A Note on the Upper-Truncated Pareto Distribution

David R. Clark, FCAS

Abstract

The Pareto distribution is widely used in modeling losses in Property and Casualty insurance. The thick-tailed nature of the distribution allows for inclusion of large events. However, in practice it may be necessary to apply an upper truncation point so as to eliminate unreasonably large loss amounts and to ensure that the first and second moments of the distribution exist.

This paper provides some background on the characteristics of the upper-truncated Pareto distribution, and suggests some diagnostics, based on order statistics, to assist in selecting the upper truncation point.

Keywords. Enterprise Risk Management, Pareto, Truncation, Order Statistics

1. INTRODUCTION

The Pareto distribution is useful as a model for losses in Property and Casualty insurance. It has a heavy right tail behavior, making it appropriate for including large events in applications such as excess-of-loss pricing and Enterprise Risk Management (ERM).

For applications in Enterprise Risk Management, however, there may be practical problems with the Pareto distribution because non-remote probabilities can still be assigned to loss amounts that are unreasonably large or even physically impossible. Further, a Pareto distribution with shape parameter $\alpha < 2$ will not have a finite variance, meaning we cannot calculate a correlation matrix between lines of business. In practice, an upper truncation point (T) is introduced and losses above that point are not included in the model. This upper truncation point may be considered the "Maximum Possible Loss" (MPL).

The difficulty for setting the upper truncation point is that the true Maximum Possible Loss for a given risk portfolio may not be easily determined. Analysts may hold different opinions as to what is possible.

In Enterprise Risk Management models, one goal is to evaluate the "tail" of the distribution, which can be very sensitive to the selection of the upper truncation point.

The goal of this paper will be to describe the characteristics of the upper-truncated Pareto and to offer some measures that may be useful in selecting the upper truncation point based on the sample of historical loss data. Some of these measures are results taken from the field of order statistics. We will not eliminate the need for the analyst to make an informed judgment when selecting the upper truncation, but we can give some objective measures to assist in making that judgment more informed.

1.1 Research Context

The literature on the Pareto distribution is vast. The text by Johnson et al. [5] provides the standard overview including historical genesis of the mathematical form, key characteristics and a comprehensive bibliography. Within the Casualty Actuarial Society literature, the 1985 paper by Stephen Philbrick is a recommended introduction and includes a brief discussion of upper truncation.

Our primary focus will be those characteristics of the Pareto distribution, particularly order statistics, that will be most useful for the Enterprise Risk Management application.

Order statistics is a branch of statistics that has grown over recent decades. It is concerned with inferences from an ordered sample of observations. In the CAS literature, an introduction to this topic related to estimating Probable Maximum Loss (PML, as distinguished from MPL) is given by Wilkinson (1982).

Extreme Value Theory (EVT) has developed as a branch from order statistics, with attention given to the distribution of the largest value of a sample. Much of EVT deals with approximations to the distribution of the largest value assuming the original distribution form is unknown.

1.2 Objective

The objective of this paper is entirely practical: given that the upper-truncated Pareto is widely used in insurance applications, we wish to supply analysts with additional information for selecting the upper truncation point.

1.3 Outline

The remainder of the paper proceeds as follows:

Section 2 will discuss the characteristics of the upper-truncated Pareto distribution itself.

Section 3 will review the maximum likelihood method for estimating the model distribution parameter.

Section 4 will introduce order statistics related to the upper-truncated Pareto and how they can be useful for selecting the upper truncation point.

Section 5 will present two brief examples to illustrate the technique of estimating the upper truncation based on the order statistics for the largest loss.

2. CHARACTERISTICS OF THE UPPER-TRUNCATED PARETO

2.1 The [Untruncated] Single Parameter Pareto

The cumulative distribution function for the Pareto distribution is given below in formula (2.1). This form represents losses that are at least as large as some lower threshold, θ, following the notation in Klugman et al. This form is sometimes referred to as the "single parameter Pareto" with shape parameter α and a lower threshold used to define the range of loss amounts supported (θ is not considered a parameter). Sometimes this form of the distribution is referred to as a "European Pareto" (see Rytgaard, 1990) to distinguish it from the two-parameter form. An alternative form uses a shift $Y = X - \theta$, representing just the portion of the excess loss above the threshold and the theta θ treated as a scale parameter. For the remainder of this paper we will only consider the single parameter or "European" form of the distribution.

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^{\alpha} \qquad \theta \leq x, \qquad \alpha > 0 \tag{2.1}$$

The moments of the unlimited Pareto distribution are given as follows.

$$E(X^k) = \frac{\alpha \cdot \theta^k}{\alpha - k} \qquad \alpha > k \tag{2.2}$$

We note that not all moments exist for the Pareto distribution. For example, when $\alpha \leq 1$ the expected value is undefined.

