Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management

FRANCIS X. DIEBOLD, TIL SCHUERMANN, AND JOHN D. STROUGHAIR

FRANCIS X. DIEBOLD is Michael Armellino Professor of Business at New York University and a member of the Oliver Wyman Institute.

TIL SCHUERMANN is Director of Research at Oliver, Wyman & Company in New York.

JOHN D. STROUGHAIR is a Director at Oliver, Wyman & Company in New York.

Financial risk management is intimately concerned with tail quantiles (e.g., the value of x such that $P(X > x) = 0.05$) and tail probabilities (e.g., $P(X > x)$, for a large value x). *Extreme* quantiles and probabilities are of particular interest because the ability to assess them accurately translates into the ability to manage extreme financial risks effectively, such as those associated with currency crises, stock market crashes, and large bond defaults.

Unfortunately, traditional parametric statistical and econometric methods, typically based on estimation of entire densities, are ill-suited to the assessment of extreme quantiles and event probabilities. These parametric methods implicitly strive to produce a good fit in regions where most of the data fall, potentially at the expense of good fit in the tails, where, by definition, few observations fall.[1] Seemingly sophisticated non-parametric methods of density estimation, such as kernel smoothing, are also well-known to perform poorly in the tails.[2]

It is common, moreover, to require estimates of quantiles and probabilities not only *near* the boundary of the range of observed data, but also *beyond* the boundary. The task of estimating such quantiles and probabilities would seem hopeless. A key idea, however, emerges from an area of probability and statistics known as "extreme value theory" (EVT): One can estimate extreme quantiles and probabilities by fitting a "model" to the empirical survival function of a set of data using only the extreme event data rather than all the data, thereby fitting the tail and only the tail.[3] The approach has a number of attractive features, including:

1. The estimation method is tailored to the object of interest, the tail of the distribution, rather than the center of the distribution.
2. An arguably reasonable functional form for the tail can be formulated from a priori considerations.

The upshot is that the methods of EVT offer hope for progress toward the elusive goal of reliable estimates of extreme quantiles and probabilities.

The concerns of EVT are in fact wide-ranging and include fat-tailed distributions, time series processes with heavy-tailed innovations, general asymptotic theory, point process theory, long memory and self-similarity, and much else. Our concerns here, alluded to above, are much more narrow: We focus primarily on estimation of extreme quantiles and probabilities, with applications to financial risk management. Specifically, we provide a reaction, from the perspective of financial risk management, to recent developments in the EVT literature, the maturation of which has resulted in optimism regarding the prospects for practical applications of the ideas to financial risk management.[4] In our view, as we hope to make clear, the optimism is partly appropriate but also partly exaggerated, and at any rate much of the potential of EVT remains latent; hence what follows is partly praise, partly criticism, and partly a wish list.

PITFALLS AND OPPORTUNITIES

Much of our discussion is related directly or indirectly to the idea of tail estimation under a power law assumption, so we begin by introducing the basic framework. We assume that returns are in the maximum domain of attraction of a Frechet distribution, so that the tail of the survival function is a power law times a slowly varying function:

$$P(X > x) = k(x)x^{-\alpha} \tag{1}$$

That family includes, for example, α-stable laws with $\alpha < 2$ (but not the Gaussian case, $\alpha = 2$). Often it is assumed that $k(x)$ is in fact constant, in which case attention is restricted to densities with tails of the form

$$P(X > x) = kx^{-\alpha} \tag{2}$$

with the parameters k and α to be estimated.

A number of estimation methods can be used. The most popular, by far, is Hill's [1975] estimator, which is based directly on the extreme values and proceeds as follows. Order the observations with $x_{(1)}$ the largest, $x_{(2)}$ the second largest, and so on, and form an estimator based on the difference between the m-th largest observation and the average of the m largest observations:

$$\hat{\alpha} = \left[\left(\frac{1}{m}\sum_{i=1}^{m}\ln(x_{(i)})\right) - \ln(x_{(m)})\right]^{-1} \tag{3}$$

It is a simple matter to convert an estimate of α to estimates of the desired quantiles and probabilities. The Hill estimator has been used in empirical financial settings; see, for example, the impressive early work of Koedijk, Schafgans, and deVries [1990]. It also has good theoretical properties. It can be shown that it is consistent and asymptotically normal, assuming the data are iid and that m grows at a suitable rate with the sample size.[5]

Selecting the Cutoff for Hill Estimation

We discuss the iid assumption more later; here we focus on choice of m. Roughly speaking, consistency and asymptotic normality of the Hill estimator require that m approach infinity with the sample size, but at a slower rate. Unfortunately, however, this asymptotic condition gives no guidance regarding choice of m for any asset return series of fixed length, and the finite sample properties of the estimator depend crucially on the choice of m.

