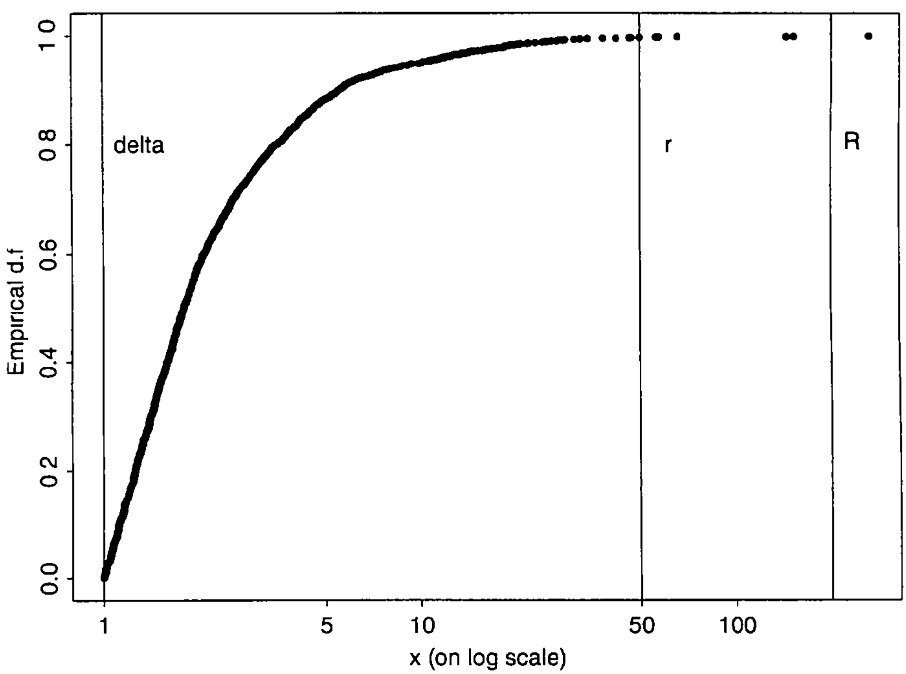


FIGURE 2 High-excess layer in relation to available data

Figure 2 shows the empirical distribution function of the Danish fire loss data evaluated at each of the data points. The empirical d f. for a sample of size n is defined by $F_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \le x\}}$, i.e. the number of observations less than or equal to x divided by n. The empirical d f. forms an approximation to the true d.f which may be quite good in the body of the distribution; however, it is not an estimate which can be successfully extrapolated beyond the data.

The full Danish data comprise 2492 losses and can be considered as being essentially all Danish fire losses over one million Danish Krone (DKK) from 1980 to 1990 plus a number of smaller losses below one million DKK We restrict our attention to the 2156 losses exceeding one million so that the effective displacement δ is one We work in units of one million and show the x-axis on a log scale to indicate the great range of the data

Suppose we are required to price a high-excess layer running from 50 to 200 In this interval we have only six observed losses. If we fit some overall parametric severity distribution to the whole dataset it may not be a particularly good fit in this tail area where the data are sparse

There are basically two options open to an insurance company Either it may choose not to insure such a layer, because of too little experience of possible losses. Or, if it wishes to insure the layer, it must obtain a good estimate of the severity distribution in the tail

To solve this problem we use the extreme value methods explained in the next section. Such methods do not predict the future with certainty, but they do offer good models for explaining the extreme events we have seen in the past. These models are not arbitrary but based on rigorous mathematical theory concerning the behaviour of extrema

3 EXTREME VALUE THEORY

In the following we summarize the results from EVT which underlie our modelling General texts on the subject of extreme values include Falk, Husler & Reiss (1994), Embrechts, Kluppelberg & Mikosch (1997) and Reiss & Thomas (1996)

3.1 The generalized extreme value distribution

Just as the normal distribution proves to be the important limiting distribution for sample sums or averages, as is made explicit in the central limit theorem, another family of distributions proves important in the study of the limiting behaviour of sample extrema. This is the family of extreme value distributions

This family can be subsumed under a single parametrization known as the generalized extreme value distribution (GEV). We define the d f of the standard GEV by

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}) & \text{if } \xi \neq 0, \\ \exp(-e^{-x}) & \text{if } \xi = 0, \end{cases}$$

where x is such that $1 + \xi x > 0$ and ξ is known as the shape parameter. Three well known distributions are special cases. if $\xi > 0$ we have the Fréchet distribution, if $\xi < 0$ we have the Weibull distribution; $\xi = 0$ gives the Gumbel distribution

If we introduce location and scale parameters μ and $\sigma > 0$ respectively we can extend the family of distributions We define the GEV $H_{\xi,\mu,\sigma}(x)$ to be $H_\xi((x-\mu)/\sigma)$ and we say that $H_{\xi,\mu,\sigma}$ is of the type H_ξ.

3.2 The Fisher-Tippett Theorem

The Fisher-Tippett theorem is the fundamental result in EVT and can be considered to have the same status in EVT as the central limit theorem has in the study of sums The theorem describes the limiting behaviour of appropriately normalized sample maxima.

Suppose we have a sequence of i.i.d random variables $X_1, X_2, \ldots$ from an unknown distribution F – perhaps a loss severity distribution. We denote the maximum of the first n observations by $M_n = \max(X_1, \ldots, X_n)$ Suppose further that we can find sequences of real numbers $a_n > 0$ and b_n such that $(M_n - b_n)/a_n$, the sequence of normalized maxima, converges in distribution

That is

$$P\{(M_n - b_n)/a_n \le x\} = F^n(a_n x + b_n) \rightarrow H(x), \text{ as } n \rightarrow \infty, \tag{1}$$

for some non-degenerate d f $H(x)$ If this condition holds we say that F is in the maximum domain of attraction of H and we write $F \in$ MDA (H)

It was shown by Fisher & Tippett (1928) that

$$F \in \text{MDA}(H) \Rightarrow H \text{ is of the type } H_\xi \text{ for some } \xi.$$

Thus, if we know that suitably normalized maxima converge in distribution, then the limit distribution must be an extreme value distribution for some value of the parameters ξ, μ and σ

The class of distributions F for which the condition (1) holds is large A variety of equivalent conditions may be derived (see Falk et al. (1994)) One such result is a condition for F to be in the domain of attraction of the heavy tailed Fréchet distribution (H_ξ where $\xi > 0$). This is of interest to us because insurance loss data are generally heavy tailed

Gnedenko (1943) showed that for $\xi > 0$, $F \in$ MDA (H_ξ) if and only if $1 - F(x) = x^{-1/\xi} L(x)$, for some slowly varying function $L(x)$ This result essentially says that if the tail of the d f. $F(x)$ decays like a power function, then the distribution is in the domain of attraction of the Fréchet The class of distributions where the tail decays like a power function is quite large and includes the Pareto, Burr, loggamma, Cauchy and t-distributions as well as various mixture models. We call distributions in this class heavy tailed distributions, these are the distributions which will be of most use in modelling loss severity data

Distributions in the maximum domain of attraction of the Gumbel MDA (H_0) include the normal, exponential, gamma and lognormal distributions. We call these distributions medium tailed distributions and they are of some interest in insurance Some insurance datasets may be best modelled by a medium tailed distribution and even when we have heavy tailed data we often compare them with a medium tailed reference distribution such as the exponential in explorative analyses

Particular mention should be made of the lognormal distribution which has a much heaver tail than the normal distribution. The lognormal has historically been a popular model for loss severity distributions; however, since it is not a member of MDA (H_ξ) for $\xi > 0$ it is not technically a heavy tailed distribution

Distributions in the domain of attraction of the Weibull (H_ξ for $\xi < 0$) are short tailed distributions such as the uniform and beta distributions. This class is generally of lesser interest in insurance applications although it is possible to imagine situations where losses of a certain type have an upper bound which may never be exceeded so that the support of the loss severity distribution is finite Under these circumstances the tail might be modelled with a short tailed distribution

The Fisher-Tippett theorem suggests the fitting of the GEV to data on sample maxima, when such data can be collected There is much literature on this topic (see Embrechts et al , 1997), particularly in hydrology where the so-called annual maxima method has a long history A well-known reference is Gumbel (1958)

3.3 The generalized Pareto distribution

An equivalent set of results in EVT describe the behaviour of large observations which exceed high thresholds, and this is the theoretical formulation which lends itself most readily to the modelling of insurance losses This theory addresses the question given an observation is extreme, how extreme might it be? The distribution which comes to the fore in these results is the generalized Pareto distribution (GPD)

The GPD is usually expressed as a two parameter distribution with d f

$$G_{\xi,\sigma}(x) = \begin{cases} 1-(1+\xi x/\sigma)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1-\exp(-x/\sigma) & \text{if } \xi = 0, \end{cases} \tag{2}$$

where $\sigma > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\sigma/\xi$ when $\xi < 0$ The GPD again subsumes other distributions under its parametrization When $\xi > 0$ we have a reparametrized version of the usual Pareto distribution, if $\xi < 0$ we have a type II Pareto distribution, $\xi = 0$ gives the exponential distribution

Again we can extend the family by adding a location parameter μ The GPD $G_{\xi,\mu,\sigma}(x)$ is defined to be $G_{\xi,\sigma}(x-\mu)$.

3.4 The Pickands-Balkema-de Haan Theorem

Consider a certain high threshold u which might, for instance, be the lower attachment point of a high-excess loss layer u will certainly be greater than any possible displacement δ associated with the data We are interested in excesses above this threshold, that is, the amount by which observations overshoot this level

Let x_0 be the finite or infinite right endpoint of the distribution F. That is to say, $x_0 = \sup\{x \in \Re : F(x) < 1\} \leq \infty$ We define the distribution function of the excesses over the high threshold u by

$$F_u(x) = P\{X - u \le x \mid X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)}$$

for $0 \le x < x_0 - u$

The theorem (Balkema & de Haan 1974, Pickands 1975) shows that under MDA conditions (1) the generalized Pareto distribution (2) is the limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint. That is, we can find a positive measurable function $\sigma(u)$ such that

$$\lim_{u \to x_0} \sup_{0 \le x \le x_0 - u} \left| F_u(x) - G_{\xi,\sigma(u)}(x) \right| = 0,$$

if and only if $F \in$ MDA (H_ξ) In this formulation we are mainly following the quoted references to Balkema & de Haan and Pickands, but we should stress the important contribution to results of this type by Gnedenko (1943)

This theorem suggests that, for sufficiently high thresholds u, the distribution function of the excesses may be approximated by $G_{\xi,\sigma}(x)$ for some values of ξ and σ Equivalently, for $x - u \ge 0$, the distribution function of the ground-up exceedances $F_u(x - u)$ (the excesses plus u) may be approximated by $G_{\xi,\sigma}(x - u) = G_{\xi,u,\sigma}(x)$

The statistical relevance of the result is that we may attempt to fit the GPD to data which exceed high thresholds. The theorem gives us theoretical grounds to expect that if we choose a high enough threshold, the data above will show generalized Pareto behaviour. This has been the approach developed in Davison (1984) and Davison & Smith (1990) The principal practical difficulty involves choosing an appropriate threshold The theory gives no guidance on this matter and the data analyst must make a decision, as will be explained shortly.

3.5 Tail fitting

If we can fit the GPD to the conditional distribution of the excesses above some high threshold u, we may also fit it to the tail of the original distribution above the high threshold (Reiss & Thomas 1996). For $x \ge u$, i.e points in the tail of the distribution,

$$F(x) = P\{X \le x\} = (1 - P\{X \le u\})F_u(x - u) + P\{X \le u\}$$

We now know that we can estimate $F_u(x - u)$ by $G_{\xi,\sigma}(x - u)$ for u large We can also estimate $P\{X \le u\}$ from the data by $F_n(u)$, the empirical distribution function evaluated at u

This means that for $x \ge u$ we can use the tail estimate

$$\hat{F}(x) = (1 - F_n(u))G_{\xi,u,\sigma}(x) + F_n(u)$$

to approximate the distribution function $F(x)$ It is easy to show that $\hat{F}(x)$ is also a generalized Pareto distribution, with the same shape parameter ξ, but with scale parameter $\tilde{\sigma} = \sigma(1 - F_n(u))^{\xi}$ and location parameter $\tilde{\mu} = \mu - \tilde{\sigma}((1 - F_n(u))^{-\xi} - 1)/\xi$.

