

Value at Risk Calculations, Extreme Events, and Tail Estimation

SALIH N. NEFTCI

SALIH N. NEFTCI is a professor at the Graduate School and University Center of the City University of New York (CUNY), and at the ISMA Centre at the University of Reading in the U.K.

The notion of extreme movements in asset prices is implicit in current risk management practices. Capital adequacy assumes a threshold that classifies observed changes in market risk factors either as extreme or ordinary. A probability is first chosen to measure the “extremeness” of events that may affect a particular portfolio. This probability then determines the proper threshold.

Underlying this approach is the central limit theorem, which yields an asymptotic normal distribution for the risk factors under consideration. If it is possible to obtain an asymptotic distribution for all possible values of risk factors, however, why not do the same for extreme observations as well? Extremal theory does exactly this, and provides a methodology that can be used to make statistical inferences on extremes only, using a new set of tools.

The value at risk calculations provided here use extreme distribution theory and yield some remarkable results. Both the in-sample and out-of-sample data show that extreme distribution theory performs surprisingly well in capturing both the rate of occurrence and the extent of extreme events in financial markets. In fact, statistical theory of extremes appears to be a more natural and robust approach to risk management calculations.

Volatility and extreme movements in asset prices are not synonymous. In fact, two asset prices with the same volatility may exhibit completely

different behavior in terms of their extremes. Volatility refers to the variance of a random variable, while extremes are a characteristic of the tails only.

By and large, current risk management practices rely, first, on calculation of appropriate volatility estimates, and then on estimation of the implied changes in value at risk (VaR). We investigate whether we can obtain a more satisfactory alternative by using extreme value theory.

The notion of extreme movements in asset prices is implicit in current risk management practices. Capital requirements of financial institutions call first for determination of a threshold that classifies observed changes in risk factors as extreme or ordinary. A probability is chosen to characterize the “extremeness” of events that may affect a particular portfolio. This probability determines a threshold.

RiskMetrics [1996] and the related literature follow this idea. Available time series are used, and a probability density for changes in the underlying risk factors is estimated.¹

Often, normality is assumed, and slowly moving standard deviations are estimated. Then, k times the estimated standard deviations yields the threshold that can be used in value at risk calculations.

But how can one choose a threshold to classify observed movements as extreme or not if the underlying distribution function is unknown? Central limit theorems have pro-

vided an answer to this type of question. Suppose we need to model the distribution of a random variable Y_n that represents some real-world phenomenon, but on which we have no other additional theoretical information. Then, the central limit theorem portrays the Y_n as a sum of a large number of sufficiently *small* random events:

$$Y_n = \varepsilon_1 + \dots + \varepsilon_n \quad (1)$$

and then shows that as n grows large, the distribution of Y_n will converge to a *known* function, which very often is the normal distribution.

A nice aspect of this result is that it depends on very weak assumptions. As a matter of fact, the main condition is that $E[Y_n] < \infty$. Hence, as long as n is large, one can approximate the unknown distribution of Y_n by the normal distribution:

$$N(\mu, \sigma^2) \quad (2)$$

where μ is an asymptotic mean, and σ^2 is an asymptotic variance.

It turns out that the same idea that leads to the central limit theorem can be applied to only extremes of risk factors, in order to obtain an *extreme value distribution*. We might well ask, if it is possible to obtain an asymptotic (i.e., approximate) distribution for all possible values of Y_n , why not obtain an asymptotic distribution only for the extremes? There may be several advantages to this approach.

First, note that a histogram for a time series can always be estimated accurately at or near the center of the distribution — there are many observations falling at the center. As a result, there is less need for a priori models that yield closed-form formulas for the underlying distributions. Yet extremes are *rare* events, and there are by definition few observations at the tails. Empirical densities will by definition not be accurate. A theory that provides approximate functional forms will be much more useful.

Second, we should remember that the distribution of asset price increments is known to be heavy-tailed and significantly non-Gaussian. Under these conditions, a non-parametric approach that approximates the tail areas *asymptotically* may be more appropriate than imposing an explicit functional form like the lognormal on the distribution.

Third, there is always the possibility that extreme movements in asset prices are caused by mechanisms that are *structurally* different from the routine functioning of

markets. For example, an extreme movement may be the result of a major default or a speculative bubble that bursts. During these periods, standard book-running practices such as dynamic hedging may not work, and the distributional characteristics of the data may shift. Such structural shifts also require separating tail estimation from estimation of the rest of the distribution. This is especially true when the remaining part of the density is not needed, as is the case for VaR calculations.

The statistical theory of extremes provides some assistance in the solution of these problems. This theory allows putting more emphasis on the tail areas of the observed frequency distributions, and hence may yield a more precise estimate for the thresholds in question. At the same time, the asymptotic nature of this theory would eliminate the need to impose restrictive and potentially wrong functional forms on the underlying distributions.²

Our estimates yield some remarkable results. Both the in-sample data and out-of-sample data show that tails estimated with extreme distribution theory perform surprisingly well in capturing the rate of occurrence and the (average) extent of extreme events.

I. VAR CALCULATIONS AND EXTREME VALUE THEORY

Value at risk calculations are performed in an environment where the stochastic process x_t , $t \in [0, T]$, represents a vector of risk factors such as sovereign and swap yields, exchange rates, equity indexes, and commodity prices. We let the arbitrage-free price of a financial asset B_t be a *known* function of x_t , t , and of the parameters θ :

$$B_t = B(x_t, t, \theta) \quad (3)$$

The stochastic variation in B_t during an infinitesimal interval dt is then given by Ito's lemma:

$$dB_t = B_x dx_t + \frac{\partial B_t}{\partial t} dt + \frac{1}{2} B_{xx} \sigma_t^2 dt \quad (4)$$

where B_x and B_{xx} are, respectively, the delta and the gamma of the asset:

$$B_x = \frac{\partial}{\partial x_t} B(x_t, t) \quad B_{xx} = \frac{\partial^2}{\partial x_t^2} B(x_t, t) \quad (5)$$

In order to calculate the VaR, one imposes a model on the stochastic differential dx_t . For most risk factors, practitioners choose the stochastic differential equation (SDE):

$$dx_t = \mu_t dt + \sigma_t dW_t \quad (6)$$

where μ_t and σ_t are, respectively, the (time-varying) drift and volatility parameters. W_t is a standard Wiener process with respect to I_t , the current information set.

Next, a finite sampling interval Δt is selected, and the continuous-time model (6) is approximated over discrete intervals. Assuming that Δt denotes the period of observation, we can write the approximation:³

$$x_{t+\Delta t} - x_t \cong \mu_t \Delta t + \sigma_t (W_{t+\Delta t} - W_t) \quad (7)$$

The critical step in calculating VaR is the estimation of the threshold point that will define what variation in x_t is to be considered “extreme.” We let this threshold be denoted by L , with α the probability that a dx_t exceeding the threshold L will occur. The threshold L will be defined by the equation:

$$P(\Delta x_t \geq L\sqrt{\Delta t}) = \alpha \quad (8)$$

where $P(\cdot)$ is the underlying probability distribution, assumed to be known.

