Douglas E. Johnston and Petar M. Djurić

The Science Behind Risk Management

A signal processing perspective

Risk management (RM) has a long and storied history in both engineering and finance. As far back as 1800 BC, inscribed in the Code of Hammurabi in ancient Babylon, there is evidence that insurance premiums were paid by farmers to cover the risk of a crop failure [7]. This was essentially an insurance policy, or a way to manage risk, which became a growth industry in Europe during the 1600s with the advent of global trade and the need to mitigate shipping risks [36].

In engineering applications, RM became a serious study with man's desire to control nature most notably in the building of dams. Because of the rare nature of potentially catastrophic events, engineers soon realized that the science of hydrology was different from other endeavors. Harold Edwin Hurst, a British hydrologist who studied the flood cycles of the Nile, created one of the first serious investigations of extremal events and long-range dependence [15]. More recently, the advent of quality control in the automotive industry was a direct result of the realization of the risk of bringing a malfunctioning product to market. This suggests a loose, albeit broad, definition of RM as the study and mitigation of rare events that have potentially devastating outcomes.

© iSTOCKPHOTO.COM/NIKADA

With that loose definition, we see that RM permeates our lives. Do I try to make it through the yellow light? What happens to my project if I have a cost overrun? How much insurance should I buy? These are all examples of how RM decisions are made almost continuously across a broad spectrum. As such, it is no wonder it has become a serious study in many areas and nowhere more noticeable than in the financial arena of insurance and banking. While finance will be our focus in this article, we believe RM and its close cousin, reliability theory, will enjoy continued focus in more traditional engineering applications, with signal processing (SP) playing a prominent role, as the need to formulate robust solutions to real-world problems arise. For example, in defense applications, one can see a need for improved methodologies that focus both on point estimation, or maximal probability events, as well as the occurrence, prediction, and cost of outliers. Whether we address events like the Gulf oil spill, global warming, or terrorism acts, it is clear that we need a better and

Digital Object Identifier 10.1109/MSP.2011.941549
Date of publication: 22 August 2011

 1053-5888/11/$26.00©2011IEEE

more scientific approach to understanding and controlling risk in many important areas.

> RISK MANAGEMENT HAS A LONG AND STORIED HISTORY IN BOTH ENGINEERING AND FINANCE.

While RM has many fields of application, probably none has been more storied and controversial than in the field of finance. And certainly, with the near-death experience of the banking sector in 2008, more focus has been given to improvement in methodologies. Prior to the recent financial crisis, RM was often viewed as a necessary evil to satisfy regulatory requirements and the "number crunchers" were often viewed as impediments to profitability. This admitted stereotype is probably more apt for the trading operation of a financial institution than in the insurance industry. The latter has a much higher regard for the study of risk since extreme events are part of the natural course of business [9]. It must be realized that RM in finance has unique features and difficulties compared to standard engineering fields. While SP engineers will find comfort in many of the familiar approaches and methodologies employed in this article, there are daunting problems that lurk in their application to finance. In this article, we present an overview of RM science and introduce current problems that are most amenable to the expertise of the SP community.

INTRODUCTION

In finance, RM can be defined as the study and mitigation of (financial) losses with emphasis on the extremes that can occur with particular attention given to (financial) ruin. In Figure 1, we show an idealized loss distribution function where positive abscissas are financial losses. This is simply a probability distribution function (pdf) that we all first encountered in elementary probability courses. SP engineers may think of potential losses as predictions and the loss distribution as the predictive distribution used to derive statements of the future.

The derivation of the loss pdf for the future is, of course, a critical component of RM as is the derivation of predictive pdfs in many SP problems. But in contrast to focusing on the mean, the median, or maximal value, our focus is on the so-called tail of the distribution, which is shaded in our chart and equivalent to an upper quantile. Of highest interest here is the size of the tail, its shape, and its various conditional moments upon exceeding a threshold. This makes the problem most fascinating in that, by definition, time-series models must be built upon a very limited set of data and, therefore, robust procedures must be utilized.