2.2 The Upper-Truncated Pareto

When we introduce an upper truncation point, T, the random variable for loss can only take on values between the lower threshold and the upper truncation point. It is also interesting to note that the shape parameter, α, can now be any real value and is no longer restricted to being strictly positive.

$$F(x) = \begin{cases} \dfrac{1-\left(\frac{\theta}{x}\right)^{\alpha}}{1-\left(\frac{\theta}{T}\right)^{\alpha}} & \theta \leq x \leq T, \qquad \alpha \neq 0 \\ \\ \dfrac{ln(x/\theta)}{ln(T/\theta)} & \theta \leq x \leq T, \qquad \alpha = 0 \end{cases} \tag{2.3}$$

For the special case of $\alpha = -1$, the distribution of losses is uniform between θ and T. This may be surprising given that most insurance applications are heavily skewed and restrict the shape parameter to positive values but it does show the flexibility of the truncated form. Negative alphas are theoretically valid but unusual in insurance applications; we will be concerned in this paper mainly with cases for $\alpha > 0$.

All moments for the upper-truncated Pareto will always exist.

$$E(X^k) = \frac{\alpha \cdot \theta^k}{\alpha - k} \cdot \frac{1-\left(\frac{\theta}{T}\right)^{\alpha-k}}{1-\left(\frac{\theta}{T}\right)^{\alpha}} \qquad \alpha \neq 0, k \tag{2.4}$$

We may note that for the values $\alpha = 0$ and $\alpha = k$, formula (2.4) does not hold directly but we can estimate the moments by making use of the limiting function below.

$$\lim_{\alpha \to 0} \frac{1-\left(\frac{\theta}{T}\right)^{\alpha}}{\alpha} = \lim_{k \to \alpha} \frac{1-\left(\frac{\theta}{T}\right)^{\alpha-k}}{\alpha - k} = ln(T/\theta) \tag{2.5}$$

To provide some additional insight into the shape of the upper-truncated form, we consider the expected values for some special cases. The value $\alpha = -1$ produces a uniform distribution for which the expected value is the mid-point or arithmetic average between θ and T. For the value $\alpha = 1/2$ the expected value is the square-root of the product of θ and T, also known as the geometric average. For the value $\alpha = 2$ the expected value is the harmonic average of θ and T, found by averaging their inverses.

Shape Parameter	$E(X\|\theta \le X \le T)$
$\alpha = -1$	$(T+\theta)/2$
$\alpha = 0$	$(T-\theta)/ln(T/\theta)$
$\alpha = 1/2$	$\sqrt{T\cdot\theta}$
$\alpha = 1$	$ln(T/\theta)/(\theta^{-1} - T^{-1})$
$\alpha = 2$	$2/(\theta^{-1} + T^{-1})$

Second moments are easily found by the recurrence relation:

$$E(X^2|\alpha) = E(X|\alpha)\cdot E(X|\alpha-1) \tag{2.6}$$

2.3 Moment-Matching to Evaluate Upper Truncation

We can make use of the first and second moments to make an estimate of the upper truncation point $\hat{T}$. The moment-matched parameters are found by solving the equations below.

$$\begin{gathered} E\left(X|\hat{\alpha},\hat{T}\right) = \ \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \\ E\left(X^2|\hat{\alpha},\hat{T}\right) - E\left(X|\hat{\alpha},\hat{T}\right)^2 \ = s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \end{gathered} \tag{2.7}$$

In other words, we want to set an upper truncation point such that standard deviation of the fitted distribution is [at least] the standard deviation of the historical large losses.

One obvious caution on this estimate, of course, is that it does not guarantee that the indicated upper truncation point $\hat{T}$ is greater than the largest loss actually observed historically. We therefore take it as only one part of our evaluation.

3. MAXIMUM LIKELIHOOD ESTIMATION (MLE)

Maximum Likelihood Estimation (MLE) is more commonly used than moment-matching for estimating parameters. When there is no upper truncation, the maximum likelihood estimator for the Pareto shape parameter α is found using a simple expression.

$$\hat{\alpha}_{MLE} = N \cdot \left(\sum_{i=1}^{N} ln\left(\frac{x_i}{\theta}\right) \right)^{-1} \tag{3.1}$$

When there is an upper truncation point, the maximum likelihood estimator for α is a bit more complicated and requires solving the equation below. We may note again that both the lower threshold θ and the upper truncation T are constraints supplied by the user and are not considered parameters to be estimated.

$$\hat{\alpha}_{MLE} = N \cdot \left(\sum_{i=1}^{N} ln\left(\frac{x_i}{\theta}\right) - \left\{ \frac{N \cdot ln\left(\frac{\theta}{T}\right) \cdot \left(\frac{\theta}{T}\right)^{\hat{\alpha}_{MLE}}}{1 - \left(\frac{\theta}{T}\right)^{\hat{\alpha}_{MLE}}} \right\} \right)^{-1} \tag{3.2}$$

If we do consider the lower and upper truncation points as parameters, then the MLE estimators are simply the smallest and largest observations respectively (see Aban et al., 2006); that is, the first and last order statistics from the sample.

$$\begin{aligned} \hat{\theta}_{MLE} &= MIN\{x_1, x_2, \cdots, x_N\} \\ \hat{T}_{MLE} &= MAX\{x_1, x_2, \cdots, x_N\} \end{aligned} \tag{3.3}$$

These MLE estimators are not as helpful for our purpose of selecting an upper truncation point. The goal of Maximum Likelihood Estimation is to find the model parameters that result in the highest probability assigned to events that we have actually observed. In the case of the upper-truncated Pareto, this goal is accomplished by assigning zero probability to values outside the range of what we have already observed. This is not helpful if we believe that worse events are possible.