We can also write L as:

$$L = \kappa \sigma_t \quad (9)$$

where σ_t is the diffusion parameter and κ is the α -threshold of the normal distribution.

The VaR at time t is then obtained from Equation (4) by letting $(\partial B_t / \partial t) = 0$, and $B_{xx} = 0$:

$$\text{VaR}(B_t, \alpha, \Delta t) = \kappa \sigma_t B_x \sqrt{\Delta t} \quad (10)$$

BIS guidelines require setting α as 0.01, and allow for the normal distribution as a choice for $P(\cdot)$. This makes κ equal 2.3.⁴

The central limit theorem justifies the use of the normal distribution in VaR calculations. According to this, if discrete increments in risk factors denoted by Δx_t are themselves the sum of a large number of “small” events (ε_i):

$$\Delta x_t = \varepsilon_1 + \dots + \varepsilon_n \quad (11)$$

the Δx_t will be normally distributed under very general assumptions. In fact, all we need for convergence to the *normal* distribution is that the variance of Δx_t is finite:

$$E(\Delta x_t - E[\Delta x_t])^2 < \infty \quad (12)$$

As long as the *mean* exists, however, even this assumption can be relaxed.⁵

Yet with unbounded variances, the convergence of the sum in (11) will not necessarily be to the normal distribution, but to the so-called class of α -stable distributions instead. The risk manager just needs to use different functional forms.⁶

Thus, central limit theorems look at the *sum* S_n of a large number of small events:

$$S_n = \varepsilon_1 + \dots + \varepsilon_n \quad (13)$$

Dividing this by n , we get some sort of *mean*. This mean must have more stable properties than the individual elements that enter the sum S_n . Note that by looking at the mean, we are in fact “normalizing” S_n by n . But we could as well subtract a true mean and then work with

$$\frac{1}{n} S_n - \frac{n}{n} \mu = \frac{1}{n} [S_n - n\mu] \quad (14)$$

This amounts to a transformation of S_n by first subtracting an n -dependent constant, and then normalizing using an n -dependent factor, which in this case consists of dividing by n . Central limit theorems work with this “normalized” and “centered” random sum.

Clearly, all these transformations can be generalized. Why not consider some function of the random variables (ε_i) other than the sum S_n ? Why not divide by some other function of n ? And finally, why not subtract some other constant?

For example, consider the function that gives the maximum of ε_i :

$$M_n = \max[\varepsilon_1, \dots, \varepsilon_n] \quad (15)$$

Is it possible that a centered and normalized version of *this* function will have a distribution that converges to a *known* function? If the sum of (ε_i) stabilizes around some known distribution, why wouldn't a sim-

ilar phenomenon occur in terms of its maximum?

The question is not as hopeless as it may sound, since it should be realized that central limit theorems also need to describe the behavior of the tails. So, implicit in the central limit theorem is already something about the behavior of extreme values. The only difference is that, with the theory of extremes, we look at the tails *only*, and hence a more refined approach can be taken.

In fact, if risk managers are interested in the way extremes occur in asset prices, it is more natural to ask what happens to the probabilities involving the M_n as n increases. To this end, consider the probability that the largest observation will exceed a (high) level L :

$$P(M_n > L) \quad (16)$$

Now, if the tails extend to infinity, given enough observations on dx_t , their maximum will exceed *any* level $L < \infty$. Thus:

$$P(M_n > L) \rightarrow 1 \quad (17)$$

as $n \rightarrow \infty$ for fixed L .

But this question is of no interest to a risk manager, since it is clear that in the infinite future, every extreme will be exceeded by another. Rather, the risk manager is interested in *how likely* such extreme movements are. Those that happen once every one thousand years may not be of much interest, but movements that may occur every ten or twenty years might. So, rather than working with raw extremes, the trick is to scale them, in the way the central limit theory scales S_n , and then obtain the relevant asymptotic probabilities.

Thus, in case of maxima we are also interested in finding normalizing constants, b_n and a_n , such that the probabilities associated with

$$\frac{1}{b_n}(M_n - a_n) \quad (18)$$

converge to a known distribution function, say, $H(\cdot)$, as n increases. This way, $H(\cdot)$ summarizes only the common characteristics of the extreme values, and can be used to develop a non-parametric model.

The main result of the classical extremes theory says that once some critical conditions are satisfied, the probabilities associated with properly normalized and centered (M_n) will converge to one of three distribution

functions. These distributions can be summarized by a single *generalized extreme value distribution*. If the elements ε_i are all identically distributed with mean μ and variance σ , this distribution is given by:

$$H(x) = e^{-\left(1+k\frac{(x-\mu)}{\sigma}\right)^{\frac{1}{k}}} \quad (19)$$

where k is a parameter that governs the shape of the tail under consideration. For example, the shapes of typical bell-shaped densities will have a k within the range $-0.25 < k < 1$. The “fatness” of the tail will then depend on the exact value assumed by k .

Thus, using extreme value distributions, we can obtain an approximate functional form for the distribution of extreme observations. This way, we can adopt an alternative approach to calculating value at risk.

This approach would be based upon an asymptotic argument, avoid ad hoc assumptions, and calculate the parameters needed to manage risk from data on extremes only. Such an approach may be more appropriate for risk management, which, after all, claims to calculate risks associated with “crisis” situations.

The risk manager who has exposure to a risk factor x_t , which changes by discrete increments of Δx_t , needs to know how much capital to put aside to cover at least the fraction $1 - \alpha$, $0 < \alpha < 1$, of daily losses during a year. In order to do this, the risk manager must first determine L , a threshold so that the event ($\Delta x_t \geq L$) has a probability α under $P(\cdot)$. The standard approach does this by using an explicit distribution that is in general the normal distribution, or by using some histogram whose convergence is guaranteed by asymptotic theory.

The alternative provided by extreme value theory is to work with $H(\cdot)$ instead of $P(\cdot)$, and then determine the level L by going backward from α to L by solving:

$$H(L) = 1 - \alpha \quad (20)$$

given the value of α .

This is similar to the BIS and RiskMetrics approaches except in the way L is calculated. The RiskMetrics approach amounts to assuming that as Δt , the observation interval, becomes smaller and smaller, the diffusion component of the increments $x_{t+\Delta t} - x_t$ will converge to a Wiener process. The use of the theory of extremes, on the other hand, assumes that the tail behavior of the distribution of Δx_t during periods of *extreme*

market volatility can be better approximated by the asymptotic distribution of the maximum of the series.

Thus, the risk manager will again be short of capital $\alpha\%$ of the time during a time interval of length T . But, with the new method the probability is calculated via the extreme value distribution $H(L)$. This approximation becomes arbitrarily good as $n \rightarrow \infty$, and, as long as some conditions are not violated.