In particular, there is an important bias-variance trade-off when varying m for fixed sample size: Increasing m, and therefore using progressively more data (moving toward the center of the distribution), reduces variance but increases bias. Increasing m reduces variance because more data are used, but it increases bias because the power law is assumed to hold only in the extreme tail.

The bias-variance trade-off regarding choice of m is precisely analogous to the bias-variance trade-off regarding choice of bandwidth in non-parametric density estimation. Unfortunately, although much of the recent non-parametric density estimation literature is devoted to "automatic" bandwidth selection rules, which optimize an objective function depending on both variance and bias, until recently no such rules had been developed for choice of m in tail estimation. Instead, m is typically chosen via ad hoc rules of thumb, sometimes supplemented with plots of the empirical survival function with the estimated tail superimposed. In our view, such plots are invaluable — and although they are used sometimes, they are not used often enough — but they need to be supplemented with complementary and rigorous rules for suggesting a choice of m.

Recent unpublished work by Danielsson and deVries [1997a] offers hope: They develop a bootstrap method for selecting m and establish optimality properties. We look forward to continued exploration of bootstrap and other methods for optimal selection of m, and to the accumulation of practical experience with the methods in financial settings.

Convenient Variations on the Hill Estimator I: Estimation by Linear Regression

Although the Hill estimator was not originally motivated by a regression framework, the basic idea — assume a power law for the tail and then base an estimator directly on the extreme observations — has an immediate regression interpretation, an idea that traces at least to Kearns and Pagan [1997]. Simply note that $P(X > x) = kx^{-\alpha}$ implies

$$\ln P(X > x) = \ln(k) - \alpha\ln(x) \tag{4}$$

so that α is the slope coefficient in a simple linear relationship between the log tail of the empirical survival function and the log extreme values. One should therefore be able to estimate α by a linear regression of the log tail of the empirical survival function on an intercept and the logs of the m most extreme values.

Note that nothing is presently known about the properties of standard estimators (e.g., least squares) in this context; more research is needed. The basic insight nevertheless strikes us as important. First, it highlights the conceptual fact that the essence of the tail estimation problem is fitting a log-linear function to a set of data. Second, it may have great practical value because it means that the extensive kit of powerful tools developed for regression can potentially be brought to bear on the tail estimation problem. Huge literatures exist, for example, on robust estimation of regression models (which could be used to guide strategies for model fitting), recursive methods for diagnosing structural change in regression models (which could be used to guide selection of m), and bootstrapping regression models (which could be used to improve finite sample inference).

Convenient Variations on the Hill Estimator II: Estimation by Non-Linear Regression

A more accurate tail expansion, used for example in Danielsson and deVries [1997b], and asymptotically equivalent to the one given above, but which may have better properties in small samples, is

$$P(X > x) = kx^{-\alpha}(1 + cx^{-d}) \tag{5}$$

where k, α, c, and d are the parameters to be estimated. Again, the relevant parameters could be estimated very simply by non-linear regression methods, with the choice of m and much else guided by powerful regression diagnostics.

Convenient Variations on the Hill Estimator III: Imposing Symmetry

In many (although not all) financial applications, we may want to impose symmetry on the distribution, which is to say, we may want to impose that the parameters governing the left and right tail behavior be identical. Regression technology again comes to the rescue. Using standard methods, one may estimate a system of two equations, one for each tail, imposing parameter equality across equations.

Finite Sample Estimation and Inference

A reading of the EVT literature reveals a sharp and unfortunate contrast between probability theory and statistical theory. Probability theory is elegant, rigorous, and voluminous, while statistical theory remains primitive and skeletal in many respects. Readers who have reached this point will already appreciate certain aspects of this anticlimax, such as the difficulty of choosing the cutoff in Hill estimation. But the situation is actually worse.

