

A Discussion of Parameter and Model Uncertainty in Insurance

Andrew J.G. Cairns
Department of Actuarial Mathematics and Statistics,
Heriot-Watt University,
Edinburgh, EH14 4AS, United Kingdom

E-mail: A.Cairns@ma.hw.ac.uk
WWW: <http://www.ma.hw.ac.uk/~andrewc/>

Abstract

In this paper we consider the process of modelling uncertainty. In particular we are concerned with making inferences about some quantity of interest which, at present, has been unobserved. Examples of such a quantity include the probability of ruin of a surplus process, the accumulation of an investment, the level or surplus or deficit in a pension fund and the future volume of new business in an insurance company.

Uncertainty in this quantity of interest, y , arises from three sources:

- uncertainty due to the stochastic nature of a given model;
- uncertainty in the values of the parameters in a given model;
- uncertainty in the model underlying what we are able to observe and determining the quantity of interest.

It is common in actuarial science to find that the first source of uncertainty is the only one which receives rigorous attention. A limited amount of research in recent years has considered the effect of parameter uncertainty, while there is still considerable scope for development of methods which deal in a balanced way with model risk.

Here we discuss a methodology which allows all three sources of uncertainty to be assessed in a more coherent fashion.

Keywords: parameter uncertainty, model risk, model selection criteria, Bayesian statistics, stochastic interest, ruin theory.

1 Parameter and Model Uncertainty

1.1 Measuring uncertainty

Suppose that we are interested in some quantity, y , related to a given random process governing investment returns. For example, y may be the accumulation of a single premium over a 25 year period, or the probability that the assets of a closed pension fund will be insufficient to meet the guaranteed liabilities. Alternatively, in Risk Theory y might represent aggregate claims over the next year, the adjustment coefficient or the probability of ruin. We need to make inferences about the value of y and this will be based primarily on a relevant set of past data, x , but also with the use of some subjective judgement. Subjective judgement arises in particular when a model or range of models is selected but can also arise if constraints or weights are placed on specific parameters (for example, a requirement that the mean rate of inflation is positive).

Uncertainty in this quantity of interest, y , arises from three principal sources (for example, see Draper, 1995, and Bernardo and Smith, 1994, Chapter 6):

- uncertainty due to the stochastic nature of a given model (that is, a stochastic model produces randomness in its output);
- uncertainty in the values of the parameters in a given model (if we have a finite set of data then we cannot estimate parameter values exactly);
- uncertainty in the model underlying what we are able to observe and which determines the quantity of interest.

A fourth source of uncertainty concerns the possibility of measurement errors in the data. As a general rule, this is not a significant issue in actuarial problems: the financial nature of most contracts means that relevant quantities should be known accurately. We will not pursue this point further.

Within the latter, three principal forms of model uncertainty can arise:

- The true model is assumed to be a fixed but unknown member of a general class of models which will be fitted to the data (if, for example, investment returns were known to be generated by an $ARMA(p, q)$ time series model).
- The class of models under consideration consists of models which are known to be approximations to a more complex reality in which the modeller has a firm basis of belief (for example, by using a Normal approximation to what is known to be a more complex distribution).

- The class of models consists of a range of models which may provide a proxy for a more complex reality about which the modeller has little or no prior knowledge.

Most actuarial problems fall into the third category.

For the sake of coherence between the three methods described in Section 1.3 we will concentrate on the use of the Bayesian paradigm. Equivalent approaches exist using classical methods but it is not the intention of this paper to discuss the merits of one approach or the other. Rather the intention of the paper is to focus attention on parameter and model uncertainty. With this in mind we concentrate on relatively general techniques rather than give a detailed and accurate solution to a very specific problem. We will, therefore, only consider two relatively simple examples.

It is common in actuarial work to concentrate solely on randomness arising from the structure of a given model with the model and parameter values taken as given. In some cases (for example, Daykin, Pentikäinen & Pesonen, 1994) model and parameter risk are acknowledged but then ignored. Studies which acknowledge and account for parameter uncertainty have begun to emerge in recent years. Most of these deal with general insurance rather than life or investment problems (for example: Klugman, 1992; McNeil, 1997; Pai, 1997; Dickson, Tedesco & Zenwirth, 1998; Scollnik, 1998). On the investment front, Harris (1999) uses Markov chain Monte Carlo methods to select a model for price inflation and to investigate parameter uncertainty. Cairns (in discussion of Parker, 1997) raises the issue of parameter and model risk in a pensions or life insurance context while Crouhy, Galai & Mark (1998) discuss in general terms the effect of model risk on derivative pricing. Other related work in investment can be found in the collection edited by Bawa, Brown and Klein (1979). and in other papers following on from those presented in that volume.

In the present context the term *model risk* is often used. Here we will reserve this for circumstances where the results and decisions emerging from an analysis are sensitive to the choice of model. As we will see later in this paper, there are examples where the results of an analysis do not depend significantly upon the choice of model: that is model risk is not an issue.

1.2 The principle of parsimony

A useful, early discussion of model uncertainty and model selection can be found in Jeffreys (1961). There he describes the *simplicity postulate* (often known also as the *principle of parsimony*). It is well known that people prefer simple models or explanations to complex ones. Jeffreys states “*that the simplest law is chosen because it is most likely to give correct predictions*”. Here the simplest law refers to the striking of a balance

between consistency of a model with the observed data and the prior degree of belief in a model. Jeffreys' view was that that the prior degree of belief in a model should be inversely related to the number of parameters. This is consistent with the various model selection criteria discussed in Sections 2 and 3 of this paper.

Later in his book Jeffreys describes how the point to stop adding parameters is reached when a modeller has a high degree of belief that the residual errors are random rather than containing any further structure. This point of view is consistent with the first form of model uncertainty (that the true model is one member out of a defined class of models). With the second or third forms of model uncertainty, we accept that the chosen model is only an approximation to reality and that there is additional structure in the data. However, with only a finite amount of data available we must choose a model which we believe gives an adequate description of the past and which will provide an adequate guide to the future *for the task in hand*.

1.3 Theoretical approaches to the assessment of uncertainty

In making inferences about y we can take one of three approaches:

Method 1: This approach finds the model, M^* , which provides the best fit according to some criterion (Bayesian or otherwise) along with the best estimate of its parameter set, θ^* . We then make inferences about y on the assumption that M^* and θ^* are the *true* model and parameter set respectively. The model selection criterion can range from “best fit” (leading to the choice of an overparametrized model) to ones which penalise models with too many parameters.

Method 2: Here also, we find the model, M^* , with the highest posterior probability (the best estimate model) and derive the posterior density, $f(\theta|x)$, for this model's parameter set. On the basis that the model is correct and that θ follows this distribution we can then derive the posterior distribution for y given the past data.

Method 3: This is a full Bayesian approach in which we estimate both the posterior model probabilities and parameter densities. The posterior distribution of y given the past data is then calculated by recognising that both the model and the parameter set are uncertain and are characterized by their respective posterior distributions.

The three methods described here are given in greater detail in Draper (1995) and Bernardo and Smith (1994) in a general context, in an actuarial context in Klugman (1992) (parameter uncertainty) and in an economic context by Min and Zellner (1993). The statistical literature contains a large number of papers which deal with methods 2 and 3. The references given here themselves contain pointers to many works which could be adapted to help solve actuarial problems.

Non-Bayesian methods for combining forecasts of alternative models are reviewed in Clemen (1989). Such methods (while they work effectively) tend to lack the internal consistency of Bayesian methods and we do not pursue them further here.

It is clear from these descriptions that Method 1 deals only with the uncertainty which arises as a result of the stochastic nature of a model. Method 2, adds in the effect of parametric uncertainty and will normally increase the level of uncertainty in y . Method 3 adds to this by incorporating model uncertainty and this can increase or decrease the level of uncertainty in y : this depends on the relationships between different but almost equally plausible models.

In this paper we will use a Bayesian framework to deal with Methods 2 and 3. Method 2 can also be treated within the framework of classical statistics. From a pure, classical viewpoint, the results from using different models can be compared but not combined in any meaningful way (for example, see Cox, in discussion of Draper, 1995). Here Method 3 describes the model averaging approach described by Draper (1995) and Chatfield (1995).