II. CONDITIONS FOR CONVERGENCE

The use of extreme value theory entails some non-trivial conditions, which themselves may be of some interest in finance.

The central limit theorem is about the convergence of *sums* of random variables. The theory of extremes, on the other hand, deals with the convergence of the distribution of a *maximum*. This latter is only one element of a sequence of random variables, and in order for the distribution of normalized and recentered maxima to become stable around some fixed distribution, significantly more restrictive conditions need to be satisfied. The asymptotics of the distribution of maxima are much more fragile.

Remember that in order for a properly normalized and centered sum of random variables to converge to a *normal* distribution, we need the existence of the second moment of Δx_t . Again we know that if the variance is infinite, the convergence still occurs, but not necessarily to the normal distribution.

This generality disappears when the issue is the convergence of the distribution of extremes, or equivalently, when we are interested in the asymptotics of tails only. It turns out that significant restrictions on tails of the distribution of each ε_t are needed in order to obtain non-trivial asymptotic extreme value probabilities. In particular, these tails must be relatively smooth, in the sense that they should not exhibit any significant jump components.

In order to get a flavor of these important conditions, consider the normalized tail:⁷

$$\frac{P(\Delta x_t > L + y)}{P(\Delta x_t > L)} \quad (21)$$

This is the probability that the extreme will exceed the level $L + y$, given that it has already exceeded L — i.e., given that it already qualifies as an extreme. By asymptotic

convergence, we mean the convergence of these probabilities to *non-trivial* limits, that is, to probabilities between 0 and 1.⁸

The conditions of extreme value theory require that, as we consider tails associated with more and more extreme events, that is, as we move farther into the tails, the behavior of these tail probabilities should be *relatively* smooth.

Mathematically, this can be expressed by assuming that as L grows large:

$$\frac{\Delta F(L)}{1 - F(L^-)} \rightarrow 0 \quad (22)$$

where $\Delta F(L)$ is any “jump” in the distribution $F(\cdot)$ at the threshold point L , and where the denominator is the tail probability, *including* this jump. This condition means that any jumps in the distributions should become *relatively* negligible as we go farther and farther into the tails.

Interestingly, this condition may be non-trivial in finance, and it may end up as quite relevant in risk management applications. Suppose we are dealing with threshold bankruptcy conditions where various borrowers default as some thresholds are exceeded successively. Then, at these thresholds, the default probabilities for the portfolio under consideration could jump, and the resulting distribution for the extremes could also exhibit jumps. If the observed data on extremes are the outcome of a sum of a large number of individual credits, the tail of the aggregate series may end up quite irregular. The ratio in (22) may never converge to zero. Then a non-trivial asymptotic extreme value distribution may not exist.

How can we apply these theoretical results to real-world data coming from markets? The requirement in (22), which may be more than just a technical assumption, is maintained for the analysis. We assume that the “true” distribution function of the incremental changes in asset prices is not known but given by $F(\Delta x_t)$.⁹

In practical applications of extreme value theory, one first picks a high level U so that all $\Delta x_t > U > 0$ are defined to be in the positive tail of the distribution.¹⁰ This U has to be selected so that the threshold needed for value at risk calculations, namely, the L , is much farther into the tail, so that $L > U$.

Using this U , next define the probabilities associated with Δx_t :

$$P(\Delta x_t \leq U) = F(U) \quad (23)$$

$$P(\Delta x_t \leq U + y_t) = F(U + y_t) \quad (24)$$

where $0 < y_t$ is an *exceedance* of the threshold U at time t . Finally, let $F_u(y_t)$ be given by:

$$F_u(y_t) = \frac{F(U + y_t) - F(U)}{1 - F(U)} \quad (25)$$

We thus obtain the F_u , the conditional distribution of *how* extreme a Δx_t is, given that it *already* qualifies as an extreme. Pickands [1975] shows that F_u will be very close to the so-called generalized Pareto distribution G , if U is a high threshold:

$$G(y_t, \sigma, k) = \begin{cases} 1 - \left(1 - \frac{ky_t}{\sigma}\right)^{1/k} & k \neq 0 \\ 1 - e^{-y_t/\sigma} & k = 0 \end{cases} \quad (26)$$

That is, the distance between G and F_u will converge to zero as U increases.

So, even though we never know the true distribution, as long as the data have enough “true” extremes, and as long as the true tails are well-behaved, we can always use G to estimate the *true* tails.

III. ESTIMATION

To calculate value at risk by these tail probabilities, assume that n upcrossings of a predetermined level U have been observed for the random variable Δx_t at time (t_i) . These observations represent extremes and are written as $(U + y_{t_1}, \dots, U + y_{t_n})$, so that the y_i represent the *exceedances*. Further, let U be a high enough level so that the generalized Pareto distribution, G with $k \neq 0$, is a good approximation for the true distribution F_u . Then, the probabilities associated with these extreme observations are given by:

$$F_u(\Delta x_t - U < y_{t_i}) = 1 - \left(1 - \frac{ky_{t_i}}{\sigma}\right)^{1/k} \quad (27)$$

From here, we can calculate the density function and then the likelihood function.¹¹

$$\mathcal{L}^n(k, \sigma) = -n \ln(\sigma) +$$

$$\sum_{i=1}^n \ln \left(1 - \frac{ky_{t_i}}{\sigma}\right) k^{-1} - \ln \left(1 - \frac{ky_{t_i}}{\sigma}\right) \quad (28)$$

The likelihood function is to be maximized with respect to the unknown parameters σ and k . Once the parameters k and σ are estimated using maximum likelihood, we can approximate the tails at each $U + y_{t_i}$ using the approximation given in (25):

$$1 - \hat{F}(U + y_{t_i}) = \frac{n}{N} \left(1 - \frac{\hat{k}y_{t_i}}{\hat{\sigma}}\right)^{1/\hat{k}} \quad (29)$$

where the \hat{k} and $\hat{\sigma}$ are the maximum likelihood estimates of k and σ ; and the n and N are the number of observations on the extremes and the number of total data points, respectively. The ratio n/N is an estimate of $1 - F(U)$, the unconditional probability that an observation will exceed U .

The critical value L that corresponds to various levels of α in (20) can be found from this estimate by first letting

$$\alpha = \frac{n}{N} \left(1 - \frac{\hat{k}L}{\hat{\sigma}}\right)^{1/\hat{k}} \quad (30)$$

and then solving for the unknown value L . The solution yields:

$$\hat{L} = -\hat{\sigma} \left(\left(\frac{\alpha N}{n} \right)^{\hat{k}} - 1 \right)^{\hat{k}^{-1}}$$

The \hat{L} obtained this way can then be used for VaR calculations. For example, instead of using Equation (10), we can let:

$$\text{VaR}(B_x, \alpha) = B_x \hat{L} \quad (31)$$

The value at risk calculated this way will be obtained by using *only* the extremes of observed data.

