Dickson/Zehnwirth 27

ISBN 0 7325 1208 5

THE UNIVERSITY OF MELBOURNE

PREDICTIVE AGGREGATE CLAIMS DISTRIBUTIONS

by
David C M Dickson and Ben Zehnwirth
The University of Melbourne

RESEARCH PAPER NUMBER 27
February 1996

Centre for Actuarial Studies
Department of Economics
The University of Melbourne
Parkville, Victoria, 3052
Australia.

Predictive Aggregate Claims Distributions

by

David C M Dickson

and

Ben Zehnwirth,

The University of Melbourne

ABSTRACT

In the collective risk model we use the objective Bayesian approach to calculate predictive aggregate claims distributions. We compare these with fitted distributions which take no account of parameter uncertainty and show that actuarial functions such as premiums can be substantially understated if parameter uncertainty is ignored. We illustrate the situation when the moments of the predictive individual claim amount distribution do not exist and we discuss ways of applying such distributions to insurance problems.

KEYWORDS

Aggregate claims; parameter uncertainty; Bayesian predictive distribution

1. Introduction

The traditional inferential actuarial problem usually involves the passage from observed past data to unobserved future outcomes. It is well known that actuarial calculations, stop loss premiums, for example, are not robust to the distributional assumptions. Fitted distributions, in contrast to predictive distributions, do not incorporate parameter estimation uncertainty, or what North American actuaries call "risk parameter uncertainty". As a result, premium calculations, ruin probability calculations, surplus calculations etc., are understated if they are based on the fitted distribution. In certain circumstances the understatement can be quite substantial. Incorporation of parameter uncertainty can make a substantial difference, even to premiums calculated by the expected value principle.

In the present paper we apply the Bayesian paradigm to compute predictive aggregate cost distributions and make comparisons with the fitted aggregate cost distributions. It is shown that both ruin probabilities and stop loss premiums can under certain circumstances be substantially understated by the fitted distribution. Accordingly, actuaries who in practice use fitted distributions run major risks.

The Bayesian paradigm is often criticised because the choice of the prior distribution is subjective and removes the aura of objectivity in any analysis. This objection is removed as we use non-informative priors. A Bayesian uses the so-called Bayesian predictive density to forecast future observations. The approach is natural in that one bases one's prediction on the conditional distribution of the future given the past.

The Bayesian predictive distribution incorporates automatically both sources of uncertainty, namely process uncertainty and parameter estimation error.

The use of predictive distributions, in place of fitted distributions, is not new. Klugman (1992) describes the (objective) Bayesian paradigm and illustrates, inter alia, the difference between a fitted Weibull and a predictive Weibull. The fitted distribution does not display as much variation as the predictive distribution. Indeed, the fitted distribution does not necessarily belong to the same family of distributions as the predictive distribution. Cairns (1995) describes the Bayesian approach as a way of incorporating parameter uncertainty in the "Modelling Process" and applies the approach to the calculation of risk theory's adjustment coefficient. Applications of predictive distributions to reinsurance have been given by Hesselager(1993) and Hürlimann(1993 and 1995).

The results contained in this paper are critical to risk based capital considerations, especially if one is concerned with measuring risk statistically.

The outline of the paper is as follows. In Sections 2 and 3 we present background results. In Section 4 we compare fitted and predictive aggregate claims distributions and in Section 5 we consider reinsurance. Finally, in Section 6 we discuss the situation when the moments of the predictive aggregate claim amount distribution do not exist.

2. Preliminaries

In this section predictive distributions for the exponential and Poisson distributions are provided. In broad outline we start with prior beliefs represented by proper prior distributions and take limits to obtain posterior beliefs and predictive distributions based on diffuse (or ignorant) priors. We will use these predictive distributions in our examples in subsequent sections.

2.1. Exponential distribution with gamma prior

Let $X_1, \ldots, X_n|\theta$ be independent and identically distributed random variables with p.d.f.

$$f(x|\theta) = \theta \exp\{-\theta x\} \qquad \text{for } x > 0 \tag{2.1}$$

where $\theta > 0$. Let $D = (X_1, \ldots, X_n)$ represent the data vector. When the prior distribution for θ is $\mathrm{G}(\alpha, \beta)$, i.e. a gamma distribution with p.d.f.

$$f(\theta) = \frac{\theta^{\alpha-1}\beta^{\alpha}e^{-\beta\theta}}{\Gamma(\alpha)} \qquad \text{for } \theta > 0$$

where $\alpha, \beta > 0$, it is well known that

$$\theta|D \sim G(\alpha + n, \beta + n\bar{X})$$

Now let $g(\theta|D)$ be the p.d.f. of $\theta|D$ and let X^* be a subsequent observation from the exponential distribution given by (2.1). Then the p.d.f. of $X^*|D$ is given by

$$f(x^*|D) = \int_0^\infty f(x^*|\theta)g(\theta|D)d\theta$$

so that

$$f(x^*|D) = \frac{(n+\alpha)(\beta + n\bar{X})^{n+\alpha}}{(\beta + n\bar{X} + x^*)^{n+\alpha+1}} \qquad \text{for } x^* > 0$$

Thus $X^*|D$ has a Pareto distribution with parameters $n + \alpha$ and $\beta + n\bar{X}$.

The diffuse prior is obtained by letting both α and β go to zero in such a way that α/β is constant. Hence, with a diffuse prior $X^*|D$ has a Pareto$(n, n\bar{X})$ distribution.

Note that when the prior is diffuse both $E(X^*|D)$ and $V(X^*|D)$ exceed the mean and variance of an exponential distribution whose parameter is the maximum likelihood estimate of θ. This may not be the case when the prior is not diffuse.

2.2. Poisson distribution with gamma prior

Let $N|\lambda \sim Poisson(\lambda)$ and let the prior distribution for λ be $G(\alpha, \beta)$. Then the unconditional distribution of N is negative binomial $NB(\alpha, \beta/(1+\beta))$ and the posterior distribution for $\lambda|N$ is $G(\alpha + N, 1 + \beta)$. It therefore follows that if N^* is a subsequent observation from the Poisson distribution then the predictive distribution for $N^*|N$ is $NB(\alpha + N, (1+\beta)/(2+\beta))$.

The diffuse prior then leads to a predictive distribution that is $NB(N, 1/2)$. Hence the predictive distribution has the same mean as the fitted Poisson(N) distribution, but has twice the variance of the fitted distribution.

In this particular case, the mean square error (MSE) of prediction using classical theory is

$$\begin{aligned} E(N^* - N)^2 &= E(N^* - \lambda + \lambda - N)^2 \\ &= E(N^* - \lambda)^2 + E(N - \lambda)^2 \\ &= \lambda + \lambda \\ &= 2\lambda \end{aligned}$$

Hence, an estimate of the MSE is given by $2\hat{\lambda} = 2N$ which is the same as the predictive variance of N^* under the Bayesian paradigm with a diffuse prior. Accordingly, the estimate of the MSE of prediction using the classical approach is equivalent to the predictive variance using the objective Bayesian approach.

2.3. Normal distribution with normal prior for mean and known variance

Let $X_1, \ldots, X_n | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$ with σ^2 known and let the prior distribution for μ be $N(\mu_0, \sigma_0^2)$. Let $D = (X_1, \ldots, X_n)$ again represent the data vector. It is well known (see, for example, Lee(1989)) that

$$\mu | D \sim N(\hat{\mu}, \hat{\sigma}^2)$$

where $\hat{\mu}$ is given by

$$\hat{\mu} = (1 - Z)\mu_0 + Z\bar{X}$$

and the posterior variance or mean square error (MSE) is given by

$$\hat{\sigma}^2 = (1 - Z)\sigma_0^2$$

where Z (the credibility factor) gives the relative precision of the two sources of information, i.e.

$$Z = \frac{n}{n + \sigma^2/\sigma_0^2}$$

Let us assume for the remainder of this section that the prior is diffuse, i.e. that $\sigma_0^2 \to \infty$. Then we find that $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \sigma^2/n$.

