



SOCIETY OF ACTUARIES

Article from:

Risk Management Newsletter

November 2004 – Issue No. 3

2004 Bowles Symposium: Extreme Value Seminar

by Steve Craighead

Our seminar at the Bowles Symposium was so packed with intense mathematical formulae and discussions of complicated material that you might feel as if you just took a 10-week graduate course in three hours!

However, in those three hours we covered the fundamentals of the use of extreme value theory in risk management. Here we discussed domains of attraction (this describes whether a statistical distribution has a light, medium or heavy tail). We examined various distributions that one would fit to model these domains of attraction such as Fréchet, Weibull, Gumbel and generalized extreme value distributions. We examined ideas of conditional tail expectation (CTE) (just think of this as an average of claims over a threshold). This discussion also included methods to estimate the required number of scenarios for different levels of CTEs. We covered many other topics as well, that are related to understanding and modeling risks.

Dr. H. N. Nagaraja, a world expert in extreme value statistics who is from Ohio State University, was the primary speaker. His presentation was excellent (notwithstanding the complexity). His paper is one of my favorite references as I create new applications using extreme value statistical techniques at my company.

I made two contributions. First, I punctuated Dr. Nagaraja's presentation with brief exposés of insurance applications of the current topic under discussion. Second, I demonstrated a spreadsheet that I have used over the past 12 years to implement some of the extreme value statistical techniques that Dr. Nagaraja presented.



Steve Craighead, ASA, MAAA, is an assistant actuary at Nationwide Financial in Columbus, Ohio. He can be reached at craighs@nationwide.com.

This sheet is fairly simple. First the user inputs the data he or she wishes to examine (starting in cell C22 of the sheet). When the user processes the sheet, it provides two results. The first is that it gives the domain of attraction (remember this is how much area is in one tail of a distribution) as reported in cell C17¹. The second result is that it allows the user to input a list of various extreme percentiles (such as 0.01 percent) starting in cell H22 and it will extrapolate the user's data to estimate these percentiles. This extrapolation is done either by a kernel method² or resampling³. A kernel method is where one fits a specific statistical distribution (the kernel) locally over a small subset of the data. This process creates something very similar to a moving weighted average of the data. The size of the local subset used is determined by a distance called a bandwidth. For the actuaries that have used graduation, this is very similar to the ideas used in the Whittaker-Henderson graduation method. Resampling is where one uses their data over and over (resampling) to determine specific statistical values, in this case the extreme percentiles.

The user chooses which kernel by entering a "1" in either cells C4, C5 or C6, which are normal, lognormal or Pareto. Note: Lognormal is most frequently the best kernel.

If the user wants to use resampling, enter a "1" in the resampling cell C7.

Note: The options of fitting the Fréchet, Weibull and Gumbel distributions are not currently active in the spreadsheet.

Next the user should specify how many elements are in the data in cell C12. The sheet does not set an express limit on the number of

¹ The algorithm is described in the article "The Selection of the Domain of Attraction of an Extreme Value Distribution from a Set of Data," by Castillo, E., Galambos, J., and Sarabia, J. M. This is a chapter in *Lecture Notes in Statistics Vol. 51: Extreme Value Theory* edited by J. Husler and R. Reiss and published by Springer-Verlag.

² Dan Heyer, FCAS, introduced me to the use of kernel methods in extreme value estimations in reinsurance.

³ This method was introduced by D. Zetterman, in the article "A Semi-parametric Bootstrap Technique for Simulating Extreme Order Statistics." This is an article in the June 1993 *Journal of the American Statistical Association* (Vol. 88, No. 422, pp. 477-485).

elements that you can input, however, you may need to revise the diagnostic graph to reflect more than 999 elements.

In cell C13, the user inputs where they believe that the tail of the distribution begins. I have stolen a reinsurance term here and call this the attachment point. This is entered as a percent. Frequently this is a lower percent, which is contained within the data provided. Most frequently reinsurance companies use 1 percent or 2 percent. If you set this value too high, your extrapolations may be too conservative.

In cell C14, enter the number of extreme percentiles that you wish to extrapolate. Recall these were input starting in cell H22.

If you are using resampling, cell C15 is where you specify how many scenarios that you wish processed.