First, notoriously little is known about the finite sample properties of the tail index estimator under various rules for cutoff selection, even under the standard maintained assumptions, most importantly that the data are iid. In our view, a serious and thorough Monte Carlo study is needed, standardizing the choice of m by a defensible automatic selection rule such as the Danielsson-deVries bootstrap, and characterizing the sampling distribution of the tail index estimator as a function of sample size for various data-generating processes satisfying the power law assumption. Such a Monte Carlo analysis would be facilitated by the simplicity of the estimator and the speed with which it can be computed, but hindered by the tedious calculations required for bootstrap cutoff selection.

Second, the problem is not just that the properties of the existing approach to estimation of the tail index have been inadequately explored. The problem is augmented by the fact that crucial variations on the theme have been left untouched. For example, in financial risk management, interest centers not on the tail index per se, but rather on the extreme quantiles and probabilities. The ability to withstand big hits, quantified by specific probabilities, translates directly into credit ratings, regulatory capital requirements, and so on. The extreme quantile and probability estimators are highly non-linear functions of the tail index; hence, poor sampling properties of the tail index estimator will likely translate into even worse properties of the quantile and probability estimators, and the standard approximation based on a Taylor series expansion ("the delta method") is likely to be poor. Moreover, often one would like an *interval* estimate of, say, an extreme event probability. One rarely sees such intervals computed in practice, and we know of no Monte Carlo work that bears on their coverage accuracy or width.

Third, various maintained assumptions may of course be violated and may worsen the (potentially already poor) performance of the estimator. Again, a systematic Monte Carlo analysis would help us to assess the severity of the induced complications. To take one example, the scant evidence available suggests the estimator is likely to perform poorly if the assumed slowly varying function k(x) is far from constant, even if the tail behavior satisfies the power law assumption (see Embrechts, Klüppelberg, and Mikosch [1997, Chapter 6]). This is unfortunate insofar as, although power law tail behavior can perhaps be justified as reasonable on a priori grounds, no justification can be given for the assumption that k(x) is constant.

Other maintained assumptions can also fail and may further degrade the performance of the estimator. We now consider two such crucial assumptions, iid data and power law tail behavior, in greater detail.

Violations of Maintained Assumptions I: Dependent Data

It is widely agreed that high-frequency financial asset returns are conditionally heteroscedastic, and hence not iid.[6] Unfortunately, the EVT literature routinely assumes iid data. Generalizations to dependent data have been attempted (e.g., Leadbetter, Lindgren, and Rootzen [1983]), but they typically require "anticlustering" conditions on the extremes, which rule out precisely the sort of volatility clustering routinely found in financial asset returns. It seems clear that Monte Carlo analyses need to be done on the performance of the tail estimator under realistic data-generating processes corresponding to routine real-world complications, such as the dependence associated with volatility dynamics. The small amount of Monte Carlo that has been done (e.g., Kearns and Pagan [1997]) indicates that the performance can be very poor.

We see at least two routes to improving the performance of the tail index estimator in the presence of dependent data. First, one can fit generalized extreme value distributions directly to a series of per period maxima. This idea is not new, but it is curiously underutilized in our view, and it has a number of advantages. Use of per period maxima naturally reduces dependence, use of an exact generalized extreme value distribution is justified if the periods are taken to be long enough, and maximum likelihood estimation is easily implemented. Moreover, the question addressed — the distribution of the maximum — is of intrinsic interest in risk management.

On the downside, however, the aggregation reduces efficiency, and a new "bandwidth" selection problem of sorts is introduced — selection of the appropriate amount of aggregation. That is, should one use weekly, monthly, quarterly, or annual maxima?

Second, one can estimate the tail of the conditional, rather than the unconditional, distribution. In the iid case, the two of course coincide, but they diverge under dependence, in which case the conditional distribution is of greater interest. One can fit the tail of the conditional distribution by first fitting a conditional volatility model, standardizing the data by the estimated conditional standard deviation, and then estimating the tail index of the standardized data. Such a procedure is similar to Engle and González-Rivera [1991], except that we fit only the tail rather than the entire density of the standardized data. The key is recognizing that, if the adopted volatility model is a good approximation to the true volatility dynamics, then the standardized residuals will be approximately iid, which takes us into the world for which EVT was designed.

