Quantifying Uncertainty in Blending Severe Thunderstorm Model Results with Experience Loss Data for Ratemaking

April 26, 2013

The AIR Institute

AIR WORLDWIDE

Copyright

Trademarks

Contact Information

If you have any questions regarding this document, contact:

Siew Mun Ha
AIR Worldwide
131 Dartmouth Street
Boston, MA 02116-5134
USA
sha@air-worldwide.com
(617) 954-1859

Table of Contents

Introduction

Evaluating catastrophe (cat) loss costs for ratemaking constitutes one of the principal applications of catastrophe models for primary insurers. Cat models, by virtue of their comprehensive stochastic event catalogs, provide a more complete view of cat risk and enable insurers to obtain superior cat-loss estimates than would otherwise be achievable from limited historical loss data alone. However, in some applications, it might be better to supplement the modeled data with historical information. For example, in the case of a high-frequency peril such as severe thunderstorm, even relatively limited historical data contain valuable information about the cat losses and should not be summarily discarded in favor of model results. An enhanced approach to cat loss estimation therefore aims to blend historical data and modeled data optimally to extract the maximum useful information from each.

A 2012 report by David Lalonde (*Blending Severe Thunderstorm Model Results with Loss Experience Data—A Balanced Approach to Ratemaking*) formulated a method to blend historical and modeled severe thunderstorm losses together with non-cat losses to produce a composite estimate of total loss cost for ratemaking. The key idea in this approach was to blend the *modeled* losses above a selected threshold with the *historical* losses below that threshold to determine the cat component of the total losses. This is based on the rationale that there is an unequal quality of information between the tail and non-tail losses in the historical data. While the tail losses are not adequately captured in the historical data due to the low frequency of extreme events, the higher frequency non-tail losses are better captured, and represent a more complete picture consistent with a company's underwriting philosophy. AIR's blending formulation was combined with standard actuarial formulas to derive the following formula for the total loss cost:

$$L = A(1+B) + C \qquad (1)$$

Where:

L = Total loss cost

A = Trended historical non-cat loss cost

$$B = \frac{\sum (\text{Historical cat loss} < T)/N}{\text{Historical non-cat loss}}$$

C = Loss cost for modeled cat loss > T

In this formulation, T is the loss threshold corresponding to a specified exceedance (EP) value and N is the historical data sample size in years.

This blending method was illustrated with a U.S. severe thunderstorm case study, showing the effect of a range of threshold values on the total loss cost estimate. The results naturally raise the practical issue of selecting an appropriate value of T that produces the "best" total loss cost estimate. Assuming that the usefulness of the estimate is inversely related to its uncertainty, one logical approach to selecting T would be to consider the criterion of minimum uncertainty in the total loss cost estimate. This uncertainty is governed by T through three principal components in (1).

The first component originates from the AB term, which represents the cost of the historical cat losses below the threshold. Statistically, the historical cat loss data may be treated as a single sample from a putative infinite population distribution of annual cat losses. AB is thus a single point *estimate* of the true, but unknown, cat loss cost, and consequently, subject to *sampling uncertainty*, which can be estimated through the method of bootstrapping. Other potential sources of uncertainty in the historical data, such as measurement error, are not included here to keep the analysis tractable.

The second and third uncertainty components are the modeled loss uncertainties arising from the C term, which represents the cost of the extreme event (tail) losses derived from the cat model. Cat models, by virtue of their design and structure, encapsulate inherent *primary*, *secondary*, and *sampling uncertainties* in their results. The analysis of primary uncertainty, which arises from the event generation process within the model, is a complex task that is beyond the scope of this paper. On the other hand, secondary uncertainty, which arises from the stochastic variation of a structure's response to a hazard, is readily quantifiable and included as standard features of AIR's Touchstone™ and CLASIC/2™ cat modeling software. The model's sampling uncertainty arises from the finite size of the stochastic catalog and can be estimated through bootstrapping in a similar manner to the estimation of the sampling uncertainty in the historical cat loss term B. For simplicity, we will therefore eliminate primary uncertainty from further consideration and define the modeled loss uncertainty to comprise of secondary and sampling uncertainty only.