Given that the asymptotic convergence is limited to tails, this concentration on extremes is likely to provide more accurate VaR estimates. More important, the functional forms selected will *not* be ad hoc, if the conditions for the existence of asymptotic distributions are satisfied. Finally, note that this approach allows for all types of asymmetries in the extremes for drops and jumps.

IV. VAR WITH EXTREME DISTRIBUTIONS

Our main objective is to obtain estimates of the thresholds L , given by Equation (30), and then compare these with the thresholds generated by the standard approach.

We use daily interest rate and exchange rate data from RiskMetrics, in order to provide estimates of L obtained from extreme value theory. The sample period varies slightly from one variable to another. The overall sample period ranges from January 1, 1990, through October 31, 1995.

We use only some of the series in this broad data set. In particular, we concentrate on the extremes of three-month, two-year, five-year, and seven-year interest rates from the U.S., and on the dollar (USD)-mark (DEM), dollar-yen (YEN), dollar-French franc (FF), and dollar-British pound (GBP) exchange rates. We work with the levels of interest rates, and with the logarithms of exchange rates.

These results are shown in Exhibit 1. The first step is to obtain the extremes from the RiskMetrics data. Here we follow standard procedures used in engineering applications.¹²

We first estimate the standard deviation, $\hat{\sigma}_n$, from the whole sample of observed increments. Then, we select all *positive* and all *negative* increments greater than $1.65\hat{\sigma}_n$ in absolute value as representing “extremes.” Thus, the extremes would belong to 5% tails if the true distribution is indeed normal.

This way, we obtain sixteen samples of extreme observations, two for each risk factor under consideration. The first subsample is made of extreme *negative* increments, the second of extreme *positive* increments.

The approach separates sudden drops from sudden jumps. This allows us to test whether the two tails are indeed symmetric.¹³ Otherwise, risk exposures of portfolios may be different, depending on whether one is long or short on a particular instrument. One advantage of the asymptotic theory of extremes is that such asymmetries in the two tails can easily be handled.¹⁴

To calculate the \hat{L} shown in the second column of Exhibit 1, the maximum likelihood estimates of k and σ must be obtained. We do this by programming, in Matlab, the first and second derivatives of the likelihood function (28). Then, the Matlab function “fsolve” is used to find the maxima. The estimates \hat{k} and $\hat{\sigma}$ obtained this way are reported in Exhibit 2. Once \hat{k} and $\hat{\sigma}$ are selected, we choose $\alpha = 0.01$, and then calculate the L using Equation (30).

Exhibit 1 prompts several observations.

- The extreme tails yield threshold points that are up to 33% higher than the thresholds one obtains with the standard approach. This result seems to be quite homogeneous across risk factors.
- Although this implies a 20% to 30% greater VaR for most risk factors than the RiskMetrics approach, the BIS multiplication factor of 3 seems to be excessive.
- For exchange rates, the extreme tails seem to be quite similar whether one looks at drops or jumps. For interest rates, however, drops seem to have tail areas farther away from normal.
- Finally, short-term interest rates have more pronounced non-normal tails than do long rates.

The estimated tail areas can be seen more closely in Exhibits 3 and 4. The same parameter estimates are also used in plotting the extreme and Gaussian tails for each of the risk factors under consideration. In the case of normality, the tails shown are those of the standard normal distribution with a cutoff point drawn at $2.3\hat{\sigma}$, which is the standard deviation estimated from the time series of that risk factor’s daily increments. In the case of extreme distributions, the tail areas are obtained by replacing the k and σ in the generalized Pareto density function $g(y, \sigma, k)$ by their maximum likelihood estimates. In this case, the initial point of the tail is at $y = 0$.

V. PERFORMANCE OF EXTREME DISTRIBUTIONS

So far, the estimates indicate that the tail areas obtained from extreme theory are quite different from and potentially more useful than the tails obtained in the standard fashion. The implied value at risk estimates are consequently different as well. Yet, the real test for a theory is not that it is different from the alternatives, but whether it is any *better* according to some specific criteria.

We provide three different sets of evidence that the

EXHIBIT 1

Estimated Thresholds

Series		Normal Distribution Threshold	Extreme Distribution Threshold	Mult. Factor
90-Day T-Bills	Positives	0.1245	0.1527	1.23
	Negatives	0.1245	0.1660	1.33
2-Year Rates	Positives	0.1426	0.1720	1.21
	Negatives	0.1426	0.1841	1.29
5-Year Rates	Positives	0.1421	0.1651	1.16
	Negatives	0.1421	0.1669	1.17
7-Year Rates	Positives	0.1396	0.1430	1.02
	Negatives	0.1396	0.1659	1.19
DEM/USD	Positives	0.0109	0.0137	1.26
	Negatives	0.0109	0.0135	1.24
FF/USD	Positives	0.0030	0.0034	1.33
	Negatives	0.0030	0.0034	1.20
GBP/USD	Positives	0.0268	0.0296	1.10
	Negatives	0.0268	0.0345	1.29
Yen/USD	Positives	0.1447	0.1790	1.24
	Negatives	0.1447	0.1674	1.16

use of extreme distributions may represent a significant improvement over the standard VaR estimates.¹⁵

In-Sample Performance

The first question that we ask is related to the in-sample performance of VaR numbers calculated from the extreme distributions. Note that we fit the parameters of the extreme distribution using the 5% tail observations, and then calculate the 1% VaRs. Now, if this technique is superior to the current alternatives, the standard VaR should actually leave more than 1% of the observations beyond the threshold, while the extreme value method should capture 1% of the observations more closely. Thus we first compare the in-sample performance of the extreme value method with the alternative of using the normal distribution.

Exhibit 5 is the result of tallying the number of observations that fall in the 1% tails for normal and for extreme distributions. The results show the normal distribution to be quite inadequate for determining a 1% threshold for market risk factors. In fact, given that we have 1,519 in-sample data points, we would expect 15 observations to fall into each 1% tail. Yet, on the average, 1% tails estimated from the normal distribution include 30.4 observations, twice the number expected.

EXHIBIT 2

Parameter Estimates

Series		k	σ
90-Day T-Bills	Positives	0.31	0.056
	Negatives	-0.32	0.043
2-Year Rates	Positives	0.57	0.063
	Negatives	0.23	0.068
5-Year Rates	Positives	0.22	0.058
	Negatives	0.09	0.053
7-Year Rates	Positives	0.29	0.044
	Negatives	0.33	0.043
DEM/USD	Positives	-0.04	0.00283
	Negatives	0.31	0.00496
FF/USD	Positives	0.13	0.00097
	Negatives	-0.03	0.00079
GBP/USD	Positives	-0.05	0.00975
	Negatives	0.04	0.00665
Yen/USD	Positives	-0.14	0.04520
	Negatives	-0.05	0.06230

According to Exhibit 5, the 1% tails from the extreme distributions capture on the average 15.1 observations, exactly as expected a priori. Also, because the extreme distribution can estimate the shape of the tails for negative and positive extremes *separately*, the 1% observations are more evenly distributed. Because the tails of the normal distribution are symmetric, the traditional approach lacks this flexibility, leading to large differences between 1% negative and positive extremes.