RM in finance can essentially be boiled down into two problems. First, there is the statistical study and theory of extreme events [27]. For example, we are interested in the limiting properties of a sequence of maxima of random variables compared to their sum, which we have all encountered in the central limit theorem. This is also where we enter the realm of nonnormal pdfs and, in particular, the study of leptokurtotic, or "fat-tailed," pdfs where we might not have the luxury of finite moments. The statistical study of extreme events is not limited to finance and, for example, it has been playing a vital role in research of telecommunication traffic where packet size distributions share similar properties to financial data [20], [37]. In related work it has been employed for investigating impulsive interference [34], and further in SP, for obtaining accurate performance measures of track-before-detect algorithms [17]. In medical screening, it has been used for detecting "abnormal" events [29], in mechanical engineering for identifying anomalous episodes in gas-turbine vibration data [4], and in oceanography for studying sea-floor data acquired by multibeam sonar systems [10].

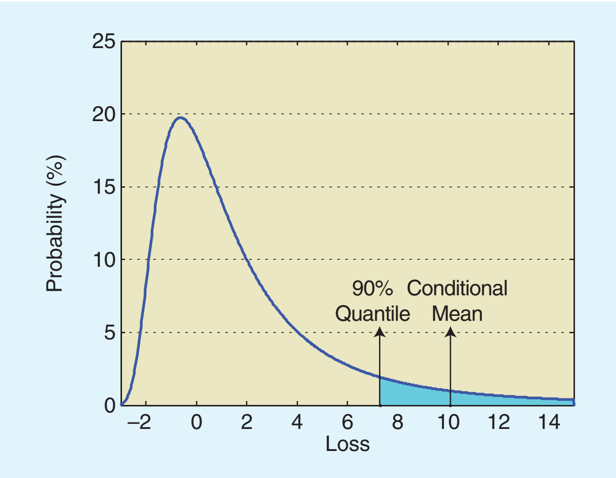
The second problem is that of modeling and predicting the future distribution of losses. We want to emphasize the second part of this problem. It is vital that the derived loss-distribution function not only reflects conditional uncertainty under the various model assumptions but also its unconditional nature. This naturally lends the problem to the Bayesian approach, which, to date, research has not particularly focused on. The other emphasis is that we are specifically interested in the far right tail of the distribution (i.e., extreme losses or errors) and would willingly sacrifice predictive accuracy elsewhere to get a better estimate of the probabilities of rare events.

For specialized terms with their meanings, see "Glossary of Terms." The words in the glossary are italicized in the text upon their first mention.

RISK MANAGEMENT TERMINOLOGY

A fundamental concept in finance is the price (or value) function that, in its simplest form, is the value of a financial claim. It could be the price of a stock, which is a claim on a company's excess cash flows, or it could be the value of a complex *portfolio* of risky *assets* [24]. This price function, denoted by $P(t)$ at time t, is a random variable (RV) defined on a suitable probability space, and we define the dollar loss ($-loss) over some time period τ as

$$L(t, t+\tau) = -(P(t+\tau) - P(t)). \qquad (1)$$



[FIG1] An example of a loss pdf.

Generally speaking, $P(t)$ is known at time t and $P(t+\tau)$ is the RV of interest. An SP analogy is that $P(t)$ is an observation or measurement, and the loss is the predicted change of the observation after a period of τ. Typically, we do not work with prices directly but rather the natural logarithm of the price which, as we will see shortly, implies the loss function is the negative of the return (percentage price change). In either case, the value function $P(t)$ is modeled by a risk-mapping function, $f(\cdot,\cdot)$, of time t and a random vector $\boldsymbol{x}(t) = [X_1(t), X_2(t), \ldots, X_N(t)]^\top$ whose N elements are called *risk factors*. As engineers we think of $\boldsymbol{x}(t)$ as the state vector of the system, and $f(\cdot,\cdot)$ as the system function. The representation, $P(t) = f(t, \boldsymbol{x}(t))$, may be exact, as we will see in a simple formulation shortly, or it may be unknown, which opens it up to system identification techniques and modeling with innovations. In general, the risk mapping is based on a (possibly empirical) relationship between the log-price function and more readily observable market variables.

IT MUST BE REALIZED THAT RM IN FINANCE HAS UNIQUE FEATURES AND DIFFICULTIES COMPARED TO STANDARD ENGINEERING FIELDS.

We rewrite our loss function in (1) in terms of the *risk factors* to get

$$L(t, t+\tau) = -(f(t+\tau, \boldsymbol{x}(t+\tau)) - f(t, \boldsymbol{x}(t))). \tag{2}$$

Since the risk factors, or states of the system, are often known at time t, it is the distribution of risk factor changes, $\boldsymbol{z}(t+\tau) = \boldsymbol{x}(t+\tau) - \boldsymbol{x}(t)$, that is the object of statistical interest.

