Article from:

Risk Management

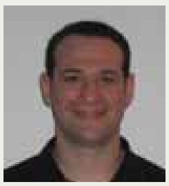
September 2009 – Issue 17

Modeling Tail Behavior with Extreme Value Theory

By Damon Levine

MODEL BACKLASH

ONE OF THE MANY THINGS to come out of the recent market turmoil is a long list of scapegoats. Experts and laymen alike have assigned varying amounts of blame to a wide variety of sources ranging from greed and conceit, to rating agencies, the government, and executive bonus plans, which rewarded excessive risk taking. Of course, they also blamed the risk models.



Damon Levine, CFA, is assistant vice president at Assurant Inc., in New York, N.Y. He can be reached at *damon.levine@assurant.com.*

The idea that a model is not meant to capture reality or have significant predictive power is such a pervasive concept that it borders on truism. That being said, it is likely that models should take some of the blame for the subprime meltdown and the subsequent crisis in the financial markets in general.

Perhaps the biggest problem was that, by design, a lot of the models could not warn of the potential for an observation significantly worse than outliers in the historical data. To use Nassim Taleb's phrase, they failed to provide information about the magnitude of potential "black swans."

Ideally, a statistical distribution that is used in a risk model should fit historical data well, both in the central portion of the data set and in the tail. But the distribution should not be "constrained by history." Rather, it should make use of previous extreme values to offer information on the probability and magnitude of potential values *more* extreme than those seen previously. Extreme Value Theory provides a theoretical basis for such a model. This theory quantifies, in a statistically sound manner, the potential black swans hinted at by historical extremes.

EXTREME VALUE THEORY

Extreme Value Theory (EVT) is a branch of statistics dealing with the extreme deviations from the median of probability distributions. Under very general conditions, EVT's main results characterize the distribution of the sample maximum or the distribution of values above a given threshold.

It is this second result, the Pickands-Balkema-de Haan (PBH) Theorem, which will be used here. This theorem describes the distribution of observations above a high threshold as a generalized Pareto distribution.

This result is particularly useful because it can be applied in a great many situations with a minimal set of assumptions about the "true" underlying distribution of an arbitrary data set.

THE DISTRIBUTION OF EXCESSES

Given a data set, choose a large threshold value u such that we have several data points larger than u. Assume for example, that in a data set of 1000 insurance claim amounts (in dollars) we choose u to be the 95th percentile, and there are 50 points above u.

For each of those 50 points, $\{p_1, p_2,\ldots, p_{50}\}$, we compute the excess above u: $\{p_1-u, p_2-u,\ldots, p_{50}-u\}$. These may be interpreted as random observations from a population with some underlying "distribution of excesses."

The PBH Theorem states that for a very large family of distributions, for a sufficiently large threshold value u, the distribution of excesses over u can be well approximated by a generalized Pareto distribution.

The generalized Pareto distribution (GPD) can be expressed as a two parameter distribution with cumulative distribution function (CDF):

$G_{s,k}(x) = 1 - (1 - kx/s)^{1/k}$ for nonzero k, and $G_{s,k}(x) = 1- \exp(-x/s)$ for k=0

Note that if there is a left tail consisting solely of negative values, below some negative threshold far less than the median, we may apply the PBH theorem by simply looking at absolute values. The "excess" of an observation in this tail is the (positive) distance from the observation to the threshold. This idea will be used in the example shown later.

> "Perhaps the biggest problem was that a lot of models failed to provide information about Black Swans."

MODELING WITH A HYBRID EMPIRICAL/ GPD MODEL

Let F represent the "true," underlying cumulative distribution function of the full set of claim data in the above example. We assume the observed data set is a random sample drawn from some population following a statistical distribution. Based on a particular choice of the threshold u, the cumulative distribution function of the excesses denoted by $F_u(y)$ is defined for non-negative y as:

$F_u(y) = P\{X - u \leq y \mid X>u\}$ = P{excess $\leq y$ for a random observation exceeding u}

It is important to realize this CDF describes the distribution of the *excess* over the threshold. It gives a probability that the excess over u, of a random observation larger than u, will be less than or equal to y. It does not refer to the magnitude of the extreme value itself, but it is straightforward to make use of F_u to do so.