Now let X^* be a subsequent observation from the $N(\mu, \sigma^2)$ distribution. Then it is straightforward to show that the distribution of $X^*|D$ is $N(\hat{\mu}, \hat{\sigma}^2 + \sigma^2)$. Hence the predictive distribution contains the two sources of uncertainty or variability, viz., $\hat{\sigma}^2 (= \sigma^2/n)$ and σ^2.

Now let $Y^* = \exp(X^*)$. Then the distribution of $Y^*|D$ is lognormal with parameters $\hat{\mu}$ and $(\sigma^2/n) + \sigma^2$, and so

$$E(Y^*|D) = \exp\left(\hat{\mu} + \frac{1}{2}\left(\frac{\sigma^2}{n} + \sigma^2\right)\right)$$

and

$$V(Y^*|D) = E^2(Y^*|D).\left(\exp\left(\frac{\sigma^2}{n} + \sigma^2\right) - 1\right)$$

Note that both $E(Y^*|D)$ and $V(Y^*|D)$ exceed the mean and variance of a lognormal distribution whose parameters are the maximum likelihood estimates of μ and σ^2. The predictive mean incorporates the component $\exp(\sigma^2/2n)$ where σ^2/n is the variance of the sample mean.

2.4. Normal distribution with gamma prior for unknown variance

Once more let $X_1, \ldots, X_n | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$. If σ^2 is unknown the diffuse prior is given by

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

Note that the prior is an improper distribution. However, both the posterior and predictive distributions are proper. We know (see, for example, Lee(1989)) that

$$\begin{aligned}\frac{1}{\sigma^2}|D &\sim G\left(\frac{n-1}{2}, \frac{S}{2}\right) \\ \mu|\frac{1}{\sigma^2}, D &\sim N\left(\bar{X}, \frac{\sigma^2}{n}\right)\end{aligned}$$

and,

$$\mu|D \sim t\left(n-1, \bar{X}, \frac{(n-1)n}{S}\right) \tag{2.2}$$

where

$$S = \sum_{i=1}^{n}(X_i - \bar{X})^2,$$

$G(\alpha, \beta)$ denotes a gamma distribution with mean $\alpha\beta$, and $t(v, r, s)$ denotes a t-distribution which is defined in Appendix 1. If we let $s = (S/(n-1))^{1/2}$ then an alternative way of writing (2.2) is

$$\sqrt{n}\frac{\bar{X}-\mu}{s}|D \sim t(n-1, 0, 1)$$

The classical statistics (or sampling theory statistics) approach leads to a similar conclusion. In this case

$$\sqrt{n}\frac{\bar{X}-\mu}{s} \sim t(n-1, 0, 1)$$

However, in the classical approach it is the data that are regarded as random and the parameters μ and σ^2 are fixed. By contrast, the objective Bayesian approach regards μ and σ^2 as random and the available data as being fixed.

If X^* denotes a subsequent observation from the $N(\mu, \sigma^2)$ distribution, the predictive distribution for $X^*|D$ is

$$X^*|D \sim t\left(n-1, \bar{X}, \frac{(n-1)n}{(n+1)S}\right)$$

This result is derived in Appendix 2. From results given in Appendix 1 it follows that $E(X^*|D) = \bar{X}$ and $V(X^*|D) = (n+1)S/(n(n-3))$. From the classical statistics standpoint, the distribution of

$$\sqrt{\frac{n}{n+1}}\frac{X^* - \bar{X}}{s}$$

is also $t(n-1, 0, 1)$. (Here we use $S/(n-3)$ as the estimator of σ^2.)

If we again define Y^* by $Y^* = \exp(X^*)$, then the predictive distribution for $Y^*|D$ is the distribution of $\exp(X^*)|D$.

3. Predictive Aggregate Claims Distributions

The traditional risk model for aggregate claims is as follows.

$$S = Y_1 + \ldots + Y_N$$

where S represents the aggregate claim amount in a fixed time period (typically one year), N represents the number of claims occurring in that period, and $Y_1, Y_2, \ldots$, represent the amounts of successive claims. We assume that $\{Y_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. random variables and that N is independent of this sequence.

In the last fifteen years much attention has been devoted to computing the aggregate claims distribution given the distributions of N and Y_i. Examples of recursive algorithms to calculate the distribution of S are given by Panjer (1981), Schroter(1991) and Sundt (1992).

The approach in practice is to fit the distributions of N and Y_i to data and, if the distribution of N belongs to the appropriate class, to apply the fitted distributions as inputs into these algorithms. There is a major drawback to this approach as the distributions do not incorporate parameter estimation error. The distributions are assumed to be known with certainty.

Consider the following statistical setting:

Data for one time period: $D = \{N; Y_1, \ldots, Y_N\}$
"Future" observations for the next period are $N^*, Y_1^*, \ldots, Y_{N^*}^*$.

We are interested in computing the predictive distribution of

$$S^* = Y_1^* + \ldots + Y_{N^*}^*$$

(conditional on the data D).

We will compare this with the fitted aggregate claims distribution defined by

$$\hat{S} = \hat{Y}_1 + \ldots + \hat{Y}_{\hat{N}}$$

where $\hat{N}$ has the fitted distribution of N based on the data D and $\hat{Y}_i$ has the fitted distribution of Y_i.

We will assume throughout that the claim count $N|\lambda$ has a Poisson distribution with mean λ. Accordingly, if $\hat{\lambda}$ is the fitted value of λ, the mean and variance of the fitted and predictive aggregate costs are

$$\begin{aligned} E(\hat{S}) &= \hat{\lambda} E(\hat{Y}_i) = N E(\hat{Y}_i) \\ V(\hat{S}) &= N E(\hat{Y}_i^2) \end{aligned}$$

for the fitted distribution and

$$\begin{aligned} E(S^*|D) &= E(N^*|D) E(Y_i^*|D) = N E(Y_i^*|D) \\ V(S^*|D) &= V(N^*|D) E^2(Y_i^*|D) + E(N^*|D) V(Y_i^*|D) \end{aligned}$$

for the predictive distribution.

In the special case when we have a diffuse prior for the Poisson parameter and the distribution of the individual claim amounts is assumed to be known we have $E(\hat{S}) = E(S^*|D)$ and

$$\begin{aligned} V(S^*|D) &= 2NE^2(Y_i) + NV(Y_i) \\ &= NE(Y_i^2) + NE^2(Y_i) \end{aligned}$$

Therefore,

$$V(S^*|D) = V(\hat{S}) + NE^2(Y_i)$$

Thus, uncertainty in the value of the Poisson parameter results in equal means but greater variability in the predictive distribution. We will see in our examples in the next section that this is not true when the parameters of the individual claim amount distribution are unknown, or when the prior for the Poisson parameter is not diffuse.

4. Moments and Percentiles of Aggregate Claims Distributions

In this section we will consider moments and percentiles of aggregate claims distributions. We will compare values for fitted and predictive distributions under three scenarios for the claim number distribution. Recall that our basic model is $N|\lambda$ has a Poisson(λ) distribution. Having observed the value of N, we will use the following three distributions for the claim number distribution in the next year:

1. Poisson($\hat{\lambda}$), where $\hat{\lambda}$ is the maximum likelihood estimate of λ (which is just equal to the observed value of N).