Modeling the risk-factor changes to potentially improve on unconditional estimates of future losses is a critical design issue in RM, as it is in many SP applications, and the many issues confronting SP engineers are similar for risk managers. Should we consider a conditional or unconditional loss distribution? The former implies a deeper model selection problem versus just pure innovations. How should we design a robust system to handle, or mitigate, large errors? There is also the choice of horizon or time period for estimation. Are we interested in one-day losses or one-year losses? They can be dramatically different depending on the underlying process. In this article, we consider the horizon to be one day for ease of exposition (thus, from here on, $\tau = 1$). Finally, and possibly unique to financial RM, we are interested in the distribution of potential *drawdown*, which includes cumulative loss and time to recovery [25]. Cumulative loss is the amount of total loss experienced over multiple days, and the time to recovery is how long it takes to recover from such loss. We will not discuss drawdown in detail here but suffice it to say it is a more complex topic [35].

As a concrete example of a risk mapping, consider the somewhat circular risk factor $X(t) = \log P(t)$, the natural log of the *security's* price. This is the most granular of representations as each security is modeled directly and independently. Clearly, $P(t) = e^{X(t)}$, and we can write the risk-factor changes as

$$\begin{aligned} Z(t+1) &= X(t+1) - X(t) \\ &= \log(P(t+1)) - \log(P(t)) \\ &\approx P(t+1)/P(t) - 1 \\ &= r(t+1), \end{aligned}$$

where $r(t+1)$ is referred to as the price return at time $t+1$. It turns out that in finance, price returns are more natural time series for investigation. So in this example, we are interested in the statistical properties of security returns and, if we can infer the distribution of future security returns, the \$-loss distribution is easily obtained. As such, most financial time-series models focus on security returns and, in the sequel, we will assume log prices are modeled and losses are interpreted as negative returns.

GLOSSARY OF TERMS

Asset
An item of economic value owned or controlled by an individual or corporation.

Derivative
A security whose price is dependent upon or derived from one or more underlying assets like stocks, bonds, commodities, currencies, and interest rates.

Drawdown
A measure of decline of a variable from a historical peak.

Option
A contract sold by one party (option writer) to another party (option holder) offering the buyer the right, but not the obligation, to buy (call option) or sell (put option) a financial asset at a specified price (the strike price) during a stated period of time or on a specific date (exercise date).

Portfolio
A collection of investments owned by an individual or organization. A portfolio can include stocks (investments in businesses), bonds (investments in debt), and mutual funds (professionally managed pools of money from investors).

Position
The amount of a security either owned (a long position) or owed (a short position) by an investor. It is a trade that the investor holds open.

Price/Earning Ratio
A measure of the price paid for a share relative to the annual net income or profit earned by the firm per share.

Risk Factor
A variable whose change can affect the value of an asset.

Security
An instrument that has a financial value, for example, a banknote, a bond, a stock, or a derivative contract.

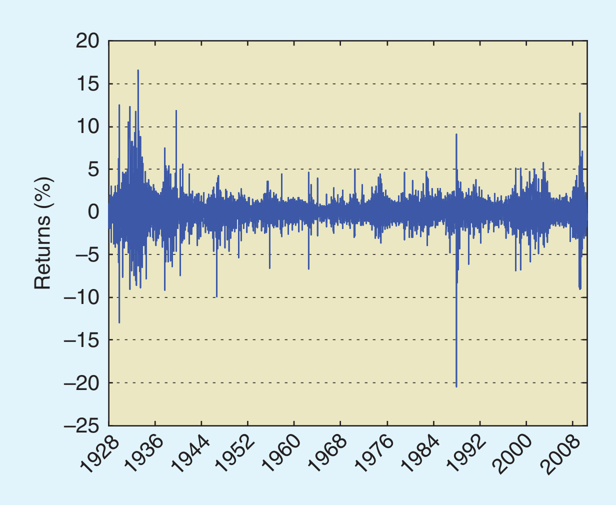
STYLIZED FACTS OF FINANCIAL TIME SERIES

Before we delve into the details of RM, it is useful to present some stylized facts about the financial time series of returns and, in general, most risk factors encountered. We will illustrate these using the daily returns of the popular Standard & Poor's (S&P) 500 stock market index from January 1928 through the end of 2009. This gives us not only plenty of observations (over 20,000), but also we have various events such as the stock market crashes of 1929 and 1987 as well as the more recent dot-com crash and the large decline post the subprime mortgage crisis. In addition, S&P 500 stock market returns are often used as a risk factor in many RM systems as a proxy for the overall market. The time series of returns is shown in Figure 2, where one can clearly see the market crash of 1987, which looks more like a bad data point than the actual 20%-plus decline witnessed on so-called Black Monday. Given that the empirical, or historical, (unconditional) standard deviation is 1.2%, one can see that really extreme events can and do occur quite often. Just the fact that the previous Black Monday in 1929 occurred a mere 58 years prior with a 13% decline should quell any desire to model the unconditional returns using a normal distribution.