3.6 Statistical Aspects

The theory makes explicit which models we should attempt to fit to historical data. However, as a first step before model fitting is undertaken, a number of exploratory graphical methods provide useful preliminary information about the data and in particular their tail. We explain these methods in the next section in the context of an analysis of the Danish data

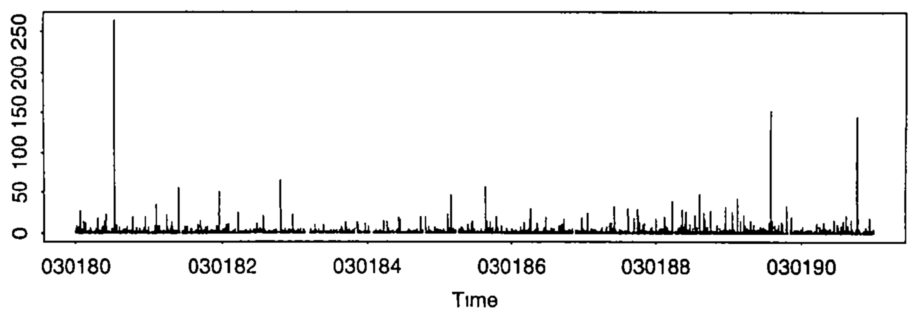
The generalized Pareto distribution can be fitted to data on excesses of high thresholds by a variety of methods including the maximum likelihood method (ML) and the method of probability weighted moments (PWM) We choose to use the ML-method For a comparison of the relative merits of the methods we refer the reader to Hosking & Wallis (1987) and Rootzén & Tajvidi (1996).

For $\xi > -0.5$ (all heavy tailed applications) it can be shown that maximum likelihood regularity conditions are fulfilled and that maximum likelihood estimates $(\hat{\xi}_{N_u}, \hat{\sigma}_{N_u})$ based on a sample of N_u excesses of a threshold u are asymptotically normally distributed (Hosking & Wallis 1987)

Specifically for a fixed threshold u we have

$$N_u^{1/2}\begin{pmatrix}\hat{\xi}_{N_u}\\ \hat{\sigma}_{N_u}\end{pmatrix} \xrightarrow{d} N\left[\begin{pmatrix}\xi\\ \sigma\end{pmatrix},\begin{pmatrix}(1+\xi)^2 & \sigma(1+\xi)\\ \sigma(1+\xi) & 2\sigma^2(1+\xi)\end{pmatrix}\right],$$

as $N_u \to \infty$. This result enables us to calculate approximate standard errors for our maximum likelihood estimates.



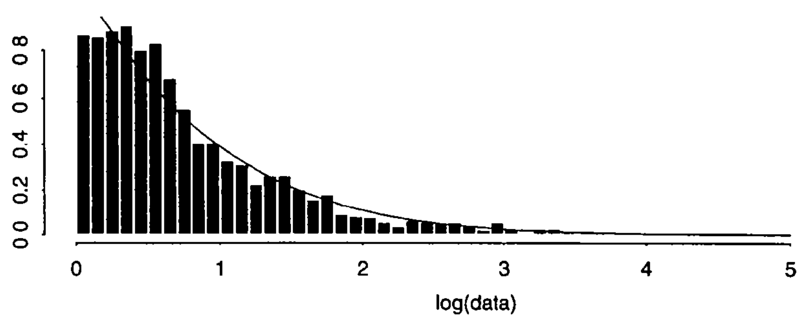


FIGURE 3 Time series and log data plots for the Danish data Sample size is 2156

4 Analysis of Danish Fire Loss Data

The Danish data consist of 2156 losses over one million Danish Krone (DKK) from the years 1980 to 1990 inclusive (plus a few smaller losses which we ignore in our analyses) The loss figure is a total loss figure for the event concerned and includes damage to buildings, damage to furniture and personal property as well as loss of profits. For these analyses the data have been adjusted for inflation to reflect 1985 values

4.1 Exploratory data analysis

The time series plot (Figure 3, top) allows us to identify the most extreme losses and their approximate times of occurrence We can also see whether there is evidence of clustering of large losses, which might cast doubt on our assumption of i.i.d data This does not appear to be the case with the Danish data

The histogram on the log scale (Figure 3, bottom) shows the wide range of the data It also allows us to see whether the data may perhaps have a lognormal right tail, which would be indicated by a familiar bell-shape in the log plot.

We have fitted a truncated lognormal distribution to the dataset using the maximum likelihood method and superimposed the resulting probability density function on the histogram. The scale of the y-axis is such that the total area under the curve and the total area of the histogram are both one The truncated lognormal appears to provide a reasonable fit but it is difficult to tell from this picture whether it is a good fit to the largest losses in the high-excess area in which we are interested

The QQ-plot against the exponential distribution (Figure 4) is a very useful guide to heavy tails It examines visually the hypothesis that the losses come from an exponential distribution, i.e from a distribution with a medium sized tail. The quantiles of the empirical distribution function on the x-axis are plotted against the quantiles of the exponential distribution function on the y-axis The plot is

$$\left\{(X_{k,n}, G_{0,1}^{-1}(\frac{n-k+1}{n+1})), k = 1, \ldots, n\right\},$$

where $X_{k,n}$ denotes the kth order statistic, and $G_{0,1}^{-1}$ is the inverse of the d f. of the exponential distribution (a special case of the GPD) The points should lie approximately along a straight line if the data are an i.i.d. sample from an exponential distribution

A concave departure from the ideal shape (as in our example) indicates a heavier tailed distribution whereas convexity indicates a shorter tailed distribution. We would expect insurance losses to show heavy tailed behaviour.

The usual caveats about the QQ-plot should be mentioned. Even data generated from an exponential distribution may sometimes show departures from typical exponential behaviour In general, the more data we have, the clearer the message of the QQ-plot. With over 2000 data points in this analysis it seems safe to conclude that the tail of the data is heavier than exponential.

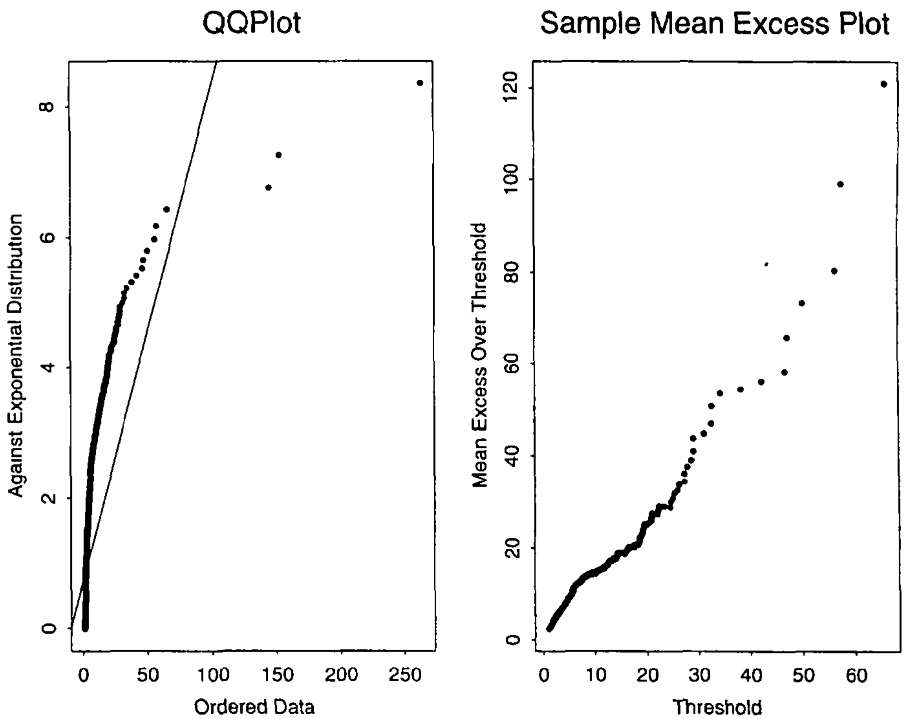


FIGURE 4 QQ-plot and sample mean excess function

A further useful graphical tool is the plot of the sample mean excess function (see again Figure 4) which is the plot.

$$\{(u, e_n(u)), X_{n,n} < u < X_{1,n}\}$$

where $X_{1,n}$ and $X_{n,n}$ are the first and nth order statistics and $e_n(u)$ is the sample mean excess function defined by

$$e_n(u) = \frac{\sum_{i=1}^{n} (X_i - u)^+}{\sum_{i=1}^{n} 1_{\{X_i > u\}}};$$

i e. the sum of the excesses over the threshold u divided by the number of data points which exceed the threshold u

The sample mean excess function $e_n(u)$ is an empirical estimate of the mean excess function which is defined as $e(u) = E[X - u \mid X > u]$ The mean excess function describes the expected overshoot of a threshold given that exceedance occurs.

In plotting the sample mean excess function we choose to end the plot at the fourth order statistic and thus omit a possible three further points, these points, being the averages of at most three observations, may be erratic.

The interpretation of the mean excess plot is explained in Beirlant, Teugels & Vynckier (1996), Embrechts et al. (1997) and Hogg & Klugman (1984). If the points show an upward trend, then this is a sign of heavy tailed behaviour. Exponentially distributed data would give an approximately horizontal line and data from a short tailed distribution would show a downward trend.

In particular, if the empirical plot seems to follow a reasonably straight line with positive gradient above a certain value of u, then this is an indication that the data follow a generalized Pareto distribution with positive shape parameter in the tail area above u

This is precisely the kind of behaviour we observe in the Danish data (Figure 4) There is evidence of a straightening out of the plot above a threshold of ten, and perhaps again above a threshold of 20. In fact the whole plot is sufficiently straight to suggest that the GPD might provide a reasonable fit to the entire dataset.

4.2 Overall fits

In this section we look at standard choices of curve fitted to the whole dataset We use two frequently used severity models – the truncated lognormal and the ordinary Pareto – as well as the GPD

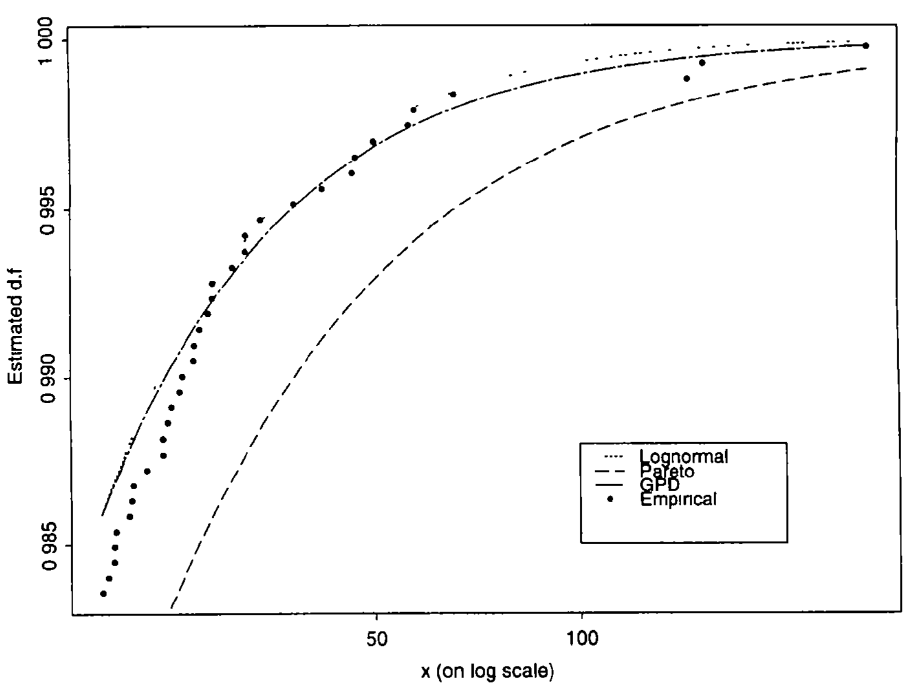


FIGURE 5 Performance of overall fits in the tail area

By ordinary Pareto we mean the distribution with d f $F(x) = 1 - (a/x)^{\alpha}$ for unknown positive parameters a and α and with support $x > a$ This distribution can be rewritten as $F(x) = 1 - (1 + (x - a)/a)^{-\alpha}$ so that it is seen to be a GPD with shape $\xi = 1/\alpha$, scale $\sigma = a/\alpha$ and location $\mu = a$. That is to say it is a GPD where the scale parameter is constrained to be the shape multiplied by the location It is thus a little less flexible than a GPD without this constraint where the scale can be freely chosen.

As discussed earlier, the lognormal distribution is not strictly speaking a heavy tailed distribution However it is moderately heavy tailed and in many applications it is quite a good loss severity model.