Out-of-Sample Performance

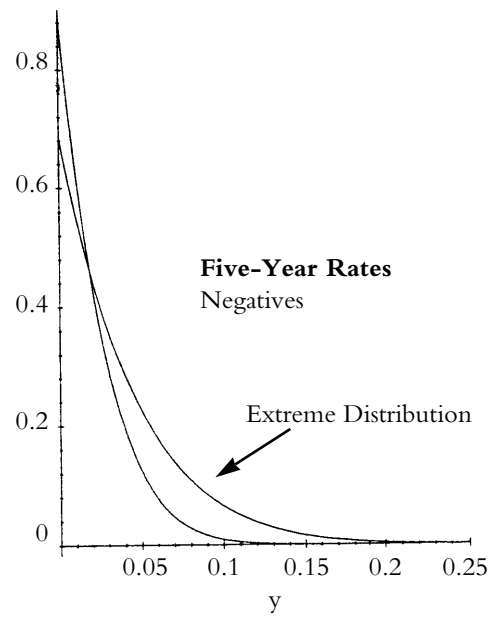
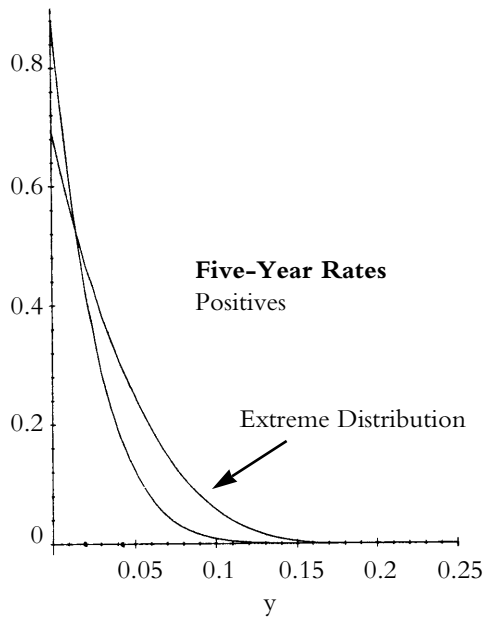
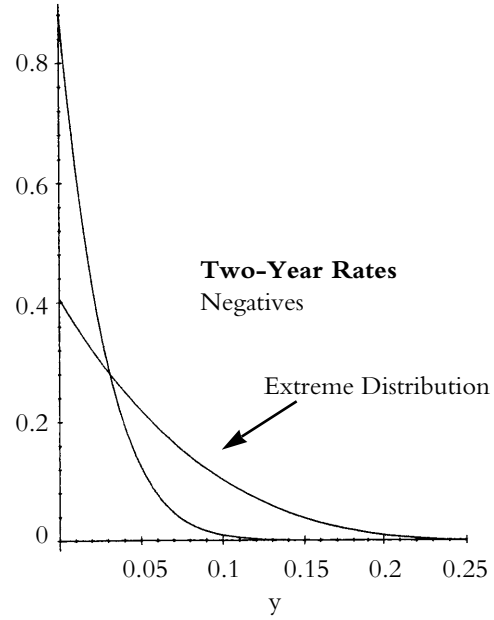
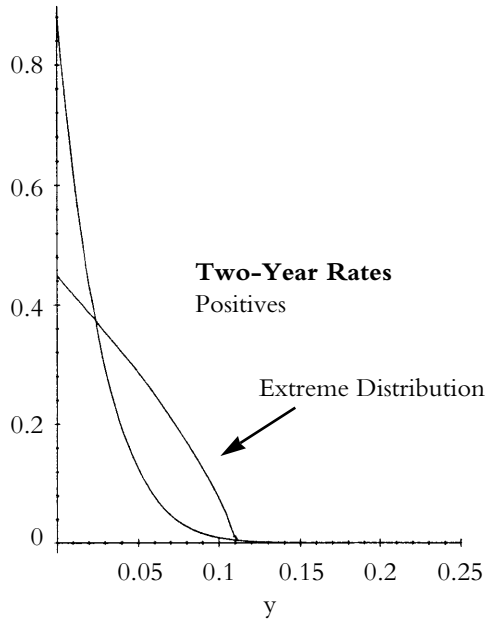
The real test of a risk management methodology is out-of-sample performance. The risk manager, by definition, obtains VaR estimates in real time and hence must use parameters obtained from an already observed sample in order to evaluate the risks associated with current and future random movements in risk factors. Hence, a true test for the extreme distribution methods is their performance outside the sample used to estimate the underlying parameters.¹⁶

To measure out-of-sample performance, we proceed as follows. Our in-sample data are for the period 1990–1995. We obtain additional data on the same variables observed since the end of 1996. This gives us two years of daily data for the period 1997–1998. All in all, this is 504 observations.