In the AIR blending formulation, the threshold T functions as a weighting factor that governs the balance between the historical and modeled data in the total loss cost estimate. A greater value of T increases the weight of historical data relative to modeled data, and vice versa. Consequently, the balance between the historical and modeled data uncertainties also varies with T. The optimum value of T is therefore that which minimizes the total uncertainty in the total loss cost estimate. In this paper, we will formulate a method to estimate each of the uncertainty components in (1) with the objective of finding the optimum T for minimum total loss cost uncertainty.

Mathematical Formulation

The statistical models for the uncertainties in AB and C can be written as:

$$AB = AB^* + \varepsilon \qquad (2)$$

$$C = C^* + \delta + \omega \qquad (3)$$

where AB* and C* are the true, but unknown, values of AB and C, respectively, and ε, δ, and ω are random variables representing the sampling uncertainty in AB*, and secondary and sampling uncertainties in C*, respectively. The statistical model for the total loss cost uncertainty is then:

$$L = A + (AB^* + \varepsilon) + (C^* + \delta + \omega)$$

$$= (A + AB^* + C^*) + (\varepsilon + \delta + \omega) \qquad (4)$$

The total uncertainty in L is therefore the linear sum of ε, δ, and ω. Denoting the uncertainties of L, ε, δ, and ω as σ_L^2, σ_ε^2, σ_δ^2, and σ_ω^2 respectively, we can thus relate the total loss uncertainty function to the threshold T through the uncertainty components in (1):

$$\sigma_L^2(T) = \sigma_\varepsilon^2(T) + \sigma_\delta^2(T) + \sigma_\omega^2(T) \qquad (5)$$

The next step is to determine the functional relationships for $\sigma_\varepsilon^2 = \sigma_\varepsilon^2(T)$, $\sigma_\delta^2 = \sigma_\delta^2(T)$, and $\sigma_\omega^2 = \sigma_\omega^2(T)$ in order to locate the value of T that minimizes $\sigma_L^2(T)$.

Since the putative population of historical annual losses is unavailable, $\sigma_\varepsilon^2(T)$ cannot be estimated by sampling directly from it. However, this difficulty can be overcome by applying bootstrapping to the sample itself to approximate the sampling distribution for ε. Denoting the i'th bootstrap sample of the AB loss cost as X_i ($i = 1 \ldots n$), the sampling uncertainty is then given by:

$$\sigma_\varepsilon^2(T) = \frac{1}{n-1}\sum_{i=1}^{n}\left[X_i(T) - \frac{1}{n}\sum_{i=1}^{n}X_i(T)\right]^2 \qquad (6)$$

The model secondary uncertainty $\sigma_\delta^2(T)$ is readily estimated by utilizing the secondary uncertainty function in AIR's CLASIC/2 software. For each aggregate annual loss, λ_i, the software computes the standard deviation σ_i of its attendant secondary uncertainty. The secondary uncertainty in C is then the sum of the annual loss variances for all losses above the threshold T divided by the earned house years:

$$\sigma_\delta^2(T) = \frac{\sum_i \sigma_i^2}{EHY} \qquad \forall\, \lambda_i > T \qquad (7)$$

The model sampling uncertainty is estimated in a similar fashion to (6):

$$\sigma_{\omega}^2(T) = \frac{1}{n-1}\sum_{i=1}^{n}\left[Y_i(T) - \frac{1}{n}\sum_{i=1}^{n} Y_i(T)\right]^2 \quad (8)$$

Where Y_i (i = 1 … n) is now the i'th bootstrap sample of the loss cost for modeled losses > T.

Equations (5)–(8) provide the framework to quantify the total loss cost uncertainty $\sigma_L^2(T)$ as a function of T, given the historical and modeled loss data for any high-frequency peril.

