General Insurance Study Group

1996 General Insurance Convention

2 - 5 October 1996

Workshop on Stochastic Error, Parameter Error & Model Error

Stephen Walker

1 OVERVIEW

The subject of error and uncertainty is and always has been a key issue in General Insurance. There is a growing interest in complex models of the claims process and computer models of both assets and liabilities. However, it is important that the degree of error in such models is recognised.

The true uncertainty about the future outcome exceeds greatly that caused purely by random chance, due to errors in the modelling process. It is important that actuaries both communicate and, where possible, quantify such possible errors.

The purpose of this workshop is to discuss some of the issues which may affect the results of a modelling process and to discuss what practical approaches exist to describe the components of uncertainty.

Approaches will be considered in the context of a simple example described on the following page.

EXAMPLE

You have been asked to assess the likely size of large claims (above £100,000) occurring from a book of motor business next year.

You have a sample of 40 such claims from the current year, and to make life easy, there is no inflation, reserving error, IBNR, Ogden, or any other complications to worry about.

The claim amounts are listed in the Appendix.

Questions:

What is the expected size of claim that will be incurred next year (assuming there is one)?

What is the range of possible sizes?

What is the expected cost to a simple excess-of-loss program (defined in the Appendix)?

How much harder is this, if all those nice assumptions (above) are removed?

2 BACKGROUND

We are faced with the usual problem:

- We want to estimate the future.
- We only have some data and our judgement about how the world works.
- People will be upset if we are "wrong".

By "wrong" I mean that the outcome is different from our estimate, and the uncertainties are not properly described to the user.

The outcome could be different for the following reasons:

• We may have been unlucky.	STOCHASTIC ERROR
• There was insufficient data.	PARAMETER ERROR
• We may have applied the wrong model.	MODEL ERROR

3 STOCHASTIC ERROR

Stochastic Error is the possibility that the outcome is not that expected, given that both the model and parameters are correct.

This error can easily be estimated via simulation or calculating the relevant mathematical distribution. As such it is the most commonly described error. However, it is not necessarily the most significant.

4 PARAMETER ERROR

Parameter Error is the possibility that the parameters used to define the model are incorrect given that the model is correct. This possibility occurs, firstly, because there is only a limited amount of data to estimate the parameters. Secondly, the parameters themselves will evolve through time, so those applicable for future events are always unknown at the present.

This error can be estimated by several approaches to be discussed in the workshop:

- Confidence intervals for the parameters.
- Bootstrap technique for the parameter estimate.
- Bayesian estimation of the parameter.

Additional considerations which occur when there are correlations between parameter estimates will be discussed.

The following questions need to be addressed:

- What is the best way to approach this component?
- What is the best way to describe this component to the reader of a report?

Combining the Parameter and Stochastic Errors

Often it is possible to estimate the combined uncertainty caused by the parameter and stochastic errors directly within the simulation process. Additional uncertainty caused by the parameter error may be included by means of a convolution of distributions.

For example, if the uncertainty on a Poisson claim rate parameter estimate is approximated by a gamma distribution, then a negative binomial distribution can be used to simulate the number of claims occurring.

If the convolution is not easy to evaluate, it is always possible to use a two-stage simulation process. The first stage is to pick randomly the parameters, and then simulate the outcome using these parameters. However, this adds greatly to the number of simulations required.

Does this approach help to illustrate indirectly the additional effect of parameter uncertainty?

5 MODEL ERROR

Model Error is the possibility that the analysis technique used is deficient. This is usually drawn in one of three situations:

- The true model could be one drawn from a known set of possible models. The error here is that the best fit model may not be the true model.
- The model chosen may be drawn from a family known to approximate the true model. The error here is the difference between the true model and the approximation chosen.
- The model may just happen to fit the (historic) data. There is no reason why the true model should be of the form chosen.

The choice of model is one of the most important steps in any analysis. Through this decision, prior knowledge is included through the model's implicit properties, for example the likelihood of extreme values. Often these properties are key to the behaviour of the entire process.

This error can only be partially investigated. The following possible techniques are to be discussed at the workshop:

- A set of possible models.
- A parameterised super-family of models.

The following questions need to be addressed:

- Is there a "best way" to approach this component?
- Is there any real difference between model and parameter error? Are models just large numbers of parameters?
- How should the possibility of model failure be communicated?

Distribution-free methods: Removing the model and parameter assumptions?

It is often desirable to eliminate the 'apparent' need to make assumptions by using a distribution-free approach such as bootstrapping. However these techniques implicitly imply quite strong assumptions and it is important to modify the process to remove any of these which are inappropriate.

In our example, a simple bootstrap for next year's claim size would assume no future claim can be greater than the maximum already observed.

Because of this feature, and the need to include as much 'judgement' information into the process as is possible, the bootstrap technique cannot be regarded as a panacea and may even be regarded as dangerous.

6 CONCLUSION

Understanding the total uncertainty in a modelling process is an extremely difficult task. It is important, as more complicated sets of models are considered, that practical methodology aimed at quantifying the potential uncertainty be developed.

I believe that within a set of models, each fitting the data, there is little difference in the total error. For each possible model the error is divided between the three components in a different manner. Only by employing our wider knowledge can the total error be reduced. The key criterion by which a set of modelling tools must be judged is how easily they enable such additional knowledge to be included.

APPENDIX

The table below displays a sample of 40 large claims simulated from a hypothetical motor account. This data will be used to illustrate various techniques as an aid to the discussion.

101	108	110	112	116
116	117	118	118	119
119	120	120	120	171
183	183	188	199	200
207	210	211	223	225
261	268	278	297	311
320	340	344	391	502
534	572	1176	1303	1650

For the excess of loss program described below, the probability of a claim hitting a layer and the expected cost of claims hitting a layer will be calculated. Again for simplicity, there is no indexation, discounting, or any other complications to worry about.

Excess of loss program:			
	250	xs	250
	500	xs	500
	1m	xs	1m
	3m	xs	2m
	5m	xs	5m
	unlimited	xs	10m