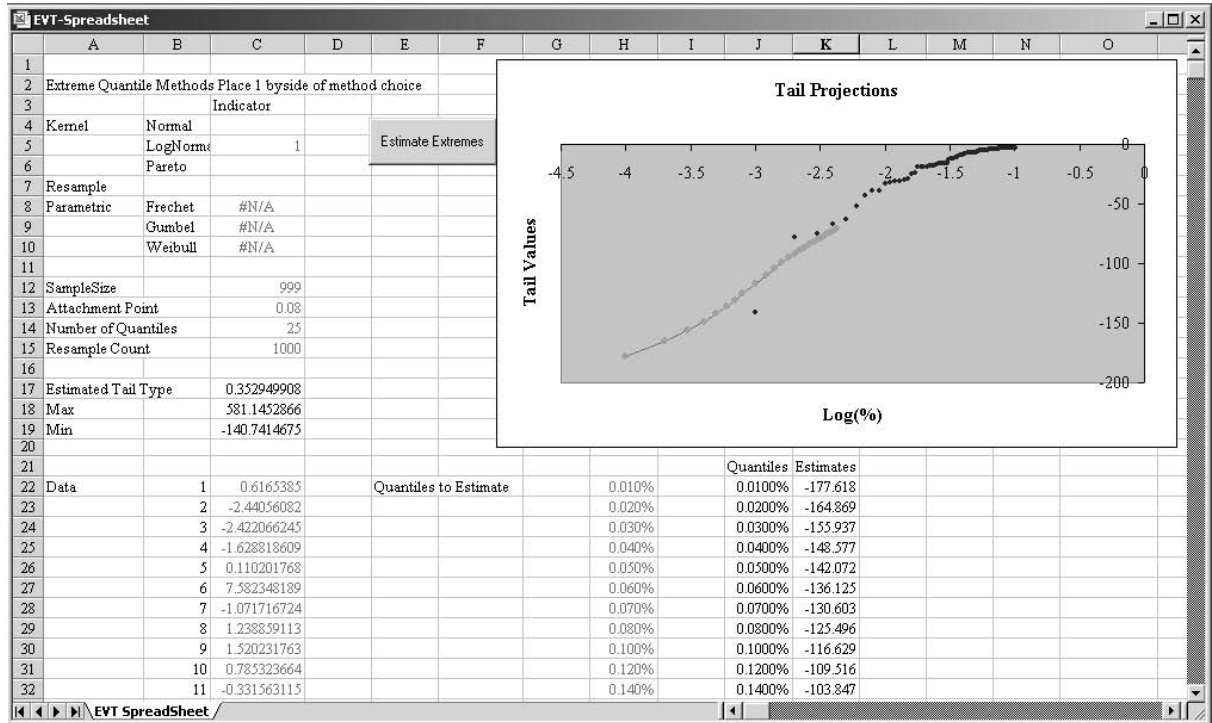
When ready, you execute by pressing the estimate extremes button. Three things then occur.

The first result is that the estimated tail type is updated. If the value is greater than 2, the tail is light. If the value is between 0.5 and 2, the tail has medium weight. If it's below 0.5, the tail is heavy.

The second result is that the extrapolated estimates are placed in cell K22 and below.

The third result is that the diagnostic graph is updated. This is a graph of the 10 percent left tail (in black dots) with the extreme percentiles (purple dots with a blue line) superimposed. The x-axis is the log of the percent.

By playing with your choice of the kernel or resampling, you can examine your own data.



I have used extreme value theory in determining the riskiness of a product. This is indicated if the Estimated Tail Type is less than 0.5.

Other areas where the extrapolation was critical were insolvency analysis and capital needs. Most of our computer financial models are highly accurate, but extremely slow to process. Here I have used extrapolation to obtain extreme percentiles that otherwise may take days or months to obtain from an existing model. The model could also be used to estimate extreme claims, if one would subtract the claim amounts from a very large number. This converts a right tail problem into a left tail problem. By then mapping the threshold by the same method and estimating many extreme percentiles, one can then average these and determine a mean claims over a threshold.

Both Dr. Nagaraja's excellent reference paper and my spreadsheet are available on the CAS 2004 Bowles Symposium Web page (look for Track E) at <http://www.casact.org/coneduc/erm/2004/handouts/>. My spreadsheet is also available on the Risk Management Task Force Extreme Value Subcommittee Web page: http://rmtf.soa.org/rmtf_evm.html. Any future updates of my worksheet will be placed on the latter Web site. ♦