Case Study Analysis

The dataset for the current case study was constructed by augmenting the AIR (2012) dataset with three additional scenarios corresponding to T = 40% EP ($22.9M), T = 60% EP ($14.6M), and T = 80% EP ($9.12M) in order to extend the analysis range over the entire EP curve. The nominal total loss cost results from AIR's case study (2012) are replicated in Table 1, together with the three new scenarios, and the attendant modeled loss EP curve and historical losses are illustrated in Figure 1 and Figure 2, respectively.

For each scenario, the three uncertainty components $\sigma_{\varepsilon}^2(T)$, $\sigma_{\delta}^2(T)$, and $\sigma_{\omega}^2(T)$ were computed from (6)–(8), and the total uncertainty $\sigma_L^2(T)$ was determined by summing the uncertainty components per (5).

Table 1. Blended loss costs from AIR's case study (2012)

Scenario	A: Non-Cat Loss Cost (AAL/EHY)	B: Experience-Based Cat Load	C: Modeled Cat Loss Cost	A(1 + B) + C: Total Loss Cost
Historical Only	454.6	0.366626	0.000	621.2
T = 5% EP	454.6	0.284259	38.43	622.2
T = 10% EP	454.6	0.232670	52.94	613.2
T = 20% EP	454.6	0.183478	68.88	606.8
T = 40% EP	454.6	0.122654	94.59	604.9
T = 60% EP	454.6	0.085943	99.11	592.7
T = 80% EP	454.6	0.057595	110.8	591.5
Model Only	454.6	0.000000	136.7	591.2

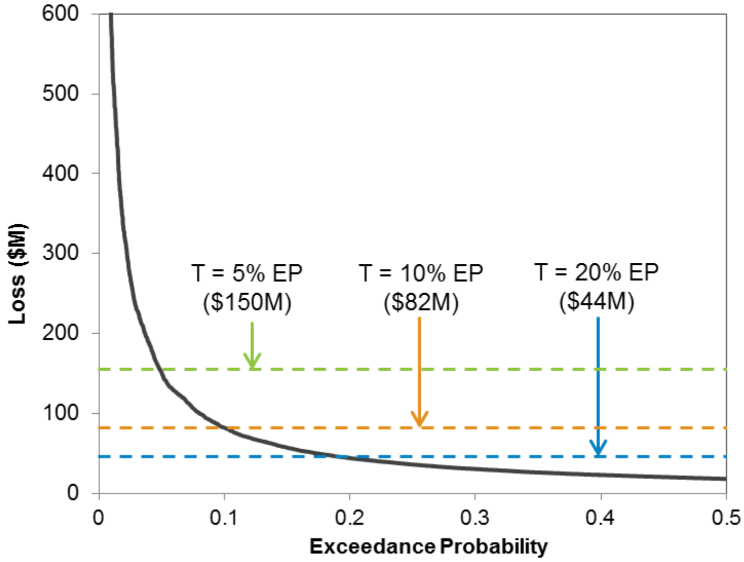

Figure 1. Severe thunderstorm modeled loss EP curve

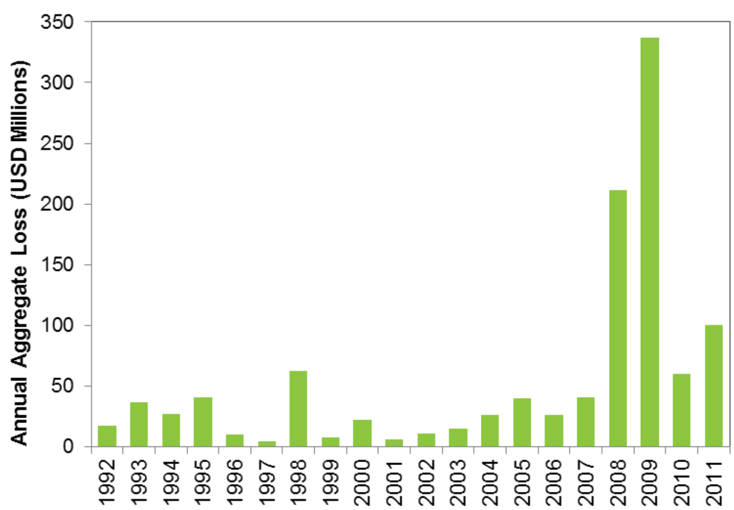

Figure 2. Historical annual aggregate losses

Case Study Results

Figure 3 shows the ε sampling distribution produced by the bootstrapping process for T = 5% EP. The sample estimate of AB*, the experience-based cat loss cost, is 129.2, and its sampling distribution ranges from 50 to 250. The sampling distribution exhibits a moderate degree of skewness but is otherwise well structured and approximately normal.

