Empirical Method-Based Aggregate Loss Distributions

C. K. "Stan" Khury

2012

INTRODUCTION

Aggregate Loss Distributions: Formulaic or Empirical

** Formulaic Distributions: Select Form and Estimate Parameters

** Empirical Distributions: Extract Implied distribution from the data

METHODS GENERALLY
< GENESIS >

Given an array of historical loss development, several approaches are available. For example, among the most common approaches:

A. Use a Loss Development Method (LDM)
B. Use a variant of LDMs: B-F, Berquist Sherman
C. Model the actual data, and work with model outputs
D. Accept the historical data (as fixed) and model the outputs
E. Use some combination of C and D
F. And there are other approaches

Regardless of the choice of approach/method, there is a universal unstated premise: *any result that one can derive is a member of a collection of many possible similarly derived outcomes*. This paper is focused on the results produced by the LDM and its first cousins (the B-F and B-S methods) and is treated using the approach described in D above.
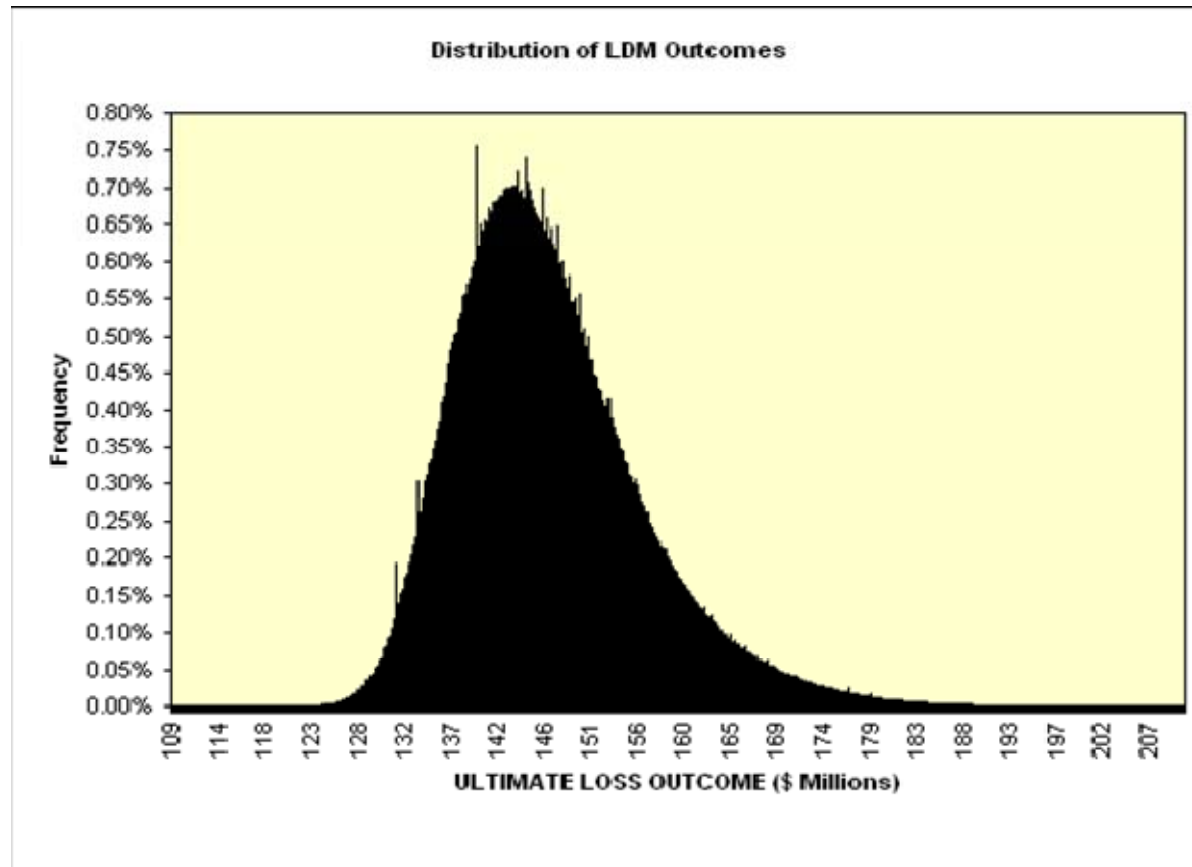
BOTTOM LINE

INPUT

A data set (an array of historical loss development values for a given cohort of claims)

OUTPUT

The unique deterministic distribution of all possible outcomes produced by the application of the Loss Development Method to the given data set

INPUT

Acc.	Number of Years of Development									
Yrs.	1	2	3	4	5	6	7	8	9	10
1996	2.08	3.65	4.88	5.35	6.38	6.88	7.09	7.16	7.19	7.20
1997	2.35	3.81	4.79	6.23	7.44	7.43	7.62	7.82	8.16	8.16
1998	2.70	5.14	6.44	8.88	9.35	9.65	10.17	11.06	11.03	11.30
1999	3.24	7.52	10.92	11.59	14.65	15.67	17.30	16.68	16.88	16.88
2000	2.84	6.36	10.62	11.89	13.56	16.90	17.40	17.96	18.02	
2001	2.40	7.01	7.82	10.58	13.04	13.86	14.36	15.03		
2002	4.26	3.96	7.71	10.70	13.27	14.06	15.02			
2003	1.78	6.07	10.03	12.06	14.02	15.06				
2004	3.25	6.09	11.03	13.56	16.32					
2005	2.59	3.89	8.05	11.13						
2006	2.76	4.03	9.58							
2007	3.15	3.88								
2008	3.25									

