Summary of

Exploring Unknown Quantities

Development and Application of a Stochastic Catastrophe Model with Output and Sensitivities

Sam Killmier

Honours student, The University of Melbourne

Author: Sam Killmier

Reference No. SM

ACKNOWLEDGEMENTS

The support of The Australian Prudential Regulation Authority (APRA) and the Reserve Bank of Australia in providing financial assistance for the writing of this paper through the Brian Gray Scholarship is greatly appreciated. APRA is also thanked for providing useful background information. The efforts of my supervisor, Dr Jules Gribble, and anonymous practitioners are acknowledged and appreciated.

AIMS and SIGNIFICANCE

Stochastic catastrophe models are now widely used to assess and manage catastrophe exposure. The models commonly used by insurers are proprietary in nature and their workings and inputs are by-and-large not publicly available information. In addition there are sources within the industry that suggest that there are significant inconsistencies in the output of models of the same phenomena.

Exploring Unknown Quantities adds to the currently limited literature regarding the details of catastrophe model design and application. The aims were twofold: firstly, to present a theoretical model and apply this model to a real-world peril (hailstorm in the Sydney region); and secondly, to present results in the form of sensitivities to model inputs and design in order to highlight potential causes of cross-model inconsistency. The paper is intended to be of interest to academics, practitioners and regulators. Its addition to the field of research will hopefully encourage other parties to report openly on their work in this area.

OUTLINE of the METHODOLOGY

An outline of the research process is given below by describing the five stages of the research.

Background Reading and Research

Background reading was extensive, particularly because the exact intention of the research was unknown at commencement. It became clear that there is a limited body of publicly available literature describing stochastic catastrophe models, indicating that a thorough presentation of all aspects of model construction would be valuable. Additionally, certain journal articles and company-sponsored articles suggest that there are significant inconsistencies between models. Anonymous practitioners who were interviewed on the matter reiterated this notion. The causes of these inconsistencies could be explored via a comprehensive account of model construction.

Model Construction

The model developed was in part inspired by readings and in part developed by the author (although there is no claim as to its originality – confirming this would be near impossible). The model design is simple and not particularly data intensive – a response to the difficultly experienced in accessing data. It has been made flexible; for example, the design may be adapted to alternative perils such as earthquake.

Data Gathering

Catastrophe models require access to vast amounts of data and other information that is difficult to access. Dozens of practitioners and academics were contacted. Broadly the assistance given was disappointing, inhibited by confidentiality agreements and the non-commercial nature of the author's requests.

The Bureau of Meteorology, The Natural Hazards Research Centre, and a major Australian general insurer were particularly helpful.

Application of the Model

The model may be easily programmed – *MATLAB* was used for this paper. The data obtained was used to estimate the parameters of the model. Small modifications were made to the model design in order to incorporate the limitations of the data. For example, the proximity of locations to one another was programmed based on the numerical proximity of postcodes, despite the fact that this simplification does not appear in the original model design.

Collation of Results and Write-up

The results produced by the model were compiled and are presented in an easy to read format so that the presence and absence of significant sensitivities may be inspected. A significant volume of background information has been written, as well as important information regarding the model itself. Given the length constraints of 10,000 words some work is not found in the paper, but rather in two further 'mini-papers':

- The Use of Normalised Catastrophe Loss Data and Extreme Value Theory to Predict Catastrophe Losses in Australia

- A Presentation of Basic Catastrophe Modelling Terminology

These papers provide a background to the submitted paper and are available on request.

DESCRIPTION of the MODEL

The following is a description of the Sydney Hail Model, the model developed for the paper. It is a stochastic catastrophe model that simulates the annual experience of an insurer to estimate a distribution of annual catastrophe losses flowing from hail events. The Zone Percentage Loss Model forms the basis of the SHM and is described first.