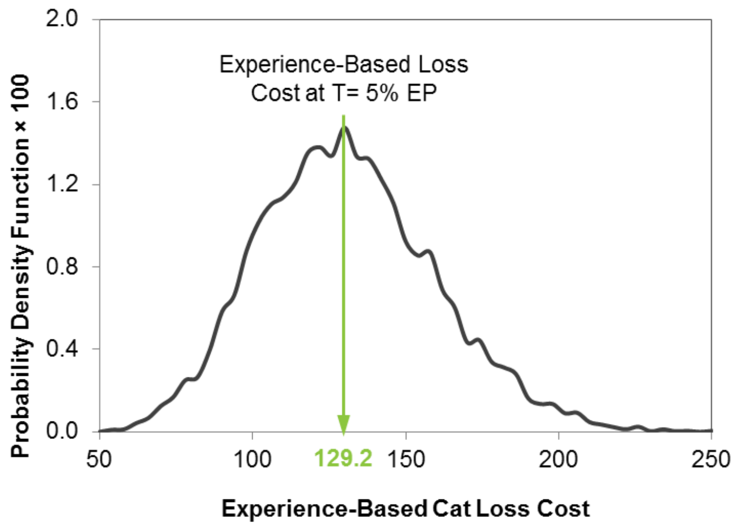

Figure 3. Sampling distribution for experienced-based cat loss cost at T = 5% EP

Table 2 and Figure 4 show the historical sampling, model and total uncertainties over the full range of T. The model uncertainty is the sum of the secondary and sampling uncertainties from the modeled loss cost for cat losses > T. Consistent with their mathematical formulations, the sampling and model uncertainties exhibit opposing trends with respect to T, with the former decreasing and the latter increasing as T increases. The relationship between the two uncertainties is characterized by three distinct intervals. At high threshold values corresponding to T ∈ [0%, 35%] EP, the sampling uncertainty is significantly greater than the model uncertainty, and consequently σ_ε^2 dominates the resultant total uncertainty. The converse is true at low threshold values corresponding to T ∈ [80%, 100%] EP where $(\sigma_\delta^2 + \sigma_\omega^2)$ dominates the result. Lastly, the two uncertainties are comparable in magnitude in the intermediate interval T ∈ [35%, 80%] EP with neither dominating the total uncertainty.

Table 2. Threshold effect on historical, model, and total uncertainty

Scenario	T ($M)	$\sigma_\varepsilon^2(T)$: Historical Sampling Uncertainty	$\sigma_\delta^2(T)$: Model Secondary Uncertainty	$\sigma_\omega^2(T)$: Model Sampling Uncertainty	$\sigma_L^2(T)$: Total Uncertainty
Historical Only	2.827	2880	0.000	0.000	2880
T = 5% EP	150	823.0	11.58	6.940	841.5
T = 10% EP	81.9	297.9	14.70	8.503	321.1
T = 20% EP	44.1	96.73	17.97	9.491	124.2
T = 40% EP	22.9	20.04	21.37	10.36	51.77
T = 60% EP	14.6	4.791	23.38	10.51	38.68
T = 80% EP	9.12	0.691	24.67	10.53	35.89
Model Only	0.630	0.000	25.35	10.75	36.10

The structure of the total uncertainty function $\sigma_L^2(T)$ is consequently shaped by the resultant interaction between the sampling and model uncertainties, the main features of which are their opposing curvatures, and their relative gradients and magnitudes over the three intervals. The result is a convex total uncertainty function (Figure 4) that decreases from $\sigma_L^2(T) = 2880$ to $\sigma_L^2(T) = 36.10$ over the interval $T \in [0\%, 100\%]$ EP with the minimum uncertainty of $\sigma_L^2(T) = 35.89$ located at T = 80% EP (USD 9.12M). The total uncertainty is therefore minimized when the total loss cost is computed with model results for cat losses > USD9.12M and historical data for cat losses < USD9.12M. However, the total uncertainty varies little from its minimum value over the interval $T \in [70\%, 100\%]$ EP, and hence any selected value of T in this range would incur only a small uncertainty penalty.