For $x \geq u$ we have:

$F(x) = P\{X \leq x\} = (1 - P\{X \leq u\})\, F_u(x-u) + P\{X \leq u\}$ [1]

Now, F_u can be estimated by some GPD, $G_{s,k}$, and $P\{X \leq u\}$ can be estimated from the data by $F_n(u)$, the empirical distribution evaluated at u. [2] So for $x \geq u$ we can approximate $F(x)$ by:

$F^*(x) = [1 - F_n(u)]\, G_{s,k}(x-u) + F_n(u)$

The two parameters of the distribution $G_{s,k}$ can be estimated by a variety of methods including maximum likelihood and the method of moments, which is used in the example shown later.

A CDF modeling the entire underlying distribution, F, can therefore be described as a hybrid empirical/GPD:

$F(x) = F_n(x)$ for $x<u$, and $F(x) = [1 - F_n(u)]\, G_{s,k}(x-u) + F_n(u)$ for $x \geq u$

If desired, one can perform simulation by regarding a random digit, r, from (0,1) as a percentile of $F(x)$, i.e. employ the mapping $r \rightarrow F^{-1}(r)$.

THE CHOICE OF THE THRESHOLD VALUE

In choosing the threshold value it is important to understand some of the technical aspects of the PBH theorem. The technical statement of theorem makes use of the notion of "right endpoint" of a distribution F. This is the smallest value, r, such that the CDF evaluated at r is equal to 1, i.e., $F(r) = 1$. In many cases r is infinite.

For our purposes, we can look at a somewhat simplified version of the theorem: for a large class of distributions, as the threshold u approaches the right endpoint of F, the excess distribution F_u approaches a GPD. The class of distributions conforming to this theorem includes *all* the common continuous distributions an actuary or statistician typically employs including the normal, lognormal, beta, exponential, F, gamma, Student t, uniform, etc. [3] Note that the only returns that are used in the parameter estimation of the GPD are those which are in the tail defined by the choice of threshold.

When fitting any distribution to a data set, a larger number of data points is ideal. By selecting a smaller value of u we can expect a fair amount of data points to exceed that value, perhaps improving the GDP fit.

Contrary to this notion is the fact that the PBH theorem states a result based on the assumption of threshold values approaching the right endpoint of the distribution F. This implies that better GPD fits are expected for larger choices of the threshold u.

One must strike a balance between choosing u large enough so that the theorem is applicable from a practical standpoint and small enough so that a sufficient number of data points can be used in estimation of the parameters of the GPD.

There is no hard and fast rule describing the "right" choice of the threshold value. Some methods for threshold selection can be found in Bensalah's "Steps in Applying Extreme Value Theory to Finance: A Review." [4]

CONTINUED ON **PAGE 16**

Modeling Tail Behavior ... | from Page 15

EXAMPLE: MODELING MONTHLY TOTAL RETURN FOR A-RATED 7-TO-10-YEAR CORPORATE BONDS

The data sample consists of monthly total returns for the A-rated 7-to 10-year corporate bond component of Citi's U.S. Broad Investment Grade Bond Index. The data was taken from Citi's Yieldbook application and consists of monthly returns from January 1980 to August 2008.

Table I

Selected Percentiles of the Return Data

Percentile	Return
99th	5.97%
90th	2.94%
80th	2.05%
70th	1.59%
60th	1.18%
50th	0.83%
40th	0.44%
30th	-0.14%
20th	-0.77%
10th	-1.54%
5th	-2.33%
1st	-3.87%

The data set consists of 344 returns ranging from a minimum of -7.16 percent to a maximum of 10.84 percent with selected percentiles as shown in Table I.

We focus on returns in the left tail; in other words, we are interested in returns less than some low value. Consider a choice of threshold u as the absolute value of some low, negative return and then, for returns less than u, define the excess to be the distance between u and the absolute value of that return. Note that a large excess is equivalent to a poor return. The application of the PBH theorem to the left tail was introduced in the Distribution of Excesses section.