The Zone Percentage Loss Model (ZPLM)

The principal assumption of the ZPLM is that the percentage loss (PL) (defined as the loss in dollar terms divided by the sum insured) is constant in each zone of the affected area of a catastrophe. The ZPLM is a general model and may be applied to multiple perils and locations. For each peril the model simulates:

- the centre of the perilous event;
- the affected area, consisting of a number of zones;
- the severity of the event – determined as a function of a number of severity factors (for example, Richter scale reading or duration of event).

The severity is used to superimpose a percentage loss curve over the affected area. PLs increase with proximity to the centre of the event. By combining PLs with sums insured in each of the zones, the loss to the insurer from the event is calculated.

The Sydney Hail Model (SHM)

The SHM is a direct application of the ZPLM – it simulates annual experience by using the ZPLM for each event in a simulation year. The severity factors used in the SHM are maximum wind-gust on storm day and maximum hailstone size. These were chosen based on the limitations of the available meteorological data and also on empirical evidence regarding the primary factors that determine damage in real storms.

For each year the SHM simulates the number of perilous events that occur and the ZPLM is then used to determine the losses arising from each of these perilous events. The sum of all losses in a year is the statistic of interest and it is the distribution of this random variable that the model seeks to estimate.

How the Model Projects Losses

The severity of an event is translated into a loss figure by comparing it with the severity of an appropriate standard storm for which the loss pattern is known. Unfortunately, the only reliable data available for this research was for the 14 April 1999 storm in Sydney (the most costly catastrophe in Australia's history). Thus all storms are related to this standard through what is called the severity ratio function. If a simulated storm has a severity that exceeds the standard then percentage losses are higher for a given proximity. The exact relationship is assumed to be linear in the first instance, so that 10% more severity leads to 10% higher percentage losses. It may be argued that the losses may diminish as a percentage of the sum insured for some perils because there is an effective limit on how much damage can be done by the peril (eg total devastation may not be possible). Exploring this matter via alternative datasets is an avenue for further work.

How the Model Differs from Alternatives

There is a broad spectrum of conceivable catastrophe models. At one end of the model spectrum are very simplest models such as the Insurance Council of Australia Zones. This model associates percentages with geographical zones that may be multiplied by the insurer's exposure to determine a figure for the probable maximum loss (Walker, 1997). On the other end of the spectrum lies highly sophisticated models, as appear to be the proprietary models offered by firms such as AIR Worldwide, RMS, and EQECAT. These models are understood to incorporate large datasets and to utilise sophisticated natural peril modelling including, for example, seasonal factors. The SHM clearly lies somewhere between these two extremes. Determining exactly where is difficult because detailed explanations of the structure of proprietary models are not publicly available. In general most stochastic catastrophe models have the following generic structure (Watson & Johnson, 2003):

(i) *The Science Module*: uses historical data and statistical modelling to generate realisations of a peril and determine the intensity of the event at each location.

(ii) *The Engineering Module*: translates intensity into property damage at each location.

(iii) *The Insurance Coverage Module*: applies the conditions of the insurer's issued policies and reinsurance arrangements to determine claims for each event.

The SHM has a science module (the severity factors and storm severity); it has an engineering module in the form of the severity ratio function and percentage loss function (which translate severity into losses) and it has a simple insurance coverage module (a simple dataset that contains the sums insured for each zone and also possibly the reinsurance arrangements of the insurer).

It is clear, however, that the model employed is *relatively* simple. An engineering module might more typically incorporate the impact that the peril has on different construction types (Kozlowski & Mathewson, 1995) and use estimates of what is actually found on the ground (for example 40% timber buildings, 40% brick buildings and so on).

Nevertheless, the simple design of the SHM has a number of key advantages:

- the model incorporates data that is not extensive and therefore the model itself avoids spurious accuracies by being appropriately simple;
- catastrophes are inherently uncertain events and so there is an argument that the most that can realistically be gained from models is an *indication* of the magnitude of potential losses;
- the model is easy to fit and may incorporate information from a number of different historical events when this information is available;
- the model might be used as one of a number of inputs into the reinsurance/pricing/reserving decisions, complementing the results of more sophisticated models – a kind of "rough check".