However, we may note that the MLE in formulas (3.1) – (3.3) includes two statistics that summarize the sample of losses:

$$Mean\ of\ Logs\ = \frac{1}{N} \cdot \sum_{j=1}^{N} ln(x_i)$$

$$x_N = MAX\{x_1, x_2, \cdots, x_N\} \tag{3.4}$$

Together these represent sufficient statistics for the model parameters α and T, informally meaning that they contain all of the information available from the sample concerning these parameters.[1]

The MLE for $\hat{T}$ is referred to as "non-regular," meaning that we cannot estimate its variance through the regular procedure using the information matrix of second derivatives of the log-likelihood function. This is not, however, a great problem because we can estimate $Var(X_N)$ using the moment functions given in section 4.3.

Finally, it is important to remember that there is a relationship between the shape parameter α and the upper truncation T. A different alpha will be estimated depending upon the selected upper truncation point. To illustrate this relationship, the table below shows how the expected value of loss severity changes based on these parameters.

		Expected Pareto Severity Subject to Upper Truncation				
		Lower Threshold (Theta):		1,000,000		
		Maximum Possible Loss (Upper Truncation)				
		10,000,000	25,000,000	50,000,000	100,000,000	999,999,999
Alpha	0.75	2,839,841	4,072,455	5,257,028	6,698,663	13,948,679
	1.05	2,507,183	3,231,920	3,793,243	4,353,690	6,137,484
	1.35	2,234,010	2,641,165	2,890,943	3,093,714	3,513,688
	1.65	2,015,287	2,236,237	2,342,509	2,412,446	2,510,008
	1.95	1,843,001	1,959,873	2,003,684	2,027,046	2,049,735

[1] Chapter 7 of Arnold et al. provides a longer discussion on order statistics and sufficiency.

4. ORDER STATISTICS

For a sample of independent losses drawn from a continuous distribution, the order statistics are simply the sample put into ascending order, with x_1 being the smallest observation and x_N being the largest observation.

4.1 The Distribution of the Largest of N Losses

The upper-truncated Pareto, the distribution of the largest of a sample of size N losses, is given in formula (4.1).

$$F_N(x_N) = F(x_N)^N = \left(\frac{1-\left(\frac{\theta}{x_N}\right)^{\alpha}}{1-\left(\frac{\theta}{T}\right)^{\alpha}}\right)^{N} \tag{4.1}$$

This distribution is unimodal, with the mode defined below in formula (4.2). The mode is not directly dependent upon the value of the upper truncation point T except for the case in which T is set below where the mode would otherwise be calculated.

$$Mode_N = MIN\left\{\theta \cdot \left(\frac{\alpha \cdot N + 1}{\alpha + 1}\right)^{1/\alpha}, T\right\} \tag{4.2}$$

Percentiles from this distribution are easily calculated, and this provides us with a hypothesis test for the upper truncation point.

4.2 Hypothesis Test for Existence of Upper Truncation

One preliminary question is whether an upper truncation is indicated at all. A simple hypothesis test can help answer this question. The null hypothesis is that there is no upper truncation point (that is $T = \infty$). We then compare the actual largest loss x_N observed in the history and ask whether it is reasonable that a Pareto with no upper truncation would have generated that loss. If the largest observed loss is "significantly" less than would have been expected, then we reject the null hypothesis and conclude that an upper truncation point should be used.

The test statistic is a p-value:

$$p = F(x_N|T = \infty)^N = \left(1 - \left(\frac{\theta}{x_N}\right)^{\alpha}\right)^N \approx exp\left(-N \cdot \left(\frac{\theta}{x_N}\right)^{\alpha}\right) \tag{4.3}$$

The shape parameter α used in this test should be based on the MLE estimate with no upper truncation as given in formula (3.1). The test is only appropriate for a sample size N large enough that the largest observation x_N does not have a significant impact on the estimate of α.

The approximation on the right side of (4.3) is the Fréchet distribution, which is a limiting case for the sample maximum and a standard result from Extreme Value Theory[2]. The hypothesis test is given in this form in Aban et al. The idea is that if the p-value is small, say $p < .05$, then we reject the null hypothesis that a Pareto with no upper truncation point is appropriate. Unfortunately, this test does not tell us what the upper truncation point should be; in fact, it does not even tell us that an upper-truncated Pareto is correct but only that an untruncated Pareto is unlikely.

Conversely, if the p-value is large, say $p > .05$, that does not necessarily mean that an untruncated Pareto should be used – but only that our sample data is not sufficient to reject it. The usefulness of the test is therefore quite limited.