Within a Bayesian context, results will not normally be overly sensitive to the choice of an uninformative prior distribution for the parameters. However, this may not be the case with respect to the prior model probabilities. In particular, when two models give similar fits but make significantly different predictions about the future a change in the prior model probabilities can have a significant impact on the posterior distribution of the quantity y of interest. The posterior distribution for y can be used in the process of minimising expected loss to obtain a point estimate \hat{y} . The above remarks suggest that variation of the prior model probabilities can lead to significantly different estimates \hat{y} . In the case of Method 3, therefore, it is not that the Bayesian approach to model uncertainty should be treated with caution (for example, see Cox, in discussion of Draper, 1995). This does not mean that model uncertainty can be disregarded. On the contrary, a good analysis of uncertainty should investigate the sensitivity of results to the choice of model: it may just be the case that it is not possible to combine the results from different models into a single statement because we are unwilling to prescribe specific prior model probabilities. A good example of an analysis of the sensitivity of the results to the choice of model can be found in Schmock (1999).

2 The Basic Approach

Let $M = \{M_1, M_2, \dots, M_N\}$ be the class of models under consideration: that is M_i represents model i , $i = 1, 2, \dots, N$, $2 \leq N \leq \infty$.

Let $p(x|M_i, \theta_i)$ be the likelihood function for model M_i , given parameters θ_i , and the set

of past data, x , and let $p(\theta_i|M_i)$ be the prior distribution for θ_i given model M_i .

The posterior distribution for θ_i given M_i and x is then $p(\theta_i|M_i, x) = c_i p(x|M_i, \theta_i) p(\theta_i|M_i)$ where c_i is some normalising constant.

Let us consider first the choice of model. In considering models with different numbers of parameters it is clear that it is always possible to improve the fit of a model by adding in further parameters (unless we already have a perfect fit). However, we may regard an improvement in the fit of a model to be so marginal that it is not felt to be worth changing to. For an increase in the number of model parameters to be acceptable we need to be satisfied that it has identified additional structure in the available data. We therefore need to quantify by how much the quality of fit should be improved by the introduction of a further parameter. This is done by modifying the likelihood function by adding a term which makes explicit allowance for differences in the number of parameters between models. In particular, such an adjustment should favour models with fewer parameters. For example, the Akaike Information Criterion (AIC) may be employed (for example, see Wei, 1990). Hence we choose the model which *maximises*:

$$\log \tilde{p}(x|M_i, \theta_i) = \log p(x|M_i, \theta_i) - A(k_i)$$

where k_i is the dimension of θ_i , and $A(k_i) = k_i$ is an increasing function of k_i . (Note that this may sometimes be specified as *minimise* $-2 \log p(x|M_i, \theta_i) + 2k_i$. For example, this is the version used in the SPlus computer package.) An alternative to the AIC is the Schwarz Bayes Criterion (SBC) in which $A(k_i) = \frac{1}{2}k_i \log n$ where n is the size of the sample, x . For all but small sample sizes the SBC introduces a stiffer penalty on the introduction of additional parameters. In the context of time series, in fact, the AIC has been shown to overestimate the order of the model underlying the data (Shibata, 1976). This has not been found to be the case when the SBC is used. As will be noted in Section 3, the SBC gives an asymptotic approximation to the posterior model odds (if the prior model probabilities are equal) although this approximation can be relatively poor with limited sample sizes. It follows that selection of a *single* model is equivalent to minimising the expected loss with a zero-one loss function on the model set, \mathcal{M} (assuming that the true model is a member of cm).

Once a model $M^* = M_{i^*}$ has been selected, we maximise the posterior density $p(\theta_{i^*}|M_{i^*}, x)$ over θ_{i^*} (or, for example, take the posterior mean $E[\theta_{i^*}|M_{i^*}, x]$) giving a best estimate of $\hat{\theta}_{i^*}$. Where the quantity of data, x , is large and the prior $p(\theta_{i^*}|M_{i^*})$ diffuse, the posterior maximum can reasonably be replaced by the maximum likelihood estimate. The analysis then proceeds as if M_{i^*} and θ_{i^*} are the true model and parameter set.

3 Incorporating Parameter and Model Uncertainty

The underlying objective in this paper is to provide a statement about the value of some quantity of interest y . In the basic approach we fixed the model and its parameter set and it is straightforward (at least in principal) to make probability statements about such a quantity. When we admit the possibility of parameter and model uncertainty the Bayesian paradigm gives us a coherent framework within which we can continue to make probability statements about y .

Here we aim to derive a posterior distribution for both M_i and $\theta_i|M_i$.

Let $Pr(M_i)$ be the prior probability for model i and $p(\theta_i|M_i)$ be the prior distribution for θ_i given model i .

Treating the pair (θ_i, M_i) as the parameter set in the simple Bayesian framework we have

$$\begin{aligned} p(\theta_i, M_i|x) &= p(\theta_i|M_i, x)Pr(M_i|x) \\ \text{where } Pr(M_i|x) &= c.Pr(M_i).p(x|M_i) \\ p(x|M_i) &= \int p(x|\theta_i, M_i)p(\theta_i|M_i)d\theta_i \end{aligned}$$

and c is some normalising constant. Thus, the joint posterior density for (θ_i, M_i) is equal to the product of the posterior density for θ_i , given model M_i is correct, times the posterior probability of model M_i . Note that it is incorrect to write

$$p(\theta_i, M_i|x) = cp(x|\theta_i, M_i)p(\theta_i|M_i)p(x|M_i)Pr(M_i)$$

but rather we should write

$$p(\theta_i, M_i|x) = c \left(c_i p(x|\theta_i, M_i)p(\theta_i|M_i) \right) p(x|M_i)Pr(M_i)$$

where the c_i are first calculated to normalise the $p(\theta_i|M_i, x)$, and c is then calculated in order to normalise the overall posterior distribution.

Suppose that we are ultimately interested in some quantity y which depends on the model and the parameter set. For example, y may be the adjustment coefficient, or the probability of ruin. Then

$$\begin{aligned} p(y|x) &= \sum_i p(y|x, M_i)Pr(M_i|x) \\ \text{where } p(y|x, M_i) &= \int p(y|x, M_i, \theta_i)p(\theta_i|M_i, x)d\theta_i \end{aligned}$$

The first of these follows the standard Bayes formula for the posterior distribution: that is,

$$p(\theta_i|M_i, x) = c_i p(x|M_i, \theta_i)p(\theta_i|M_i)$$

where c_i is a normalising constant specific to model i .

An asymptotic analysis of $p(\theta_i|M_i, x)$ (for example, see Bernardo and Smith, 1994) shows that

$$\log p(\theta_i|M_i, x) = c - \frac{1}{2}(\theta_i - \hat{\theta}_i)^T H_i(\theta_i - \hat{\theta}_i) + o(|\theta_i - \hat{\theta}_i|^2)$$

where $\hat{\theta}_i$ is a weighted mean of the maximum likelihood estimate for θ_i and the mode of the prior distribution for θ_i , and H_i is the sum of the hessian matrices at the respective maxima of the likelihood and prior density functions.

When the prior distribution is very diffuse and n is large we can make some appealing approximations. First, H_i can be well approximated by $n\hat{B}_i$, where $\hat{B}_i = B_i(\hat{\theta}_i)$ is the information matrix for a single observation given $\hat{\theta}_i$ is the true value for θ_i . Second, $\hat{\theta}_i$ can be well approximated by the maximum likelihood estimate for θ_i . (Those wishing to take a pragmatic approach can use this result as a justification for using maximum likelihood (for example, see Jeffreys, 1961, Chapter 4: *Approximate Methods and Simplifications*.)

Our treatment of the posterior distribution over the collection of models is rather more complex.

Again we start with the Bayes formula for the posterior distribution for M_i :

$$Pr(M_i|x) = cPr(M_i)p(x|M_i)$$

The components $p(x|M_1), p(x|M_2), \dots$ (up to some multiplicative constant) are called Bayes Factors. Calculation of the Bayes factors often requires complex numerical techniques. Kass and Raftery (1995) give a useful description of some computational procedures to evaluate these quantities.

Since the Bayes factors can be difficult to work with, Draper (1995) simplifies the expressions by applying a similar asymptotic argument to that above, finding that

$$\log p(x|M_i) = c + \frac{1}{2}k_i \log 2\pi - \frac{1}{2} \log |\hat{I}_i| + \log p(x|\hat{\theta}_i, M_i) + \log p(\hat{\theta}_i|M_i) + O(n^{-1})$$

where \hat{I}_i is the information matrix for the dataset x evaluated at $\hat{\theta}_i$ under model i , k_i is the dimension of θ_i and $p(\theta_i|M_i)$ is the prior distribution for θ_i .