Violations of Maintained Assumptions II: Assessing the Power Law Assumption

As we have seen, most of the tail estimation literature assumes that returns are in the maximum domain of attraction of a Frechet distribution, so that the tail of the survival function follows a power law. But how do we check that assumption, and what is the evidence? Inoue [1998] provides some key results that help answer such questions. He develops a general framework for consistent testing of hypotheses involving conditional distributions of time series, which include conditional tail behavior restrictions of the form

$$P(X_t > x \mid \Omega_{t-1}) = kx^{-\alpha} \qquad (6)$$

for sufficiently large x.[7] The hypothesized conditional tail behavior restriction means that k and α are independent of the information set, which simplifies the assessment of the conditional probabilities.

CONCLUDING REMARKS

EVT is not needed for estimating routine quantities in financial risk management, such as 10%, 5%, and perhaps even 1% value at risk. Rather, EVT is concerned with *extreme* (e.g., one-tenth of 1%) tail behavior. So EVT has its place, but it is important that it be kept in its place, and that tools be matched to tasks.[8] EVT promises exciting opportunities for certain subareas of risk management, and for helping to fill some serious gaps in our current capabilities, but it will not revolutionize the discipline.

Even when EVT is appropriate, we argue that caution is needed, because estimation of aspects of very low-frequency events from short historical samples is fraught with pitfalls. This has been recognized in other fields, albeit often after long lags, including the empirical macroeconomics literature on long-run mean reversion in real output growth rates (e.g., Diebold and Senhadji [1996]) and the empirical finance literature on long-run mean reversion in asset returns (e.g., Campbell, Lo, and MacKinlay [1997]).

The problem for the present context — applications of EVT in financial risk management — is that for estimating objects such as a "once every hundred years" quantile, the relevant measure of sample size is likely much better approximated by the number of non-overlapping hundred-year intervals than by the number of data points. From that perspective, our data samples are terribly small relative to the demands we place on them.

We emphasize, however, that the situation is *not* hopeless. EVT is here to stay, but we believe that best-practice applications of EVT to financial risk management will benefit from awareness of its limitations — as well as its strengths. When the smoke clears, the contribution of EVT remains basic and useful: It helps draw smooth curves through the extreme tails of empirical survival functions in a way that is guided by powerful theory and hence provides a rigorous complement to alternatives such as graphical analysis of empirical survival functions.[9] Our point is simply that we shouldn't ask more of the theory than it can deliver.

Finally, we have also tried to highlight a number of areas in which additional research will lead to stronger theory and more compelling applications, including strategies for automatic cutoff selection, linear and non-linear regression implementations, thorough Monte Carlo analysis of estimator performance under ideal and non-ideal conditions, focus on interval as well as point estimation, strategies for handling dependent data, and tests of the power law assumption.

We look forward to vigorous exploration of the role of EVT in risk management. In particular, we await progress on the multivariate front, since the risk manager almost always has a portfolio of assets or securities to worry about, not an individual asset, and to date the EVT literature describes a *univariate* theory. While we could simply treat the extreme value problem directly at the portfolio level, the top-down approach, it is much more useful to take the bottom-up view, where we explicitly model the joint stochastic structure of the portfolio elements.

Current risk management is based on confidence intervals rather than tail probabilities. Clearly, the two are dual, but the former is easier to estimate than the latter. Perhaps unsurprisingly, risk managers at the leading financial institutions are often concerned about tail probabilities, in particular the severity and frequency of really big hits. Meanwhile, note that the regulatory agencies, the Fed and the BIS, for instance, assign capital, the currency in which punishment and reward are denominated, on the basis of confidence interval performance. Successful characterization of the tail would therefore enliven considerably the current regulatory discussion.

ENDNOTES

The authors thank Casper deVries for thoughtful comments. Diebold thanks the Oliver Wyman Institute for its hospitality and inspiration. All errors, however, are ours alone. We gratefully acknowledge support from the National Science Foundation and the University of Pennsylvania Research Foundation.