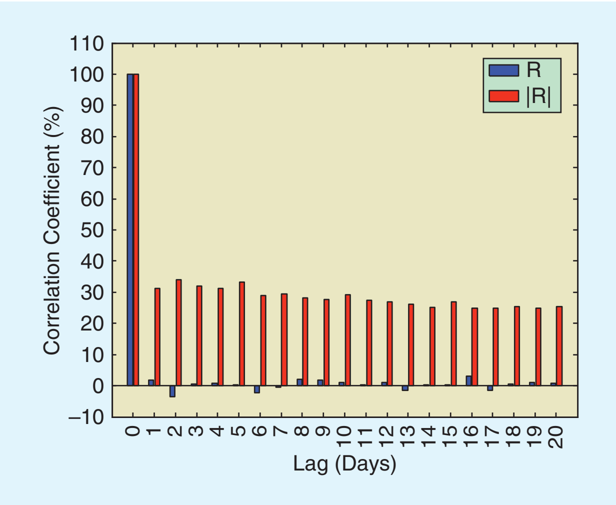
The first stylized fact is that returns show little serial correlation but absolute or squared returns show significant correlation. This is seen in Figure 3, which shows the correlation coefficient as a function of lag for the raw and absolute returns out to a lag of 20 days or about one month. We avoid analyzing squared returns since financial time series tend to be modeled by processes with potentially infinite higher moments. While the serial correlations are insignificant, statistical tests on the absolute return correlations, such as the Ljung-Box test, easily reject the strict white noise assumption. This is a bit of a shock since much of the foundation of modern finance is based on modeling the return process as Brownian motion, or a Weiner process. For example, the assumption of Brownian motion underlies the celebrated Black-Scholes *option* pricing theory [16]. So, from an SP perspective, we are immediately presented with the challenge of designing an appropriate model, as unconditional approaches will have an inherent misspecification.

The presence of serial dependence in returns makes one question whether there is predictability in the time series. But our second stylized fact, that conditional expected returns are close to zero, puts a lid on that. Simply stated, and as most seasoned market participants know, it is difficult if not impossible to predict the market in the long run based purely on historical prices. A result of the first two observations—that returns are serially dependent but lack predictability—can be paraphrased with another stylized fact that volatility, or the standard deviation of returns, varies over time and appears to have a predictable component. Thus, models that incorporate volatility dynamics enjoy success in analyzing financial time series.

The next group of stylized facts have to do with the extreme values typically observed starting with the observation that returns are leptokurtic with heavy or fat tails particularly to the left (to the right for the loss distribution). This is



[FIG2] S&P 500 stock market returns: From 1 January 1928 to 31 December 2009 (data source: Bloomberg, L.P., and S&P indices).



[FIG3] Correlation coefficient versus lag for S&P 500 daily returns, raw and absolute values.

observed for the S&P 500 data set in Figure 4, which is a QQ plot of the return data versus a standard normal reference. Recall that a QQ plot is a plot of ordered statistics of empirical data against a reference distribution. We only show the losses, or negative returns, which are positive numbers in the chart. If the data were truly from a normal distribution then one would expect a linear relationship, which is clearly rejected. Any number of tests of normality, such as the Jarque-Bera test, can be run on the stock market data with all of them being easily rejected [32]. The exponential shape of the curve highlights the extreme outliers observed in the data and suggests that the conditional mean, upon exceeding a threshold, grows with the threshold. More colloquially, conditioned on observing a "bad" return, the chances of observing a "really bad" return increases with the level of "badness." This is a critical feature of financial time series and one that will drive the distributional assumptions used for RM [13]. In particular, we will find that distributions whose tails decay according to a power law do a much better job of modeling financial data, which is why we use many results from extreme value theory (EVT) in modeling tail probabilities [11].