Setting u=1.54 percent, corresponding to the 10th percentile, we may then approximate the excess distribution F_u as a GPD with parameters s and k. We determine these parameters by the method of moments.[5] Note that the only returns that are used in the parameter estimation of the GPD are those above the threshold.

Recording this threshold value and its corresponding pair of parameters for the GPD, we will then choose a value of u farther out in that left tail and find the resulting pair of parameters for the associated GPD. This process will continue so that we have sequence of threshold candidates $u1$, $u2$,... , etc. moving further and further into this tail of poor returns.

Our choice of threshold will be the first of these candidate values for which there is stability in the GPD parameter estimates from that point on.[6] If no such stability is seen then the fitting of a GPD to the tail may not be pragmatic.

Table II

Threshold Candidates: Associated Tails and GPD Parameter Estimates

Percentile	Raw Data Value	Thereshold Candidate* = abs(return value)	Numbers of Returns Worse than Threshold**	*GPD Parameter Estimates* s*	k*
10th	-1.54%	1.54%	35	0.01	-0.12
8th	-1.68%	1.68%	28	0.01	-0.06
6th	-1.93%	1.93%	21	0.01	0.03
5th	-2.33%	2.33%	18	0.01	-0.12
4th	-2.52%	2.52%	14	0.01	-0.07
3rd	-2.82%	2.82%	11	0.01	-0.07

* these are threshold candidates in the left tail of rhe return data; in all cases the value prior to taking absolute value is negative
** e.g. worse than the 5th percentile means less than -2.33% or a negative return whose absolute value exceeds 2.33%

"The alternative to well-founded methods like EVT may be reliance on intuition or fitting of distributions that are dangerously misleading."

We begin with the 10th as a threshold candidate. The results are summarized in Table II.

Taking incrementally larger values of a threshold choice u, we begin to see stability in the parameter estimates with the threshold set to the (absolute value of the) 4^{th} percentile of the data set and this is our choice for the threshold.

As a result, for returns below -2.52% we may model the distribution of distances below that threshold as a GPD with s=.01 and k=-.07. The CDF can be written as:

$G(x) = 1 – (1 + .07x/.01)^{1/-.07} = 1 – (1+7x)^{-14.286}$ for x>0

Assume we are interested in the probability of a monthly return being less that -5% assuming it is less than -2.52%. This is equivalent to probability that the excess is greater than or equal to 2.48% (i.e. 5% - 2.52%) and can be found as 1 - G(2.48%) = 1 - .898 = .102. This is, of course, a conditional probability. The unconditional probability that a monthly return is less that -5% is .102 multiplied by the probability of being in this tail: .102*.04 or .004.

Keeping in mind that we've applied the PBH theorem to a left tail, the CDF for the entire underlying distribution, F, can be written as a hybrid empirical/GPD:

$F(x) = F_n(x)$ for $x \geq u$, and $F(x) = F_n(u) * [1-G(u-x)]$ for $x<u$

KEY MODEL RESULTS

The model implies that for a *return in the tail,* the monthly probability of the return being at least 8.29 percent less than the threshold of -2.52 percent (i.e., less than or equal to -10.81 percent) is about .001. The unconditional probability is then .001*.04 or .00004. Because this is a monthly probability, the probability of seeing a return at least this low in a year is $1- (1 - .00004)^{12}$ or .00048. Over a period of 30 years the probability of seeing a return at least this low is $1- (1 - .00004)^{30*12}$ or 1.4 percent.

This is certainly not a large likelihood but we are, after all, talking about a black swan. As of August 2008, the idea of a monthly return worse than -10.81 percent was generally thought to be out of the realm of possibility. At that time, the worse monthly return since 1980 had been -7.16 percent and this occurred back in February of 1980! So the methods of EVT allow us to say:

P(at least one monthly return ≤ -10.81 percent over a 30-year period) = 1.4 percent.