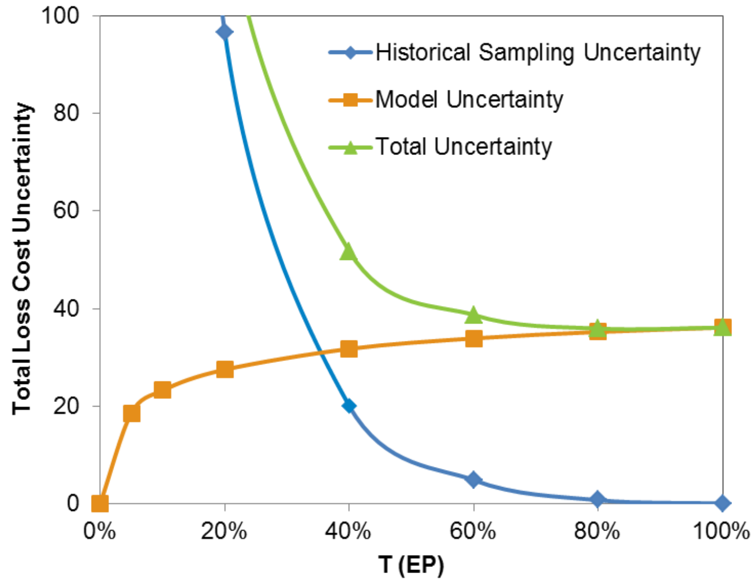

Figure 4. Threshold effect on historical, model, and total uncertainty

Aside from informing the selection of T, the total uncertainty results can also be utilized to compute confidence intervals around the total loss cost estimate to adjust the estimate to a desired level of conservatism relative to the risk of underestimating or overestimating the total loss cost. For example, at T = 5% EP, the standard deviation of the total loss cost uncertainty is $\sigma_L = \sqrt{\sigma_L^2} = \sqrt{841.5} = 29.01$, and a 95% confidence interval for L is $L \pm 1.96\sigma_L \rightarrow 622.2 \pm 56.86$.

Therefore, to be 97.5% confident that the total loss cost is not underestimated, the estimate should be adjusted from the nominal value of L = 622.2 to L = 622.2 + 56.86 = 679.1. Similarly, an adjusted total loss cost of L = 622.2 - 56.86 = 565.3 ensures that there is a 97.5% probability that the total loss cost is not overestimated.

Figure 5 summarizes the global analysis process flow to compute the uncertainties, select T, compute the confidence interval, and adjust the total loss cost estimate.