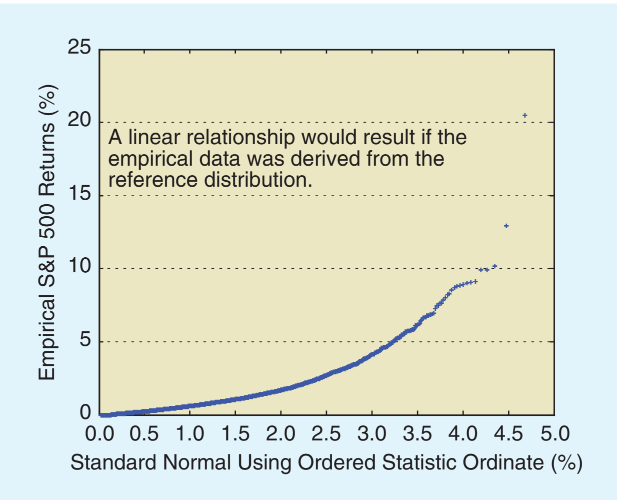
THE PRESENCE OF SERIAL DEPENDENCE IN RETURNS MAKES ONE QUESTION WHETHER THERE IS PREDICTABILITY IN THE TIME SERIES.

Another stylized fact of financial time series is that extreme values tend to be clustered in time [33]. To some degree, this is a reflection of the serial dependence and dynamic volatility of the data. If the data were truly independent identically distributed (i.i.d.), theory suggests the largest values will tend to occur as in a homogeneous Poisson process with i.i.d. exponential interarrival times. It turns out that the Poisson assumption is reasonable but the arrival rate, or intensity, is not constant, which results in too many short interarrival times. Intuitively, when there is a large shock to the market leading to liquidations and a reduction in risk, the ability, or desire, of the market to absorb subsequent shocks is diminished. In other words, the market gets skittish and volatility tends to persist.

Finally, a multivariate stylized fact is that correlations between series of financial data vary over time and exhibit tail dependency. The latter is a formal way of stating the old market adage that "everything is correlated when the market takes a dive." Two time series that appear to be locally uncorrelated for modest market moves can be highly correlated in market stress. This was witnessed most recently in late 2008 to early 2009 when all stocks were seemingly phase locked.

THE RISK MANAGEMENT PROBLEM

There are many types of risk but, in our discussion, it is the quantitative management of financial market risk that is of



[FIG4] Quantile-quantile (QQ) plot of the negative S&P 500 returns against the standard normal reference distribution.

interest [26]. To begin with, the simplest and still often quoted quantification is called gross notional risk, which is the sum of the absolute dollar amounts of the portfolio *positions*. The benefit of this approach is its simplicity and the difficulty in tweaking its value based on a model parameter. There are problems, however. First, there is no standardization of risk before aggregation. For example, consider US$1 worth of Microsoft (MSFT) stock and US$1 of U.S. Treasury bills. Is it fair to add these together and say I have US$2 worth of notional risk? Clearly, they do not bear the same risk with the former being inherently riskier. So the question of how to put securities on equal footing comes into play. The other main issue with this approach is that correlations are not accounted for and, in particular, negatively correlated securities are not distinguished. For example, if I own US$1 of MSFT and also own a contract to sell it tomorrow (a forward sale), then I have a gross notional exposure of US$2 even though the position is virtually riskless. On the other hand, correlations are also the reason risk managers need to examine (gross) notional exposures as they are not subject to estimation error and so-called model risk. Finally, there is difficulty determining the notional amount of some *derivative* contracts, such as options, where the notional exposure can easily change or is ill defined.

The second approach to quantifying market risk is through the factor sensitivities of a portfolio. Typically, this is reported as the gradient of the risk-mapping function with respect to the risk factors, where the gradient is given by

$$\nabla f(t, \boldsymbol{x}(t)) = [df/dt, df/dX_1, \cdots, df/dX_N]^\top.$$

As mentioned before, in some cases the risk mapping is known and its linearization can be analytically obtained or numerically computed. In many cases, due to the nonlinearity of the risk mapping, the second derivative or Hessian matrix also needs examination. The issue with this approach to RM is that it does not tell us the amount of risk, in terms of potential loss, unless we specify assumptions on the risk factor changes. In other words, and as we discuss shortly, unless we specify the distributional assumptions for risk factor changes, we cannot aggregate the sensitivities into a single metric. Nonetheless, the factor sensitivities are an important part of a risk manager's view into a portfolio, and therefore, limits for each sensitivity are often imposed on a portfolio.