4.3 Evaluating Moments for the Largest Loss

The calculation for the moments of the distribution of the largest loss is not trivial but can be accomplished with a careful strategy. Using a binomial series expansion of the distribution of the largest loss, the moments can be written as follows.

$$E_N\left(X_N^k\right) = \theta^k \cdot \sum_{j=1}^{N} \binom{N}{j} \cdot (-1)^{j-1} \cdot \frac{\alpha \cdot j \cdot \theta^k}{\alpha \cdot j - k} \cdot \frac{\left(1 - \left(\frac{\theta}{T}\right)^{\alpha \cdot j - k}\right)}{\left(1 - \left(\frac{\theta}{T}\right)^{\alpha}\right)^N} \qquad \alpha \neq k \tag{4.4}$$

As discussed above, the terms when $\alpha j = k$ can be evaluated using the limit formula (2.5). The difficulty with this form is that for large sample sizes, say $N > 30$, the factorial functions become extremely large, making the calculation numerically unstable. An alternative form that works for

[2] This is a result of the Gnedenko, Fisher-Tippett Theorem. The Fréchet distribution is given in Loss Models by Klugman et al. as the "Inverse Weibull" distribution.

larger values of N makes use of the incomplete beta distribution.

$$E_N\left(X_N^k\right) = \theta^k \cdot \frac{\Gamma(N+1)\cdot\Gamma(1-k/\alpha)}{\Gamma(N+1-k/\alpha)} \cdot \frac{\beta\left(1-\left(\frac{\theta}{T}\right)^{\alpha}; N, 1-k/\alpha\right)}{\beta\left(1-\left(\frac{\theta}{T}\right)^{\alpha}; N, 1\right)} \qquad \alpha > k \tag{4.5}$$

This form makes use of the incomplete beta function, defined below.

$$\beta(y; N, b) = \int_0^y \frac{t^{N-1}\cdot(1-t)^{b-1}}{B(N,b)}\,dt \qquad N, b > 0 \quad 0 < y < 1 \tag{4.6}$$

The incomplete beta function cannot be used directly when $\alpha \leq k$. Klugman et al gives a recursive form that can be used for small values of α when N is large. This form will not work when k/α is exactly equal to an integer (e.g., the cases $\alpha = 1$ or $1/2$). A third alternative is needed for those cases.

Another form is written in terms of an infinite series. Formulas (4.7) and (4.8) provide two infinite series that converge to the expected dollar moments.

$$E_N\left(X_N^k\right) = \theta^k \cdot \sum_{j=0}^{\infty}\left(\frac{N}{N+j}\right)\cdot\left(\frac{\Gamma(j+k/\alpha)}{j!\cdot\Gamma(k/\alpha)}\right)\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.7}$$

$$E_N\left(X_N^k\right) = T^k - \theta^k \cdot \sum_{j=0}^{\infty}\left(\frac{j}{N+j}\right)\cdot\left(\frac{\Gamma(j+k/\alpha)}{j!\cdot\Gamma(k/\alpha)}\right)\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.8}$$

These series may be slow to converge when $\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)$ is close to 1.00, so this may not be an optimal formula for evaluating the moments. However, they do not have the numerical instability of formulas (4.4) or (4.5). Further, each term in the summation is a positive value, so the first series converges from below whereas the second series converges from above. The use of these two series together therefore lets us calculate moments to within any desired degree of accuracy.

We may note also that there are various recurrence relationships between moments of order statistics, for example as given by Khurana & Jha (1987), that can produce other methods for calculating the moments. However, these do not seem to offer more numerical stability than the formulas given above.

Just as we estimated $\hat{\alpha}$ and $\hat{T}$ by matching the mean and standard deviation of historical losses, we can alternatively estimate them by matching to the mean and largest value of the historical losses.[3]

$$E\left(X|\hat{\alpha},\hat{T}\right) = \bar{x}$$
$$E_N\left(X_N|\hat{\alpha},\hat{T}\right) = x_N \tag{4.9}$$

However, we can make a better estimate by using the order statistics of the logarithms of the losses, instead of the losses themselves.

4.4 Evaluating Moments for the Largest Ln(Loss)

Where we had used $E_N\left(X_N|\hat{\alpha},\hat{T}\right)$ to represent the expected value of the largest loss in a sample of N, we now define $E_N\left(ln(X_N/\theta)|\hat{\alpha},\hat{T}\right)$ as the expected value of the logarithm of the largest loss, relative to the lower threshold.

The transformed variable $ln(X/\theta)$ follows an exponential distribution, and this allows for simpler calculations of the order statistic moments.

This form will turn out to have some advantages over working with the order statistics of the loss dollars themselves.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \int_{\theta}^{T} ln\left(\frac{x}{\theta}\right) \cdot \frac{N \cdot \left(1-\left(\frac{\theta}{x}\right)^{\alpha}\right)^{N-1} \cdot \alpha \cdot \theta^{\alpha} \cdot x^{-\alpha-1}}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{N}}\, dx \qquad \alpha \neq 0 \tag{4.10}$$

This integral can be evaluated as follows.

---

[3] This procedure is essentially the same as the recommendation in Cooke (1979).

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \frac{\left\{\frac{1}{\alpha}\sum_{j=1}^{N}\frac{1}{j}\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j}\right\} - ln\left(\frac{T}{\theta}\right)\cdot\left(1-\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{N}\right)}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{N}} \qquad \alpha \neq 0 \qquad (4.11)$$

We may also note that for the untruncated Pareto as $T \to \infty$ this expression simplifies to:

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)|T \to \infty\right) = \frac{1}{\alpha}\sum_{j=1}^{N}\frac{1}{j} \qquad \alpha \neq 0 \qquad (4.12)$$

These formulas can be re-written as a simple recurrence relationship between different sample sizes is given as follows.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = ln\left(\frac{T}{\theta}\right) + \frac{1}{\alpha\cdot N} - \frac{ln\left(\frac{T}{\theta}\right) - E_{N-1}\left(ln\left(\frac{X_{N-1}}{\theta}\right)\right)}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)} \qquad \alpha \neq 0 \qquad (4.13)$$