Now since the observations are assumed to be independent, $\hat{I}_i = n\hat{B}_i$ where \hat{B}_i is the information matrix for a single observation given $\hat{\theta}_i, M_i$, so

$$\log |\hat{I}_i| = k_i \log n + \log |\hat{B}_i|$$

Note that as n increases, the second term will remain roughly constant for each model. If k_i is the same for each model then the $\log |\hat{B}_i|$ term will be the only term of relevance,

whereas if k_i varies between models then, as n increases, the $k_i \log n$ term becomes much more significant than the $\log |\hat{B}_i|$ term. Draper (1995) notes that the effect of the prior $p(\hat{\theta}_i|M_i)$ is negligible, and this is particularly so for a diffuse prior. We are therefore able to drop this term.

If we define \hat{l}_i to be equal to the maximum log-likelihood for model i , $\log p(x|\hat{\theta}_i, M_i)$, then we have

$$\log p(x|M_i) \approx \hat{l}_i - \frac{1}{2}k_i \log n - \frac{1}{2} \log |\hat{B}_i| + \frac{1}{2}k_i \log 2\pi + c$$

Hence

$$\log Pr(M_i|x) \approx \log Pr(M_i) + \hat{l}_i - \frac{1}{2}k_i \log n + \frac{1}{2}k_i \log 2\pi - \frac{1}{2} \log |\hat{B}_i| + c$$

where c is some normalising constant which ensures that $\sum_i Pr(M_i|x) = 1$.

Note that the expression for the model likelihood $\log p(x|M_i)$ is very similar to $\hat{l}_i - \frac{1}{2}k_i \log n$ which is the Schwarz Bayes Criterion (for example, see Schwarz, 1978, or Wei, 1990), the remaining terms $-\frac{1}{2} \log |\hat{B}_i| + \frac{1}{2}k_i \log 2\pi$ being of a lower order.

The analysis above gives us the framework for the application of Method 3 described in Section 2, and therefore accounts for model uncertainty. Method 2 follows a similar path, but, once the posterior model probabilities have been computed, we concentrate upon the model with the highest posterior probability, treating it as though it were the correct model. This is consistent with the use of a zero-one loss function (that is, a loss of 1 if the wrong model is chosen) which would result in the selection of the model with the highest posterior probability. However, this loss function only really makes sense when the true model is known to be a member of the class M of models under consideration. We will not, therefore, press this point too firmly. In any event it is perhaps more relevant to use loss functions which apply to the success or otherwise of the decisions taken as a result of the modelling process.

Method 2 combined with the use of the Schwarz Bayes Criterion rather than the posterior model probabilities described in this section has been used in an actuarial context by Klugman (1991). He considers how the predictive quantities from a mortality study and for pure premiums in casualty insurance can be affected by the acknowledgement of parameter uncertainty.

Harris (1999) considers an example in which the dataset is sufficiently small that the asymptotic approximations are relatively poor. This is done by comparing the results of applying Markov chain Monte Carlo techniques which converge to the exact posterior model probabilities with the results of applying similar formulae to the above.

4 Application to Stochastic Interest Models

In this section we consider a very simple stochastic interest rate problem: that of the accumulation of a single payment made at time 0 up to some time t under the action of random rates of interest. The payment will be invested in equities.

4.1 Theoretical development

Suppose that $\delta(t)$ represents the force of interest between times $t - 1$ and t : that is, an investment of 1 made at time 0 will accumulate to $e^{\delta(1)}$ at time 1, and $\exp(\delta(1) + \delta(2) + \dots + \delta(t))$ at time t . For convenience we define $y(t) = \sum_{s=1}^t \delta(s)$. It is therefore of interest to know what the distributions of $y(t)$ and $F(t) = \exp y(t)$ look like.

In making inferences about $y(t)$ and $F(t)$ we will consider two simple models for the process $\delta(t)$:

Model 1

$\delta(t)$ is an independent and identically distributed series with $\delta(t) \sim N(\mu, \sigma^2)$. The model parameters μ and σ^2 are not known and must be estimated from a set of n past observations $\delta = (\delta(-n+1), \delta(-n+2), \dots, \delta(0))$.

Under Method 1 of inference we take

$$\hat{\mu} = \frac{1}{n} \sum_{s=-n+1}^0 \delta(s)$$
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{s=-n+1}^0 (\delta(s) - \hat{\mu})^2$$

and these are assumed to be the true values in our analysis of uncertainty. Thus

$$y(t) | \delta \sim N(\hat{\mu}t, \hat{\sigma}^2 t)$$

Under Method 2 we retain the assumption of independence of returns but take into account parameter uncertainty. In the Bayesian analysis we use the Jeffreys prior distribution $f(\mu, \sigma^2) = \sigma^{-3}$ described by O'Hagan (1994, Chapter 5). It is often the case that we have no prior information about the magnitude of a particular parameter. This immediately suggests the use of a non-informative or diffuse prior. Jeffreys considered it appropriate

that the outcome of an analysis should not be sensitive to the particular way in which a model is parametrised (for example, here taking σ as a parameter or σ^2). Invariance of the results relative to such a transformation led to the introduction of the Jeffreys prior. (There exist alternative non-informative priors to that proposed by Jeffreys. Brief discussions of this can be found in Bernardo & Smith, 1994, Section 5.6, or Dawid, 1983. Jeffreys, himself, preferred to use $f(\mu, \sigma^2) = \sigma^{-2}$.) Using the Jeffreys prior we are able to derive the posterior distributions given $\delta = (\delta(-n+1), \dots, \delta(0))$:

$$\begin{aligned}\mu|\sigma^2, \delta &\sim N\left(\hat{\mu}, \frac{\sigma^2}{n}\right) \\ \frac{n\hat{\sigma}^2}{\sigma^2}|\delta &\sim \chi_n^2\end{aligned}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are as defined above under Method 1.

From this we can derive the posterior law of $y(t)$ given δ (for example, see O'Hagan, 1994, Chapters 2 and 9)

$$y(t)|\delta = \hat{\mu}t + \sqrt{\frac{(t^2 + nt)\hat{\sigma}^2}{n}}T_n$$

where T_n has a standard t-distribution with n degrees of freedom. Hence

$$\begin{aligned}E[y(t)|\delta] &= \hat{\mu}t \text{ (as above)} \\ \text{Var}[y(t)|\delta] &= \frac{t^2 + nt}{n-2}\hat{\sigma}^2\end{aligned}$$

(This is derived straightforwardly from the properties of the t-distribution.)

Although $F(t) = \exp y(t)$ the thick tails of the t-distribution dictate that $F(t)$ has *infinite* mean and variance (although, of course, we could characterise the distribution through its median and other quantiles). This is quite obviously different from the finite values obtained under Method 1.

Now consider the formula for $\text{Var}[y(t)|\delta]$. The second part is $nt\hat{\sigma}^2/(n-2)$ which closely resembles (but exceeds) the variance predicted by Method 1 but this is added to a further component of variance $t^2\hat{\sigma}^2/(n-2)$ due to uncertainty in the mean trajectory μt . As t increases this second component becomes dominant: for example, if t equals n then the variance is doubled by uncertainty in μ . This is a point which is invariably overlooked when questions of long-term risk-management are being considered. (For example, the

folklore surrounding the use of equities as the asset of choice for long-term investment is significantly compromised by the existence of parameter uncertainty.)

Model 2

$\delta(t)$ is modelled by an AR(1) time series model: that is,

$$\delta(t) - \mu = \alpha(\delta(t-1) - \mu) + \sigma Z(t)$$

where $Z(t)$ is a sequence of independent and identically distributed standard normal random variables.