2. Negative binomial with parameters $\alpha + N$ and $(\beta + 1)/(\beta + 2)$. This is the predictive distribution resulting from a Gamma(α, β) prior for λ.

3. Negative binomial with parameters N and $1/2$. This is the predictive distribution for N resulting from a diffuse prior for λ.

Hence case (i) results in a fitted aggregate claims distribution and cases (ii) and (iii) in predictive aggregate claims distributions. Note that all subsequent results are conditional on the observed data D.

4.1. Known claim amount distribution

Let us first assume that the parameter of the exponential individual claim amount distribution is known and that the distribution has mean 1. Thus, we are initially considering the effect of uncertainty about the claim number distribution on the aggregate claims distribution.

Example 1: Let the observed value of N be 106. (This was in fact obtained as a simulation from the Poisson(100) distribution.). Let the parameters of the prior distribution in (ii) $\alpha = 4$ and $\beta = 0.04$ so that the prior distribution has mean 100 and standard deviation 50. Table 1 shows the mean, variance and coefficient of skewness of the aggregate claims distribution for each of the three cases. Formulae for calculating these quantities can be found in Panjer and Willmot (1992).

Table 1

Case	Mean	Variance	Skewness
(i)	106.00	212.00	0.2060
(ii)	105.77	313.24	0.2598
(iii)	106.00	318.00	0.2617

The pattern of figures in Table 1 is much as expected. As the variability of the counting distribution increases from case (i) through to case (iii) the variance and skewness of the aggregate claims distribution both increase. The mean for case (ii) is less than for the other two cases. Thus, parameter variability in the counting distribution alone impacts on all insurance calculations, such as setting premiums or surplus requirements, which hinge on the moments of the aggregate claims distribution.

Table 2 shows percentiles of the aggregate claims distribution in each case. These distributions were calculated according to Panjer's (1981) recursion formula and the exponential distribution was discretised on intervals of 0.05 using the method of Goovaerts and de Vylder (1988). We will use this discretisation interval in all our examples. Percentiles are denoted by C_x and the tabulated values show the least value of z such that the probability that the aggregate claim amount is no more than z is at least x. It is a consequence of our discretisation method that the percentiles are integer multiples of 0.05.

Table 2

Case	$C_{0.90}$	$C_{0.95}$	$C_{0.99}$	$C_{0.995}$
(i)	124.95	130.80	142.05	146.30
(ii)	128.90	136.15	150.25	155.60
(iii)	129.30	136.60	150.85	156.25

The figures in Table 2 confirm the findings from Table 1. As variability in the claim number distribution increases, percentiles of the aggregate claims distribution increase.

As a simple illustration of the effect of parameter uncertainty, consider the following problem. Suppose that the insurer calculates the premium, P, for a risk by the expected value principle using a premium loading factor of 10%, and wants to find the surplus U such that

$$\Pr(U + P > S) = \pi$$

where S denotes aggregate claims from the risk. Table 3 shows values of U for different values of π when the distribution of S is given by each of cases (i), (ii) and (iii).

Table 3

Case	$\pi = 0.1$	$\pi = 0.05$	$\pi = 0.01$	$\pi = 0.005$
(i)	8.35	14.20	25.45	29.70
(ii)	12.55	19.80	33.90	39.25
(iii)	12.70	20.00	34.25	39.65

Table 3 shows that by applying a fitted distribution instead of a predictive distribution, the insurer can set a surplus level that is quite inadequate - the worst case in Table 3 shows a surplus under the fitted distribution that is about $2/3rds$ of that required using the predictive distribution with a diffuse prior. This is a substantial margin, bearing in mind that the individual claim amount distribution is assumed known in this example!

4.2. Unknown claim amount distribution

Let us now assume that the parameter of the exponential individual claim amount distribution is unknown. In the following example we have used the same number of claims as in Example 1, then simulated this number of observations from an exponential distribution with mean 1.

Example 2: The maximum likelihood estimate of the parameter of the exponential distribution based on 106 simulated individual claim amounts is $\hat{\theta} = 1.0113$. We will use the same three counting distributions as before. The individual claim amount distributions will be:

- Exponential with mean 0.9888 for case (i), i.e. the fitted exponential distribution.
- Pareto(110,108.81) for case (ii), i.e. the predictive distribution based on a gamma prior for the exponential parameter with mean 1 and variance 0.25.
- Pareto(106,104.81) for case (iii), i.e. the predictive distribution based on the diffuse prior.

Tables 4 and 5 show the same quantities as Tables 1 and 2 respectively.

Table 4

Case	Mean	Variance	Skewness
(i)	104.81	207.28	0.2060
(ii)	105.59	314.12	0.2616
(iii)	105.81	318.89	0.2635

Table 5

Case	$C_{0.90}$	$C_{0.95}$	$C_{0.99}$	$C_{0.995}$
(i)	123.55	129.30	140.45	144.65
(ii)	128.75	136.00	150.15	155.55
(iii)	129.15	136.45	150.75	156.15

Comparing Tables 1 and 4 we see that both the variance and skewness of each aggregate claim amount distribution are slightly increased when we introduce parameter uncertainty to the individual claim amount distribution. However, these increases are not of a huge magnitude. Comparing Tables 2 and 5, we see that the same pattern is present in each table. The slightly smaller values in Table 5 simply reflect the lower means in Table 4. Figure 1 shows the three aggregate claims distributions.

These tables suggest that uncertainty in the claim number distribution is of much greater significance than uncertainty in the individual claim amount distribution. This is confirmed in Example 3 where we have considered a larger portfolio. This is not particularly surprising. In each case the coefficient of variation of the estimate of the parameter λ of the claim number distribution is very much greater than that of the estimate of the parameter θ of the individual claim amount distribution.

Example 3: The maximum likelihood estimate of the parameter of the exponential distribution based on 515 simulated individual claim amounts is $\hat{\theta} = 1.0137$. We will consider the three cases used in Example 2. For case (ii) we have adopted the same prior distribution for the exponential parameter. The prior for the Poisson parameter has mean 500 and variance 2500. Table 6 shows percentiles of the fitted and predictive aggregate claims distributions. In addition we have shown percentiles when the individual claim amount distribution is assumed to be known (and has mean 1). These cases are denoted by K in the table.

Table 6

Case	$C_{0.90}$	$C_{0.95}$	$C_{0.99}$	$C_{0.995}$
(i)K	556.45	568.65	591.85	600.45
(ii)K	561.80	576.40	604.35	614.75
(iii)K	565.85	580.95	609.85	620.60
(i)	548.95	560.95	583.85	592.35
(ii)	555.35	569.80	597.45	607.70
(iii)	559.30	574.25	602.80	613.45

This table shows that apart from a small change in location, the use of fitted and predictive individual claim amount distributions has little impact on the percentiles of the aggregate claims distribution calculated with the known individual claim amount distribution.

5. Reinsurance premiums

In this section we make a comparison between the pure premiums for excess of loss reinsurance and for stop loss reinsurance for the three cases described in Example 2.