The primary disadvantage of using an oversimplified model is that the model may overlook subtleties of the portfolio and peril that if included would have a significant impact on results.

OUTPUT and SENSITIVITIES

Results are presented in the form of output and sensitivities in order to demonstrate potential causes of cross-model inconsistencies.

The Base Model

The Base Model uses a set of assumptions as follows:

- a maximum PL of 1;
- "normal" sensitivity of insured properties to horizontal wind-damage; that is, the sensitivity to horizontally damage is the same as the sensitivity to vertically driven damage ($\theta = 1$);
- relationship between the severity of the storm and the PLs are linearly related; that is, a severity 10% higher than the standard storm leads to PLs that are 10% higher for a given proximity;
- correlation between hailstone size and affected area, Corr(m, A), equal to 0.56, as indicated by the data; and
- other parameters as estimated using the available data.

Some Results

Making one adjustment to the Base Model at a time and running the simulator under the altered design allows sensitivities to be explored. A number of key quantities can be compared across the different adjustments. The results are given in *Table 1* using 100,000 simulations. Estimates are based on the natural statistical estimator of the given quantity. For example, the estimate of the 250-year PML is the 99,600th element of the ordered observations.

Table 1 All quantities are in millions of dollars for $10 billion sum insured

Quantity	Base Model	(a)	(b)	(c)	(d)	(e)	(f)
Avg ann. loss	8.95	6.81	2.63	8.43	9.43	7.80	11.42
Median ann. loss	0.4504	0.5111	0.6975	0.4867	0.4223	0.3530	0.6315
250-year PML*	544	364	57	508	567	477	680
500-year PML	932	535	84	879	1,056	835	1,165
Max. Simulated loss	2,571	2,065	335	2,392	2,845	2,258	3,236

*250-year PML (or Probable Maximum Loss) is, in untechnical language, the greatest loss that is expected to occur over a 250-year period. The loss is exceeded over the coming year with a probability of 1/250.

Adjustments to the base model:

(a) Extreme Value Theory used in the distributions of hailstone size and number of affected postcodes (as in the base model), was replaced by an exponential distribution and distribution based on sample frequencies instead.

(b) Correlation between hailstone size and the area affected, Corr(m, A), equal to zero.

(c) Corr(m, A) = 0.48.

(d) Corr(m, A) = 0.61.

(e) Low sensitivity of insured properties to horizontal wind-damage ($\theta = 0.5$).

(f) High sensitivity of insured properties to horizontal wind-damage ($\theta = 2$).

The quantities given in Table 1 are intended to describe the shape of the distribution of annual losses. For the Base Model we can see that the distribution is highly positively skewed: the average is nearly twenty times larger than the median; and while 50% of observations lie at or below $450,400, there is a 0.4% chance that the observed loss will be greater than $544,000,000. Broadly, a similar pattern in the quantities appears under each adjustment. However, the extent to which the distribution is skewed is clearly affected by the adjustments made. Under the Base Model, for example, approximately 1% of observations lay beyond $335,000,000, while under adjustment (b) there are no observations that exceed this value. This is a highly significant difference considering that it is this very tail of the distribution that is being modelled.

Using the Output of the Model

There are many uses for model output. The insurer may use output to determine the amount of reinsurance cover required, the minimum capital requirements, the price of a product, the locations in which it is optimal to promote sales, and so on. As an example, the insurer's maximum event retention (MER) under the Base Model of the portfolio underlying Table 1 might simply be calculated as the excess of $544 million over the reinsurance cover obtained plus one reinstatement premium of the reinsurance arrangement.[1] The average annual loss might be incorporated as a loading in premiums to cover the long-term cost of catastrophic hail events. The maximum simulated loss gives the insurer an idea of the very greatest loss that could occur, which may or may not be of any use given the limitations imposed by the costs of acquiring reinsurance cover. Running the model with the same assumptions using different portfolios allows the insurer to analyse the impact of changing the geographical location of its exposures. Simulations under different portfolios are presented in the paper proper.