In Figure 5 we see the fit of these models in the tail area above a threshold of 20. The lognormal is a reasonable fit, although its tail is just a little too thin to capture the behaviour of the very highest observed losses. The Pareto, on the other hand, seems to overestimate the probabilities of large losses. This, at first sight, may seem a desirable, conservative modelling feature But it may be the case, that this d f is so conservative, that if we use it to answer our attachment point and premium calculation problems, we will arrive at answers that are unrealistically high

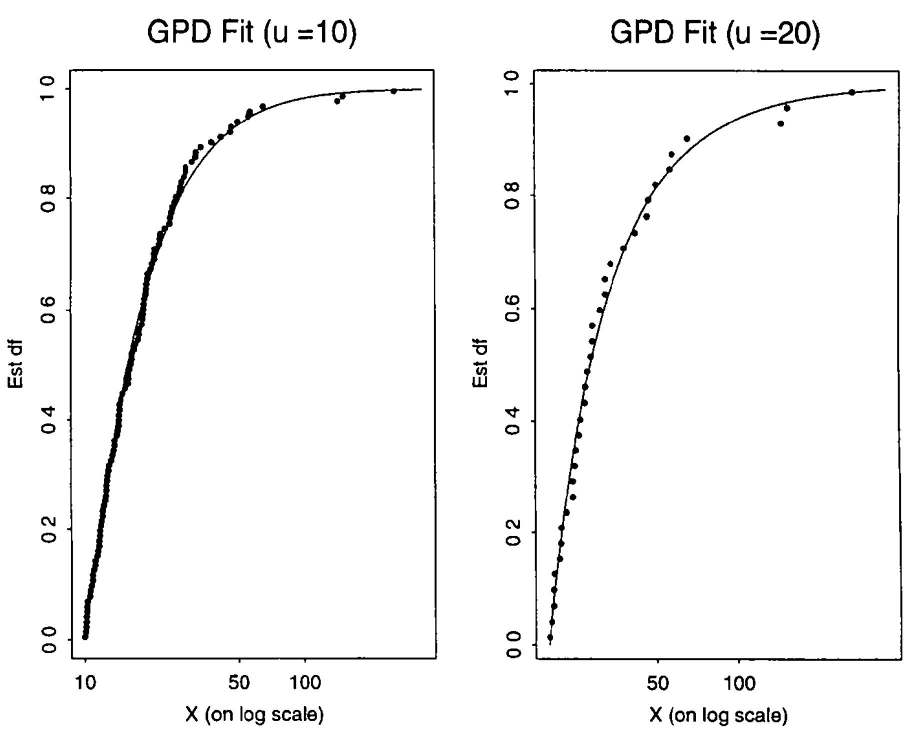


FIGURE 6 In left plot GPD is fitted to 109 exceedances of the threshold 10 The parameter estimates are $\xi = 0\,497$, $\mu = 10$ and $\sigma = 6\,98$ In right plot GPD is fitted to 36 exceedances of the threshold 20 The parameter estimates are $\xi = 0\,684$, $\mu = 20$ and $\sigma = 9\,63$

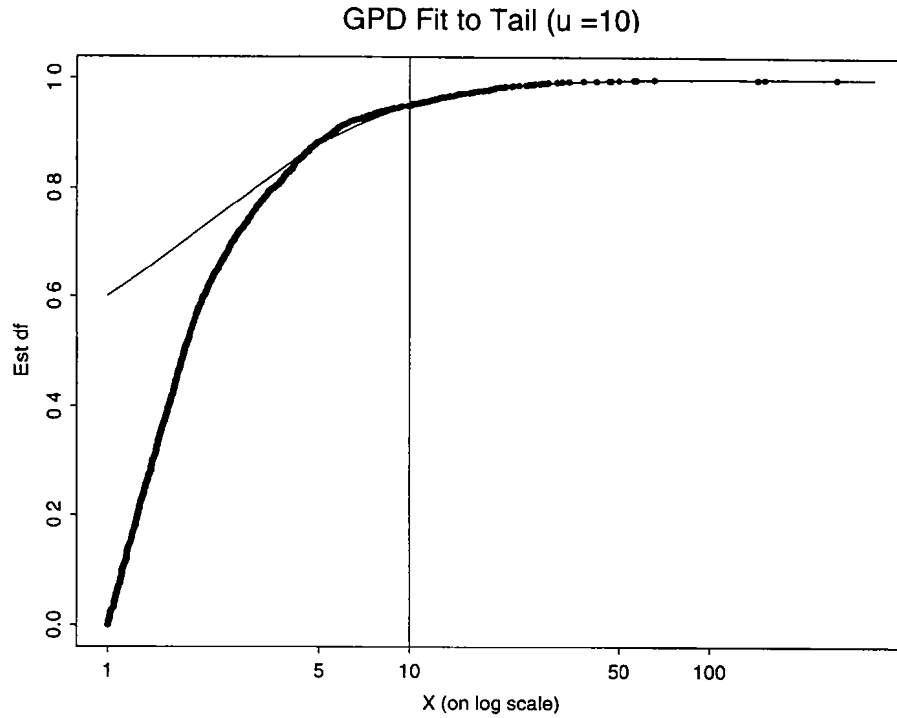


FIGURE 7 Fitting the GPD to tail of severity distribution above threshold 10
The parameter estimates are $\xi = 0.497$, $\mu = -0.845$ and $\sigma = 1.59$

The GPD is somewhere between the lognormal and Pareto in the tail area and actually seems to be quite a good explanatory model for the highest losses The data are of course truncated at 1 M DKK, and it seems that, even above this low threshold, the GPD is not a bad fit to the data By raising the threshold we can, however, find models which are even better fits to the larger losses

Estimates of high quantiles and layer prices based on these three fitted curves are given in table 1

4.3 Fitting to data on exceedances of high thresholds

The sample mean excess function for the Danish data suggests we may have success fitting the GPD to those data points which exceed high thresholds of ten or 20, in Figure 6 we do precisely this. We use the three parameter form of the GPD with the location parameter set to the threshold value We obtain maximum likelihood estimates for the shape and scale parameters and plot the corresponding GPD curve superimposed on the empirical distribution function of the exceedances The resulting fits seem reasonable to the naked eye

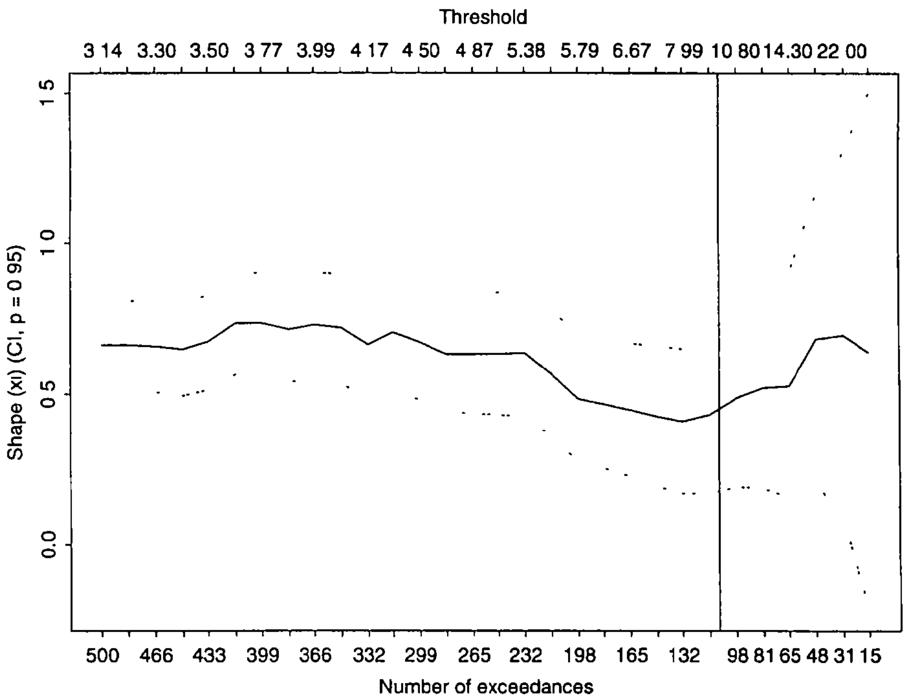


FIGURE 8 Estimates of shape by increasing threshold on the upper x-axis and decreasing number of exceedances on the lower x-axis, in total 30 models are fitted

The estimates we obtain are estimates of the conditional distribution of the losses, given that they exceed the threshold Quantile estimates derived from these curves are conditional quantile estimates which indicate the scale of losses which could be experienced if the threshold were to be exceeded.

As described in section 3 5, we can transform scale and location parameters to obtain a GPD model which fits the severity distribution itself in the tail area above the threshold. Since our data are truncated at the displacement of one million we actually obtain a fit for the tail of the truncated severity distribution $F_X^{\delta}(x)$ This is shown for a threshold of ten in Figure 7 Quantile estimates derived from this curve are quantile estimates conditional on exceedance of the displacement of one million.

So far we have considered two arbitrary thresholds. In the next sections we consider the question of optimizing the choice of threshold by investigating the different estimates we get for model parameters, high quantiles and prices of high-excess layers.

4.4 Shape and quantile estimates

As far as pricing of layers or estimation of high quantiles using a GPD model is concerned, the crucial parameter is ξ, the tail index. Roughly speaking, the higher the value of ξ the heavier the tail and the higher the prices and quantile estimates we de-

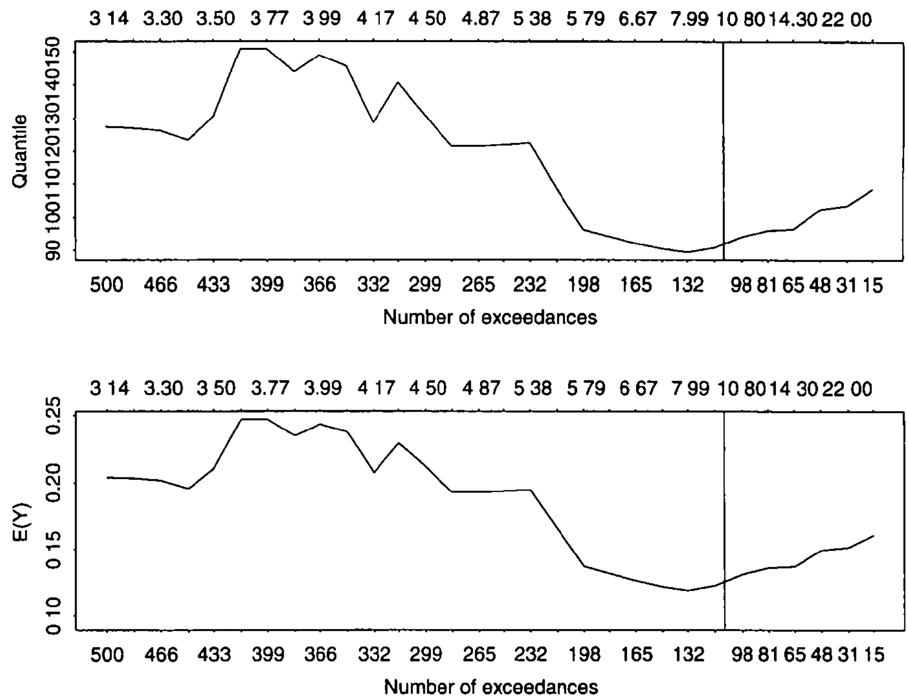


FIGURE 9 999 quantile estimates (upper picture) and price indications for a (50,200) layer (lower picture) for increasing thresholds and decreasing numbers of exceedances

rive. For a three-parameter GPD model $G_{\xi,\mu,\sigma}$ the pth quantile can be calculated to be $\mu + \sigma/\xi((1-p)^{\xi} - 1)$

In Figure 8 we fit GPD models with different thresholds to obtain maximum likelihood estimates of ξ, as well as asymptotic confidence intervals for the parameter estimates. On the lower x-axis the number of data points exceeding the threshold is plotted; on the upper x-axis the threshold itself The shape estimate is plotted on the y-axis. A vertical line marks the location of our first model with a threshold at ten

In using this picture to choose an optimal threshold we are confronted with a bias-variance tradeoff Since our modelling approach is based on a limit theorem which applies above high thresholds, if we choose too low a threshold we may get biased estimates because the theorem does not apply On the other hand, if we set too high a threshold we will have few data points and our estimates will be prone to high standard errors. So a sensible choice will lie somewhere in the centre of the plot, perhaps a threshold between four and ten in our example

The ideal situation would be that shape estimates in this central range were stable. In our experience with several loss severity datasets this is sometimes the case so that the data conform very well to a particular generalized Pareto distribution in the tail and inference is not too sensitive to choice of threshold. In our present example the shape estimates vary somewhat and to choose a threshold we should conduct further investigations.