OUTPUT

Distribution of LDM Outcomes

The Loss Development Method
Standard Application

- Given a data set
- Calculate historical LDFs
- Select an LDF for each development period
- Calculate a loss development pattern
- Apply loss development pattern to all years
- Produce a single estimate
- Many answers are possible

The Natural Solution

A. Obtain the distribution by calculating all possible outcomes

B. Create a histogram and associated frequency distribution

This is easier said than done. For example, take an array shaped like a parallelogram, with n values to each side, this array yields $(n-1)^{[n(n-1)/2]}$. These values grow very rapidly as n increases:

When n is 10, the number of outputs is 8.7×10^{42} (2.8×10^{26} YRS)

When n is 15, the number of outputs is 2.2×10^{120} (7.0×10^{103} YRS)

There just is not enough time to do these calculations. It is literally an impossible task. *__This paper presents an algorithm to approximate the distribution of outcomes to within any given error tolerance, ε.__*

FRAMEWORK

A. Set max and min values of the distribution for each component years.

$$\text{Max Value} = V * \Pi(\text{Max \{LDF\}})$$
$$\text{Min Value} = V * \Pi(\text{Min \{LDF\}})$$

B. Add all MIN values for all years to obtain the overall MIN of distribution.

C. For each year, define a number, $N(i)$, of equal subintervals spanning the (MIN, MAX) interval for year i.

D. Set the width of the subinterval such that the radius of the subinterval when divided by the lower bound of the (MIN, MAX) interval is less than ε.

E. This construction guarantees that any one ultimate value of year i, when placed along the interval (MIN, MAX) range is within ε of the midpoint of the subinterval in which it falls.

F. Once the $N(i)$ value has been determined for each year, select the maximum element in the set $\{N(i)\}$. This value, simply designated by N, is the number of subintervals that will be used in the construction of the ultimate all-years-combined histogram.

SUBINTERVAL CONSTRUCTION for YEAR i ASSURING THE ERROR CONDITION IS MET

← R A D I U S →

MIN I_____ … _____I_____V_____I____…_____I MAX
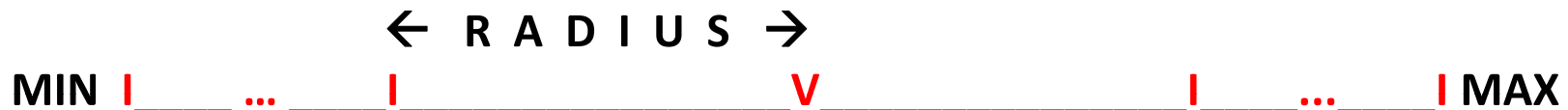
Actual formula that delivers this construction is derived and documented in the article itself.

With all subintervals constructed, now calculate all outcomes for any one year. For every value thus produced, substitute the midpoint of the subinterval which contains the actual value for the actual value. The result is a histogram (frequency distribution) of all values produced by the LDM for any individual year. (NOTE: Process takes less than ten minutes on a desktop computer. A 10X10 array will generate approximately 400 million outcomes.)

SUBINTERVAL CONSTRUCTION FOR ALL YEARS COMBINED

A. From the previous construction, we have in hand a set of subintervals for each open year, along with an associated frequency. Thus we have in hand every midpoint of the subintervals associated with every single year's distribution.

B. Construct a new set of midpoints: the nth new midpoint is equal to the sum of the nth midpoints of all the component subintervals of the underlying distributions.

C. Construct the convolution distribution of all the underlying distributions, and do so iteratively, using the first two distributions, adding a third, a fourth, and so on until all distributions have been accounted for.

D. Each outcome of this convolution distribution is replaced by the midpoint of the subinterval (of the overall distribution) in which it falls.

E. The mathematical demonstration that such substitution actually meets the requirement that any actual value of the convolution distribution is within ε of the midpoint is shown in the published article.

SUMMATION

A. We start with an array of loss development data (paid, incurred, counts, amounts, etc.)

B. Construct the master overall interval that contains all possible values produceable by the LDM that assures the error condition is met.

C. Apply the LDM in its most general form (i.e., permute all possible LDFs) to create all possible outcomes to within a pre-assigned tolerance ε.

D. The end result is a frequency distribution of all possible outcomes indicated by the history.

E. The statistics of the distributions are easily calculated (mean, variance, SD, etc.)

EXTENSIONS

A. The most immediate obvious extensions are:

    1. <u>B-F method</u> (Two types)

        a. IEL is a single value.

        b. IEL is drawn from a LR distribution.

    2. <u>The B-S methods</u>.

B. <u>Tail Factors</u>. One or more tail factors can be added as a final diagonal of values.

C. <u>Weighting the LDFs</u>.

D. <u>Outliers</u>. Clear outlier LDF values can be managed easily.

APPLICATIONS, COMMENTARY, & LIMITATIONS

1. <u>Statistics of the Distribution</u>. Directly derivable.
2. <u>The Reserve Decision</u>. Role of the actuary.
3. <u>Benchmarking</u>:
   - A. Flash benchmarking
   - B. Longitudinal benchmarking
4. <u>Bootstrapping</u>. Results consistent with outputs of bootstrapping methods. Advantage in communications.
5. <u>Limitations</u>. A number of limitations and cautions apply when using this methodology:
   - A. Main objective is to identify inherent variability.
   - B. Secondary purpose is to identify a default reserve indication.
   - C. The output is a conditional distribution.
   - D. No recognition of model risk.
   - E. No recognition that the data set is a sample.
   - F. All assumptions underlying the LDM carry forward intact.