Use of the Adjustments

Suppose the Base Model represents the insurer's best estimate of each of the model components. One use of the adjustments is to see how results differ when the evidence (for example, weather data) leads to alternative estimates. That is, if insurers were all using the same model, but had different approaches to some of its components, the results show how their results would differ. This ignores any variation due to completely different model structure, which is another source of inconsistency between the output of different models.

---

1 Technically all that is required by the relevant APRA Guidance Note (GGN 110.5) is an estimate of the loss to flow from a single event that occurs no more often than 1 in 250 years and involves the peril that produces the highest loss of the perils to which the portfolio is exposed. Supposing the insurer believed that this peril was hailstorm then they might simulate 250-year periods and take an average of the most costly single event that occurs across the simulations. What has been done above is different because the model is producing an estimate of the annual loss distribution. The highest annual loss in 250 years must be at least as great as the highest single event loss in 250 years. Thus we know that the method used here will satisfy the APRA requirements. Note that Blong et al (2001) suggests that hail PMLs based on single event losses are about 10% lower than those based on annual loss figures.

As an example, if two different insurers estimate the distributions of hailstone size and affected area, with one insurer using extreme value theory and the other regular statistical distributions, then they will reach markedly different results as seen in (a). The insurers may want to closely assess the need for the use of EVT by comparing with the available datasets and using judgement where this information is not available.

Another use of output resulting from the adjustments is to assess the need to incorporate elements of conservatism or margins to take account of the risk of model estimation error. Where there is some significant risk that the assumption used is wrong, the modeller may wish to use a more conservative assumption. An example might be adopting a higher correlation between storm affected area and hailstone size to reduce the risk that this quantity has been underestimated.

A third use of the adjustments is to explore the relative importance of assumptions and parameters. For example, moderate changes in θ are not vital to output. The insurer may place less importance on their estimation and avoid expending large amounts of resources on accurately estimating such a parameter.

Plausibility of Output

Because the output of commercial models is by in large confidential information, it is not possible to know exactly how the results in Table 1 relate to real-world results. Some publicised results from the HailAUS model for a residential portfolio of the same sum insured as above indicate 250-year PMLs of around $117 million (Leigh & Kuhnel, 2001). The portfolio used in the paper is a commercial property portfolio and while this would seem to indicate that the figures in Table 1 are broadly too high, it is not possible to draw strong conclusions because of the inherent differences in the nature of the underlying portfolios. Importantly, there is no way of knowing whether the output from a model is correct. All that one can do is compare results and note that, if they are similar to other models, then they are plausible. The 250-year PML being within an order of magnitude of the HailAUS result indicates some level of reasonableness.

Choosing Between Models

The main differences between models are:

- structure and level of complexity;
- assumptions; and
- data differences.

Each of these will be different when modellers are acting in isolation. Until such time as a proper and open assessment of model alternatives can be performed, then there is little guidance available on model choice. In practice it is likely that the following occurs. For the insurer that does not develop its own internal model, the choice is essentially between different “black boxes” because they are unable to make an assessment of the soundness of models. For those insurers that do develop their own model, then the model will be used because of the costs associated with development. Where results from multiple models are available, then some insurers will inevitably adopt those that best suit its agenda, regardless of whether they are based on the soundest modelling. In other cases, some kind of results averaging might be employed.

