A Stable Method for Determining Tail Factors Using Age to Age Development Factors

In traditional loss development calculations, one difficulty is in selecting a factor to develop losses from the end of the triangle to ultimate, referred to as the tail factor. Some methods typically used to estimate this factor are:

- Study more extended loss development triangles (e.g. use a twenty year triangle instead of a ten year triangle).
- Fit parametric curves to loss development data.
- Repeat the final age to age factor as the tail factor (assumes each age to age factor after a particular point is half the size (relative to no development of the preceding factors).
- Generalize the previous method to use a decay factor other than one half, applied to the last age to age factor.

The method outlined in this paper is related to the method of selecting a decay in age to age factors, but generalizes further to use more than the last age to age factor when determining the tail factor. It relies on run tests to determine which age to age factors to use in the calculation.

The method summarized is as follows:

- Calculate the logarithm of each age to age factor
- Calculate the decay of the logarithm from each period to the next
- Use run tests to ascertain changes in rate of decay.
- Reduce the data (eliminating earlier age to age factors) until run test does not reject the null of no change in rate of decay
- Select median of remaining decays.
- Use median decay with each of the remaining age to age factors to estimate a tail factor (logarithm)
- Select the median of these tail factor estimates.
- Calculate tail factor exp(log)

Formal Description of the Algorithm

$A_i \sim$ Development factor from age i to age i +1.

$A_t \sim$ Tail factor (to be determined)

$$L_i = \ln A_i$$

$$D_i = \frac{L_{i+1}}{L_i}$$

Let j denote the smallest number for which a run test on the ordered set of decay factors $\{D_i : i > j\}$ does not reject the hypothesis of randomness.

$$\widehat{D} = \underset{i>j}{\text{median}}\, D_i$$

$$T_i = L_i * \frac{\widehat{D}^{t-i}}{1-\widehat{D}} \quad \forall i > j$$

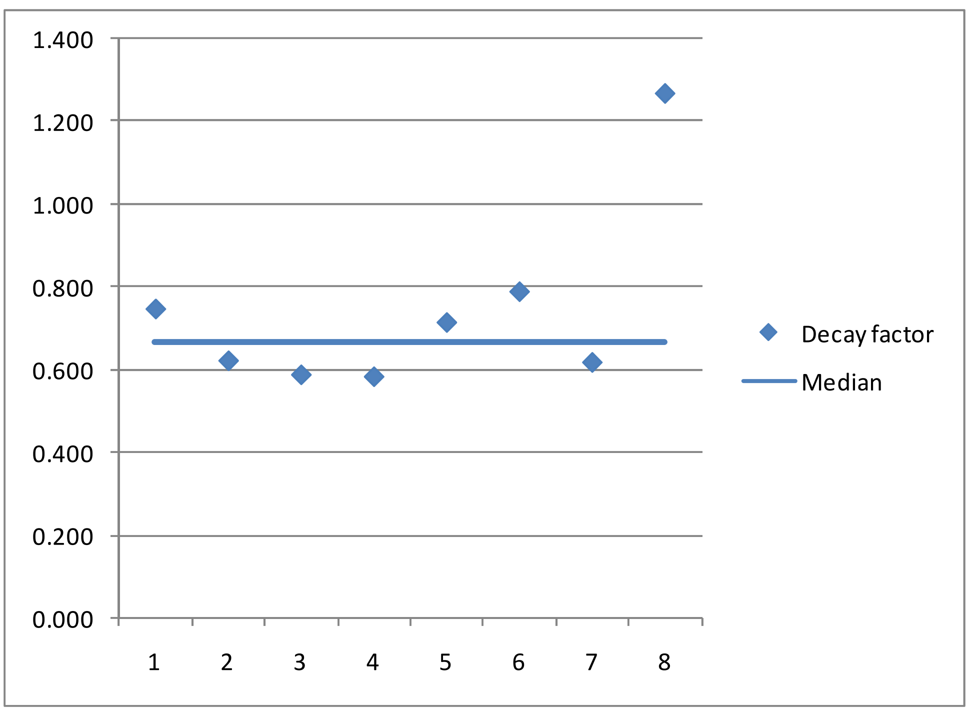
$$\hat{T} = \underset{i>j}{\text{median}}(T_i)$$

$$A_t = e^{\hat{\hat{T}}}$$

Example 1:

Age	Age i to Age i+1 Factor (A)	Ln (A)	D	T
1	1.802	0.589	0.746	0.047
2	1.552	0.440	0.621	0.052
3	1.314	0.273	0.587	0.048
4	1.174	0.160	0.583	0.043
5	1.098	0.093	0.714	0.037
6	1.069	0.067	0.788	0.040
7	1.054	0.053	0.617	0.047
8	1.033	0.032	1.267	0.044
9	1.042	0.041		0.083
		Median	0.667	0.047
		Tail Factor		1.048