The sequence is starting by using the expected $E\big(ln(X/\theta)\big)$ for a single loss.

$$E_1\left(ln\left(\frac{X_1}{\theta}\right)\right) = E\left(ln\left(\frac{X}{\theta}\right)\right) = \frac{1}{\alpha} - \frac{ln\left(\frac{T}{\theta}\right)\cdot\left(\frac{\theta}{T}\right)^{\alpha}}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)} \qquad \alpha \neq 0 \qquad (4.14)$$

It can also be quickly recognized that if the expected value $E\big(ln(X)\big)$ is replaced by the mean of the logarithms of the sample of observed losses, then formula (4.14) is equivalent to the MLE formula (3.2). Matching the first moment of the logs is the same as performing a maximum likelihood estimate for the shape parameter α. This is a very useful result because it means that anyone currently using MLE to estimate the shape parameter will be able to use this moment matching strategy as an enhancement to their existing model.

Formulas (4.13) and (4.14) are not valid when $\alpha = 0$ but the moments for that special case are easily calculated.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \left(\frac{N}{N+1}\right)\cdot ln\left(\frac{T}{\theta}\right) \qquad \alpha = 0 \tag{4.15}$$

We can also evaluate the expected value of the largest log-loss using infinite series similar to those in formulas (4.7) and (4.8). As with those earlier expressions, we have a series that converges from below and a second that converges from above. These series are also somewhat faster to converge than those for the dollar moments.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \frac{1}{\alpha}\cdot\sum_{j=0}^{\infty}\left(\frac{N}{N+j}\right)\cdot\frac{1}{j}\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.16}$$

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = ln\left(\frac{T}{\theta}\right) - \frac{1}{\alpha}\cdot\sum_{j=0}^{\infty}\left(\frac{1}{N+j}\right)\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.17}$$

With these formulas, we are able to match the moments:

$$\begin{gathered} E\left(ln(X)|\hat{\alpha},\hat{T}\right) = \frac{1}{N}\cdot\sum_{j=1}^{N} ln(x_i) \\ E_N\left(ln(X_N)|\hat{\alpha},\hat{T}\right) = ln(x_N) \end{gathered} \tag{4.18}$$

These moment-matching equations make use of the same sufficient statistics identified in formula (3.4). The parameters $\hat{\alpha}$ and $\hat{T}$ are solved for numerically.

We may note a few advantages to the use of the estimators in formula (4.18):

1) The estimated $\hat{\alpha}$ is equivalent to the MLE estimate conditional upon $\hat{T}$
2) The estimates rely upon sufficient statistics, meaning they make use of all of the information about the truncated Pareto parameters contained in the sample

3) The recurrence formula is easily calculated
4) The estimate of $\hat{T}$ based on log-order statistics is slightly more conservative than the estimate based directly on the order statistics of the loss dollars. This is due to Jensen's Inequality:

$$E\big(ln(X)\big) \leq ln\big(E(X)\big)$$

We now go on to show how this procedure can be applied in real-world examples.

5. TWO ILLUSTRATIVE EXAMPLES

Having outlined a basic approach for estimating an upper truncation point, we will now look at two examples to illustrate the approach. The examples are not intended for use as actual pricing factors but just to show the thought process.

The numbers used in these examples are historical statistics related to natural disasters, and the samples are shown in Appendix A. The fact that these examples are from natural disasters does not mean that the same techniques could not be used for casualty events.

5.1 Earthquake Fatalities 1900-2011

The earthquake statistics are the estimated number of deaths for events from 1900 to 2011 as published by the U.S. Geological Survey (USGS). In many cases, these numbers are rough estimates. For this example, we look at the 21 earthquakes with 20,000 or more deaths. None of these figures has been adjusted for population changes or other factors.

The numbers can be summarized by the following statistics:

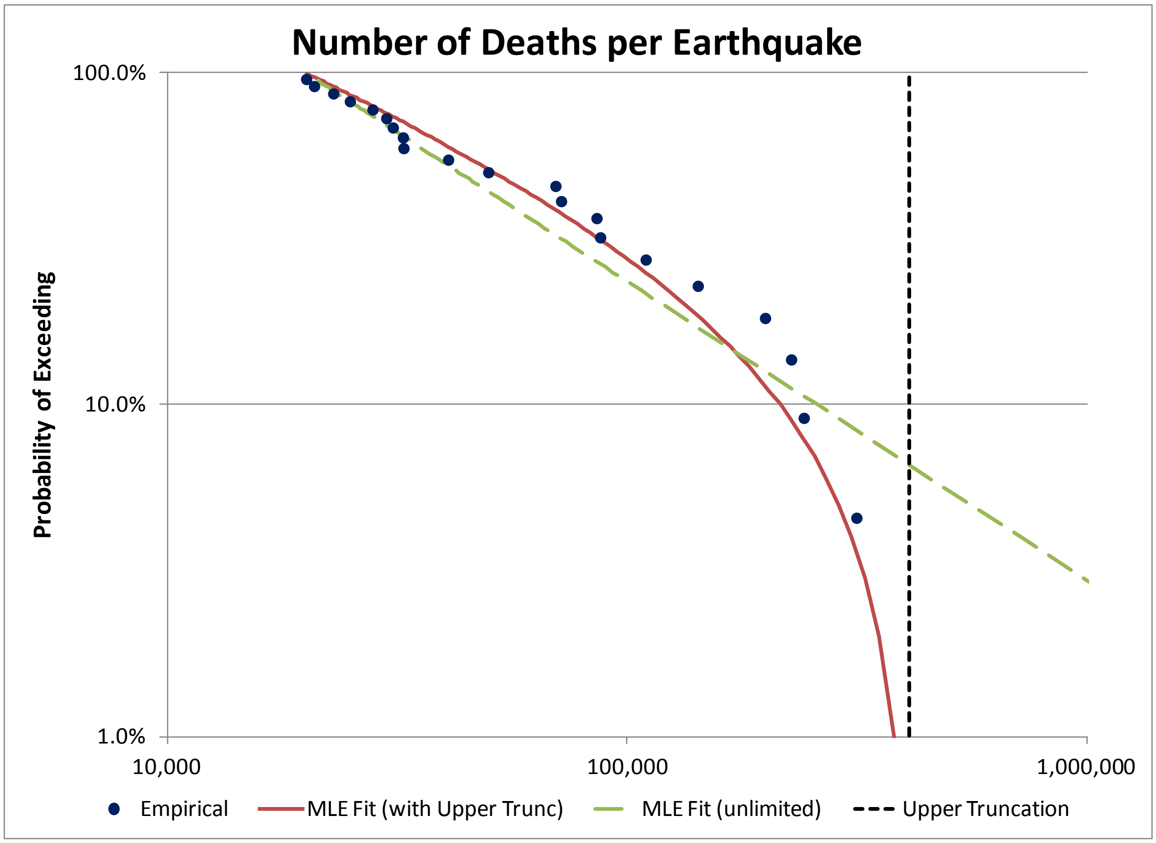
Number of Events	21
Lower Threshold θ	20,000
Average # Deaths	89,964
Standard Deviation of # Deaths	86,416
Largest # Deaths	316,000
Pareto Shape Parameter α (from MLE)	0.89993

The largest earthquake event, in terms of number of deaths, occurred in 2010 in Haiti, with 316,000 fatalities.

The p-value from this data is .173, meaning that there is a 17.3% chance that the largest event in a sample of 21 would be 316,000 or less from an untruncated Pareto. This is not strong evidence for rejecting the untruncated Pareto but does not rule out the possibility of including an upper truncation point.

A Pareto fit with no upper truncation indicates a shape parameter of 0.89993. Because this is less than 1.00000, the expected value would be undefined (infinite). This would create a serious problem in modeling the events, as simulation results could be chaotic. It is desirable to include an upper truncation.

We can select an indicated upper truncation point by matching the expected values to the average and largest amounts in our sample. As the graph below shows, this is an improved fit to the data also. The empirical points on the log-log graph show a downward curving shape, rather than a pure linear relationship that would indicate an untruncated curve.



The values from this moment-matching calculation are listed below:

Lower Threshold θ	20,000
Estimated Upper Truncation T	437,171
Pareto Shape Parameter α conditional on T	0.57122
Expected Value of # Fatalities	88,563
Expected Standard Deviation	88,334
Expected Largest of 21 Events $E_N(X_N)$	326,681

The key output from this analysis is the estimated upper truncation point as 437,171. This implies that the maximum possible number of deaths from an earthquake is 437,171 or about 38% higher than the worst event seen in the history.

The standard deviation and actual observed largest loss the actual data are slightly lower than would have been predicted by the model. This means our estimate of the upper truncation point is slightly higher than what would be needed to exactly match the sample; this conservatism is desirable since our goal is to select an upper truncation point that represents the largest possible loss.

We can also re-fit the model with different lower thresholds to include more or fewer losses to evaluate the sensitivity of the calculation.

Most importantly, we want to compare the moment-matching indication to what is known about the physical world that might create an upper bound on the possible number of deaths. Factors such as population density, construction of buildings and the possible intensities of earthquakes should be considered. Catastrophe models attempt to estimate the probability distribution based on these factors, and output from these models should be compared.

5.2 Large U.S. Weather Losses 1980-2011

The weather statistics come from the National Climatic Data Center (NCDC), a part of the National Oceanic and Atmospheric Administration (NOAA). The dollars are listed in thousands, and have been adjusted (by NCDC) to 2012 cost levels using the CPI. The losses represent estimates of total damages, not limited to just the insured portion. The sample in Appendix A are those events that caused $5 billion or more in 2012 dollars.