Conditional on $\delta(-n+1)$, the likelihood for the observed series is

$$f(\delta|\mu, \sigma^2, \alpha) = \prod_{s=-n+2}^0 \left\{ (2\pi\sigma^2)^{-1/2} \exp \left[-\frac{1}{2\sigma^2} (\delta(s) - \mu - \alpha(\delta(s-1) - \mu))^2 \right] \right\}$$

Using the conditional likelihood, we can derive the following maximum likelihood estimates:

$$\begin{aligned} \hat{\mu} &= \frac{1}{n-1} \left(\sum_{s=-n+1}^{-1} \delta(s) + \frac{\delta(0) - \delta(-n+1)}{1 - \hat{\alpha}} \right) \\ \hat{\alpha} &= \frac{\sum_{s=-n+2}^0 (\delta(s) - \hat{\mu})(\delta(s-1) - \hat{\mu})}{\sum_{s=-n+2}^0 (\delta(s-1) - \hat{\mu})^2} \\ \hat{\sigma}^2 &= \frac{1}{n-1} \sum_{s=-n+2}^0 [\delta(s) - \hat{\mu} - \hat{\alpha}(\delta(s-1) - \hat{\mu})]^2 \end{aligned}$$

We can rewrite the conditional likelihood as

$$f(\delta|\mu, \sigma^2, \alpha) \propto (\sigma^2)^{-(n-1)/2} \exp \left[-\frac{1}{2\sigma^2} \left\{ \phi_1 + \phi_2(\mu - \hat{\mu})^2 + \phi_3(\alpha - \hat{\alpha})^2 \right. \right. \\ \left. \left. + \phi_4(\mu - \hat{\mu})(\alpha - \hat{\alpha}) + \phi_5(\mu - \hat{\mu})(\alpha - \hat{\alpha})^2 \right. \right. \\ \left. \left. + \phi_6(\mu - \hat{\mu})^2(\alpha - \hat{\alpha}) + \phi_7(\mu - \hat{\mu})^2(\alpha - \hat{\alpha})^2 \right\} \right]$$

where

$$\begin{aligned} \phi_1 &= (n-1)\hat{\sigma}^2 \\ \phi_2 &= (n-1)(1-\hat{\alpha})^2 \\ \phi_3 &= \frac{(n-1)\hat{\sigma}^2 + (\delta(-n+1) - \hat{\mu})^2 - (\delta(0) - \hat{\mu})^2}{1-\hat{\alpha}^2} \approx \frac{(n-1)\hat{\sigma}^2}{1-\hat{\alpha}^2} \\ \phi_4 &= 2(\delta(-n+1) - \delta(0)) \approx 0 \\ \phi_5 &= \frac{2(\delta(-n+1) - \delta(0))}{1-\hat{\alpha}} \approx 0 \\ \phi_6 &= -2(n-1)(1-\hat{\alpha}) \\ \phi_7 &= (n-1) \end{aligned}$$

(The three approximations take out the terms which have a lower order than $n-1$.)

We add to the likelihood the Jeffreys prior distribution for (μ, σ^2, α) of $f(\mu, \sigma^2, \alpha) = \sigma^{-3}(1-\alpha)^{1/2}(1+\alpha)^{-1/2}$. (Again, some authors dislike the form of this prior due to its asymmetry and the singularity at $\alpha = -1$. For example, if one believes strongly that $\alpha \neq \pm 1$ but have no strong beliefs about α otherwise, then a prior of the form $\sigma^{-3}(1-\alpha)^{3/2}(1+\alpha)^{1/2}$ would be required to ensure that the posterior $f(\alpha|\delta)$ derived below tends to zero as α tends to ± 1 . We have not investigated here the effect of this change of prior or of the change in the model 1 prior to σ^{-2} .) Thus, we obtain the following posterior distributions (which, given the above adjustments, are approximate):

$$\mu|\delta, \sigma^2, \alpha \sim N\left(\hat{\mu}, \frac{\sigma^2/(n-1)}{(1-\alpha)^2}\right)$$

$$\frac{(n-1)\hat{\sigma}^2 \left(1 + \frac{(\alpha-\hat{\alpha})^2}{1-\hat{\alpha}^2}\right)}{\sigma^2} \Bigg| \alpha, \delta \sim \chi_{n-1}^2$$

and $\alpha|\delta$ has the posterior density function

$$f(\alpha|\delta) = k(\hat{\alpha}, n) [1 - \hat{\alpha}^2 + (\alpha - \hat{\alpha})^2]^{-(n-1)/2} [1 - \alpha^2]^{-1/2} \quad \text{for } -1 < \alpha < 1$$

Now given μ , σ^2 , α and δ

$$\begin{aligned}\delta(s) - \mu &= \alpha(\delta(s-1) - \mu) + \sigma Z(s) \\ &= \sigma(Z(s) + \alpha Z(s-1) + \dots + \alpha^{s-1} Z(1)) + \alpha^s(\delta(0) - \mu) \\ \text{where } Z(s) &\sim N(0, 1)\end{aligned}$$

and it can be shown that

$$\begin{aligned}y(t)|\mu, \sigma^2, \alpha, \delta &= \mu t + (\delta(0) - \mu)M(\alpha, t) + \sigma V(\alpha, t)^{1/2} Z \\ \text{where } M(\alpha, t) &= \frac{\alpha(1 - \alpha^t)}{1 - \alpha} \\ V(\alpha, t) &= \frac{1}{(1 - \alpha)^2} \left(t - \frac{2\alpha(1 - \alpha^t)}{1 - \alpha} + \frac{\alpha^2(1 - \alpha^{2t})}{1 - \alpha^2} \right) \\ \text{and } Z &\sim N(0, 1). \\ \text{Hence } y(t)|\alpha, \delta &= \hat{\mu}(t - M(\alpha, t)) + \delta(0)M(\alpha, t) + \sigma \tilde{V}(\alpha, t)^{1/2} Z' \\ \text{where } Z' &\sim N(0, 1) \\ \tilde{V}(\alpha, t) &= \frac{(t - M(\alpha, t))^2}{(1 - \alpha)^2(n - 1)} + V(\alpha, t) \\ \Rightarrow y(t)|\alpha, \delta &= \hat{\mu}(t - M(\alpha, t)) + \delta(0)M(\alpha, t) \\ &\quad + \sqrt{\tilde{V}(\alpha, t)} \sqrt{\hat{\sigma}^2 \left(1 + \frac{(\alpha - \hat{\alpha})^2}{1 - \hat{\alpha}^2} \right)} T_{n-1}\end{aligned}$$

where T_{n-1} has a t-distribution with $n - 1$ degrees of freedom.

The form of the posterior distribution for α is such that a simple expression for $y(t)|\delta$ cannot be found. Instead, the posterior distribution for $y(t)$ must be found by combining the above expression for $y(t)|\alpha, \delta$ and the posterior distribution for α using straightforward numerical techniques.

4.2 Data and results

The data to which these models were applied were observed annual (log) returns on equities in the UK (including reinvestment of dividends) over the last 30 years. Model uncertainty here is of the third form described in Section 1.1. In order to keep the example simple we have restricted ourselves to two possible models. We could, of course,

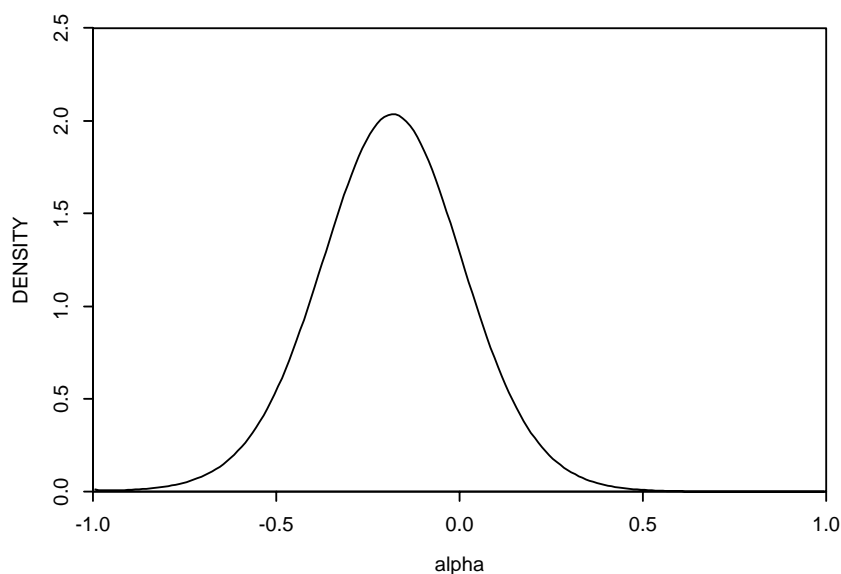


Figure 1: Posterior density for α given Model 2

extend the analysis to include all $AR(p)$ times series models, models with fatter tails than Normal *etc.*. We have restricted ourselves to two models here just to keep the presentation simple and for illustration purposes only.