TABLE 1

COMPARISON OF SHAPE AND QUANTILE ESTIMATES FOR VARIOUS MODELS

Model	*u*	*Excesses*	ξ	*s.e.*	*.995*	*.999*	*.9999*	*P*
GPD	3	532	0 67	0 07	44 0	129	603	0 21
GPD	4	362	0 72	0 09	46 3	147	770	0 24
GPD	5	254	0 63	0 10	43 4	122	524	0 19
GPD	10	109	0 50	0 14	40 4	95	306	0 13
GPD	20	36	0 68	0 28	38 4	103	477	0 15
MODELS FITTED TO WHOLE DATASET								
GPD	all	data	0 60	0 04	38 0	101	410	0 15
Pareto	all	data			66 0	235	1453	0 10
Lognormal	all	data			35 6	82	239	0 41
SCENARIO MODELS								
GPD	10	109 − 1	0 39	0 13	37 1	77	201	0 09
GPD	10	109 + 1	0 60	0 15	44 2	118	469	0 19

Figure 9 (upper panel) is a similar plot showing how quantile estimates depend on the choice of threshold We have chosen to plot estimates of the .999th quantile. Roughly speaking, if the model is a good one, one in every thousand losses which exceed one million DKK might be expected to exceed this quantile; such losses are rare but threatening to the insurer In a dataset of 2156 losses the chances are we have only seen two or three losses of this magnitude so that this is a difficult quantile estimation problem involving model-based interpolation in the tail

We have tabulated quantile estimates for some selected thresholds in table 1 and give the corresponding estimates of the shape parameter. Using the model with a threshold at ten the 999th quantile is estimated to be 95. But is we push the threshold back to four the quantile estimate goes up to 147. There is clearly a considerable difference these two estimates and if we attempt to estimate higher quantiles such as the .9999th this difference becomes more pronounced Estimating the .9999th quantile is equivalent to estimating the size of a one in 10000 loss event In our dataset it is likely that we have not yet seen a loss of this magnitude so that this is an extremely difficult problem entailing extrapolation of the model beyond the data.

Estimating the .995th quantile is a slightly easier tail estimation problem. We have perhaps already seen around ten or 11 losses of this magnitude For thresholds at ten and four the estimates are 40.4 and 46 3 respectively, so that the discrepancy is not so large.

Thus the sensitivity of quantile estimation may not be too severe at moderately high quantiles within the range of the data but increases at more distant quantiles. This is not surprising since estimation of quantiles at the margins of the data or beyond the data is an inherently difficult problem which represents a challenge for any method. It should be noted that although the estimates obtained by the GPD method often span a wide range, the estimates obtained by the naive method of fitting ordinary Pareto or lognormal to the whole dataset are even more extreme (see table) To our knowledge the GPD estimates are as good as we can get using parametric models

4.5 Calculating price indications

In considering the issue of the best choice of threshold we can also investigate how price of a layer varies with threshold To give an indication of the prices we get from our model we calculate $P = E[Y_i \mid X_i > \delta]$ for a layer running from 50 to 200 million (as in Figure 2) It is easily seen that, for a general layer (r, R), P is given by

$$P = \int_r^R (x-r) f_{X^\delta}(x)dx + (R-r)(1 - F_{X^\delta}(R)), \tag{3}$$

where $f_X{}^\delta(x) = dF_X{}^\delta(x)/dx$ denotes the density function for the losses truncated at δ Picking a high threshold u ($< r$) and fitting a GPD model to the excesses, we can estimate $F_X{}^\delta(x)$ for $x > u$ using the tail estimation procedure We have the estimate

$$\hat{F}_{X^\delta}(x) = (1 - F_n(u))G_{\hat{\xi},u,\hat{\sigma}}(x) + F_n(u),$$

where $\hat{\xi}$ and $\hat{\sigma}$ are maximum-likelihood parameter estimates and $F_n(u)$ is an estimate of $P\{X^\delta \le u\}$ based on the empirical distribution function of the data We can estimate the density function of the δ-truncated losses using the derivative of the above expression and the integral in (3) has an easy closed form

In Figure 9 (lower picture) we show the dependence of P on the choice of threshold The plot seems to show very similar behaviour to that of the 999[th] percentile estimate, with low thresholds leading to higher prices The question of which threshold is ultimately best depends on the use to which the results are to be put If we are trying to answer the optimal attachment point problem or to price a high layer we may want to err on the side of conservatism and arrive at answers which are too high rather than too low. In the case of the Danish data we might set a threshold lower than ten, perhaps at four The GPD model may not fit the data quite so well above this lower threshold as it does above the high threshold of ten, but it might be safer to use the low threshold to make calculations

On the other hand there may be business reasons for trying to keep the attachment point or premium low. There may be competition to sell high excess policies and this may mean that basing calculations only on the highest observed losses is favoured, since this will lead to more attractive products (as well as a better fitting model)

In other insurance datasets the effect of varying the threshold may be different Inference about quantiles might be quite robust to changes in threshold or elevation of the threshold might result in higher quantile estimates Every dataset is unique and the data analyst must consider what the data mean at every step The process cannot and should not be fully automated

4.6 Sensitivity of Results to the Data

We have seen that inference about the tail of the severity distribution may be sensitive to the choice of threshold It is also sensitive to the largest losses we have in our dataset. We show this by considering two scenarios in Table 1.

In the first scenario we remove the largest observation from the dataset. If we return to our first model with a threshold at ten we now have only 108 exceedances and the estimate of the .999th quantile is reduced from 95 to 77 whilst the shape parameter falls from 0.50 to 0 39 Thus omission of this data point has a profound effect on the estimated quantiles. The estimates of the .999th and 9999th quantiles are now smaller than any previous estimates

In the second scenario we introduce a new largest loss of 350 to the dataset (the previous largest being 263) The shape estimate goes up to 0.60 and the estimate of the 999th quantile increases to 118 This is also a large change, although in this case it is not as severe as the change caused by leaving the dataset unchanged and reducing the threshold from ten to five or four

The message of these two scenarios is that we should be careful to check the accuracy of the largest data points in a dataset and we should be careful that no large data points are deemed to be outliers and omitted if we wish to make inference about the tail of a distribution. Adding or deleting losses of lower magnitude from the dataset has much less effect

5. DISCUSSION

We hope to have shown that fitting the generalized Pareto distribution to insurance losses which exceed high thresholds is a useful method for estimating the tails of loss severity distributions In our experience with several insurance datasets we have found consistently that the generalized Pareto distribution is a good approximation in the tail

This is not altogether surprising As we have explained, the method has solid foundations in the mathematical theory of the behaviour of extremes; it is not simply a question of ad hoc curve fitting It may well be that, by trial and error, some other distribution can be found which fits the available data even better in the tail But such a distribution would be an arbitrary choice, and we would have less confidence in extrapolating it beyond the data

It is our belief that any practitioner who routinely fits curves to loss severity data should know about extreme value methods There are, however, an number of caveats to our endorsement of these methods We should be aware of various layers of uncertainty which are present in any data analysis, but which are perhaps magnified in an extreme value analysis

On one level, there is parameter uncertainty. Even when we have abundant, good-quality data to work with and a good model, our parameter estimates are still subject to a standard error We obtain a range of parameter estimates which are compatible with our assumptions As we have already noted, inference is sensitive to small changes in the parameters, particularly the shape parameter

Model uncertainty is also present – we may have good data but a poor model. Using extreme value methods we are at least working with a good class of models, but they are applicable over high thresholds and we must decide where to set the threshold If we set the threshold too high we have few data and we introduce more parameter uncertainty If we set the threshold too low we lose our theoretical justification for the model. In the analysis presented in this paper inference was very sensitive to the threshold choice (although this is not always the case).

Equally as serious as parameter and model uncertainty may be data uncertainty. In a sense, it is never possible to have enough data in an extreme value analysis Whilst a sample of 1000 data points may be ample to make inference about the mean of a distribution using the central limit theorem, our inference about the tail of the distribution is less certain, since only a few points enter the tail region. As we have seen, inference is very sensitive to the largest observed losses and the introduction of new extreme losses to the dataset may have a substantial impact For this reason, there may still be a role for stress scenarios in loss severity analyses, whereby historical loss data are enriched by hypothetical losses to investigate the consequences of unobserved, adverse events

Another aspect of data uncertainly is that of dependent data. In this paper we have made the familiar assumption of independent, identically distributed data In practice we may be confronted with clustering, trends, seasonal effects and other kinds of dependencies. When we consider fire losses in Denmark it may seem a plausible first assumption that individual losses are independent of one another, however, it is also possible to imagine that circumstances conducive or inhibitive to fire outbreaks generate dependencies in observed losses. Destructive fires may be greatly more common in the summer months, buildings of a particular vintage and building standard may succumb easily to fires and cause high losses Even after ajustment for inflation there may be a general trend of increasing or decreasing losses over time, due to an increasing number of increasingly large and expensive buildings, or due to increasingly good safety measures

These issues lead to a number of interesting statistical questions in what is very much an active research area. Papers by Davison (1984) and Davison & Smith (1990) discuss clustering and seasonality problems in environmental data and make suggestions concerning the modelling of trends using regression models built into the extreme value modelling framework The modelling of trends is also discussed in Rootzén & Tajvidi (1996).

We have developed software to fit the generalized Pareto distribution to exceedances of high thresholds and to produce the kinds of graphical output presented in this paper It is writen in Splus and is available over the World Wide Web at http://www math ethz ch/~mcneil

6 ACKNOWLEDGMENTS

Much of the work in this paper came about through a collaborative project with Swiss Re Zurich. I gratefully acknowledge Swiss Re for their financial support and many fruitful discussions[1]

I thank Mette Rytgaard of Copenhagen Re for making the Danish fire loss data available and Richard Smith for advice on software and algorithms for fitting the GPD to data I thank an anonymous reviewer for helpful comments on an earlier version of this paper

Special thanks are due to Paul Embrechts for introducing me to the theory of extremes and its many interesting applications.

REFERENCES

BALKEMA, A and DE HAAN, L (1974), 'Residual life time at great age', *Annals of Probability*, **2**, 792-804

BEIRLANT, J and TEUGELS, J (1992), 'Modelling large claims in non-life insurance', *Insurance Mathematics and Economics*, **11**, 17-29

BEIRLANT, J, TEUGELS, J and VYNCKIER, P (1996), *Practical analysis of extreme values*, Leuven University Press, Leuven

DAVISON, A (1984), Modelling excesses over high thresholds, with an application, in J de Oliveira, ed, 'Statistical Extremes and Applications', D Reidel, 461-482

DAVISON, A and SMITH, R (1990), 'Models for exceedances over high thresholds (with discussion)', *Journal of the Royal Statistical Society, Series B*, **52**, 393-442

DE HAAN, L (1990), 'Fighting the arch-enemy with mathematics', *Statistica Neerlandica*, **44**, 45-68

EMBRECHTS, P, KLUPPELBERG, C (1983), *'Some aspects of insurance mathematics'*, Theory of Probability and its Applications **38**, 262-295

EMBRECHTS, P, KLUPPELBERG, C and MIKOSCH, T (1997), *Modelling extremal events for insurance and finance*, Springer Verlag, Berlin To appear

FALK, M, HÜSLER, J and REISS, R (1994), *Laws of Small numbers extremes and rare events*, Birkhauser, Basel

FISHER, R and TIPPETT, L (1928), 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Proceedings of the Cambridge Philosophical Society*, **24**, 180-190

GNEDENKO, B (1943), 'Sur la distribution limite du terme maximum d'une série aleatoire', *Annals of Mathematics*, **44**, 423-453

GUMBEL, E (1958), *Statistics of Extremes*, Columbia University Press, New York

HOGG, R and KLUGMAN, S (1984), *Loss Distributions*, Wiley, New York

HOSKING, J and WALLIS, J (1987), 'Parameter and quantile estimation for the generalized Pareto distribution', *Technometrics*, **29**, 339-349

PICKANDS, J (1975), 'Statistical inference using extreme order statistics', *The Annals of Statistics* **3**, 119-131

REISS, R and THOMAS, M (1996), 'Statistical analysis of extreme values' Documentation for XTREMES software package

ROOTZÉN, H and TAJVIDI, N (1996), 'Extreme value statistics and wind storm losses a case study' To appear in Scandinavian Actuarial Journal

SMITH, R (1989), 'Extreme value analysis of environmental time series an application to trend detection in ground-level ozone', *Statistical Science*, **4**, 367-393

[1] The author is supported by Swiss Re as a research fellow at ETH Zurich

DISCUSSION OF THE DANISH DATA ON LARGE FIRE INSURANCE LOSSES

SIDNEY I. RESNICK

Cornell University

ABSTRACT

Alexander McNeil's (1996) study of the Danish data on large fire insurance losses provides an excellent example of the use of extreme value theory in an important application context. We point out how several alternate statistical techniques and plotting devices can buttress McNeil's conclusions and provide flexible tools for other studies

KEYWORDS

Heavy tails, regular variation, Hill estimator, Poisson processes, linear programming, parameter estimation weak convergence, consistency, estimation, independence, autocorrelations.