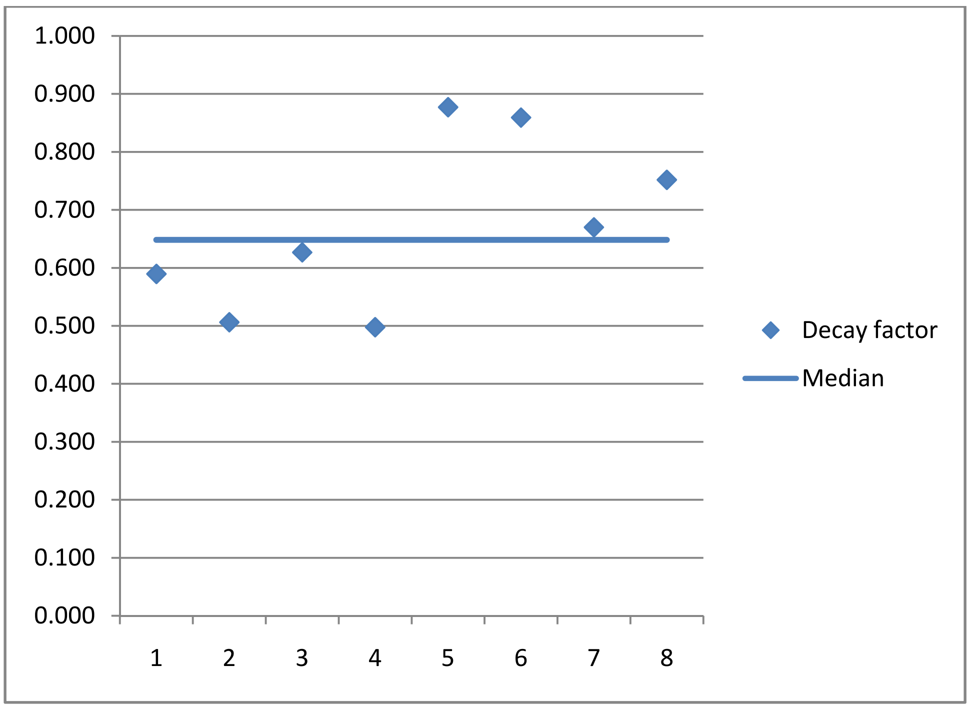
A graph of all decay factors D, together with the median of these factors is shown below:



In this example, there were five runs above and below the median (+, - - - ,++,-,+). This easily passes a run test for randomness with a 62.86% probability of five or fewer runs, and therefore no data points were discarded (see probability table at the end of this paper).