The two models were fitted with the following results:

Model 1: $\hat{\mu} = 0.1472$, $\hat{\sigma}^2 = 0.05907$

Model 2: $\hat{\mu} = 0.1457$, $\hat{\sigma}^2 = 0.05724$, $\hat{\alpha} = -0.1758$

Each model was set a prior probability of 0.5 since it is difficult to argue in advance of looking at a set of investment data that one model is more suitable than the other.

The posterior model probabilities were found to be $Pr(M_1|\delta) = 0.63$ and $Pr(M_2|\delta) = 0.37$, so that model 1 is therefore the best estimate model.

The posterior density function under Model 2 for α is plotted in Figure 1. It can be seen that with only 30 years of data there is considerable uncertainty in the value of α .

By means of illustration we now consider accumulations up to time $t = 25$. The density functions for $y(25)$ are plotted in Figure 2 and for $F(25)$ in Figure 3 for Methods 1, 2 and 3.

In the description of the different approaches to this problem we noted that there would be a marked increase in the variance of $y(t)$. This is clearly found to be true from looking at Figure 2: Method 1 produces a much narrower predictive distribution for $y(25)$ than

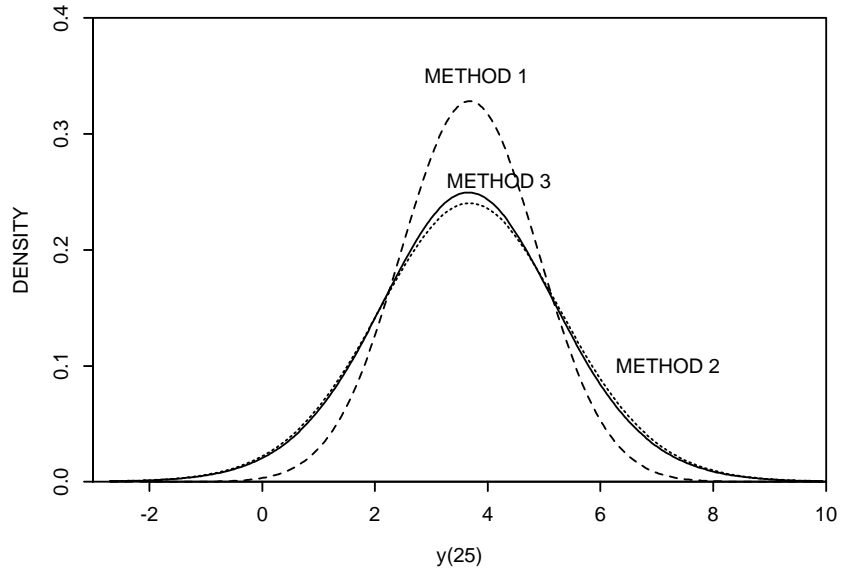


Figure 2: Posterior densities for $y(25)$: Dashed curve – Method 1, Model 1; Dotted curve – Method 2, Model 1; Solid curve – Method 3.

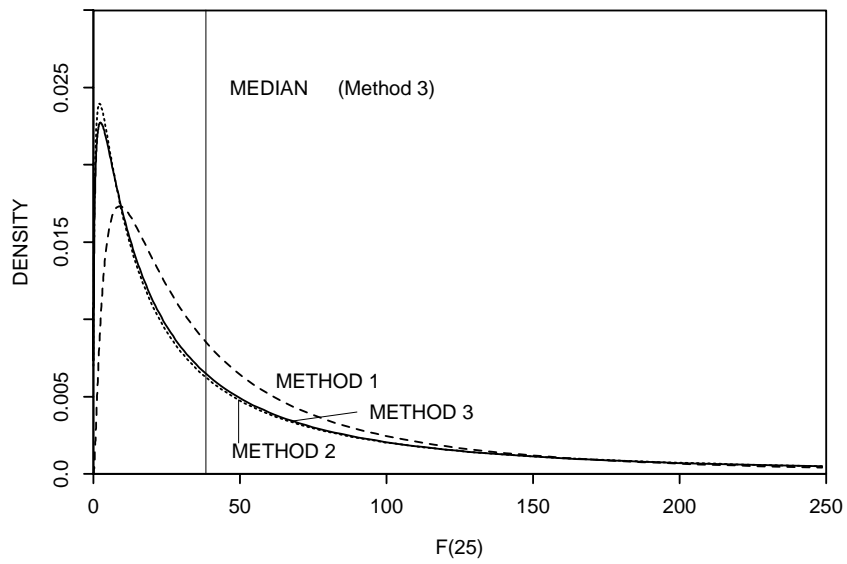


Figure 3: Posterior densities for $F(25)$: Dashed curve – Method 1, Model 1; Dotted curve – Method 2, Model 1; Solid curve – Method 3.

Methods 2 and 3. Clearly parameter uncertainty is important here.

If we compare the solid and dotted curves in Figure 2 we find that the inclusion of model uncertainty makes a smaller but still significant change to the distribution of $y(25)$. (Experimentation with simulated sets of data suggest that differences between Methods 2 and 3 become more significant when returns in successive years are much more highly correlated than was observed. This, for example, might be the case for fixed interest investments.)

It should be noted that the density under Method 3 is narrower than that for Method 2. This is, perhaps, counterintuitive. However, since $\hat{\alpha} < 0$, the posterior density for Model 2 is narrower than that for Model 1. Therefore, model averaging in Method 3 narrows, rather than widens, the distribution.

The densities plotted in Figure 3 provide us with the same conclusions about the accumulation function $F(25)$, although the high degree of skewness has tended to make these differences appear less. However, it should be recalled that whereas the dashed density curve (Method 1) provides us with finite moments, the dotted and solid curves (Methods 2 and 3) both represent random variables with infinite moments.

4.3 Further Remarks

First, it has been noted that in the presence of parameter and model uncertainty that the posterior mean and variance of $F(t)$ is infinite. Similarly we would find that the mean and variance of the (random) present value function $V(t) = 1/F(t)$ is infinite.

Consider, then, a savings policy which guarantees to pay out a sum of 1, say, at time t in return for a single premium S made at time 0. The above remarks just serve as a reminder that to use the expected value of $V(t)$ makes no sense. (Even when $E[V(t)]$ is finite this is not necessarily the correct way to set a premium rate.) Instead we must set a single premium at a level which, for example, ensures that the probability that the issuing office has sufficient cash to pay off the liability at time t is at least 0.9.

Second, it should be noted that with only 30 years of data on equity returns, the estimated mean rate of return of 14.6% has a standard error of 4.5%. This, when added to the inherent randomness of investment returns, calls into question the emphasis which practitioners place not only on deterministic methods but also on the use of a much more restricted range of assumptions.

Third, the approach taken here and in the next section is to choose which models out of the family M fit the data best. It is possible that M has been poorly specified so that none of the models investigated give an adequate description of the data. It is, therefore appropriate to carry out a supplementary analysis of the residual errors for each model

fitted, checking, for example, for outliers, autocorrelation and general lack of adherence to the assumed distribution (for example see Zellner, 1975).

5 Application to Ruin Theory

Let $U(t) = u + ct - S(t)$ be a surplus process, where u is the initial surplus, c is the rate of premium income and $S(t)$ is the aggregate claims between times 0 and t (for example, see Panjer and Willmot, 1992). We assume that $S(t)$ is a Compound Poisson Process so that

$$S(t) = \sum_{i=1}^{N(t)} X_i$$

where $N(t) \sim Po(\lambda t)$ is a Poisson Process, X_i is the amount of claim i , $i = 1, 2, \dots$, X_1, X_2, \dots are independent and identically distributed, positive random variables, and $\{N(t) : t \geq 0\}$ and X_1, X_2, \dots are independent. (In addition we assume that the X_i are positive almost surely and that, given M_i and θ_i there exists $\gamma = \gamma(M_i, \theta_i) > 0$ such that $E[\exp \gamma(M_i, \theta_i)X_i | M_i, \theta_i] < \infty$.)

Let $\psi(u)$ be the probability of ruin given an initial surplus of u . Lundberg's inequality states that

$$\psi(u) \leq e^{-Ru}$$

where R , the adjustment coefficient, is the unique positive solution of the equation

$$\lambda + cr = \lambda M_X(r)$$

λ is the *known* claims rate, c is the *known* rate of premium income, and $M_X(r)$ is the moment generating function of the *known* claims distribution.

We look at two problems:

- What is the value of the adjustment coefficient, R ?
- How much surplus, u , should we have in order that the probability of ruin is no more than 0.01?