We then provide two different sets of measures for

EXHIBIT 3

Normal and Extreme Tail Areas for Interest Rates

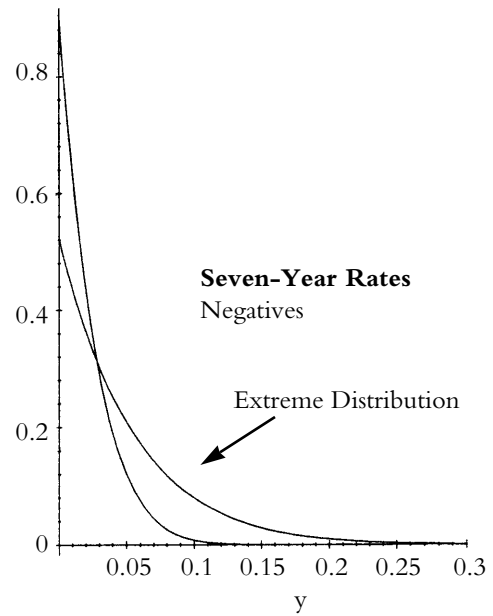
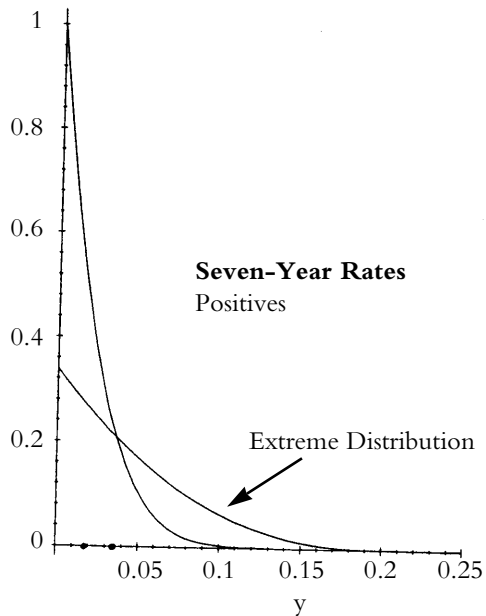
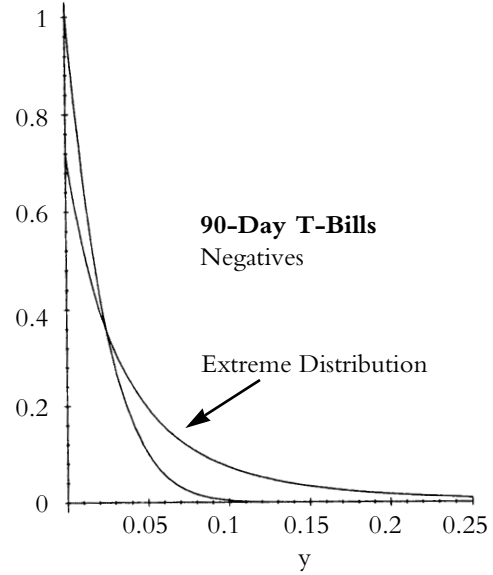
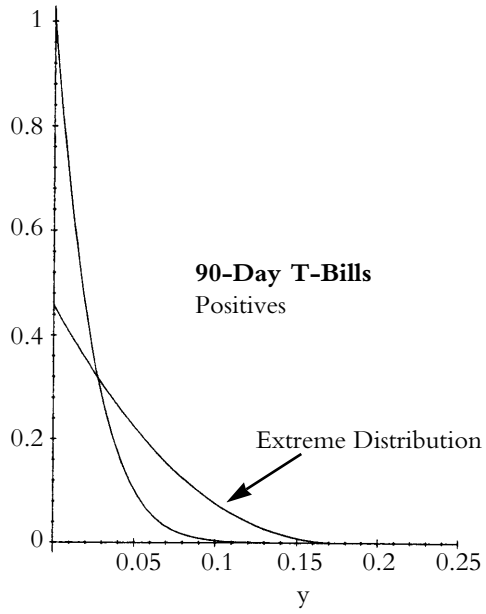


the performance of extreme value tails. First, we recalculate Exhibit 5 with the new out-of-sample data. To be more precise, we try to see how many of the out-of-sample observations fall in the 1% tails estimated using normal and extremal distributions. With 504 data points,

we expect the two 1% tails to capture around five observations each.

The out-of-sample results are displayed in Exhibit 6. We see that the VaRs for the normal distribution have on the average nine observations falling in the 1% tails.

EXHIBIT 3
Continued



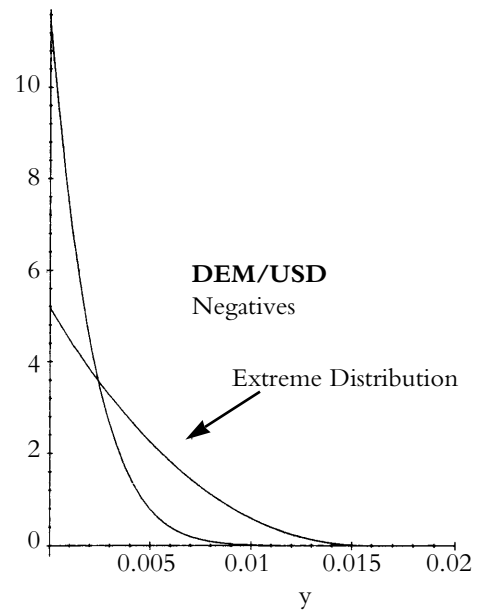
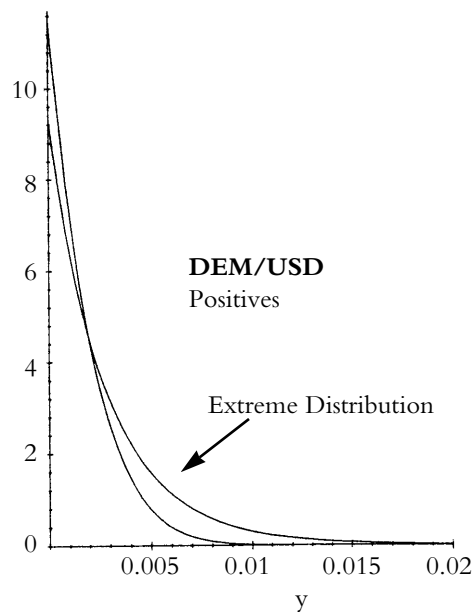
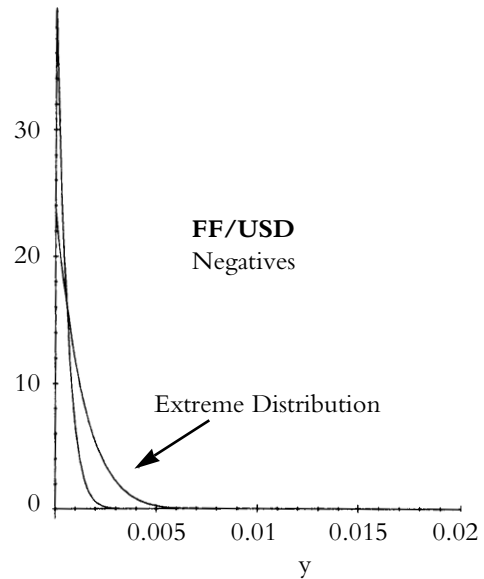
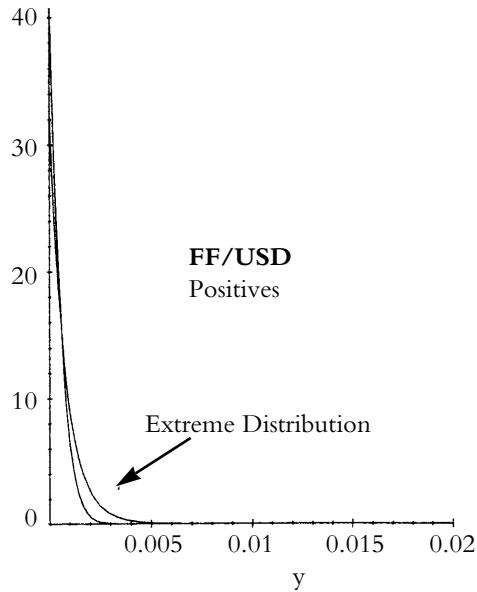
This is again twice the correct number. In the case of the extreme distributions, the tails contain 4.5 observations each for interest rates. This is quite a remarkable result in that the performance of extreme tails is quite robust even when the data are from a relatively turbulent period.