5.1. Excess of loss reinsurance

Let $S_R(M)$ denote the reinsurer's aggregate claim amount under an excess of loss reinsurance arrangement with retention level M. Figure 2 shows the pure excess of loss premium, $E(S_R(M))$, as a function of the retention level for each of the three cases. These functions were calculated from the following formulae. For case (i)

$$E(S_R(M)) = (\lambda/\theta)\exp\{-\theta M\}$$

where λ is the parameter of the fitted Poisson distribution and θ is the parameter of the fitted exponential distribution. For cases (ii) and (iii)

$$E(S_R(M)) = \frac{k(1-p)}{p}\left(\frac{\delta}{\delta+M}\right)^{\alpha}\frac{\delta+M}{\alpha-1}$$

where k and p are the parameters of the predictive negative binomial claim number distribution and α and δ are the parameters of the predictive Pareto individual claim amount distribution. We can see that for some values of M there is not a great deal of difference between the pure premiums. However, the difference can be significant. For example, when $M = 2$ the value of $E(S_R(M))$ under case

(iii) is about 5% greater than under case (i). Figure 3 shows that there are much greater differences between the variances of the aggregate claims distributions for the reinsurer. Again considering the case $M = 2$, the variance under case (iii) is 16% greater than under case (i). The formulae underlying Figure 3 are

$$V(S_R(M)) = (2\lambda/\theta^2)\exp\{-\theta M\}$$

for case (i) and

$$V(S_R(M)) = \frac{k(1-p)}{p}\left(\frac{\delta}{\delta+M}\right)^{\alpha}\frac{(\delta+M)^2}{\alpha-1}\left(\frac{2}{\alpha-2}+\frac{1-p}{p(\alpha-1)}\left(\frac{\delta}{\delta+M}\right)^{\alpha}\right)$$

for cases (ii) and (iii).

In both Figures 2 and 3 the values under case (ii) are very close to those under case (iii). Our experiments with other parameter values for the prior distribution for θ in case (ii) indicate that the functions $E(S_R(M))$ and $V(S_R(M))$ are not particularly sensitive to the parameters of this prior distribution. Thus, the main reason for differences in values between the fitted and predictive aggregate claims distributions is the difference in the claim number distributions.

5.2. Stop loss reinsurance

Figure 4 shows the pure stop loss premiums as a function of the retention level, denoted d, for each of the three cases. These have been calculated recursively from the discrete aggregate claim amount distribution. (See, for example Bowers *et al (1986).)* This figure shows that the premium calculated from the fitted aggregate claims distribution always understates that calculated from the predictive distribution. For example, when $d = 120$, the premium calculated from the fitted distribution is about 50% of that calculated from the predictive distribution of case (iii).

6. Distributions whose moments do not exist

In section 2 we noted that when the model for individual claim amounts is lognormal with parameters μ and σ, with a diffuse prior, the predictive distribution is that of $\exp(X^*)|D$ where $X^*|D \sim t(n-1, \bar{X}, n(n-1)/(n+1)S)$. An immediate problem that arises with applying this predictive distribution to insurance problems is that its moments do not exist. Klugman (1992, p.21) notes this problem in relation to a Weibull distribution. It is possible to calculate the aggregate claims distribution with this predictive individual claim amount distribution, but it does not seem to be a useful model, especially as insurance claim amounts are finite in practice. In this section we consider two pragmatic approaches to the problem of predictive individual claim amount distributions whose moments do not exist. Each approach approximates the predictive distribution by a distribution whose moments exist. The first is specific to the lognormal model, the second is more generally applicable.

Our first approach is to approximate the $t(n-1, \bar{X}, n(n-1)/(n+1)S)$ distribution by a $N(\bar{X}, (n+1)S/n(n-3))$ distribution. This is a well-known approximation, and as the value of n increases, the quality of the approximation improves. An immediate consequence of this approximation is that the predictive individual claim amount distribution is lognormal and hence the moments of the predictive distribution exist. If we fit a lognormal distribution to data, then the maximum likelihood estimates of the parameters are $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = S/n$. Hence the moments of our (approximate) predictive lognormal distribution exceed those of the fitted lognormal distribution.

The second, and more general, approach is to assume that there is a fixed amount, say ω, which is the maximum possible claim. Thus, if $Y^*|D = \exp(X^*)|D$ has distribution function $F(x)$ we will approximate this distribution over the interval $(0, \omega)$ by $F(x)/F(\omega)$. There is of course an element of subjectivity in this approach, namely in the choice of ω. However, the advantage of this approach is that once again all the moments of the individual claim amount distribution exist. We refer to this distribution below as the truncated predictive distribution.

To illustrate these ideas, we consider a set of 100 observations which were simulated from a lognormal distribution with mean 1 and variance 3. These observations gave $\bar{X} = -0.6889$ and $S = 142.36$. Table 7 below shows the first three moments of the three individual claim amount distributions. Case (i) is the fitted lognormal distribution, case (ii) is the (approximating) predictive lognormal distribution and case (iii) is the truncated predictive distribution. For case (iii) the moments are actually the moments of the discretised distribution used to calculate the aggregate claims distribution. For this case only, the individual claim amount distribution was discretised using the method of crude rounding. (See, for example, Panjer and Willmot (1992).) We assumed that $\omega = 300$. Under the true distribution of $Y^*|D$ the probability of observing a claim in excess of 300 is less than 10^{-6}.

Table 7

Case	1st moment	2nd moment	3rd moment
(i)	1.0232	4.3469	76.6781
(ii)	1.0537	4.8884	99.8625
(iii)	1.0598	5.3427	135.6334

As expected, the moments of both the approximate and the truncated predictive distributions exceed those of the fitted distribution. Table 8 shows moments of the aggregate claims distributions. For case (i) we have used a fitted Poisson(100) distribution for the claim number distribution, whereas for cases (ii) and (iii) we used a predictive $NB(100, 0.5)$ distribution.

Table 8

Case	Mean	Variance	Skewness
(i)	102.32	434.69	0.8461
(ii)	105.37	599.86	0.8008
(iii)	105.98	646.59	0.9427

Whilst the means of the three distributions are relatively close there is a considerable increase in variance and third central moment going from case (i) through to case (iii). Strangely the coefficient of skewness is smaller in case (ii) than in case (i).

Finally, to get an idea of how appropriate these approaches are, let us consider percentiles of the aggregate claims distributions. Table 9 shows percentiles for four aggregate claims distributions. Cases (i) to (iii) represent the situations covered in Table 8. Case (iv) represents the true predictive distribution, using the true distribution of $Y^*|D$ and a $NB(100, 0.5)$ counting distribution.

Table 9

Case	$C_{0.90}$	$C_{0.95}$	$C_{0.99}$	$C_{0.995}$
(i)	129.10	139.10	161.70	171.90
(ii)	136.95	148.60	174.50	186.00
(iii)	138.35	150.75	179.50	193.20
(iv)	138.35	150.75	179.50	193.20

The ordering of values in Table 9 is what we would expect from Table 8. Two features stand out. First, there is quite a difference between values in cases (ii) and (iv). Second, case (iii) proves to be a very good approximation to case (iv), in terms of percentiles at least.

Each of the above two approaches has its advantages and disadvantages. The first approach has the advantage that the predictive individual claim amount distribution is easy to deal with, particularly if we wish to calculate moments. It also has greater variability than the fitted lognormal distribution. However, both the moments and percentiles are smaller than in case (iii). The second approach has the advantage that its percentiles provide a good match for those of the true predictive distribution. (It is largely a feature of our discretisation interval that there is exact correspondence in the figures given in Table 9.) The disadvantage is the subjective element introduced by ω. Although there are major disadvantages to fitting parameters to distributions by matching percentiles, it does seem like a possible way of determining a suitable value for ω. However, our experience has been that the moments of the predictive aggregate claims distribution are more sensitive to the value of ω than the percentiles are. Nevertheless, we would suggest that truncated predictive distribution provides a good solution to the problem of predictive distributions whose moments do not exist.

7. Conclusions

The main conclusions to be drawn from the examples in Section 4 is that parameter uncertainty has a major impact on moments and percentiles of aggregate claims distributions. In particular, parameter uncertainty in the claim number distribution seems to be of more importance than in the individual claim amount distribution when the moments of the predictive individual claim amount distribution exist.