1 INTRODUCTION

McNeil's (1996) interesting study of large fire insurance losses provides an excellent case history illustrating a variety of extreme value techniques The goal of my remarks is to show additional techniques and plotting strategies which can be employed for similar data.

Our remarks concentrate on the following:

- Diagnostics for assessing the appropriateness of heavy tailed models
- Diagnostics for testing for independence.

It is customary in many insurance studies involving heavy tailed phenomena to assume independence without actually statistically checking this important fact so some attention is given to this issue

2 APPROPRIATENESS OF HEAVY TAILED MODELS

Given a particular data set, there are various methods of checking that a heavy tailed model is appropriate. The methods given below (these are also reviewed in Resnick 1995, 1996, Feigin and Resnick, 1996) supplement the techniques discussed by McNeil such as mean excess plots and QQ-plots against exponential quantiles. Unlike the mean excess plot, the following methods do not depend on existence of a finite mean for the marginal distribution of the stationary time series This is important since it is becoming clear that it is not difficult to find examples of heavy tailed data which require infinite mean models for adequate fits (See for example the teletraffic examples in Resnick (1995, 1996)).

For the discussion that follows, we suppose $\{X_n, n \geq 1\}$ is a stationary sequence and that

$$P[X_1 > x] = x^{-\alpha} L(x), \quad x \to \infty \tag{2 1}$$

where L is slowly varying and $\alpha > 0$ Consider the following techniques

(1) The Hill plot. Let

$$X_{(1)} > X_{(2)} > \dots > X_{(n)}$$

be the order statistics of the sample $X_1, \dots, X_n$ We pick $k < n$ and define the Hill estimator (Hill, 1975) to be

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}}$$

Note k is the number of upper order statistics used in the estimation The Hill plot is the plot of

$$((k, H_{k,n}^{-1}), 1 \leq k < n)$$

and if the $\{X_n\}$ process is iid or a linear moving average or satisfies certain mixing conditions then since $H_{k,n} \xrightarrow{P} \alpha^{-1}$ as $n \to \infty$, $k/n \to 0$ the Hill plot should have a stable regime sitting at height roughly α See Mason (1982), Hsing (1991), Resnick and Stărică (1995, 1996a), Rootzen et al (1990), Rootzen (1996). In the iid case, under a second order regular variation condition, $H_{k,n}$ is asymptotically normal with asymptotic variance $1/\alpha^2$ (See de Haan and Resnick, 1996)

(2) The smooHill Plot The Hill Plot often exhibits extreme volatility which makes finding a stable regime in the plot more guesswork than science and to counteract this, Resnick and Stărică (1996a) developed a smoothing technique yielding the smooHill plot Pick an integer u (usually 2 or 3) and define

$$smooH_{k,n} = \frac{1}{(u-1)k} \sum_{j=k+1}^{uk} H_{j,n}$$

In the iid case when a second order regular variation condition holds, the asymptotic variance of $smooH_{k,n}$ is less than that of the Hill estimator, namely.

$$\frac{1}{\alpha^2} \frac{2}{u} \left(1 - \frac{\log u}{u}\right)$$

The sensitivity of the Hill estimate to the choice of k corresponds in McNeil's work to the sensitivity of the fit of the generalized Pareto to the data to the choice of threshold Perhaps some comparable smoothing technique would help in GPD fitting.

(3) Alt plotting, Changing the scale. As an alternative to the Hill plot, it is sometimes useful to display the infrormation provided by the Hill or smooHill estimation as

$$\{(\theta, H^{-1}_{\lceil n^{\theta} \rceil, n}), 0 \leq \theta \leq 1\}$$

and similarly for the smooHill plot where we write $\lceil y \rceil$ for the smallest integer greater or equal to $y \geq 0$. We call such plots the *alternative Hill plot* abbreviated AltHill and the *alternative smoothed Hill plot* abbreviated AltsmooHill. The alternative display is sometimes revealing since the initial order statistics get shown more clearly and cover a bigger portion of the displayed space. However, when the data is Pareto or nearly Pareto, this alternate plotting device is less useful since in the Pareto case, the Hill estimator applied to the full data set is the maximum likelihood estimator and hence the correct answer is usually found at the right end of the Hill plot.

(4) Dynamic and static QQ-plots. As we did for the Hill plots, pick k upper order statistics

$$X_{(1)} > X_{(2)} > \dots > X_{(k)}$$

and neglect the rest. Plot

$$\{(-\log(1 - \frac{j}{k+1}), \log X_{(j)}), 1 \leq j \leq k\}. \tag{2.2}$$

If the data are approximately Pareto or even if the marginal tail is only regularly varying, this should be approximately a straight line with slope $1/\alpha$. The slope of the least squares line through the points is an estimator called the QQ-estimator (Kratz and Resnick, 1996). Computing the slope we find that the QQ-estimator is given by

$$\widehat{\alpha^{-1}}_{k,n} = \frac{\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))\log(\frac{X_{(i)}}{X_{(k+1)}}) - \frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))H_{k,n}}{\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))^2 - (\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))^2} \tag{2.3}$$

There are two different plots one can make based on the QQ-estimator. There is the dynamic QQ-plot obtained from plotting $\{k, 1/\widehat{\alpha^{-1}}_{k,n}, 1 \leq k \leq n\}$ which is similar to the Hill plot. Another plot, the static QQ-plot, is obtained by choosing and fixing k, plotting the points in (3.2) and putting the least squares line through the points while computing the slope as the estimate of α^{-1}.

The QQ-estimator is consistent for the iid model if $k \to \infty$ and $k/n \to 0$ and under a second order regular variation condition and further restriction on $k(n)$, it is asymptotically normal with asymptotic variance $2/\alpha^2$. This is larger than the asymptotic variance of the Hill estimator but the volatility of the QQ-plot always seems to be less than that of the Hill estimator.

(5) De Haan's moment estimator. McNeil discusses the extreme value distributions (see also Resnick, 1987; de Haan, 1970, Leadbetter et al, 1983, Castillo, 1988, Embrechts et al 1997) which can be parameterized as a one parameter family

$$G_\xi(x) = \exp\{-(1+\xi x)^{-\xi^{-1}}\}, \quad \xi \in \Re, 1+\xi x > 0$$

When $\xi = 0$, we interpret G_0 as the Gumbel distribution

$$G_0(x) = \exp\{-e^{-x}\}, \quad x \in \Re.$$

A distribution whose sample maxima when properly centered and scaled converges in distribution to G_ξ is said to be in the *domain of attraction* of G_ξ which in McNeil's notation is written $F\in MDA(G_\xi)$ If $\xi > 0$ and $F\in MDA(G_\xi)$ then $1 - F\in RV_{-1/\xi}$ De Haan's moment estimator $\hat{\xi}_{k,n}$ (Dekker's, Einmahl, de Haan, 1989, de Haan, 1991, Dekkers and de Haan, 1991; Resnick and Stărică, 1996b) is designed to estimate $\xi = 1/\alpha$ Note that $\hat{\xi}_{k,n}$, like the Hill estimator, is based on the k-largest order statistics Since most common densities such as the exponential, normal, gamma and Weibull densities and many others are in the $MDA(G_0)$, the domain of attraction of the Gumbel distribution, this provides another method of deciding when a distribution is heavy tailed or not If $\hat{\xi}_{k,n}$ is negative or very close to zero, there is considerable doubt that heavy tailed analysis should be applied and the moment estimator is usually much more reliable in these circumstances than the Hill estimator In particular, when $\xi = 0$, the Hill estimator is not usually informative and the moment estimator does a much better job of identifying exponentially bounded tails Smoothed versions of the moment estimator can also be devised (Resnick and Stărică, 1996b) which overcome volatility in the plot of $\{k, \hat{\xi}_{k,n}, 1 \le k \le n\}$

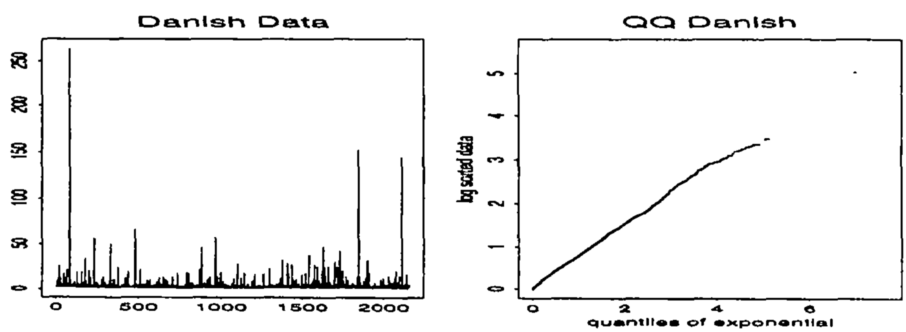


FIGURE 2 1 Tsplot and QQ plot of Danish data

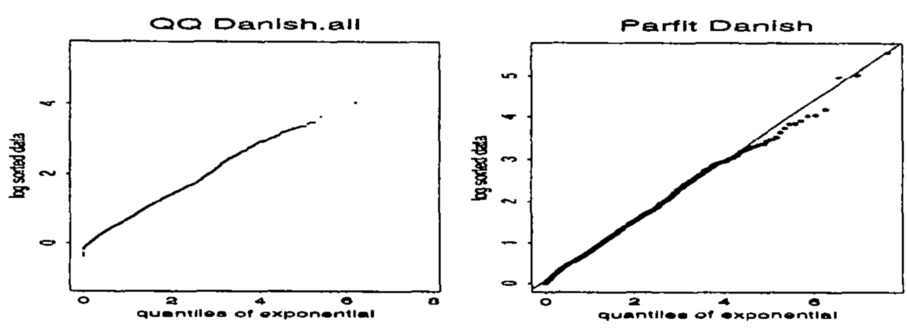


FIGURE 2 2 QQ plot of Danish all data and parameter estimate

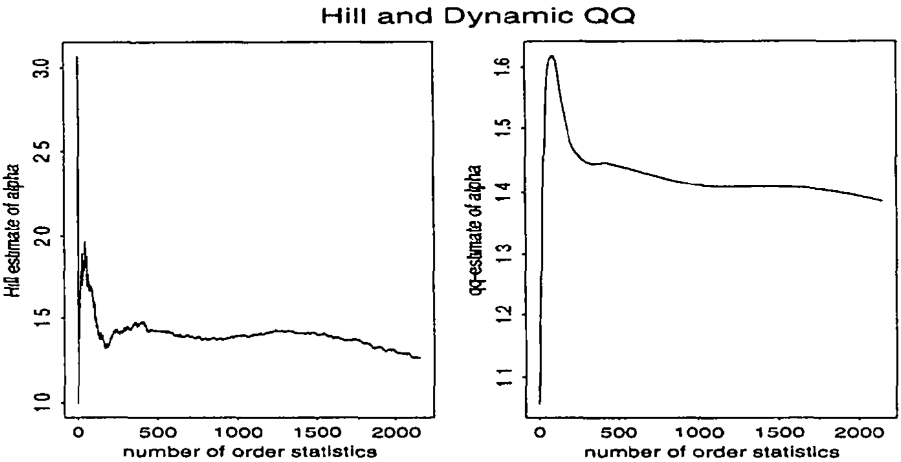


FIGURE 2 3 Hill and QQ-plot of Danish data

Figure 2 1 gives a time series plot of the 2156 Danish data consisting of losses over one million Danish Krone (DKK) and the right hand plot is the QQ-plot (2 2) of this data yielding a remarkebly straight plot Figure 2.2 gives the QQ-plot of all of the 2492 losses recorded in the data set labeled danish.all and shows why McNeil was statistically wise to drop losses below one million DKK (In the left hand plot the data is scaled to have a range of (0.3134041, 263 2503660) and the dots below height 0 represent the 325 values which are less than 1 in the scaled data.) The right hand plot in Figure 2 2 puts a line through the QQ-plot of the losses above one million and yields an estimate of α = 1 386 Using only the largest 1500 order statistics and then estimating α from the slope of the LS line produces an estimate of α = 1 4

We next attempted to estimate α by means of the Hill plot Figure 2 3 shows a Hill plot side by side with the dynamic QQ-plot. Because the plot in the right side of Figure 2 1 is so straight, we tend to trust the Hill plot near the right end of the plot This is because the straight plot in Figure 2 1 indicates the underlying distribution is close to Pareto and for the Pareto distribution the maximum likelihood estimator of the shape parameter is the Hill estimator calculated using all the data This analysis is confirmed by the excellent fit achieved by McNeil using a GPD with ξ = 0 684 or α = 1.46 corresponding to losses exceeding a threshold of 20 million DKK. Such a GPD is a shifted Pareto

On the other hand, examining the altHill and altsmooHill plots in Figure 2 4 makes it seem unlikely that α could be as large as 2 01 which is what is given in McNeil's Figure 7. This corresponds to a ξ = 0 497. Our methods indicate a likely value of α = 1 45

In Figure 2 5 we present four views of the moment estimator $\hat{\xi}_{k,n}$ of $\xi = 1/\alpha$. The upper right graph and the lower two graphs are in alt scale where k, $1 \le k \le n$ is replaced by $\lceil n^{\theta} \rceil$, $0 \le \theta \le 1$ Interestingly, we see here and in the four views of the Hill plot, that when the data are very close to Pareto, the alt scale is not advantageous

When the data is close to Pareto, the reliable part of the graph is toward the end and this is the part of the graph under emphasized by the alt scale The situation is very different for something like stable data (Resnick, 1995) where the traditional Hill plot is incapable of identifying the correct value of α but the alt plot does a superior job.