The essential conclusion of this research with regard to choosing between models is that model selection ought to be driven by co-operation and consensus. There is no simple gauge for what the best model is because the actual answers are unknown. There is, however, potential to compare and contrast the designs and assumptions of different models to reconcile the opinions and resources of modellers. A model should be selected based on how well it is believed the assumptions and structure reflect reality and this is a matter on which the consensus of many opinions should be sought. In the modern commercial world such co-operation occurring without regulatory intervention is an unrealistic expectation. Suggestions for what forms this might take are given in the next section.

IMPLICATIONS for PRUDENTIAL SUPERVISION

The interest that APRA has in catastrophe modelling is driven by the organisation’s objective of ensuring that in all reasonable circumstances the promises of financial institutions are met. Insurers’ practice of using catastrophe models to determine their Maximum Event Retention (MER), forming part of the minimum capital requirements, is of prime concern to APRA. In addition, improved modelling is of interest because it should increase the probability of firm survival and result in more efficient allocations of capital. These issues are highly relevant to APRA’s mission statement.

The implications of the research in regard to APRA are as follows:

- the presence of significant cross-model inconsistencies suggests that insurers facing the same risks may have substantially different minimum capital requirements, which is clearly an unfair and undesirable outcome;
- the apparent near absence of cross-party co-operation in the industry in regard to model design and datasets suggests that regulatory guidance or assistance could be warranted; and
- the position of APRA as an impartial body in which firms may confide makes it an attractive mediator through which firms may better share and communicate.

Approaches that APRA may take to improve the interaction between firms:

1. Sponsor conferences at which providers of modelling services, insurers and so on, share ideas and have the opportunity to arrange exchanges of data or other information; and/or
2. Manage a central data exchange that may be contributed to anonymously by parties. Data would be fully documented and checked and then available to other donors; and/or
3. Obtain authorisation to inspect the workings and inputs of models to determine whether the methodology used is sound and consistent with broad industry practice (much like the practices adopted by the Florida Commission on Hurricane Loss Projection Methodology in the United States).

While these suggestions may seem more interventionist than the stance that APRA has traditionally taken on these 'internal-to-the-firm' issues, it is believed that the added improvement to industry modelling may more than outweigh the disadvantages of any constraints on individual firm behaviour – an issue that would need further exploration.

CONCLUSIONS

Exploring Unknown Quantities is a double-edged title. On one hand the phenomena that parties seek to model are inherently uncertain and therefore the 'true' answers are unknown. However, more pertinently for this paper, the models themselves are unknown quantities because of the closed manner in which practitioners conduct themselves. This paper attempts to contribute to a new era in catastrophe modelling by being open and suggesting that more desirable outcomes might occur if co-operation was more commonplace. Further work by willing parties will ensure that catastrophe models are no longer unknown quantities that need to be 'explored'. This in turn will hopefully lead to more consistent regulation of Australian insurers.

Additional copies of the paper are available on request by telephoning 0421 550 415.

References

Blong, R. et al (2001). Hail and Flood Hazards – Modelling to Understand the Risk. In N. Britton & J. Oliver (Eds.), Enhancing Shareholder Value Through Capital Risk Management, Brisbane, Queensland: Aon Group Australia Ltd and Griffith University.

Kozlowski, R. T. & Mathewson, S. B. (1995). Measuring and Managing Catastrophe Risk. Journal of Actuarial Practice, 3 (2), 211-241.

Leigh, R. & Kuhnel, I. (2001). Hailstorm Loss Modelling and Risk Assessment in the Sydney Region, Australia. Natural Hazards, 24, 171-185.

Walker, G. R. (1997). Current Developments in Catastrophe Modelling. In N. Britton & J. Oliver (Eds.) Financial Risk Management for Natural Catastrophes, Brisbane, Queensland: Aon Group Australia Ltd and Griffith University.

Watson, C. C. & Johnson, M. E. (2003). An Assessment of Computer Based Estimates of Hurricane Loss Costs in North Carolina. Kinetic Analysis Corporation University of Central Florida. Retrieved 20 September 2004, from www.methaz.com/ncdoi/ncdoi_final.pdf