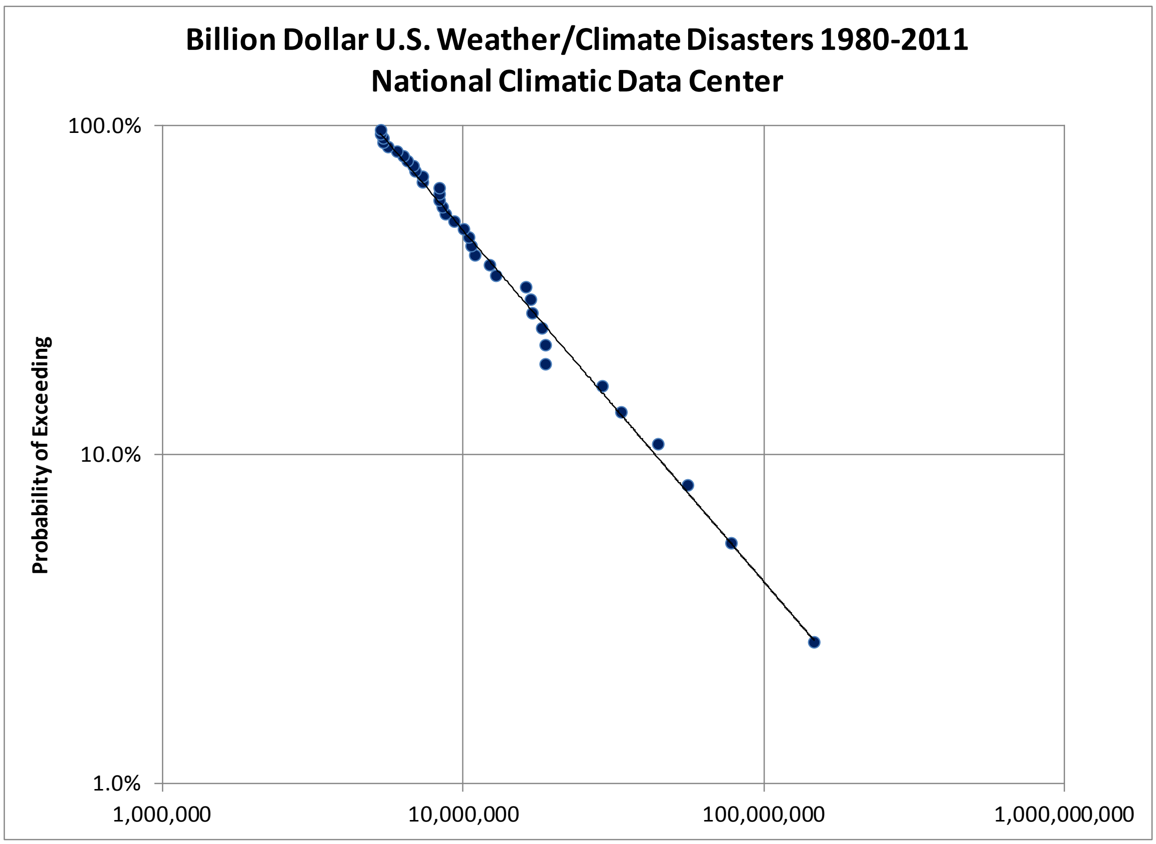
The numbers can be summarized by the following statistics:

Number of Events	36
Lower Threshold θ	5,000,000
Average Loss > Threshold	18,994,444
Standard Deviation of Losses	26,701,171
Largest Loss Damage	146,300,000
Pareto Shape Parameter α (from MLE)	1.11299

The largest weather event in this sample was Hurricane Katrina in 2005, estimated to be $146 billion in 2012 dollars.

The p-value from this data is .432, meaning we fail to reject the null hypothesis that the losses came from an untruncated Pareto. In practice, this implies that if we want to include an upper truncation point, it should be well above the largest order statistic.

The graph below shows the log-log graph of damage amount (in thousands) compared to the empirical survival probabilities (probability of exceeding the dollar amount). The historical amounts line up pretty closely along a straight line indicating, again, that if there is an upper truncation point then it must be much larger than the largest historical point.



If we calculate an upper truncation point so as to match the average and largest of the historical events, we find the following:

Lower Threshold θ	5,000,000
Estimated Upper Truncation T	480,073,321
Pareto Shape Parameter α conditional on T	1.07182
Expected Amount of Damage	21,014,276
Expected Standard Deviation	39,261,964
Expected Largest of 36 Events $E_N(X_N)$	178,675,516

The estimated upper truncation point of 480,073,321 is more than three times the largest observed historical event. The conclusion is that the largest possible hurricane damage is significantly larger than 2005's Hurricane Katrina. This indication is itself subject to estimation uncertainty but it does provide one more piece of information for use in modeling loss exposure.

5.3 Discussion of the Examples

These two numerical examples illustrate some of the assumptions and limitations of this estimation process.

First, we may note that the estimation is dependent upon the truncated Pareto being the "true" distribution for the phenomenon. Our estimate does not reflect the possibility that some other distributional model might be correct. If a different model would have been better, then it is possible that a higher upper truncation point would have been estimated.

Second, we are assuming that the sample we have observed is representative, and that future events will be of the same kind as those that have taken place historically. Events that are qualitatively different (not just bigger events of the same kind) need to be modeled separately. It is common to talk of events that have never been observed as "black swans" and we should recognize that a model that is parameterized based on past observations cannot account for these.

Third, we note that in both of the examples above the amounts observed were only estimates of the actual values, and include estimating error in themselves. An exact count of deaths from the Haiti earthquake was not made, so the upper truncation point is also less exact. This estimation error is compounded by uncertainty in inflation or demographic trends. The number of earthquake deaths or losses from weather events was gathered from a variety of sources, including newspaper reports. This type of uncertainty is also an issue in insurance losses where claim values may include case reserves rather than actual ultimate payments.