For exchange rates, the results again favor extreme

distribution theory over the normal distributions, although here the difference between the two methods is somewhat smaller. The normal distribution tails contain on the average 7.4 out-of-sample observations, while the extreme distribution tails contain 3.25. When we look at individual exchange rates, we can see that it is the dollar-mark rate

EXHIBIT 4

Normal and Extreme Tail Areas for Exchange Rates



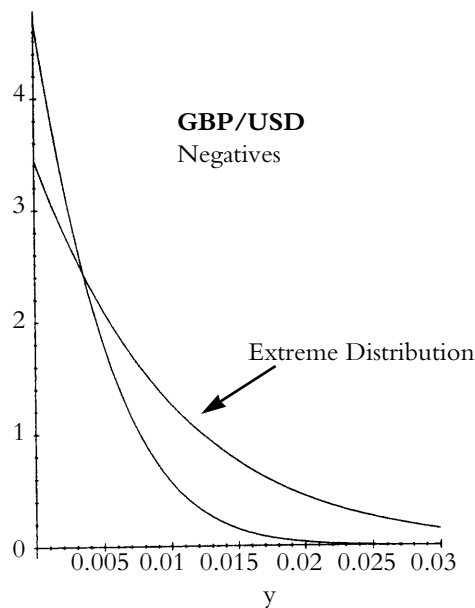
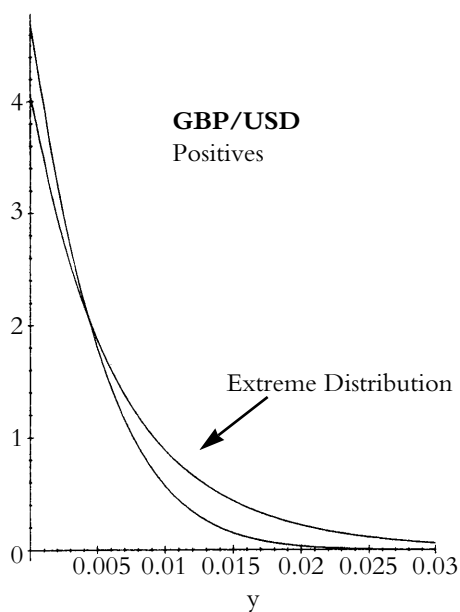
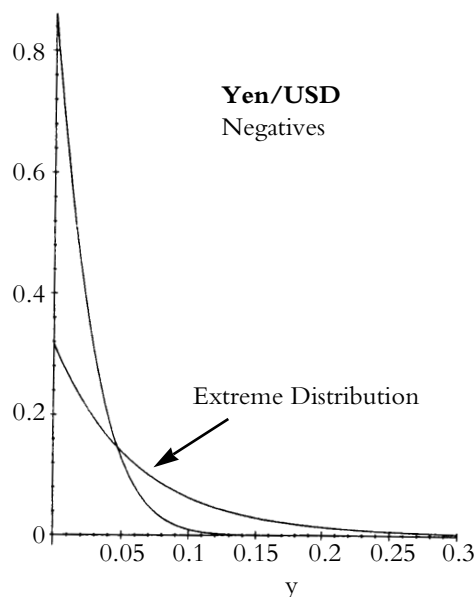
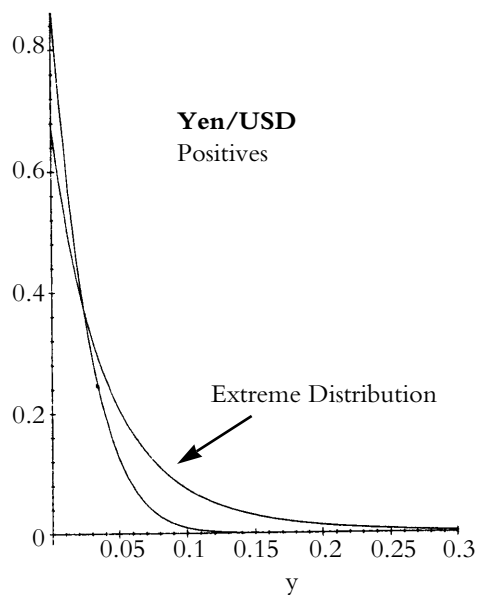
that has been less volatile than others. If we exclude this rate, the extreme tails contain on the average four observations at each tail, very close to the critical number of five.

Hence the out-of-sample performance indicates that the extreme tail estimates would provide a much more reliable method for estimating VaRs.

Mean Excesses

Traditional approaches to VaR estimation give us a tail cutoff, and this is used in capital allocation. But, VaR does not tell us how extreme a typical observation is, *conditional* on it being in the tail.

EXHIBIT 4
Continued



In other words, let Δx_t represent the increment of an observed risk factor, and let the \hat{L} denote the threshold calculated using the extreme distribution. Consider now the expectation:

$$E^P[\Delta x_t - \hat{L} \mid \Delta x_t > \hat{L}] \quad (32)$$

calculated using the probability distribution P . This expectation is called the *mean excess function* of the level \hat{L} and measures how extreme the extremes of a series get on average.

We can use the notion of mean excess function in evaluating the performance of extreme distributions. To

EXHIBIT 5

Relative Performance of VaR and Extreme Distribution — Within-Sample Data

Series		Normal Distribution Count	Extreme Distribution Count
90-Day T-Bills	Positives	37	14
	Negatives	46	14
2-Year Rates	Positives	27	12
	Negatives	28	14
5-Year Rates	Positives	27	16
	Negatives	24	15
7-Year Rates	Positives	26	19
	Negatives	28	17
Average		30.4	15.1
DEM-USD	Positives	25	9
	Negatives	29	14
FF-USD	Positives	23	13
	Negatives	29	19
GBP-USD	Positives	22	17
	Negatives	35	14
Yen-USD	Positives	32	13
	Negatives	20	14
Average		26.9	14.1

do this, we consider interest rate processes only, and calculate the mean excesses in two different ways.¹⁷

First, we calculate the actual out-of-sample mean excesses for the observed 1% extremes. Then, we calculate the mean excess predicted by the formula:

$$E^P[\Delta x_t - \hat{L} | \Delta x_t > \hat{L}] = \frac{\hat{\sigma} + \hat{k}U}{1 - \hat{k}} \quad (33)$$

where P is the generalized Pareto distribution, and $\hat{\sigma}$, \hat{k} , and U are as defined in Equation (27).¹⁸ Given the estimates $\hat{\sigma}$, \hat{k} , and \hat{U} , Equation (33) yields the mean excess predicted by the extreme distribution function.

The results are shown in Exhibit 7. The first column shows the out-of-sample estimates for mean excesses over the thresholds \hat{L} for that particular series. In other words, this is the observed mean of how much an extreme exceeds the 1% threshold.