Based on an amalgam of the QQ, Hill and moment plots, we settle on an estimate of $\alpha = 1.4$ or $\xi = .71$

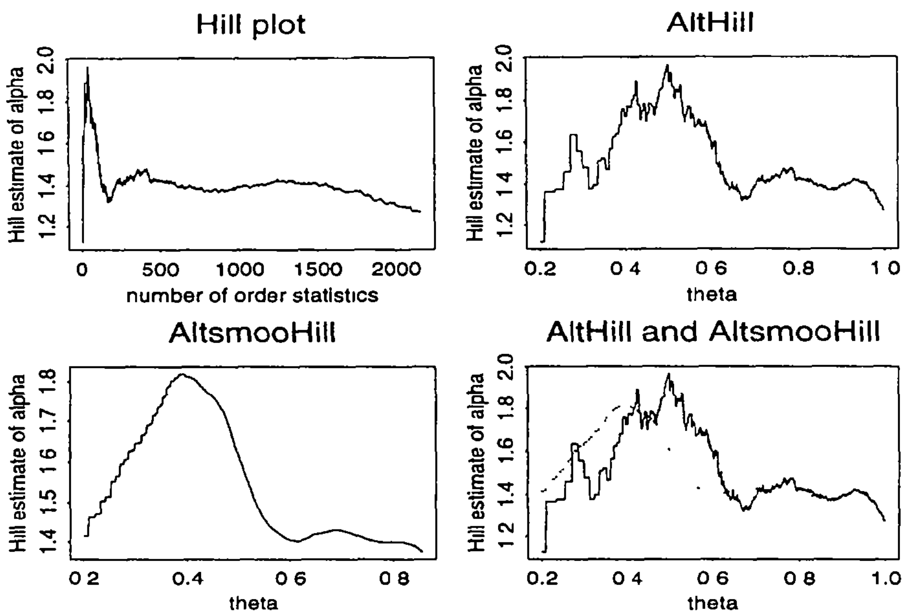


FIGURE 2.4 Hill and smooHill plots for Danish data

3 TESTING FOR INDEPENDENCE

We outline several tests for independence which can help reassure the analyst that an iid model is adequate and that it is not necessary to try to fit a stationary time series with dependencies to the data. Some of our tests are motivated by our experience trying to fit autoregressive processes to heavy tailed data

Here is a survey of several methods which can be used to test independence Some of these are based on asymptotic methods using heavy tailed analysis and the rest are standard time series tests of homogeneity

(1) Method based on sample acf. An exploratory, informal method for testing for independence can be based on the sample autocorrelation function $\hat{\rho}(h)$ where for h any positive integer

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h}(X_t - \overline{X})(X_{t+h} - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

In many studies of heavy tailed data, the centering by the sample mean is omitted since if mathematical expectation does not exist, there is no advantage or sense to centering by the sample mean However, since our chosen value of $\alpha = 1\ 4$ implies $E|X_1| < \infty$, we have decided to include the centering From Davis and Resnick (1985a), if $\{X_t\}$ are iid with regularly varying tail probabilities, then

$$\lim_{n\to\infty} \hat{\rho}(h) = \begin{cases} 1, & \text{if } h = 0, \\ 0, & \text{if } h \neq 0. \end{cases}$$

Thus, if upon graphing $\hat{\rho}(h)$, $h = 0, \ldots, n-h$ we get only small values for $h \neq 0$ there is no evidence against independence The limit distribution of $\hat{\rho}(h)$, $h = 1, \ldots, q$ is known (Davis and Resnick, 1985b, 1986 Corollary 1) but it is somewhat difficult to work with and the percentiles must be calculated by simulation It is important to realize that the 95% confidence bands drawn by a typical statistics package like Splus are drawn using Bartlett's formula (Brockwell and Davis, 1991) on the assumption that the data is Gaussian or at least has finite fourth moment This assumption is totally inappropriate for heavy tailed data and the confidence band must be drawn taking into account the heavy tailed limit distribution for $\hat{\rho}(h)$, $h = 1, \ldots, l$

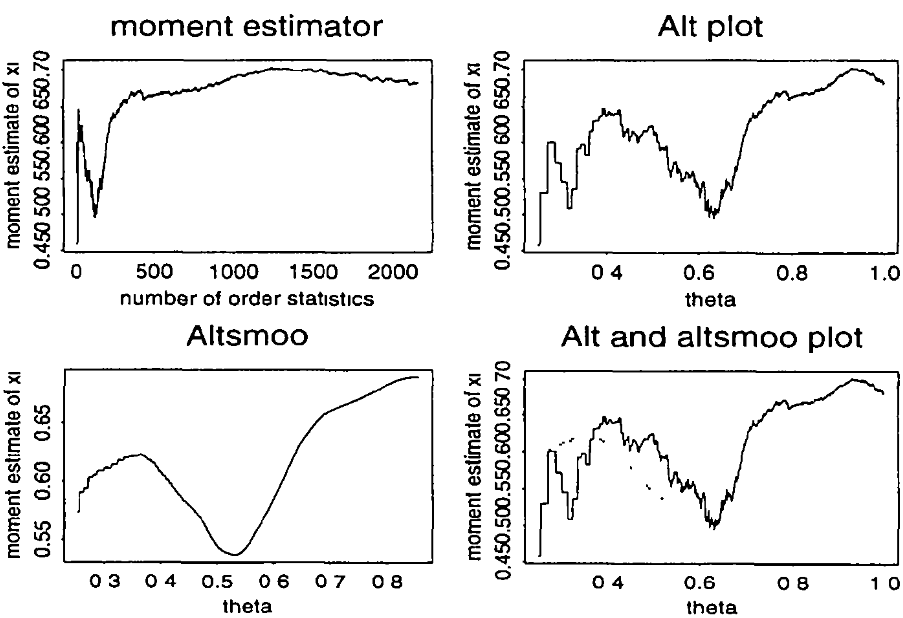


FIGURE 2 5 Moment estimator plots for Danish data

We discuss implementation of the acf based procedure when $1 < \alpha < 2$ since in the case of the Danish loss data we have settled on an estimate of $\alpha = 1.4$ Suppose $\{Y_1, \ldots, Y_n\}$ are iid non-negative random variables satisfying

$$P[Y_1 > x] \sim x^{-\alpha} L(x), \quad x \to \infty, 1 < \alpha < 2$$

where L is slowly varying From Corollary 1, page 553 of Davis and Resnick (1986), if we set $\hat{\rho}_Y(h)$ to be the lag h sample acf for $Y_1, \ldots, Y_n$, then we have

$$\lim_{n\to\infty} P[\hat{b}_n^{-1} b_n^2 \hat{\rho}_Y(h) \le x] = P[U_h / V_0 \le x]$$

where U_h is a one sided stable random variable with index $\alpha = 1.4$ and V_0 is a positive stable random variable with index $\alpha/2 = 0.7$ and b_n is the solution to

$$P[Y_1 > x] = 1/n$$

and $\hat{b}_n$ is the solution to

$$P[Y_1 Y_2 > x] = 1/n$$

Thus an approximate symmetric 95% confidence window for the sample correlations of the Y's would be placed at $\pm l\hat{b}_n / b_n^2$ where l satisfies

$$P\left[\left|U_h / V_0\right| \le l\right] = .95.$$

We estimate the 95%-quantile of $|U_h/U_0|$ by simulation and if we assume the distribution of Y_t's is Pareto from some point on, we find

$$l\frac{\hat{b}_n}{b_n^2} = l\left(\frac{n}{\log n}\right)^{-1/\alpha}$$

The assumption of a Pareto distribution seems mild in view of Figure 2 2 and the good fit found by McNeil of the GPD with positive shape parameter

Figure 3 1 presents this technique applied to the Danish loss data. No spike is protruding from the band and hence this acf based technique does not provide any evidence against the assumption of independence.

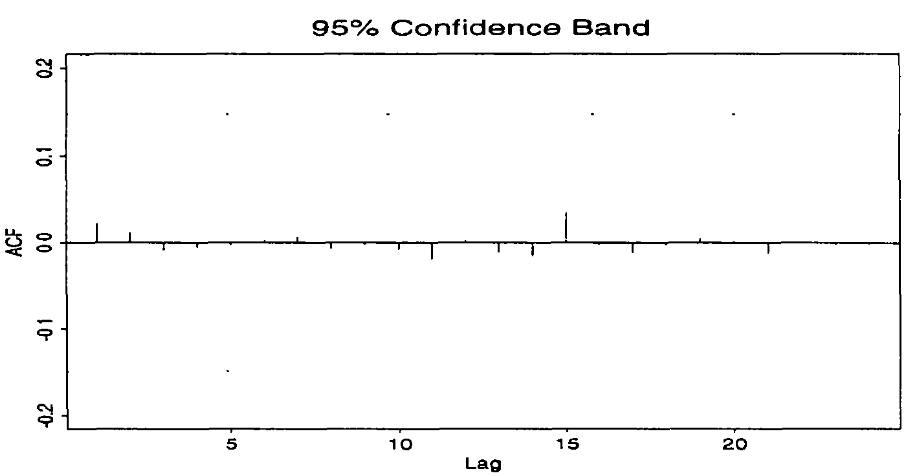


FIGURE 3 1 95% confidence band for the acf of the Danish loss data

(2) Tests based on asymptotic theory Estimators of autoregressive coefficients for heavy tailed time series can be used to fashion tests for independence against autoregressive alternatives If the autoregression is described as

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t, \quad t = 0, 1, \ldots$$

where $\{Z_t\}$ are iid heavy tailed residuals, then we test if

$$\phi_1 = \ldots = \phi_p = 0,$$

that is independence, by rejecting when the maximal estimated coefficient

$$\bigvee_{i=1}^{p} |\hat{\phi}_i(n)|$$

is too large This procedure has been implemented by Feigin, Resnick and Stărică (1996) based on *linear programming* (LP) estimators under the assumption that the iid heavy tailed residuals $\{Z_t\}$ are non-negative. See also Feigin and Resnick (1993)

It would not be possible to fix the size of the LP test if the limit distribution of the LP estimator did not considerably simplify Fortunately it does under the null hypothesis of independence and we then have

$$b_n(\hat{\phi}_1(n), \ldots, \hat{\phi}_p(n)) \Rightarrow \boldsymbol{L} \equiv (V_1^{-1}, \ldots, V_p^{-1})$$

where for $x_i \geq 0$, $i = 1, \ldots, p$ we have that

$$P[V_i \leq x_i, i = 1, \ldots, p] = \exp\{-\int_{(y_1, \ldots, y_p) \in [0,\infty]^p} \left(\bigwedge_{l=1}^{p} y_l x_l \right)^{-\alpha} F(dy_1) \ldots F(dy_p)\} \quad (3.2)$$

This means that if we want a 0.05 level rejection region, we should reject when $\vee_{i=1}^{p} |\hat{\phi}_i(n)| > K(.05)$ where $K(.05)$ is defined by

$$P\left[\bigvee_{i=1}^{p} |\hat{\phi}_i(n)| > K(.05) \right] = .05$$

and to find an approximate value of $K(.05)$ we write

$$P\left[\bigvee_{i=1}^{p} |\hat{\phi}_i(n)| > K(.05) \right] \approx P\left[\bigvee_{i=1}^{p} L_i > b_n K(.05) \right] \leq pP\left[L_1 > b_n K(.05)\right] = pe^{-c(b_n K(.05))^{\alpha}}, \quad (3.3)$$

where $c = E(Z_1^{-\alpha})$ This yields

$$K(.05) \approx \frac{\left(\frac{-\log(.05/p)}{c} \right)^{1/\alpha}}{b_n} = \frac{\left(\frac{\log(20p)}{c} \right)^{1/\alpha}}{b_n}$$