In the statement of Lundberg's inequality, it was emphasized that each component was known with certainty. In reality, each of these components has some uncertainty attached to it, and this is what we now investigate. Under this uncertainty $\psi(u)$ should now be defined as the probability of ruin given an initial surplus of u *based on our (limited) past*

claims experience. From this definition it is clear that $\psi(u)$ must take account of the inherent parameter and model uncertainty.

We will, however, assume that the claims rate λ is known (and, without loss of generality this is equal to 1). This is a rather artificial assumption which simplifies the problem somewhat but it does not compromise the purpose of the exercise which is to illustrate the effects of incorporating parametric and, more importantly, model uncertainty.

Suppose we have a set of data $x = (x_1, x_2, \dots, x_n)$, these being the amounts of the last n claims with which to make inferences about the claims distribution.

Approaching the problems with a finite data set it is clear that neither the claims model nor the parameter values will be known with certainty. As a result of this, the adjustment coefficient itself will not be known with certainty. Instead, R will have a level of uncertainty which is consistent with the level of uncertainty in the parameters and in the model.

Before looking at the two problems in more detail we will put the three methods described in Sections 1 to 3 into the present context.

Method 1: The best estimate model is chosen using the AIC or SBC and then its parameter values are set to be equal to their maximum likelihood estimates. By doing this R is also fixed and $\psi(u)$ has the usual exponential upper bound.

Methods 2 and 3: These derive posterior parameter and model distributions which allow us to calculate a posterior distribution for R and to assess solvency requirements. Under these circumstances the upper bound for $\psi(u|x)$ is no longer exponential. Instead

$$\begin{aligned}\psi(u|x) &= E[I|x] \text{ where } I = \begin{cases} 1 & \text{if ruin occurs} \\ 0 & \text{otherwise} \end{cases} \\ &= E[E(I|R)|x] \\ &\leq E[e^{-Ru}|x]\end{aligned}$$

where, of course, $R = 0$ in the event that $c \leq \lambda E[X_i|\theta_i, M_i]$.

[Note that $E[e^{-Ru}|x] \geq e^{-\mu_R u}$ where $\mu_R = E(R|x)$ by Jensen's inequality (for example, see Williams, 1991). However, this is, of course, a lower bound so its usefulness is limited.]

Given only a finite amount of claims data, we can never say with certainty that the premium rate will exceed the expected claims rate in the future.

Now even in the most simple cases the posterior distributions for each model and parameter set are difficult to work with. So we now extend some of the asymptotic techniques described in Section 3. The development which follows is not intended to be completely rigorous: rather it is given here to illustrate in general terms, the order of the effect parameter and model uncertainty.

Under Model i (with parameter set θ_i) we have, for large n ,

$$\theta_i \approx N(\hat{\theta}_i, H_i^{-1})$$

where $\hat{\theta}_i$ is the posterior mode and H_i is the Hessian matrix of the posterior density evaluated at $\hat{\theta}_i$ for model i , with $H_i^{-1} = O(n^{-1})$ as $n \rightarrow \infty$.

Now given θ_i (and $\lambda = 1$) we solve

$$1 + cR = M_i(R; \theta_i)$$

where $M_i(R; \theta_i)$ is the moment generating function for claims model i and parameter set θ_i . Let $R(\theta_i)$ be the unique positive solution of this equation or 0 if no positive solution exists, and let $\hat{R}_i = R(\hat{\theta}_i)$. Expanding $M_i(R; \theta_i)$ about $(\hat{R}_i; \hat{\theta}_i)$ we have

$$M_i(R; \theta_i) = M_i(\hat{R}_i; \hat{\theta}_i) + \frac{\partial M_i(\hat{R}_i; \hat{\theta}_i)}{\partial R} (R - \hat{R}_i) + \sum_{j=1}^{k_i} \frac{\partial M_i(\hat{R}_i; \hat{\theta}_i)}{\partial \theta_{ij}} (\theta_{ij} - \hat{\theta}_{ij}) + o(|(R - \hat{R}_i; \theta_i - \hat{\theta}_i)|)$$

where k_i is the dimension of θ_i in model i .

Now $M_i(\hat{R}_i; \hat{\theta}_i) = 1 + c\hat{R}_i$ and ignoring the smaller order terms we solve

$$\begin{aligned} 1 + cR &\approx 1 + c\hat{R}_i + m_{ir}(R - \hat{R}_i) + \sum_{j=1}^{k_i} m_{ij}(\theta_{ij} - \hat{\theta}_{ij}) \\ \text{where } m_{ir} &= \frac{\partial M_i(\hat{R}_i; \hat{\theta}_i)}{\partial R} \\ \text{and } m_{ij} &= \frac{\partial M_i(\hat{R}_i; \hat{\theta}_i)}{\partial \theta_{ij}} \quad j = 1, 2, \dots, k_i \\ \Rightarrow (c - m_{ir})R &\approx (c - m_{ir})\hat{R}_i + \sum_{j=1}^{k_i} m_{ij}(\theta_{ij} - \hat{\theta}_{ij}) \\ \Rightarrow R(\theta_i) &\approx \hat{R}_i + \frac{1}{c - m_{ir}} m_i^T (\theta_i - \hat{\theta}_i) \\ \text{where } m_i^T &= (m_{i1}, m_{i2}, \dots, m_{ik_i}) \end{aligned}$$

[In fact this relationship is exact in the case of the exponential distribution.]

Hence, given model i

$$R \approx N(\hat{R}_i, \sigma_i^2)$$

$$\text{where } \sigma_i^2 = \frac{1}{(c - m_{ir})^2} m_i^T H_i^{-1} m_i$$

[Since $H_i^{-1} = O(n^{-1})$ as $n \rightarrow \infty$, it follows that $\sigma_i^2 = O(n^{-1})$ as $n \rightarrow \infty$: that is, the Normal approximation becomes more accurate as $n \rightarrow \infty$.]

Let

$$Pr[R \leq r | x, M_i] = \begin{cases} 0 & \text{if } r < 0 \\ \Phi\left(-\frac{\hat{R}_i}{\sigma_i}\right) & \text{if } r = 0 \\ \Phi\left(\frac{r - \hat{R}_i}{\sigma_i}\right) & \text{if } r > 0 \end{cases} \quad (1)$$

be the approximate posterior cumulative distribution function for R given model i , and let $p_i = Pr(M_i|x)$ be the posterior probability of model i ($i = 1, 2, \dots$). (This approximation is a crude acknowledgement of the fact that R cannot be negative.)

Let i^* be the value of i which maximises p_i , so that M_{i^*} is the best estimate model. Under Method 2 the posterior distribution for R is approximately $f_{i^*}(r|x)$. Under Method 3 the posterior distribution for R is approximately $f(r|x) = \sum_i p_i f_i(r|x)$.

Applying these posteriors for R we are now able to calculate the now approximate upper bound for the probability of ruin.

Recall that $\psi(u) \leq E[e^{-Ru}|x]$.

Using the distribution defined in equation (1), it is straightforward to show that

$$E[e^{-Ru}|x, M_i] = \exp\left[-\hat{R}_i u + \frac{1}{2} \sigma_i^2 u^2\right] \left\{ 1 - \Phi\left(\frac{-\hat{R}_i + u \sigma_i^2}{\sigma_i}\right) \right\} + \Phi\left(\frac{-\hat{R}_i}{\sigma_i}\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Normal distribution.