The objective Bayesian approach may lead to predictive distributions for which moments do no exist. However, we have shown in Section 6 that given a Bayesian predictive distribution we can modify this distribution in such a way that it is suitable for insurance purposes.

Acknowledgment
The authors gratefully acknowledge financial support for this project from an Institute of Actuaries of Australia Research Grant and a Faculty of Economics and Commerce Research Grant (University of Melbourne). We are also grateful to Leanna Tedesco for assistance in this work.

8. References

Bowers, N.J., Gerber, H.U., Hickman, J.C. Jones, D.A. and Nesbitt, C.J. (1986) *Actuarial Mathematics.* Society of Actuaries, Itasca, IL.

Cairns, A.J.G. (1995) Uncertainty in the modelling process. *Transactions of the XXV International Congress of Actuaries Vol. 1, 67-94.*

De Vylder, F. and Goovaerts, M.J. (1988) Recursive calculation of finite-time ruin probabilities. *Insurance: Mathematics & Economics 7, 1-8.*

Hesselager, O. (1993) A class of conjugate priors with applications to excess of loss reinsurance. *ASTIN Bulletin 23, 77-95.*

Hürlimann, W. (1993) Predictive stop-loss premiums. *ASTIN Bulletin 23, 55-76.*

Hürlimann, W. (1995) Predictive stop-loss premiums and Student's t-distribution. *Insurance: Mathematics & Economics 16, 151-159.*

Klugman, S.A. (1992) *Bayesian statistics in actuarial science.* Kluwer Academic Publishers, Norwell, MA..

Lee, P.M. (1989) *Bayesian statistics: an introduction.* Oxford University Press, New York.

Panjer, H.H. (1981) Recursive evaluation of a family of compound distributions. *ASTIN Bulletin 12, 21-26.*

Schroter, K.J. (1991) On a family of counting distributions and recursions for related compound distributions. *Scandinavian Actuarial Journal, 161-175.*

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

Panjer, H.H. and Willmot, G.E. (1992) *Insurance risk models.* Society of Actuaries, Schaumberg, IL.

Appendix 1

A random variable X is said to have a t-distribution with parameters ν, μ and p, denoted $t(\nu, \mu, p)$, if the density of X is

$$f(x) = \frac{\Gamma(\frac{1}{2}(\upsilon+1))\sqrt{p}}{\sqrt{\pi\upsilon}\Gamma(\frac{\upsilon}{2})}\left(1+\frac{p(x-\mu)^2}{\nu}\right)^{-(\nu+1)/2}$$

for $-\infty < x < \infty$, where $-\infty < \mu < \infty$, $\nu > 0$ and $p > 0$.

The quantity $(X-\mu)\sqrt{p}$ has a Student t-distribution with ν degrees of freedom denoted $t(\nu, 0, 1)$ or t_ν. It is well known that if $Y \sim t_\nu$ then $E(Y) = 0$ and $V(Y) = \nu/(\nu-2)$ provided that $\nu > 2$. It therefore follows that $E(X) = \mu$ and $V(X) = \nu/(\nu-2)p$ provided that $\nu > 2$.

Appendix 2

Suppose $X_1, \ldots, X_n|\mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$ and let $\tau = 1/\sigma^2$ and $S = \sum_{i=1}^n (X_i - \bar{X})^2$. We have

$$\mu|\tau, D \sim N(\bar{X}, \frac{\sigma^2}{n})$$

$$\tau|D \sim G\left(\frac{\nu}{2}, \frac{S}{2}\right)$$

where $\nu = n-1$ and

$$\mu|D \sim t(\nu, \bar{X}, n(n-1)/S)$$

Suppose X^* is the next observation from the $N(\mu, \sigma^2)$ distribution. Then the predictive density for X^* is given by

$$\begin{aligned}
f(x^*|D) &= \int_0^\infty \int_{-\infty}^\infty f(x^*|\mu,\tau,D) f(\mu|\tau,D) f(\tau|D) d\mu d\tau \\
&\alpha \int_0^\infty \int_{-\infty}^\infty \tau^{\frac{1}{2}} \exp(-\frac{\tau}{2}(x^*-\mu)^2).\tau^{\frac{1}{2}} \exp(-\frac{\tau n}{2}(\bar{x}-\mu)^2).\tau^{\frac{n-3}{2}} \exp(-\frac{\tau S}{2})\, d\mu d\tau \\
&= \int_0^\infty \tau^{\frac{n-1}{2}} \exp(-\frac{\tau S}{2}) \left(\int_{-\infty}^\infty \exp(-\frac{\tau}{2}(x^*-\mu)^2 - \frac{\tau n}{2}(\bar{x}-\mu)^2)\right) d\mu d\tau
\end{aligned}$$

Consider the exponent in the inner integral. We have

$$\begin{aligned}
& -\frac{\tau}{2}((x^*-\mu)^2 + n(\bar{x}-\mu)^2) \\
&= -\frac{\tau}{2}(x^{*2} - 2x^*\mu + \mu^2 + n\bar{x}^2 - 2n\bar{x}\mu + n\mu^2) \\
&= -\frac{\tau}{2}((1+n)\mu^2 - 2\mu(x^* + n\bar{x}) + x^{*2} + n\bar{x}^2) \\
&= -\frac{\tau(1+n)}{2}\left((\mu - (\frac{x^*+n\bar{x}}{1+n}))^2 + \frac{x^{*2}+n\bar{x}^2}{1+n} - (\frac{x^*+n\bar{x}}{1+n})^2\right)
\end{aligned}$$

Therefore, the inner integral becomes

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \exp(-\frac{\tau}{2}(x^* - \mu)^2 - \frac{\tau n}{2}(\bar{x} - \mu)^2)\, d\mu \\
\alpha \quad & \tau^{-\frac{1}{2}} \exp(-\frac{\tau(1+n)}{2}\left(\frac{x^{*2} + n\bar{x}^2}{1+n} - (\frac{x^* + n\bar{x}}{1+n})^2\right)) \\
= \quad & \tau^{-\frac{1}{2}} \exp\left(-\frac{\tau}{2(1+n)}((1+n)(x^{*2} + n\bar{x}^2) - x^{*2} - 2n\bar{x}x^* - n^2\bar{x}^2)\right) \\
= \quad & \tau^{-\frac{1}{2}} \exp\left(-\frac{\tau}{2(1+n)}(n\bar{x}^2 + nx^{*2} - 2n\bar{x}x^*)\right) \\
= \quad & \tau^{-\frac{1}{2}} \exp\left(-\frac{\tau n}{2(1+n)}(x^* - \bar{x})^2\right)
\end{aligned}
$$