We need to estimate α, c and b_n One way to do this is to use the QQ-plot (Feigin, Resnick and Stărică, 1996; Kratz and Resnick, 1996) which yields both $\hat{b}_n$ (as the intercept of the fitted line) and $\hat{\alpha}$ (as the reciprocal of the slope of the fitted line) and then we can get

$$\hat{c} = n^{-1}\sum_{i=1}^{n} X_i^{-\hat{\alpha}}.$$

The asymptotic test is implemented and shown in Figure 3 2 None of the estimated coefficient values extend above the bar representing $K(.05)$ so this method provides no evidence against the hypothesis of independence

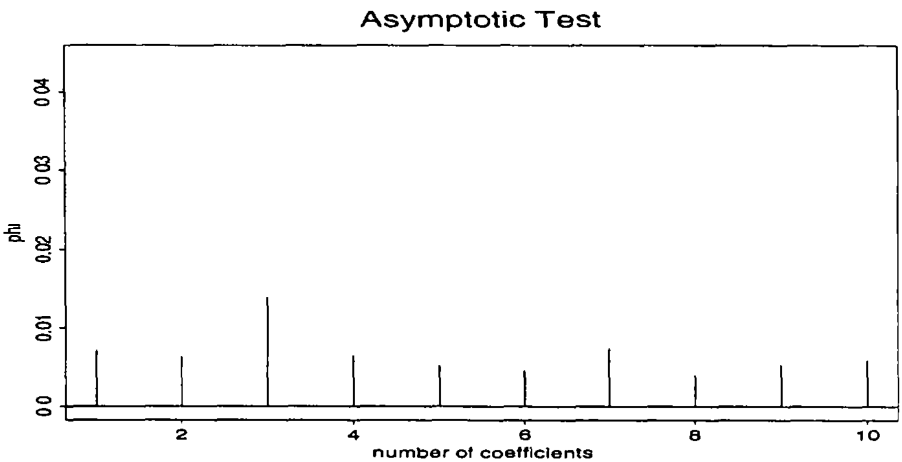


FIGURE 3 2 Asymptotic test for independence for the Danish loss data

(3) Standard tests of randomness. There are several standard time series tests of randomness (Brockwell and Davis, 1991, Section 9 4) which are non-parametric and can be employed in the present context. We give some examples below We use the notation

$$\chi_n \sim AN(\mu_n, \sigma_n^2)$$

as shorthand to mean that

$$(\chi_n - \mu_n)/\sigma_n \Rightarrow N(0,1)$$

(1) Turning point test If T is the number of turning points among $X_1, \ldots, X_n$ then under the null hypothesis that the random variables are iid we have

$$T \sim AN(2(n-2)/3, (16n-29)/90)$$

and this can be used as the basis of a test

(2) Difference-sign test Let S be the number of $i = 2, \ldots, n$ such that $X_i - X_{i-1}$ is positive Under the null hypothesis that the random variables $X_1, \ldots, X_n$ are iid we have

$$S \sim AN(\frac{1}{2}(n-1), (n+1)/12).$$

(3) Rank test Let P be the number of pairs (i, j) such that $X_j > X_i$ for $j > i$ and $i = 1, \ldots, n-1$ Under the null hypothesis that the random variables $X_1, \ldots, X_n$ are iid we have

$$P \sim AN(\frac{1}{4}n(n-1), n(n-1)(2n+5)/8).$$

We would reject the iid hypothesis at the 0 05 level if any of these standardized variables had an absolute value greather than 1 96. All of these tests are implemented in the Brockwell and Davis (1991) package ITSM. Data can easily be imported into their program and tested within the package for randomness.

We carried out these tests on the Danish loss data using ITSM and achieved the following results

Turning points	1409	AN (1436 00, $19\ 57^2$)
Difference-sign	1079	AN (1077.50, 13.41^2)
Rank test	1055894	AN (1161545, $50071\ 90^2$)

The rank test rejects the hypothesis of independence at the 5% level The turning points and difference-sign tests fail to reject.

(4) Stability testing on subsets of the data An informal but useful technique is to take a statistic, such as the sample acf, and compute it relative to different subsets of the sample If the data is iid, the values of the statistic should be similar across different subsets.

For the sample acf, if the graphs of $\hat{\rho}_{II}$ (h), $h = 1, \ldots, q$ look different for different subsets, then one should be skeptical of the correctness of the iid assumption Often it is enough to split the sample into halves or thirds to generate some skepticism One could make acf subset plots for the Danish data but since the acf values are not significantly different from 0, there seems little point to pursuing this diagnostic in this case

(5) Permutation test for independence. Another approach to testing for independence in time series analysis is based on permutation tests. Here we can use any desired statistic that is designed to measure some form of dependence between successive data This statistic might be a maximum autocorrelation or partial autocorrelation, or it may be a maximal autoregressive coefficient estimated by the linear programming paradigm

The permutation test is based on comparing the observed value of the statistic with the permutation distribution of that statistic — that is with the distribution of values of the statistic under all the possible permutations of the time series data If there is no dependence structure in the data, then the observed value should be a typical value for this reference permutation distribution. If there is some dependence of the type to which the statistic is sensitive, then the observed value should be extreme with respect to this reference distribution

This approach allows one to perform tests without relying on the asymptotic theory for the particular statistic. As we have seen earlier, the asymptotic distribution for

$$\bigvee_{i=1}^{p}\left|\hat{\phi}_i(n)\right|$$

involves various parameters that have to be estimated Moreover, the fact that we are not sure of the rate of convergence to the asymptotic distribution, also suggests the precautionary tactic of using a permutation test

In the implementation we use below, we approximate the *p-value* of the actually observed statistic This is achieved by generating 99 permutations of the time series, computing the statistic for each one, and counting the number (C) of these that are greater than or equal to the actually observed statistic The p-value is approximated by $(1+C)\%$. The statistics considered are the maximum absolute autocorrelation (macf), the maximum absolute partial autocorrelation (mpacf), and the maximum absolute linear programming coefficient estimate (mphi) In each case, one must specify the value of p, the order over which the maximum is taken

For the Danish loss data, we took the order to be 10 and ran the tests yielding the following p-values

maximum autocorrelation	0.52
maximum partial autocorrelation	0.51
maximum LP coefficient	0 22

and thus at a reasonable level, none of these tests would reject independence

4 CONCLUDING REMARKS

There is very little evidence arguing against the hypothesis of independence and it seems McNeil's presumption that the data were independent was a safe assumption to make for this data set Independence is not that common among teletraffic of finance data in my experience and thus should be treasured in the present insurance context Fitting dependent data with a heavy tailed stationary time series model can be a frustrating business (see Resnick, 1996b, Feigin and Resnick, 1996) so when one concludes the data can be modelled as iid, a loud sigh of relief is heard

The sensitivity of the estimation and fitting methods to the choice of threshold or the choice of the number of order statistics used in estimation is a persistent and troubling theme in McNeil's and my remarks. This seems inherent in the heavy tail and extreme value methods It is not clear at this point how much the techniques can be improved to reduce sensitivity to choice of k or threshold Smoothing techniques and alternate plotting help but are not a universal panacea.

It is encouraging to see the accumulating mass of theoretical and software tools which can be used to analyze such data sets

REFERENCES

BROCKWELL, P and DAVIS, R , *Time Series Theory and Methods, 2nd edition*, Springer-Verlag, New York 1991

BROCKWELL, P and DAVIS, R *ITSM An Interactive Time Series Modelling Package for the PC*, Springer-Verlag, New York 1991

CASTILLO, ENRIQUE, *Extreme Value Theory in Engineering*, Academic Press, San Diego, California, 1988

DAVIS, R and RESNICK, S , Limit theory for moving averages of random variables with regularly varying tail probabilities , Ann Probability **13** (1985a), 179-195

DAVIS, R and RESNICK, S , *More limit theory for the sample correlation function of moving averages*, Stochastic Processes and their Applications, **20** (1985b), 257-279

DAVIS, R and RESNICK, S , Limit theory for the sample covariance and correlation functions of moving averages, Ann Statist **14** (1989), 533-558

DEKKERS, A , EINMAHL, J , and HAAN, L DE, *A moment estimator for the index of an extreme value distribution*, Ann Statist **17** (1989), 1833-1855

DEKKERS, A and HAAN, L DE, On the estimation of the extreme value index and large quantile estimation, Ann Statist **17** (1989), 1795-1832

EMBRECHTS, P , KLUPPELBERG, C and MIKOSCH, T , *Modelling Extremal Events for Insurance and Finance*, To appear, Springer-Verlag, Heidelberg, 1997

FEIGIN, P and RESNICK, S , *Limit distributions for linear programming time series estimators*, Stochastic Processes and their Applications **51** (1994), 135-166

FEIGIN, P and RESNICK, S , *Pitfalls of fitting autoregressive models for heavy-tailed time series*, Available at http /www orie cornell edu/trlist/trlist htlml as TR1163 ps Z (1996)

FEIGIN, P , RESNICK S , and Stărică, C , *Testing for independence in heavy tailed and positive innovation time series*, Stochastic Models **11** (1995), 587-612

HAAN, L DE, *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Mathematical Centre, Amsterdam, Holland, 1970

HAAN, L DE, *Extreme Value Statistics*, Lecture Notes, Econometric Institute, Erasmus University, Rotterdam (1991)

HAAN, L DE and RESNICK, S , *On asymptotic normality of the Hill estimator*, TR1155 ps Z available at http /www orie cornell edu/trlist/trlist htlml (1996)

HILL, B , A simple approach to inference about the tail of a distribution, Ann Statist 3 (1975), 1163-1174

HSING, T , Extreme value theory for suprema of random variables with regularly varying tail probabilities , Stoch Proc and their Appl **22** (1986), 51-57

KRATZ, M and RESNICK, S , *The qq-estimator and heavy tails*, Stochastic Models 12 (1996), 699-724

LEADBETTER, M , LINDGREN, G and ROOTZEN, H , *Extremes and Related Properties of Random Sequences and Processes*, Spring Verlag, New York, 1983

MASON, D , Laws of large numbers for sums of extreme values, Ann Probability 10 (1982), 754-764

MCNEIL, A , Estimating the tails of loss severity distributions using extreme value theory, Preprint Dept Mathematics, ETH Zentrum, CH-8092 Zurich (1966)

RESNICK, S , *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York, 1987

RESNICK, S , *Heavy tail modelling and teletraffic data*, Available as TR1134 ps Z at http /www orie cornell edu/trlist/trlist htlml, Ann Statist (1995), (to appear)

RESNICK, S , *Why non-linearities can ruin the heavy tailed modeler's day*, Available as TR1157 ps Z at http /www orie cornell edu/trlist/trlist htlml, A PRACTICAL GUIDE TO HEAVY TAILS Statistical Techniques for Analysing Heavy Tailed Distributions (Robert Adler, Raisa Feldman, Murad S Taqqu, ed), Birkhauser, Boston 1996, (to appear)

RESNICK, S and Stărică, C , *Consistency of Hill's estimator for dependent data*, J Applied Probability **32** (1995), 139-167

RESNICK, S and Stărică, C , *Smoothing the Hill estimator*, To appear J Applied Probability (1996a)

RESNICK, S and Stărică, C , *Smoothing the moment estimator of the extreme value parameter*, Available as TR1158 ps Z at http /www orie cornell edu/trlist/trlist htlml, Preprint (1996b)

ROOTZEN, H , LEADBETTER, M and DE HAAN, L , *Tail and quantile estimation for strongly mixing stationary sequences*, Technical Report 292, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260 (1990)

ROOTZEN H *The tail empirical process for stationary sequences*, Preprint 1995 9 ISSN 1100-2255, Studies in Statistical Quality Control and Reliablity, Chalmers University of Technology (1995)

SIDNEY I RESNICK
Cornell University,
School of Operations Research and Industrial Engineering
Rhodes Hall 223
Ithaca, NY 14853 USA
E-mail: sid@orie cornell.edu

BOOK REVIEW

JAN BEIRLANT, JOZEF L TEUGELS and PETRA VYNCKIER (1996) · *Practical Analysis of Extreme Values.* Leuven University Press ISBN 90 6181 768 1

This short book aims to introduce the reader to some of the practical methods of handling extreme value statistics, with a particular leaning towards actuarial applications The emphasis is on graphical methods of fitting and comparing different types of distribution, and the estimation of extreme value index parameters.