Hence, under Method 2

$$\psi(u) \leq \exp\left[-\hat{R}_{i^*} u + \frac{1}{2} \sigma_{i^*}^2 u^2\right] \left\{ 1 - \Phi\left(\frac{-\hat{R}_{i^*} + u \sigma_{i^*}^2}{\sigma_{i^*}}\right) \right\} + \Phi\left(\frac{-\hat{R}_{i^*}}{\sigma_{i^*}}\right)$$

and under Method 3

$$\psi(u) \leq \sum_i p_i \exp\left[-\hat{R}_i u + \frac{1}{2} \sigma_i^2 u^2\right] \left\{ 1 - \Phi\left(\frac{-\hat{R}_i + u \sigma_i^2}{\sigma_i}\right) \right\} + \Phi\left(\frac{-\hat{R}_i}{\sigma_i}\right)$$

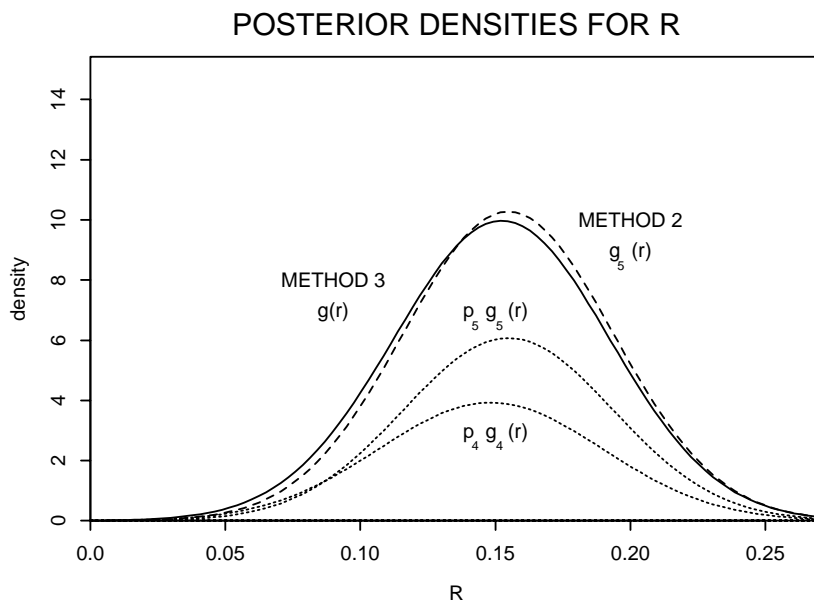


Figure 4: Experiment 1: Approximate posterior distributions for R : dashed curve $f_5(r|x)$ (Method 2); dotted curves $p_i f_i(r|x)$ for $i = 4, 5$ from left to right; solid curve $f(r|x) = \sum_i p_i f_i(r|x)$ (Method 3)

We now compare the three methods by considering some illustrative datasets.

In each of the experiments the (artificial) set of possible models is $\{M_1, M_2, \dots, M_{10}\}$ where M_i is the Gamma distribution with shape parameter $\alpha = i$. (Here, if Y has a Gamma distribution with shape parameter α and scaling parameter γ then Y has density function $f(y) = \gamma^\alpha y^{\alpha-1} e^{-\gamma y} / \Gamma(\alpha)$ and Y has mean α/γ and variance α/γ^2 .) Under each model there is one parameter, γ (the scaling parameter), to estimate.

In the absence of any prior information we choose prior distributions which are diffuse. Thus the prior model probabilities are $Pr(M_i) = 0.1$ ($i = 1, 2, \dots, 10$) and the prior distributions for γ are all improper, uniform distributions on $(0, \infty)$. [In the current context, some authors feel that it is not appropriate to use improper priors for γ since it is not then possible to compare models where one or more has such a prior (for example, see Bernardo and Smith, 1994, Chapter 6, and Lindley in discussion of O'Hagan, 1994). This problem can be avoided by choosing a proper prior for γ which is nevertheless very diffuse compared to the data (for example, an exponential prior with very high mean). In practice this would make a negligible difference to the outcome of the analysis.]

Experiment 1

400 observations were simulated from a Log-Normal distribution with parameters $\mu = 0$

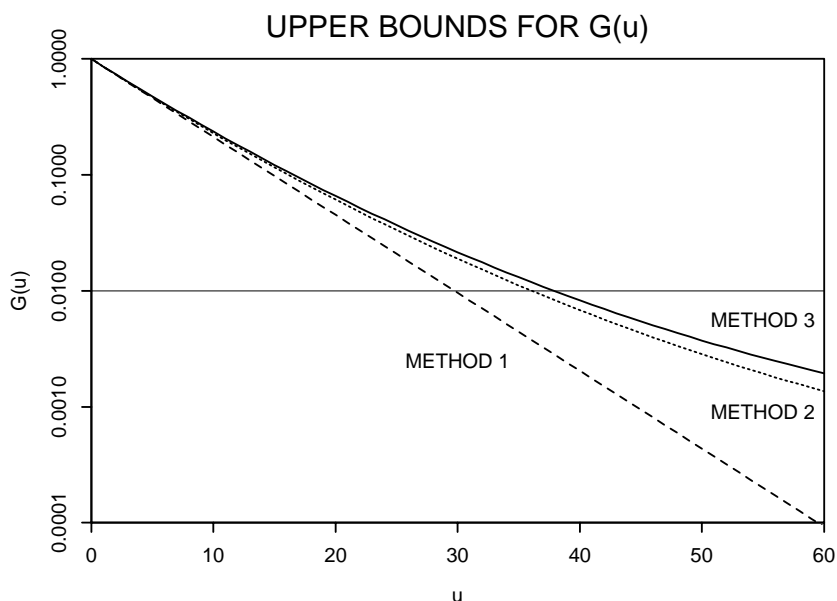


Figure 5: Experiment 1: Posterior ruin probabilities for Methods 1, 2 and 3

and $\sigma^2 = 0.25$. The data were normalised so that they had a mean of 1 (that is, the unit of currency was set equal to the mean of the sample). Based on a premium rate of $c = 1.1$ the results of the analysis were as follows.

The adjustment coefficient: Figure 4 illustrates the different results which can be obtained using Methods 2 and 3. The dashed curve is the posterior density, $f_5(r|x)$, for R given M_5 , the best estimate model. The solid curve is the posterior density, $f(r|x)$ for R over all models. The solid curve is the sum of the dotted curves which represent the $p_i f_i(r|x)$ for $i = 4, 5$. In this example, the solid curve and the dashed curve are quite similar indicating that the addition of model uncertainty has not had a significant impact on the results of our analysis.

Minimum reserves: Here the problem is to determine what reserve, u , is required to ensure that the probability of ruin is no more than 0.01.

Method 1: $R = 0.1548$. Solve $G_1(u) = e^{-Ru} = 0.01 \Rightarrow u = 29.8$.

Method 2: $i^* = 5$. Solve $G_2(u) = E[e^{-Ru}|x, M_5] = 0.01 \Rightarrow u = 36.0$.

Note that $G_2(29.7) = 0.0194$, which is significantly higher than the required probability of 0.01.

Method 3: Solve $G_3(u) = E[e^{-Ru}|x] = 0.01 \Rightarrow u = 37.9$.

Note that $G_3(29.7) = 0.0221$ and $G_3(36.1) = 0.0119$.

i	$p_i = Pr(M_i x)$	$\hat{\gamma}_i$	\hat{R}_i	σ_i
1	0.000	–	–	–
2	0.000	–	–	–
3	0.000	–	–	–
4	0.409	4	0.1483	0.0415
5	0.591	5	0.1548	0.0388
6	0.000	–	–	–
⋮				
10	0.000	–	–	–

Table 1: Experiment 1 estimates

We see that an acknowledgement of parameter uncertainty results in a large increase in the required reserve. There is a small difference between Methods 2 and 3. However, the main point to note here is that ignoring parameter and model uncertainty may significantly underestimate the required reserve.

We can, of course, look at the ruin curves in general rather than picking a specific ruin probability. The differences between the three methods are illustrated in Figure 5. The three curves are, from left to right, $G_1(u)$, $G_2(u)$ and $G_3(u)$ corresponding to Methods 1, 2 and 3 respectively. Since e^{-Ru} is convex, Jensen's inequality will always ensure that the curve $G_2(u)$ will always lie above $G_1(u)$. The relative positions of $G_2(u)$ and $G_3(u)$ will depend on the relative weights of the different models under consideration and in the degrees of pessimism predicted by each model.

Experiment 2

400 observations were simulated from a Gamma(1.3,1.3) distribution. Based on a premium rate of $c = 1.5$ the results of the analysis were as follows.

The adjustment coefficient: Figure 6 illustrates the results of the three methods which can be obtained using Methods 2 and 3. In this experiment we find that the solid curve is quite different from the dashed curve, so, in this case, it seems sensible to take model uncertainty into account. Upon considering the reserving problem, however, the need to consider model uncertainty is less clear.

Minimum reserves:

Method 1: $R = 0.333$. Solve $G_1(u) = e^{-Ru} = 0.01 \Rightarrow u = 13.83$.

Method 2: $i^* = 1$. Solve $G_2(u) = E[e^{-Ru}|x, M_1] = 0.01 \Rightarrow u = 14.54$.