Hence we have

$$
\begin{aligned}
f(x^*|D) \quad \alpha \quad & \int_0^{\infty} \tau^{\frac{n}{2}-1} \exp\left(-\frac{\tau}{2}(S + \frac{n(x^* - \bar{x})^2}{1+n}\right) d\tau \\
\alpha \quad & \left(S + \frac{n(x^* - \bar{x})^2}{1+n}\right)^{-\frac{n}{2}} \\
\alpha \quad & \left(1 + \frac{n(x^* - \bar{x})^2}{(1+n)S}\right)^{-\frac{n}{2}} \\
= \quad & \left(1 + \frac{(n-1)n(x^* - \bar{x})^2}{(1+n)(n-1)S}\right)^{-(\frac{n-1+1}{2})}
\end{aligned}
$$

and so

$$
X^*|D \sim t\left(n-1, \bar{X}, \frac{(n-1)n}{(n+1)S}\right)
$$

Figure 1: Fitted and predictive aggregate claims distributions

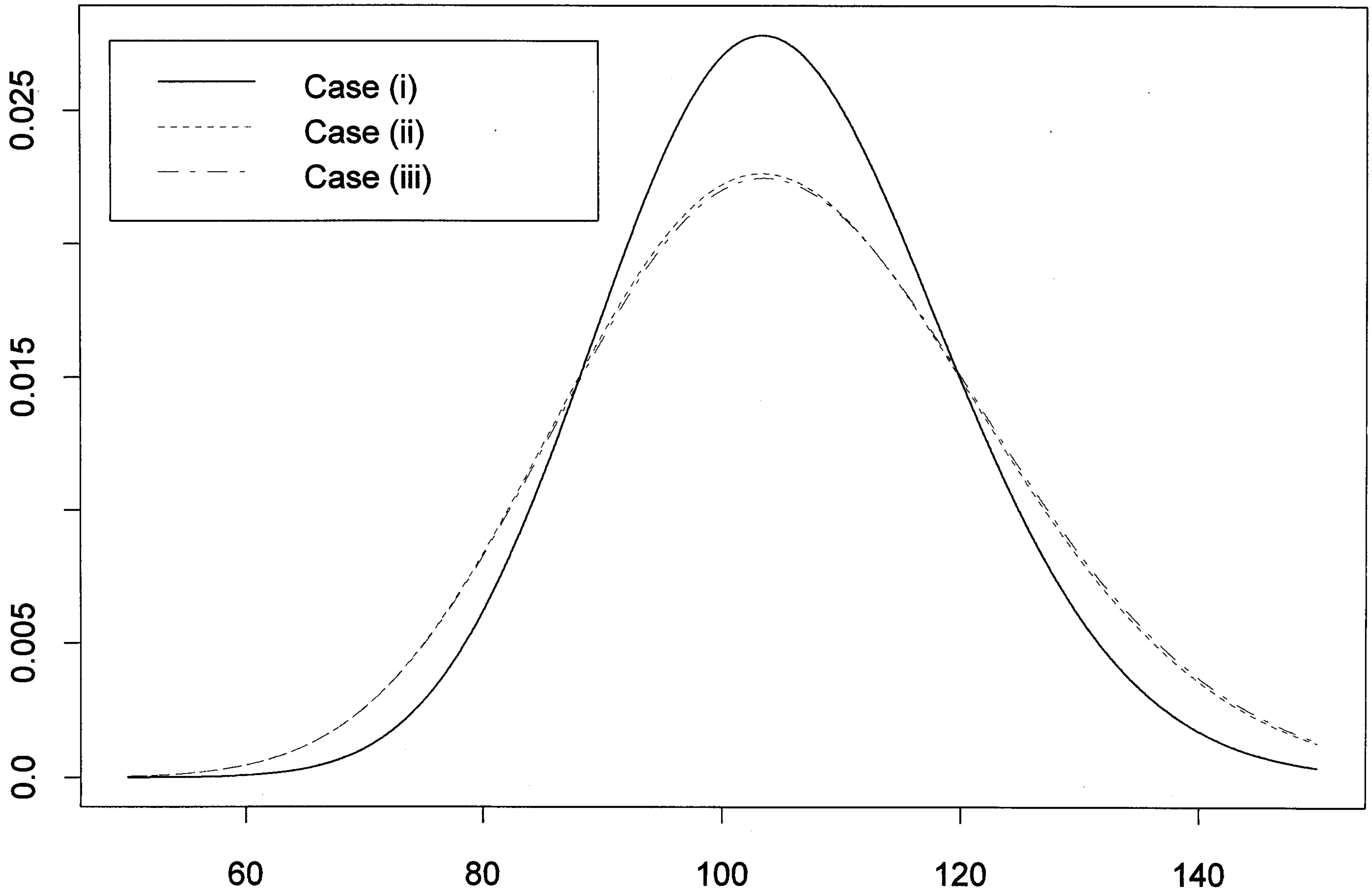


Figure 2: Reinsurer's expected aggregate claim amount

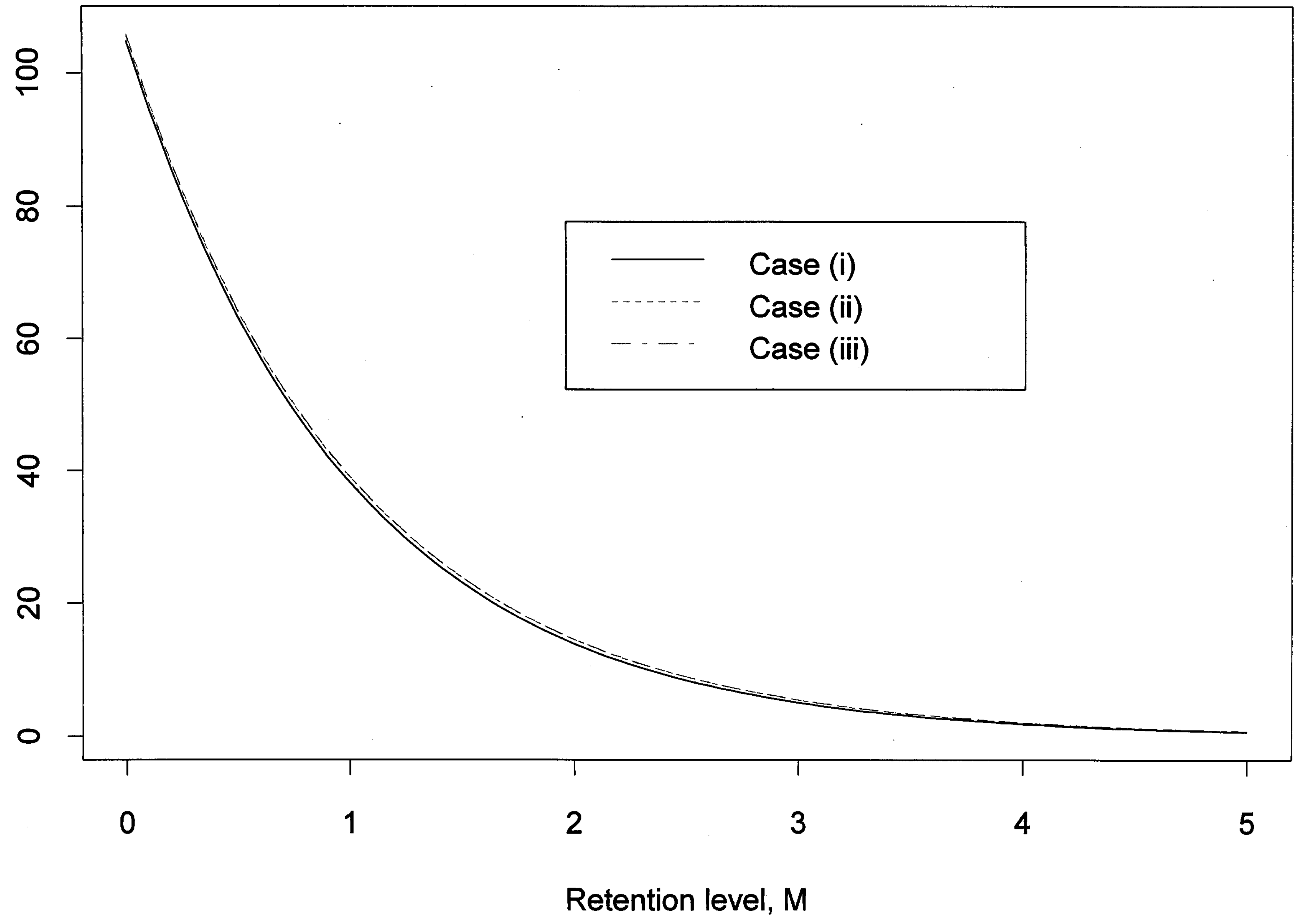


Figure 3: Variance of reinsurer's aggregate claim amount

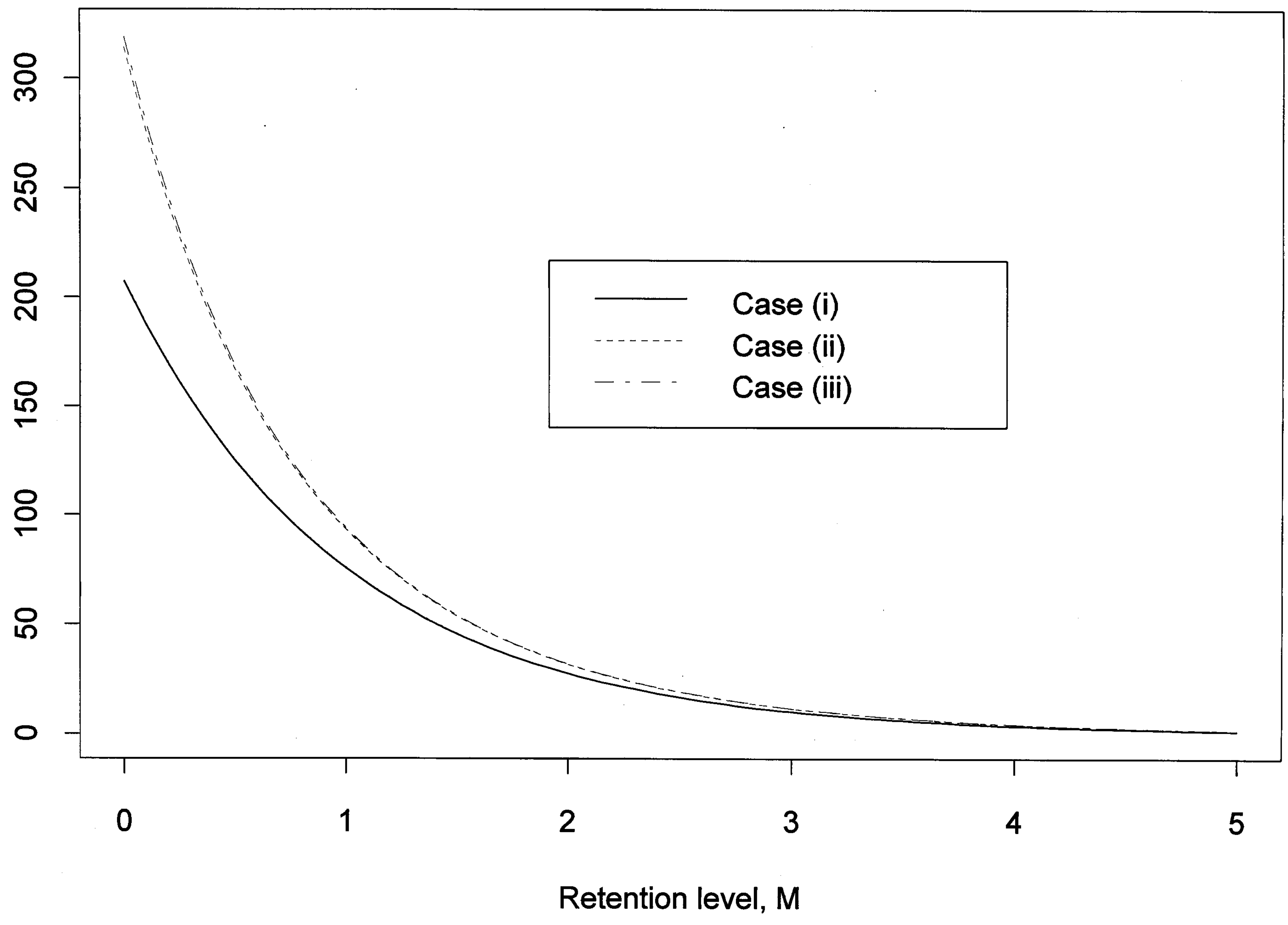
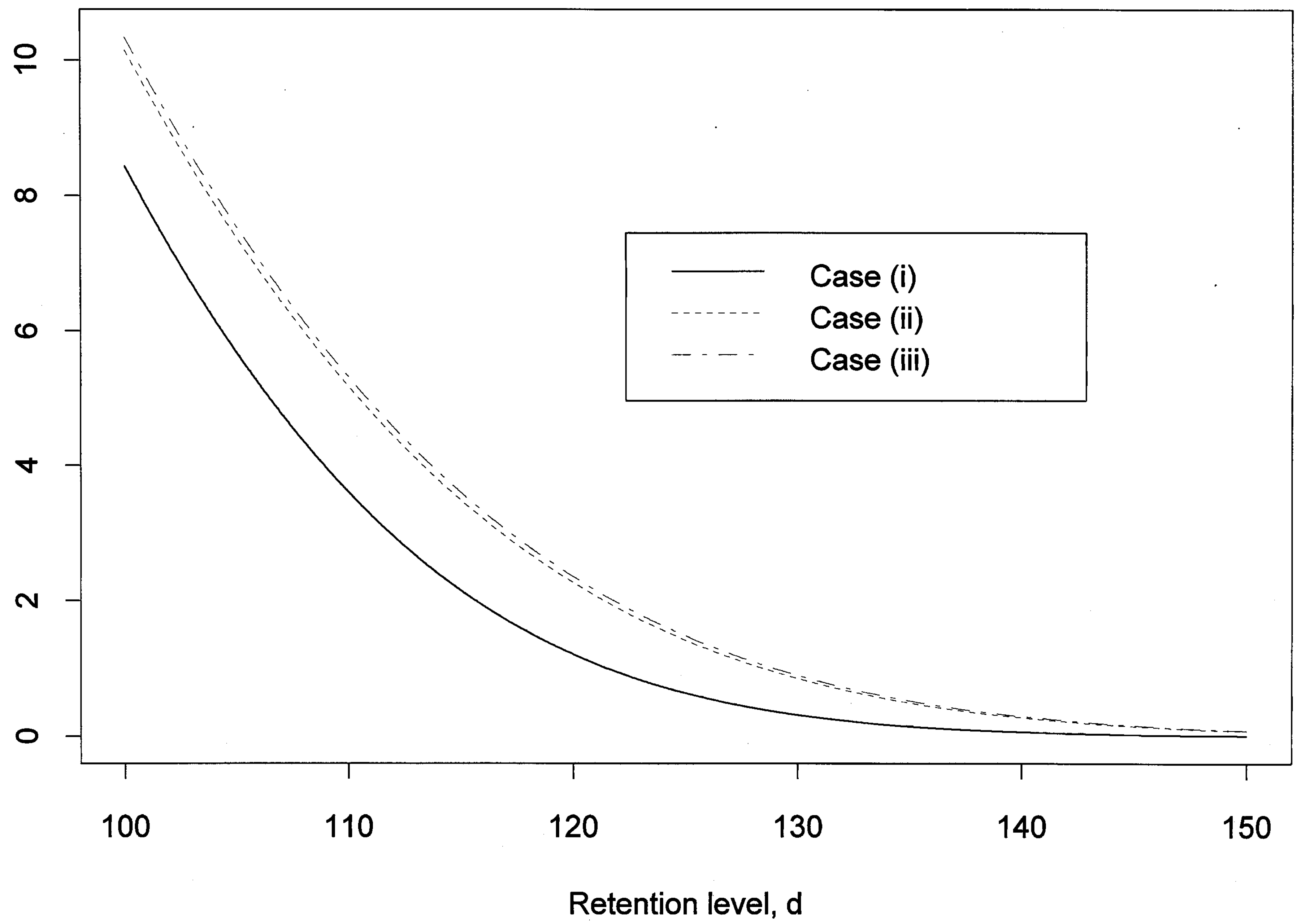