These factors are common to many statistical estimation problems. In the case of estimating an upper truncation point, we have the further difficulty that we are necessarily extrapolating beyond the range represented in our sample data. Given this level of uncertainty, our final reality check needs to be to ask if the upper truncation point corresponds to some true physical limit on the size of the loss; and if not to consider it a lower bound on the MPL.

6. CONCLUSIONS

The selection of an upper truncation point for the Pareto can be difficult in insurance applications. It represents, in theory, the Maximum Possible Loss (MPL) that could occur on the exposures written by the insurance company. This amount is generally selected based upon management's judgment about possible loss events. The use of order statistics allows us to squeeze some additional information out of the observed historical losses.

At the very least, we are able to calculate statistics such as standard deviation and expected largest loss for the upper-truncated Pareto, and compare these to the historical loss data. This provides more information to inform the judgment being made.

Appendix A - Data Sets for Examples

The data sets below are used as examples in Section 5. The earthquake statistics come from the U.S. National Geological Survey and represent estimated fatalities for international earthquakes since 1900. The U.S. Weather/Climate Disasters come from the National Climatic Data Center and represent total economic damages from weather events in the United States for 1980-2011, adjusted to 2012 dollars.

Earthqake Deaths Since 1900

Rank	# Deaths
1	316,000
2	242,769
3	227,898
4	200,000
5	142,800
6	110,000
7	87,587
8	86,000
9	72,000
10	70,000
11	50,000
12	40,900
13	32,700
14	32,610
15	31,000
16	30,000
17	28,000
18	25,000
19	23,000
20	20,896
21	20,085

NOAA Weather Losses 1980-2011

Rank	$ Damage
1	146,300,000
2	77,600,000
3	55,600,000
4	44,300,000
5	33,400,000
6	28,900,000
7	18,700,000
8	18,700,000
9	18,200,000
10	16,900,000
11	16,700,000
12	16,100,000
13	12,800,000
14	12,200,000
15	10,900,000
16	10,600,000
17	10,400,000
18	10,000,000
19	9,300,000
20	8,700,000
21	8,500,000
22	8,300,000
23	8,300,000
24	8,300,000
25	7,300,000
26	7,300,000
27	6,900,000
28	6,800,000
29	6,500,000
30	6,300,000
31	6,000,000
32	5,600,000
33	5,400,000
34	5,400,000
35	5,300,000
36	5,300,000

7. REFERENCES

[1] I. Aban, M. Meerschaert, A. Panorska, "Parameter Estimation for the Truncated Pareto Distribution," *Journal of the American Statistical Association* **2006**, Vol. 101, 270-277.
[2] Arnold, Barry C., N. Balakrishnan and H.N. Nagaraja, "A First Course in Order Statistics," Society for Industrial and Applied Mathematics (SIAM), 2008.
[3] Cooke, Peter, "Statistical inference for bounds of random variables," *Biometrika* **1979**, 66, 2, pp. 367-74.
[4] Johnson, Norman L., Samuel Kotz and Adrienne W. Kemp, "Univariate Discrete Distributions," Second Edition, John Wiley & Sons, 1993.
[5] Johnson, Norman L., Samuel Kotz and N. Balakrishnan, "Continuous Univariate Distributions," Second Edition, John Wiley & Sons, 1994
[6] A.P. Khurana and V.D. Jha, "Recurrence Relations Between Moments of Order Statistics from a Doubly Truncated Pareto Distribution," *Sankhya: The Indian Journal of Statistics* **1987**, Volume 53, Series B, Pt.1, 11-16.
[7] Klugman, Stuart A., Harry H. Panjer and Gordon E. Willmot, "Loss Models: From Data to Decisions," John Wiley & Sons, 1998.
[8] Philbrick, Stephen, "A Practical Guide to the Single Parameter Pareto," *PCAS* **1985**, *Vol. LXXII, 44-84*.
[9] Robson, D.S. and J.H. Whitlock, "Estimation of a Truncation Point," *Biometrika* **1964**, 51, 1 and 2, p. 33.
[10] Rytgaard, Mette, "Estimation in the Pareto Distribution," *ASTIN Bulletin* **1990**, Vol. 20, No. 2, 201-216.
[11] Wilkinson, Margret E., "Estimating Probable Maximum Loss with Order Statistics," *PCAS* **1982,** Vol. LXIX, 195-209.

Abbreviations and notations

N, number of large losses observed in a sample
$E_N(X_N)$, expected value of largest loss in a sample size of N
X, random variable representing a single loss amount; $\theta \leq X \leq T$
x_N, largest observed loss in a sample size of N
θ, Theta, representing the lower threshold of losses
T, Upper truncation point – loss amount above this are not considered possible
α, Alpha, representing the "shape parameter" of the Pareto distribution

Biography of the Author

David R. Clark is a Senior Actuary with Munich Reinsurance America, and a Fellow of the Casualty Actuarial Society (FCAS). His prior papers include "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach" which received the 2003 Best Reserve Call Paper prize, and "Insurance Applications of Bivariate Distributions" co-written with David Homer which received the 2004 Dorweiler Prize.

Acknowledgments

The author gratefully acknowledges the valuable review given by Michael Fackler, David Homer, Hou-Wen Jeng and James Sandor.