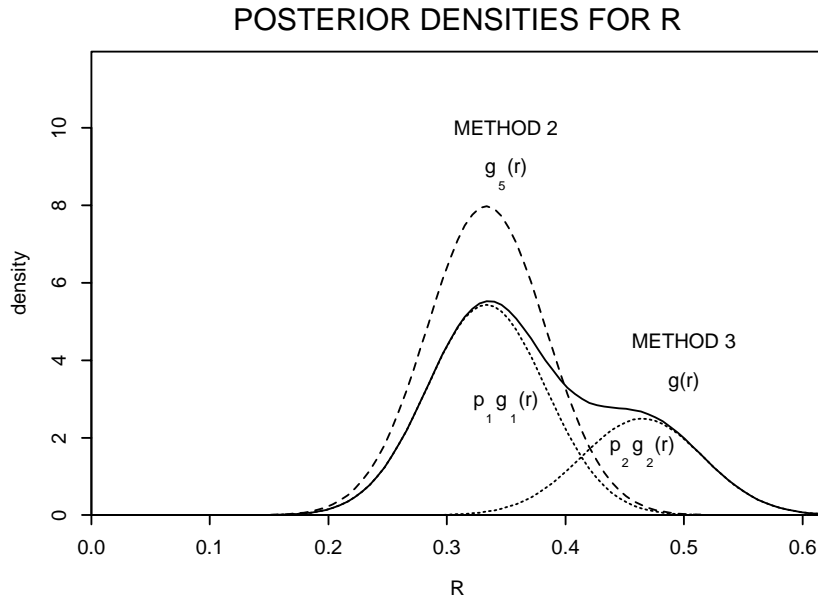


Figure 6: Posterior distributions for R : dashed curve $f_1(r|x)$ (Method 2); dotted curves $p_i f_i(r|x)$ for $i = 1, 2$ from left to right; solid curve $f(r|x) = \sum_i p_i f_i(r|x)$ (Method 3)

i	$p_i = Pr(M_i x)$	$\hat{\gamma}_i$	\hat{R}_i	σ_i
1	0.681	1	0.333	0.0500
2	0.319	2	0.465	0.0511
3	0.000	—	—	—
\vdots				
10	0.000	—	—	—

Table 2: Experiment 2 estimates

Note that $G_2(13.83) = 0.013$, which is significantly higher than the specified maximum.

Method 3: Solve $G_3(u) = E[e^{-Ru}|x] = 0.01 \Rightarrow u = 13.59$.

Note that $G_3(13.83) = 0.009$ and $G_3(14.54) = 0.007$, so Methods 1 and 2 are overestimating the amount of the required reserve. However, this is a result of the fact that the alternative Model 2 has a less risky claims distribution than Model 1 which was used by Methods 1 and 2.

In this experiment, the addition of model uncertainty has clearly had an effect on the setting of a reserve.

6 More complex examples: MCMC

Exact evaluation of uncertainty can only be achieved in a limited number of cases: for example, the Geometric Random Walk model described above. In the majority of practical applications this is not possible and it is, therefore, necessary to resort to the use of numerical techniques.

One approach is to locate the mode of the posterior distribution for the set of parameters, θ , and to use a suitable approximation (for example, Normal) around this. One can then generate the distribution of the quantity of interest, y , by taking a representative range of values of θ weighted according to its posterior distribution (either deterministically or at random), and averaging the distributions for y given θ .

Markov chain Monte Carlo (MCMC) is a more efficient class of methods which allow us to deal with more complex models with many parameters, such as the Wilkie (1995) model, where it is difficult to write down the posterior distribution for y .

The method involves taking a random (but not independent) sequence of values $\theta_1, \theta_2, \dots, \theta_N$ from the posterior distribution for $\theta|M, x$. A posterior distribution for $y|\theta_i, M, x$ can then be calculated either analytically or by further simulation.

This is a relatively new method in statistics which requires powerful computing resources. Several years on the method is only just beginning to be applied to actuarial and investment problems (for example, Harris, 1999, and Scollnik, 1998). More advanced papers on MCMC and further references can be found in the collection of papers edited by Gilks *et al.* (1995).

7 Conclusions

In this paper we have discussed an approach to the incorporation of parameter and model uncertainty in the process of making inferences about some quantity of interest, y . The approach taken is Bayesian, which provides a coherent framework allowing us to make inferences about y in the presence of parameter and model uncertainty.

In the examples given it was found that the inclusion of parameter uncertainty made a significant contribution to the outcome of the modelling exercise. Furthermore, the inclusion of model uncertainty also modified the final outcome, but to a lesser extent.

It should be noted that the quantity of data will have an impact on the significance of parameter and model uncertainty. If we obtain more data then this may reduce the need to incorporate model and perhaps even parameter uncertainty. With some actuarial applications (for example stochastic insurance or reinsurance), however, it is not realistic to assume that we can find a significant amount of additional data so that the problem under discussion here is not one which we can ignore.

From the literature it appears that researchers in non-life insurance are leading the way in addressing the problems of parameter uncertainty. However, there is still considerable scope for the development of methods which identify not just one but a group of models which merit further investigation and a fuller comparison.

In general it is not clear whether model uncertainty will alter our conclusions significantly or not. It is therefore our duty to test the sensitivity of our conclusions to the choice of model. If these conclusions are found to be sensitive then it will be essential to conduct an analysis which properly accounts for model uncertainty.

References

- Bawa, V., Brown, S., and Klein, R. (editors) (1979) *Estimation Risk and Optimal Portfolio Choice*. North Holland, Amsterdam.
- Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory*. Wiley, Chichester.
- Chatfield, C. (1995) Model uncertainty, data mining and statistical inference (with discussion). *Journal of the Royal Statistical Society, Series A* **158**, 419-466.
- Clemen, R.T. (1989) Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* **5**, 559-583.
- Crouhy, M., Galai, D., and Mark, R. (1998) Model risk. *Journal of Financial Engineering* **7**, 267-288.

- Dawid, A.P. (1983) Invariant prior distributions. *Encyclopaedia of Statistical Sciences* **4**, 228-236.
- Daykin, C.D., Pentikäinen, T., and Pesonen, M. (1994) *Practical Risk Theory for Actuaries*. Chapman and Hall, London.
- Dickson, D.C.M., Tedesco, L.M., and Zehnwirth, B. (1998) Predictive aggregate claims distributions. *Journal of Risk and Insurance* **65**, 689-709.
- Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* **57**, 45-97.
- Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1995) *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- Harris, G.R. (1999) Markov chain Monte Carlo estimation of regime switching vector autoregressions. *ASTIN Bulletin* **29**, 47-79.
- Jeffreys, H. (1961) *Theory of Probability, third edition*. Oxford University Press, Oxford.
- Kass, R.E., and Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.
- Klugman, S. (1992) *Bayesian Statistics in Actuarial Science*. Kluwer, Boston.
- McNeil, A.J. (1997) Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* **27**, 117-137.
- Min, C.-k., and Zellner, A. (1993) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics* **56**, 89-118.
- O'Hagan, A. (1994b) *Kendall's Advanced Theory of Statistics Volume 2B: Bayesian Inference*. Halstead Press, London.
- O'Hagan, A. (1995) Fractional Bayes factors factors for model comparison. (with discussion) **Journal of the Royal Statistical Society, Series B**, 5799-118.
- Pai, J. (1997) Bayesian analysis of compound loss distributions. *Journal of Econometrics* **79**, 129-146.
- Panjer, H.H., and Willmot, G.E. (1992) *Insurance Risk Models*. Society of Actuaries Schaumburg, .

- Parker, G. (1997) Stochastic analysis of the interaction between investment and insurance risks (with discussion). *North American Actuarial Journal* **1**, 55-84.
- Schmock, U. (1999) Estimating the value of the Wincat coupons of the Winterthur Insurance convertible bond. *ASTIN Bulletin* **29**, 100-164.
- Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.
- Scollnik, D.P.M. (1998) On the analysis of the truncated generalised Poisson distribution using a Bayesian method. *ASTIN Bulletin* **28**, 135-152.
- Shibata, R. (1976) Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**, 117-126.
- Wei, W.W.S. (1990) *Time Series Analysis*. Addison-Wesley, Redwood City.
- Wilkie, A.D. (1995) More on a stochastic asset model for actuarial use (with discussion). *British Actuarial Journal* **1**, 777-964.
- Williams, D. (1991) *Probability with Martingales*. Cambridge University Press, Cambridge.
- Zellner, A. (1975) Bayesian analysis of regression error terms. *Journal of the American Statistical Association* **70**, 138-144.