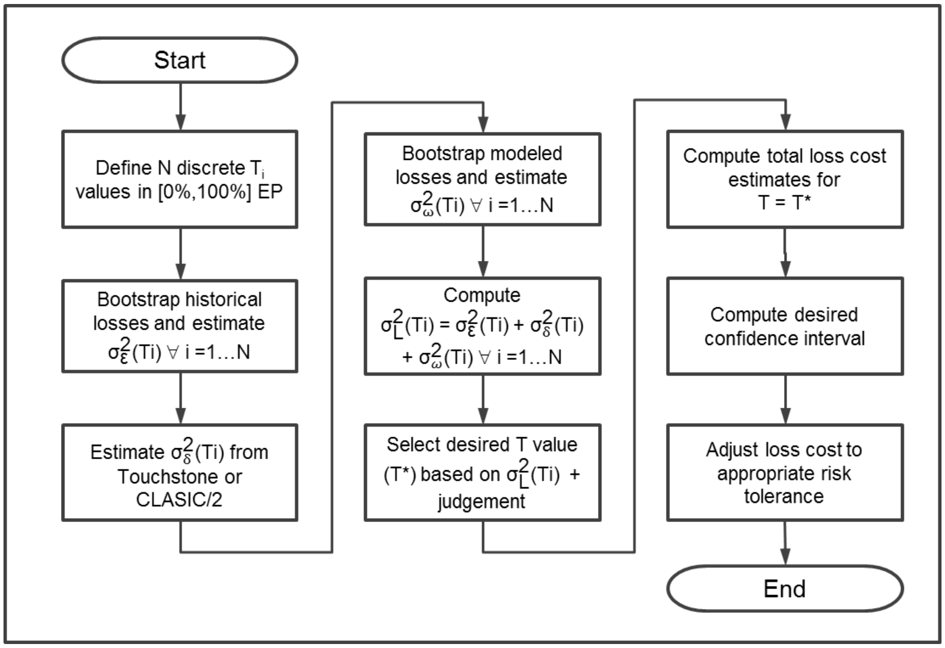

Figure 5. Global analysis process flow

Conclusion

We have developed a framework to quantify the three principal uncertainty components in the AIR model blending method as a function of the threshold parameter. This extends the utility of the AIR method by providing an estimate of total uncertainty, as well as an objective basis to determine a range of suitable threshold values with low total uncertainties. The method was tested with an augmented set of the AIR (2012) case study data and produced satisfactory results. For the current dataset, small total loss cost uncertainties were obtained with T ∈ [70%, 100%] EP with the minimum total uncertainty at T = 80% EP or USD 9.12M. These results will vary in general with different datasets.

Aside from informing the selection of a suitable threshold T for the blending process, the results can also be deployed to compute total loss cost confidence intervals so that the nominal estimate can be suitably adjusted to achieve an appropriate level of risk tolerance.

Nomenclature

A	Trended historical non-cat loss cost
AB*	True value of AB
B	Estimated experience-based cat load = $\frac{\sum(\text{Historical cat loss}<T)/N}{\text{Historical non-cat loss}}$
C	Estimated loss cost for modeled cat loss > T
C*	True value of C
EHY	Earned house years
L	Estimated total loss cost
T	Loss threshold corresponding to a specified EP value
T*	Selected value of T
N	Historical data sample size (years)
N*	Count of selected discrete T values within [0% 100%] EP
n	Count of bootstrap samples
N_C	Catalog size
X_i	i'th bootstrap sample of loss cost for historical cat loss < T
Y_i	i'th bootstrap sample of loss cost for historical cat loss > T
δ	Model secondary uncertainty random error
ε	Historical cat loss sampling random error
λ_i	Modeled annual loss for year i
σ_i	Standard deviation of secondary uncertainty for year i
$\sigma_L^2(T)$	Total uncertainty
$\sigma_\delta^2(T)$	Model secondary uncertainty
$\sigma_\varepsilon^2(T)$	Historical cat loss sampling uncertainty
$\sigma_\omega^2(T)$	Model sampling uncertainty
ω	Model sampling random error

References

Lalonde, D. (2012), *Blending Severe Thunderstorm Data with Loss Experience Results – A Balanced Approach to Ratemaking*, AIR Currents (http://www.air-worldwide.com/Publications/AIR-Currents/2012/Blending-Severe-Thunderstorm-Model-Results-with-Loss-Experience-Data%E2%80%94A-Balanced-Approach-to-Ratemaking/).

About AIR Worldwide

AIR Worldwide (AIR) is the scientific leader and most respected provider of risk modeling software and consulting services. AIR founded the catastrophe modeling industry in 1987 and today models the risk from natural catastrophes and terrorism in more than 90 countries. More than 400 insurance, reinsurance, financial, corporate, and government clients rely on AIR software and services for catastrophe risk management, insurance-linked securities, detailed site-specific wind and seismic engineering analyses, and agricultural risk management. AIR is a member of the Verisk Insurance Solutions group at Verisk Analytics (Nasdaq:VRSK) and is headquartered in Boston with additional offices in North America, Europe, and Asia. For more information, please visit www.air-worldwide.com.