Numbers in the second column are estimated using the mean excesses predicted by the estimated

EXHIBIT 6

Relative Performance of VaR and Extreme Distribution — Out-of-Sample Data

Series		Normal Distribution Count	Extreme Distribution Count
90-Day T-Bills	Positives	7	4
	Negatives	6	2
2-Year Rates	Positives	11	7
	Negatives	10	1
5-Year Rates	Positives	10	9
	Negatives	10	4
7-Year Rates	Positives	10	5
	Negatives	8	4
Average		9	4.5
DEM-USD	Positives	8	2
	Negatives	7	0
FF-USD	Positives	6	4
	Negatives	5	4
GBP-USD	Positives	11	4
	Negatives	5	3
Yen-USD	Positives	12	6
	Negatives	6	3
Average		7.4	3.25

EXHIBIT 7

Predicted and Actual Mean Excesses

Series		Actual Mean Excess	Mean Excess Predicted by Extreme Distribution	Mean Excess Predicted by Lognormal Distribution*
90-Day T-Bills	Positives	0.21	0.17	0.06
	Negatives	-0.19	-0.16	-0.06
2-Year Rates	Positives	0.28	0.33	0.04
	Negatives	-0.20	-0.14	-0.04
5-Year Rates	Positives	0.23	0.16	0.04
	Negatives	-0.20	-0.11	-0.04
7-Year Rates	Positives	0.21	0.17	0.04
	Negatives	-0.21	-0.22	-0.04
Mean Absolute Value		0.19	0.16	0.04

*The mean excess of a lognormal distribution over a 1% tail.

extreme densities, i.e., using Equation (33).

Finally, the third column displays the mean excesses for the lognormal distribution. We use the sample mean and variance of *changes* in interest rates to calculate these averages. Then, we consider the 1% tail of the distribution and obtain the expected excess over the 1% threshold. The results show that mean excesses provided by the normal distribution are, in all cases, quite inadequate in capturing the extremeness of out-of-sample data.

The estimates in Exhibit 7 are again remarkable. We can draw two conclusions. First, the out-of-sample performance of the mean excesses predicted by the extreme distributions is on the average 84% of the actual mean excesses observed for each series. Second, the mean positive excesses are slightly better predicted than the negatives.

VI. CONCLUSIONS

The theory of extremes has some advantages over the standard approach to risk management. We have shown that the statistical theory of extremes provides a more precise approach for risk management and value at risk calculations.

The results, applied to eight major risk factors, show that the statistical theory of extremes and the implied tail estimation are indeed useful for VaR calculations. The implied VaR would be 20% to 30% greater if one used the extreme tails rather than following the standard approach.

The most remarkable results are in evaluation of the out-of-sample performance of the extreme tails. The empirical results show that the VaRs calculated using tails of extreme distributions are significantly more precise than the standard approach. VaR tails are more robust and yield more accurate estimates of the rate of occurrence and the size of extreme observations.

ENDNOTES

¹This may be accomplished either parametrically or non-parametrically.

²These issues are closely related to a controversial item in the Bank for International Settlements guidelines, namely, the so-called multiplication factor. Institutions are required by BIS to calculate a 99% upper bound for their daily losses due to market risk, and then are required to multiply this figure by at least 3. This *multiplication factor* is justified essentially by the observed non-Gaussian nature of frequency distributions. Our approach, which is non-parametric and potentially more pre-

cise, yields a more satisfactory estimate for this factor.

³The approach of Ait-Sahalia [1996] does not require discretizing the SDE. Ait-Sahalia provides clever ways to estimate the σ_t without any discretization bias. Clearly, this approach is superior to the present market practice, as long as the required assumptions are met.

⁴The assumption $B_{xx} = 0$ is justified for linear products such as Eurocurrency futures. For forward rate agreements (FRAs), bonds, and swaps, the assumption can be only an approximation. Yet the industry standard is to proceed with the linearity assumption for all instruments that do not have optionality.

⁵To say that the mean exists indicates that $E[\Delta x_t] < \infty$.

⁶The case of infinite variance may look like a special case, and the α -stable distributions may appear a small class, but in finance their relative weights may be important. In fact, heavy tails are the rule rather than the exception in finance, and the convergence of sums may be to the so-called α -stable class. A typical α -stable random variable X can be represented using the corresponding characteristic function:

$$g(t) = e^{iat - c|t|^\alpha [1 - i\beta \text{sign}(t)Z(t, \alpha)]}$$

with $Z(\cdot)$ given by

$$Z(t, \alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \text{if } \alpha \neq 1 \\ \frac{-2}{\pi} \ln|t| & \text{if } \alpha = 1 \end{cases}$$

where α plays the role of a mean, $c > 0$ is a parameter that depends on the variance σ^2 , the $\alpha \in (0, 2]$ determines the basic properties of the stable distributions, and, finally, the $\beta \in [-1, 1]$ is related to the skewness of the distribution.

⁷A normalized tail has the advantage that the area under it sums to one. Hence it can be used as a “tail distribution.”

⁸Note that if we want to make statements about probabilities associated with extreme values at various levels, probabilities of 0 or 1 will hardly be useful as tail distributions. What we need instead are probabilities that indicate how much more likely one extreme value is *relative* to another. Thus we need asymptotic probabilities *between* 0 and 1.

⁹We follow Pickands [1975], Smith [1987], and Embrechts et al. [1997], and use the asymptotic theory of extremes in approximating the tail areas of $F(\Delta x_t)$.

¹⁰Here we work with upcrossings of U only. The case of downcrossings of D , that is, the “negative” tail, is similar.

¹¹The *true* density function of Δx_t can be approximated at an arbitrary observation point y by the generalized Pareto density $g(\cdot)$:

$$g(\sigma, k, y) = \left(\frac{\sigma - ky}{\sigma}\right)^{1/k} (\sigma - ky)^{-1}$$

¹²For a useful survey and the selection of extreme observations, see Davison and Smith [1990].

¹³Of course, in the case of normality, the two tails will be identical by assumption.

¹⁴The use of non-parametric density estimation in RiskMetrics will have the same effect.

¹⁵Thanks to the editor for suggesting this line of research, which yields surprisingly strong results.

¹⁶This would also be a test for our estimation method as well.

¹⁷For exchange rates we have relatively few out-of-sample observations that cross the extreme thresholds. The implied means have high standard errors.

¹⁸Embrechts et al. [1997] discuss mean excess functions for a large number of distributions. See Theorem 3.4.13, for example.

REFERENCES

Ait-Sahalia, Y. "Non-Parametric Pricing of Interest Rate Derivative Securities." *Econometrica*, 1996.

Davison, A.C., and R.L. Smith. "Models for Exceedences of High Thresholds." *Journal of the Royal Statistical Society*, 1990.

Embrechts, P., C. Kluppelberg, and T. Mikosh. *Modelling Extremal Events*. Springer, 1997.

Pickands, J. "Statistical Inference Using Extreme Order Statistics." *Annals of Statistics*, 1975.

RiskMetrics. J.P. Morgan, Technical Document, 3rd ed., 1996.

Smith, R.L. "Estimating Tails of Probability Distributions." *Annals of Statistics*, 1987.