What would the estimate of this probability be if we fit a normal distribution to the data? Based on the sample mean and standard deviation of 0.77 percent and 1.98 percent respectively, we are talking about a return of -10.81 percent, which corresponds to a Z-score of -5.858 and a probability of 2.34×10^{-9}. From this we have, under the normal distribution fit,

P_{NORM} (at least one monthly return ≤ -10.81 percent over a 30-year period) = 8.42×10^{-7}

The difference is clear: EVT points to low likelihood but puts the result on the radar screen. The probability of the event using EVT is found to be more than 16,000 times greater than had it been calculated according to a normal distribution assumption!

In September 2008 the monthly return for this bond index was -10.94 percent. A risk manager or bond trader who has worked with EVT might think the event was surprising in that there was only about a 1 percent chance of its occurrence over a 30-year time period. Had other methods been used the result might have seemed on par with a flipped coin landing on its edge! For all intents and purposes it would have been considered impossible.

The key point is that such an unimaginably bad result would have shown up, before the fact, in the analysis based on EVT.

CONCLUSIONS

Models based on EVT, like other risk models, work best in concert with subjective tools such as intuition, judgment and common sense derived from experience. A risk quantification approach that incorporates both the Delphi method and EVT may very well be the best approach to making decisions under uncertainty.

CONTINUED ON **PAGE 18**

Modeling Tail Behavior … | from Page 17

Of course, EVT is no panacea. It is, however, a scientific approach that allows the modeler to make the best use of a small number of prior extremes. The alternative to using well founded methods like EVT may be full reliance on potentially flawed intuition or the fitting of distributions that are both dangerously misleading and possibly costly. ✦

NOTES

1 For $x \geq u$, we have $P\{X \leq x\} = P\{X \leq u\} + P\{u \leq X \leq x\}$ and $P\{u \leq X \leq x\}$ can be expressed as $(1 - P\{X \leq u\})\, F_u(x - u)$.

2 An empirical distribution fitted to a data set defines a CDF consistent with percentiles directly observed in the data set. In other words, it defines a CDF, F(x), such that F(x) is equal to the proportion of data points in the set less than or equal to *x*.

3 See below in References: McNeil's "Estimating the Tails of Loss Severity Distributions using Extreme Value Theory" (pp. 7-8).

4 See below in References: Bensalah's "Steps in Applying Extreme Value Theory to Finance: A Review"

5 Let $\overline{x} = (1/n)\sum (x_i - u)$, $w = (1/n)\sum (x_i - u)^2$, where the summations extend over those *n* values, $\{x`, x_2, \ldots, x_n\}$, that exceed the threshold *u*. So $\overline{x}$ is the mean of the excesses, and *w* is the mean of the *squared* excesses.

Also, define $A = \overline{x}^2 / (w - \overline{x}^2)$. Then for the GPD described by: $G_{s,k}(x) = 1 - (1 - kx/s)^{1/k}$ (for nonzero *k*), the Method of Moments parameter estimates of s and k are: $s^* = .5\,\overline{x}\,(A + 1)$ and $k^* = .5(A - 1)$

6 The stability will begin to wane as the threshold become large enough to significantly shrink the count of data values in the associated tail; so there is a sort of "window of stability."

REFERENCES

Bensalah, Younes. 2000. "Steps in Applying Extreme Value Theory to Finance: A Review." Working paper. Bank of Canada.

Krvavych, Yuriy and Mergel, Victor. "Large Loss Distributions: probabilistic properties, EVT tools, maximum entropy characterization." *http://www.geocities.com/krvavych/ASTIN.pdf*

McNeil, Alexander. 1996. "Estimating the Tails of Loss Severity Distributions using Extreme Value Theory" Departement Mathematik, ETH Zentrum, CH-8092 Zurich.

McNeil, Alexander. 1999. "Extreme Value Theory for Risk Managers," Departement Mathematik, ETH Zentrum, CH-8092 Zurich.

Sarma, Mandira. 2005. "Characterisation of the tail behaviour of financial returns: studies from India." EURANDOM, the Netherlands.