Figure 4: Pure stop loss premiums



RESEARCH PAPER SERIES

No.	Date	Subject	Author
1	MAR 93	AUSTRALIAN SUPERANNUATION : THE FACTS, THE FICTION, THE FUTURE	David M Knox
2	APR 93	AN EXPONENTIAL BOUND FOR RUIN PROBABILITIES	David C M Dickson
3	APR 93	SOME COMMENTS ON THE COMPOUND BINOMIAL MODEL	David C M Dickson
4	AUG 93	RUIN PROBLEMS AND DUAL EVENTS	David CM Dickson Alfredo D Egidio dos Reis
5	SEP 93	CONTEMPORARY ISSUES IN AUSTRALIAN SUPERANNUATION - A CONFERENCE SUMMARY	David M Knox John Piggott
6	SEP 93	AN ANALYSIS OF THE EQUITY INVESTMENTS OF AUSTRALIAN SUPERANNUATION FUNDS	David M Knox
7	OCT 93	A CRITIQUE OF DEFINED CONTRIBUTION USING A SIMULATION APPROACH	David M Knox
8	JAN 94	REINSURANCE AND RUIN	David C M Dickson Howard R Waters
9	MAR 94	LIFETIME INCOME, TAXATION, EXPENDITURE AND SUPERANNUATION (LITES): A LIFE-CYCLE SIMULATION MODEL	Margaret E Atkinson John Creedy David M Knox
10	FEB 94	SUPERANNUATION FUNDS AND THE PROVISION OF DEVELOPMENT/VENTURE CAPITAL: THE PERFECT MATCH? YES OR NO	David M Knox
11	JUNE 94	RUIN PROBLEMS: SIMULATION OR CALCULATION?	David C M Dickson Howard R Waters
12	JUNE 94	THE RELATIONSHIP BETWEEN THE AGE PENSION AND SUPERANNUATION BENEFITS, PARTICULARLY FOR WOMEN	David M Knox
13	JUNE 94	THE COST AND EQUITY IMPLICATIONS OF THE INSTITUTE OF ACTUARIES OF AUSTRALIA PROPOSED RETIREMENT INCOMES STRATEGY	Margaret E Atkinson John Creedy David M Knox Chris Haberecht
14	SEPT 94	PROBLEMS AND PROSPECTS FOR THE LIFE INSURANCE AND PENSIONS SECTOR IN INDONESIA	Catherine Prime David M Knox

15	OCT 94	PRESENT PROBLEMS AND PROSPECTIVE PRESSURES IN AUSTRALIA'S SUPERANNUATION SYSTEM	David M Knox
16	DEC 94	PLANNING RETIREMENT INCOME IN AUSTRALIA: ROUTES THROUGH THE MAZE	Margaret E Atkinson John Creedy David M Knox
17	JAN 95	ON THE DISTRIBUTION OF THE DURATION OF NEGATIVE SURPLUS	David C M Dickson Alfredo D Egidio dos Reis
18	FEB 95	OUTSTANDING CLAIM LIABILITIES: ARE THEY PREDICTABLE?	Ben Zehnwirth
19	MAY 95	SOME STABLE ALGORITHMS IN RUIN THEORY AND THEIR APPLICATIONS	David C M Dickson Alfredo D Egidio dos Reis Howard R Waters
20	JUN 95	SOME FINANCIAL CONSEQUENCES OF THE SIZE OF AUSTRALIA'S SUPERANNUATION INDUSTRY IN THE NEXT THREE DECADES	David M Knox
21	JUN 95	MODELLING OPTIMAL RETIREMENT IN DECISIONS IN AUSTRALIA	Margaret E Atkinson John Creedy
22	JUN 95	AN EQUITY ANALYSIS OF SOME RADICAL SUGGESTIONS FOR AUSTRALIA'S RETIREMENT INCOME SYSTEM	Margaret E Atkinson John Creedy David M Knox
23	SEP 95	EARLY RETIREMENT AND THE OPTIMAL RETIREMENT AGE	Angela Ryan
24	OCT 95	APPROXIMATE CALCULATION OF MOMENTS OF RUIN RELATED DISTRIBUTIONS	David C M Dickson
25	DEC 95	CONTEMPORARY ISSUES IN THE ONGOING REFORM OF THE AUSTRALIAN RETIREMENT INCOME SYSTEM	David M Knox
26	FEB 96	THE CHOICE OF EARLY RETIREMENT AGE AND THE AUSTRALIAN SUPERANNUATION SYSTEM	Margaret E Atkinson John Creedy
27	FEB 96	PREDICTIVE AGGREGATE CLAIMS DISTRIBUTIONS	David C M Dickson Ben Zehnwirth
28	FEB 96	THE AUSTRALIAN GOVERNMENT SUPERANNUATION CO-CONTRIBUTIONS: ANALYSIS AND COMPARISON	Margaret E Atkinson
29	MAR 96	A SURVEY OF VALUATION ASSUMPTIONS AND FUNDING METHODS USED BY AUSTRALIAN ACTUARIES IN DEFINED BENEFIT SUPERANNUATION FUND VALUATIONS	Des Welch Shauna Ferris
30	MAR 96	THE EFFECT OF INTEREST ON NEGATIVE SURPLUS	David C M Dickson Alfred D Egídio dos Reis

31	MAR 96	RESERVING CONSECUTIVE LAYERS OF INWARDS EXCESS-OF-LOSS REINSURANCE	Greg Taylor
32	AUG 96	EFFECTIVE AND ETHICAL INSTITUTIONAL INVESTMENT	Anthony Asher
33	AUG 96	STOCHASTIC INVESTMENT MODELS: UNIT ROOTS, COINTEGRATION, STATE SPACE AND GARCH MODELS FOR AUSTRALIA	Michael Sherris Leanna Tedesco Ben Zehnwirth
34	AUG 96	THREE POWERFUL DIAGNOSTIC MODELS FOR LOSS RESERVING	Ben Zehnwirth
35	SEPT 96	KALMAN FILTERS WITH APPLICATIONS TO LOSS RESERVING	Ben Zehnwirth
36	OCT 96	RELATIVE REINSURANCE RETENTION LEVELS	David C M Dickson Howard R Waters
37	OCT 96	SMOOTHNESS CRITERIA FOR MULTI-DIMENSIONAL WHITTAKER GRADUATION	Greg Taylor
38	OCT 96	GEOGRAPHIC PREMIUM RATING BY WHITTAKER SPATIAL SMOOTHING	Greg Taylor
39	OCT 96	RISK, CAPITAL AND PROFIT IN INSURANCE	Greg Taylor
40	OCT 96	SETTING A BONUS-MALUS SCALE IN THE PRESENCE OF OTHER RATING FACTORS	Greg Taylor
41	NOV 96	CALCULATIONS AND DIAGNOSTICS FOR LINK RATION TECHNIQUES	Ben Zehnwirth Glen Barnett
42	DEC 96	VIDEO CONFERENCING IN ACTUARIAL STUDIES - A THREE YEAR CASE STUDY	David M Knox
43	DEC 96	ALTERNATIVE RETIREMENT INCOME ARRANGEMENTS AND LIFETIME INCOME INEQUALITY: LESSONS FROM AUSTRALIA	Margaret E Atkinson John Creedy David M Knox
44	JAN 97	AN ANALYSIS OF PENSIONER MORTALITY BY PRE-RETIREMENT INCOME	David M Knox Andrew Tomlin