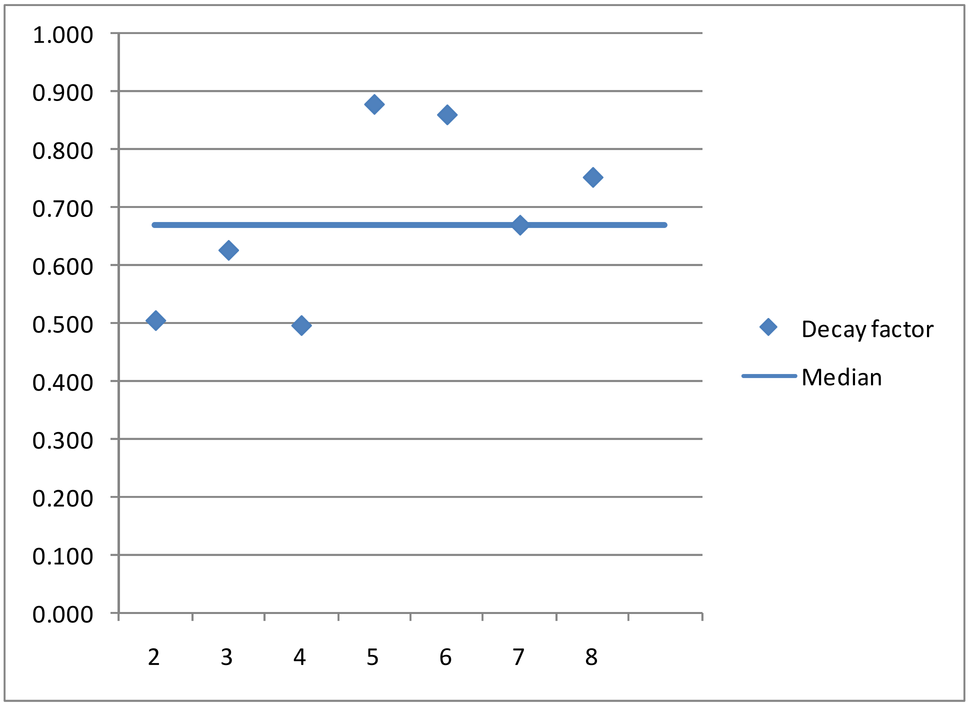
Example 2

Age	Age i to Age i+1 Factor (A)	Ln (A)	D	T
1	1.524	0.421	0.590	0.024
2	1.282	0.248	0.506	0.022
3	1.134	0.126	0.627	0.017
4	1.082	0.079	0.498	0.017
5	1.040	0.039	0.877	0.013
6	1.035	0.034	0.859	0.017
7	1.030	0.030	0.670	0.023
8	1.020	0.020	0.752	0.024
9	1.015	0.015		0.027



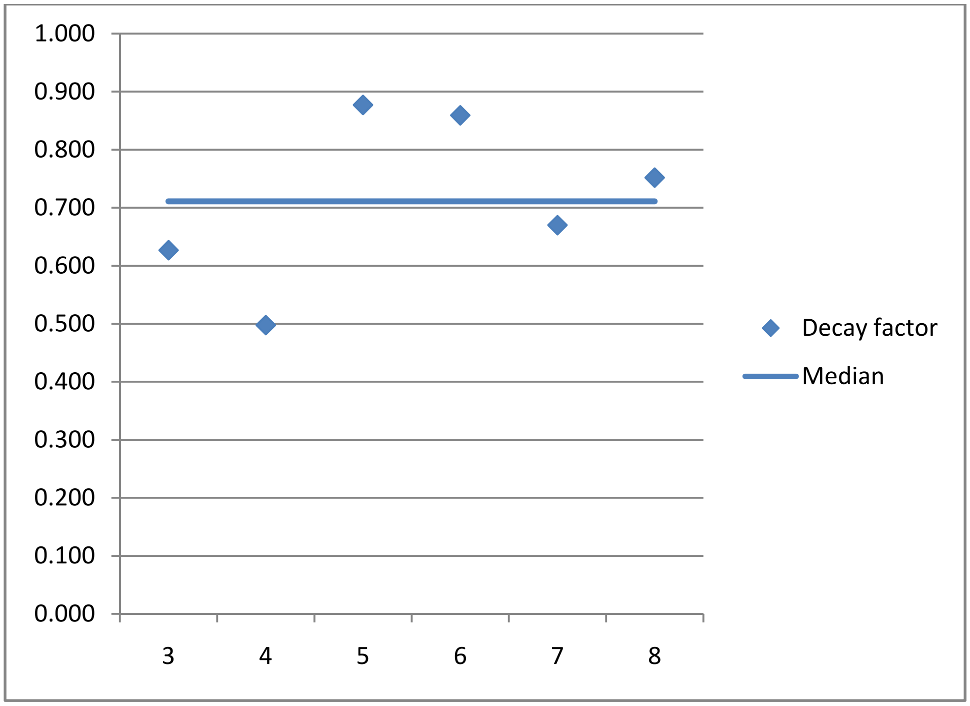
In this case, the decay factors have two runs (----,++++). With eight data points, the probability of observing only two runs with randomly distributed data is only 2.86%.

If we remove the first data point (furthest from the tail factor) and recalculate the median and the run test we have the following:



At this point we still have only two runs (---,+++). (The point exactly equal to the median is discarded for this test.) However, the probably of observing two runs with randomly distributed values in this case is higher (10%), since we have a smaller number of points. Whether or not this hypothesis of randomness is rejected with a 10% significance is a question of how reactive to such differences the analyst would like to be, perhaps factoring in subjective beliefs about reasons why the decay rate would be changing.

Assuming that the randomness hypothesis is still rejected, we would eliminate yet another point from the beginning of the series:



With the elimination of the second data point, we now have 4 runs (--,++,-,+). With 6 data points, there is a 70% probability of observing 4 or fewer runs, so we will stop here.

Using the median decay factor of .711 and each of the remaining data points, we have the following:

Age	Age i to Age i+1 Factor (A)	Ln (A)	D	T
3	1.134	0.126	0.627	0.040
4	1.082	0.079	0.498	0.035
5	1.040	0.039	0.877	0.025
6	1.035	0.034	0.859	0.030
7	1.030	0.030	0.670	0.037
8	1.020	0.020	0.752	0.035
9	1.015	0.015		0.037

The median of the values for T is 0.035, with a resulting tail factor estimate of 1.036.