The first chapter begins with elementary introductions to such concepts as density and distribution functions, and lists some of the numerous parametric distributions applied to non-life insurance data. In general this is accurate and informative, though the reader should be cautioned that the authors' definition of the "generalized Pareto" distribution is not the same as the one adopted by other writers on extreme value theory The latter part of the chapter describes a number of graphical methods for choosing among distributional families

The next three chapters concentrate on methods of estimating three different definitions of the extreme value index the Pareto index (chapter 2), the index of the general extreme value distribution (chapter 3) and Weibull indices (chapter 4) The main method of chapter 2 is the so-called Hill estimator, applied to the largest order statistics of a sample The most important practical issue with this estimator is how many of the largest order statistics to include, and the authors provide a good discussion of the mathematical principles underlining this choice I am less convinced of their proposed practical solution to the problem · it is based on a method only recently introduced by the authors themselves, and it seems to me that more experience is needed before recommending it to practising actuaries Chapters 3 and 4 are written in similar style, though I really feel that the authors should have made it clear that the general form of extreme value distribution is due, modulo changes of notation, to the original foundational papers of Fisher and Tippett (1928) and Gendenko (1943), and not, as the text implies, to a 1995 paper by two of the present three authors!

The final chapter 5 is a nice survey of the actuarial applications of extreme value theory There are also a number of data sets reproduced in an Appendix.

I feel that this book provides a useful survey of statistical techniques which will be accessible to readers without much background in statistics The desirable background in mathematics is somewhat greater, though the reader who does not feel at home in the language of regularly varying functions or Tauberian theorems can skip over those sections without losing much of the statistical thread. The book's main weakness is that it hardly gives any hint of the vast array of probabilistic and statistical extreme value theory which lies outside the rather narrow boundaries to which the authors have confined themselves here.

RICHARD SMITH

THE 6th AFIR INTERNATIONAL COLLOQUIUM

Nurnberg, Germany, 1996

The 6th AFIR International Colloquium was held at the Hotel Maritim in Nurnberg, Germany from 1 to 3 October, 1996 with about 190 participants from 17 different countries. Although most participants were from European countries there were a significant number from other countries including Australia, Israel, Japan, Taiwan, and USA. The organisation of the Colloquium was superb and the quality of the presented papers very high. There were almost 70 contributed papers. The Scientific Committee, chaired by Peter Albrecht, and the Organization Committee, chaired by Peter Burghard, are to be congratulated for an excellent meeting Invited lectures in Plenary sessions began both the morning and afternoon program Parallel sessions were then used to allow the authors of the contributed papers a reasonable time to present the main ideas in their papers. This meeting format worked well allowing participants to attend sessions in their area of interest

The social program for accompanying persons included bus tours to Bamberg, Rothenberg, a walk through "Romantic Nürnberg" and a guided tour of the court room of the "Nurnberg Trials". All of this looked enticing but most of us were there for the business side of the meeting

On completion of the Opening formalities on Tuesday 1 October the first invited lecture was by Hans Foellmer from Humboldt-University of Berlin on "Recent Developments in Option Pricing Theory" Option Pricing has been a theme of past AFIR Colloquia and this presentation was most appropriate It covered developments in stochastic mathematics and issues of incomplete markets. There followed parallel sessions with contributed papers on Option Pricing and on Asset Liability Management The area of asset-liability management has also been a common theme of previous colloquia.

After lunch, which provided the opportunity for further discussion and networking, the invited lecture was by Paul Embrechts of ETH Zurich with an advertised topic of "Methodological Issues Underlying Value at Risk Estimation". Paul's lecture emphasised modelling extreme values and the use of the generalised extreme value distributions including the Wiebul, Fréchet (Pareto related) and Gumbel (double exponential) cases Moreover, the generalised Pareto distributions are useful models for excess distributions. He mentioned that software for extreme value modelling was available from the World-Wide-Web site http://www.math.ethz.ch/~mcneil/software html and Paul also referred to a forthcoming book by Embrechts, Kluppelberg and Mikosh on "Modelling Extremal Events for Insurance and Finance" to be published by Springer in 1997.

One of the afternoon parallel sessions was on the topic of Risk Measurement and Risk Control and the other was on Asset-Liability Management. The Risk Measurement and Risk Control papers covered the areas of Value at Risk, Derivatives and reporting and supervision. The asset-liability session covered papers

ASTIN BULLETIN, Vol 27, No 1, 1997, pp 155-157

on pension fund and life insurance asset liability modelling and asset allocation including optimal asset allocation strategies In the evening the participants and accompanying persons adjourned to the Germanisches Nationalmuseum for a performance of the opera "The Abduction from the Seraglio" by Wolfgang Amadeus Mozart followed by a stand up reception. This excellent performance was especially presented for the AFIR Colloquium and the evening was most enjoyable.

Wednesday 2 October commenced with an invited lecture by Wolfgang Buehler from the University of Mannheim on "An Empirical Comparison of Valuation Models for Interest Rate Derivatives". The area of term structure models and their use in finance and actuarial applications has been an area of rapid theoretical development and understanding the different models and when they are most appropriate is an important topic. I am sure there will be more contributions to this area as actuaries increase their use of term structure models

The two parallel sessions following included one on Applications of Options in Investment Management and Insurance and one on Bond Valuation and Bond Management. The options session covered a wide range of topics including shortfall risks and the pricing of the new forms of guaranteed index-linked life insurance policies These policies have been recently introduced in Germany and are also popular now in North America They demonstrate the potential of exotic options for product design in life insurance and will be an area of much future interest as these products become more popular internationally The bond valuation session looked interesting but I chose to attend the options session.

The afternoon of Wednesday was free and participants had the choice of a tour of the city or a special guided tour of the Germanisches Nationalmuseum In the evening the social activities were "Frolics at the Imperial Castle" Europe is rich in history and, as these events testified, Nurnberg is no exception

The final day of the Colloquium was a holiday in Germany (German Unity Day) It opened with an invited lecture by David Wilkie on The European Single Currency. For both European and overseas participants this was a most interesting lecture. The intricacies involved in moving to a common currency range from deciding on a name for the currency to adjusting computer programs The following parallel sessions covered Applications of Numerical and Econometrical Methods in Finance and Portfolio-Capital Market Theory and Investment Management The Numerical and Econometric presentations included topics on Neural Networks, Genetic algorithms, and error correction models.

The final invited lecture was by Gerhard Rupprecht of Allianz Lebensversicherungs-AG who spoke on "The European Monetary Union from the Perspective of a German Life Insurer" providing another perspective on this topic to that given by David Wilkie in the morning lecture The following parallel sessions were on Current Problems in Insurance and Finance and covered a wide range of interesting topics.

The scientific program finished with a closing session summing up the Colloquium and with Catherine Prime from Australia inviting everyone to the 7th International AFIR Colloquium to be held in Cairns Australia from 13-15 August 1997 with a joint day with the ASTIN Colloquium on 13 August. We are all looking forward to next year and we have been inspired by the organisation of the Nurnberg Colloquium and intend adopting a similar structure with invited lectures and parallel sessions. Already arrangements are well in hand and those who wish to submit a paper should notify the Chair of the Scientific Committee (Mike Sherris) by email (msherris@efs mq.edu.au) or by Fax (+61 2 9850 8572) as soon as possible. Final papers are due by 1 March 1997. The call for papers can be viewed at http://www.ocs.mq edu.au/~msherris/afir97.html which includes instructions for authors.

For those who did not attend the Colloquium I can recommend that you obtain the Proceedings There were many topics covered and you will no doubt find some new ideas

The Colloquium concluded with a Gala night at the Hotel Maritim with entertainment, fine food and, most of all, fine company.

MIKE SHERRIS
School of Economic and Financial Studies
Macquarie University
Sydney NSW
Australia 2109

ASTIN COLLOQUIUM, CAIRNS, AUGUST 1997

The Call for papers made some months ago requested submissions by 1 March 1997. Many were received but the flow has now ebbed considerably.

This may reflect a belief on the part of prospective authors that further submissions are now too late to be accepted. This not the case.

Printing arrangements have now been re-negotiated to allow submissions received up to the end of May 1997 to be included in the volume of preprints circulated prior to the Colloquium Indeed, the Scientific Committee remain willing to receive papers up to the commencement of the Colloquium Those received after May will remain eligible for inclusion in the Colloquium program but will not be circulated in the preprint volume

Any further papers should be forwarded (3 copies + 1 electronic copy) to

Greg Taylor
Tillinghast-Towers Perrin
GPO Box 3279
Sydney NSW 2001
Australia

ASTIN Scientific Committee, April 1997

FACULTY APPOINTMENTS IN ACTUARIAL SCIENCE & INSURANCE

Nanyang Technological University, Singapore,
School of Accountancy and Business

Applications are invited for faculty positions in Actuarial Science in the School of Accountancy and Business. The School offers undergraduate degrees in Accountancy and Business, MBA degrees, and Master's and Doctoral degrees by research, and the latter by research and coursework.

Applicants should be experienced and qualified actuarial professionals with a strong interest in education, scholarship and research. Besides a professional actuarial qualification, they should also hold a postgraduate degree. In addition, they should be able to demonstrate academic and research achievement and potential.

The person appointed would be expected to teach in the B Bus (Actuarial Science) programme as well as actuarial subjects at a postgraduate level. In particular, he/she should be able to teach the following subjects: probability and statistics, life contingencies, mathematics of finance, applied actuarial statistics, mortality investigations, social security and pension funds, actuarial management, and actuarial aspects of general insurance.

The person appointed would also be expected to contribute actively to the School's research programme, to supervise research students and to take the lead on research projects.

Gross annual emoluments (for 12 months) range as follows

Professor:	**S$ 150,000 -S$ 202,110**	**Senior Lecturer:**	**S$ 67,940 - S$ 138,000**
Associate Professor:	**S$ 122,460 - S$ 170,000**	**Lecturer:**	**S$ 58,390 - S$ 74,300**

The commencing salary will depend on the candidate's qualifications, experience and the level of appointment offered.

In addition to the above, the University may decide to pay an annual variable component/allowance which has, in the past years ranged from 1 month' to 3 months' salary.

Leave and medical benefits will be provided. Other benefits, depending on the type of contract offered, include provident fund benefits or an end-of-contract gratuity, settling-in allowance, subsidised housing, children's education allowance, passage assistance and baggage allowance for transportation of personal effects to Singapore. Staff members may undertake consultation work subject to the approval of the University, and retain consultation fees up to a maximum of 60% of their gross annual emoluments in a calendar year.

Applicants should send their detailed curriculum vitae, including their areas of research interest, publications list and the names and addresses (internet and fax, if any) of three referees to:

Director of personnel, Nanyang Technological University
Nanyang Avenue, Singapore 639798
Telefax (65) 791340 — Internet: CCLIM@ntu.edu.sg

GUIDELINES TO AUTHORS

1. Papers for publication should be sent in quadruplicate to one of the Editors

   Paul Embrechts,
   Department of Mathematics, ETH-Zentrum,
   CH-8092 Zurich, Switzerland

   D Harry Reid,
   Eagle Star Insurance Company Ltd,
   The Grange, Bishop's Cleeve
   Cheltenham Glos GL52 4XX, United Kingdom

   or to one of the Co-Editors

   René Schnieper,
   Zurich Insurance Company,
   P O Box, CH-8022 Zurich, Switzerland

   Andrew Cairns
   Department of Actuarial Mathematics and Statistics
   Heriot-Watt University
   Edinburgh EH 14 4AS, United Kingdom

   Submission of a paper is held to imply that it contains original unpublished work and is not being submitted for publication elsewhere
   Receipt of the paper will be confirmed and followed by a refereeing process, which will take about three months

2. Manuscripts should be typewritten on one side of the paper, double-spaced with wide margins The basic elements of the journal's style have been agreed by the Editors and Publishers and should be clear from checking a recent issue of *ASTIN BULLETIN* If variations are felt necessary they should be clearly indicated on the manuscript

3. Papers should be written in English or in French Authors intending to submit longer papers (e g exceeding 30 pages) are advised to consider splitting their contribution into two or more shorter contributions

4. The first page of each paper should start with the title, the name(s) of the author(s), and an abstract of the paper as well as some major keywords An institutional affiliation can be placed between the name(s) of the author(s) and the abstract

5. Footnotes should be avoided as far as possible

6. Upon acceptance of a paper, any figures should be drawn in black ink on white paper in a form suitable for photographic reproduction with lettering of uniform size and sufficiently large to be legible when reduced to the final size

7. References should be arranged alphabetically, and for the same author chronologically Use a, b, c, etc to separate publications of the same author in the same year For journal references give 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