This article was previously published in *Decision Technologies for Computational Finance*, A.-P.N. Refenes, A.N. Burgess, and J.E. Moody, eds., Kluwer Academic Publishers, Boston, 1998.

[1]This includes, for example, the fitting of stable distributions, as in McCulloch [1996].

[2]See Silverman [1986].

[3]The survival function is simply 1 minus the cumulative density function, $1 - F(x)$. Note, in particular, that because $F(x)$ approaches 1 as x grows, the survival function approaches 0.

[4]Embrechts, Klüppelberg, and Mikosch [1997], for example, provide both a masterful summary of the literature and an optimistic assessment of its potential. See also the papers introduced by Paul-Choudhury [1998].

[5]Other estimators are available, but none are demonstrably superior, and certainly none are as popular. Hence we focus on the Hill estimator throughout this article.

[6]See, for example, the surveys by Bollerslev, Chou, and Kroner [1992] and Diebold and Lopez [1995].

[7]Here X_t is centered around its conditional mean and scaled by its conditional standard deviation.

[8]Einmahl [1990], for example, stresses the success of the empirical survival function for analyzing all but the most extreme risks.

[9]See Danielsson and deVries [1997c] for just such a blend of rigorous EVT and intuitive graphics. Note that the conditional heteroscedasticity diminishes as the periodicity increases from daily to weekly and higher periodicity (lower frequency). Therefore, many of the EVT tools may in fact be applicable to lower-frequency financial data.

REFERENCES

Bollerslev, T., R.Y. Chou, and K.F. Kroner. "ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence." *Journal of Econometrics*, 52 (1992), pp. 5-59.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. *The Economics of Financial Markets*. Princeton: Princeton University Press, 1997.

Danielsson, J., and C.G. deVries. "Beyond the Sample: Extreme Quantile and Probability Estimation." Manuscript, Tinbergen Institute Rotterdam, 1997a.

——. "Tail Index and Quantile Estimation with Very High Frequency Data." *Journal of Empirical Finance*, 4 (1997b), pp. 241-257.

——. "Extreme Returns, Tail Estimation, and Value-at-Risk." Manuscript, Tinbergen Institute Rotterdam, 1997c.

Diebold, F.X., and J. Lopez. "Modeling Volatility Dynamics." In Kevin Hoover, ed., *Macroeconometrics: Developments, Tensions and Prospects*. Boston: Kluwer Academic Press, 1995, pp. 427-472.

Diebold, F.X., and A. Senhadji. "The Uncertain Unit Root in U.S. GNP: Comment." *American Economic Review*, 86 (1996), pp. 1291-1298.

Einmahl, J.H.J. "The Empirical Distribution Function as a Tail Estimator," *Statistica Neerlandica*, 44 (1990), pp. 79-82.

Embrechts, P., C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events*. New York: Springer-Verlag, 1997.

Engle, R.F., and G. González-Rivera. "Semiparametric ARCH Models." *Journal of Business and Economic Statistics*, 9 (1991), pp. 345-359.

Hill, B.M. "A Simple General Approach to Inference About the Tail of a Distribution." *Annals of Statistics*, 3 (1975), pp. 1163-1174.

Inoue, A. "A Conditional Goodness -of-Fit Test for Time Series." Manuscript, Department of Economics, University of Pennsylvania, 1998.

Kearns, P., and A.R. Pagan. "Estimating the Tail Density Index for Financial Time Series." *Review of Economics and Statistics*, 79 (1997), pp. 171-175.

Koedijk, K.G., M.A. Schafgans, and C.G. deVries. "The Tail Index of Exchange Rate Returns." *Journal of International Economics*, 29 (1990), pp. 93-108.

Leadbetter, M.R., G. Lindgren, and H. Rootzen. *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag, 1983.

McCulloch, J.H. "Financial Applications of Stable Distributions." *Handbook of Statistics*, Vol. 14 (1996), pp. 393-425.

Paul-Choudhury, S. "Beyond Basle." *Risk*, 11 (1998), p. 89. (Introduction to a symposium on new methods of assessing capital adequacy, *Risk*, 11, pp. 90-107).

Silverman, B.W. *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall, 1986.