We often model the risk mapping as a linear system and the factor sensitivities, or system coefficients, are estimated and predicted from empirical data. An example would be modeling the sensitivity of a stock's return to an economic risk factor, such as unemployment, and a stock specific factor, like its *price/earning* (P/E) *ratio*. These system coefficients would need to be estimated using historical data along with a potential time-series model for their evolution. In addition, there are cases where the state variable, or risk factor itself, must be estimated. For example, a mortgage-backed security's value is a function of future prepayments and this state variable is unobservable. Further, in a multivariate setting, we may consider unobservable, or latent, risk factors using techniques such as principle component analysis.

THE FIRST STYLIZED FACT IS THAT RETURNS SHOW LITTLE SERIAL CORRELATION BUT ABSOLUTE OR SQUARED RETURNS SHOW SIGNIFICANT CORRELATION.

In a general multivariate setting, we write the log-price function as

$$\boldsymbol{p}(t) = \boldsymbol{f}(t, \boldsymbol{x}_0(t)) + \boldsymbol{F}_1(t)\boldsymbol{x}_1(t) + \boldsymbol{F}_2(t)\boldsymbol{x}_2(t),$$

where $\boldsymbol{p}(t)$ is now a $K \times 1$ vector of log prices and, once again, losses are negative returns. We have altered our original risk-mapping definition to include a vector-valued known function, $\boldsymbol{f}(\cdot\,,\,\cdot)$, with $N \times 1$ risk factor input $\boldsymbol{x}_0(t)$, as well as two additional terms. The first of them includes a $K \times M$ matrix, $F_1(t)$, whose columns would typically include characteristics of the underlying securities that are observable at time t. This known factor loading matrix is post-multiplied by an $M \times 1$ vector of risk factors, $\boldsymbol{x}_1(t)$, whose dynamics would typically be modeled and estimated from historical data. The second additional term is similar except that the $K \times J$ factor loading matrix, $\boldsymbol{F}_2(t)$, is unknown and needs to be jointly estimated along with the $J \times 1$ risk factor vector, $x_2(t)$. Note that a pure noise term (e.g., Weiner process) is easily accommodated as part of one of the risk factors.

We have suppressed any model parameters in the equation to keep things simple but they would, in general, need to be included. We also did not show that the risk-mapping function and loading matrices may depend on past prices. If the function $\boldsymbol{f}(\cdot\,,\,\cdot)$ can be suitably linearized and the factor loading matrices are slowly time varying, we can write the vector loss function in a compact form as

$$\boldsymbol{L}(t, t+1) \approx -\boldsymbol{F}(t)\boldsymbol{z}(t+1) \tag{3}$$

with the risk factor changes, $\boldsymbol{z}(t+1)$, as defined previously and where the risk-mapping matrix $\boldsymbol{F}(t)$ is given by

$$\boldsymbol{F}(t) = [\nabla \boldsymbol{f}(t, \boldsymbol{x}(t)) \quad \boldsymbol{F}_1(t) \quad \boldsymbol{F}_2(t)]. \tag{4}$$

Thus, similar to the previous section, the focus will be on the risk-factor changes $z(t)$ in a multivariate setting. The (i, j)th element of the loading matrix $\boldsymbol{F}(t)$ is the sensitivity of the ith security's return to the jth risk factor. As a simple example, if the only risk-factors are $x_1(t)$ being the log price of the S&P 500 index and $\boldsymbol{x}_2(t)$ as a $K \times 1$ vector $(J = K)$ of uncorrelated error terms with $\boldsymbol{F}_2(t) = \mathrm{Diag}(\sigma_1, \ldots, \sigma_K)$ then (3) is an expression of the classic single-factor model for stock returns with the $K \times 1$ vector $\boldsymbol{F}_1(t)$ being the so-called stock betas.

The true crux of RM is in the development of risk measures from the loss distribution [5]. While it is inherently risky to summarize a distribution by a few variables, except under stylized assumptions, we are not left with much choice. A risk manager needs to communicate his or her findings to a wide audience using simple, concise, and consistent language. Examples of such measures include value-at-risk (VaR) and expected shortfall (ES), which we flesh out in the next section.

One approach is to assume the risk factor changes, $\boldsymbol{z}(t)$, are represented by a random vector with known distribution, which can be unconditional or conditioned on a model assumption (e.g., a vector AR(1) model). The classic example would be a normal assumption, $\boldsymbol{z}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t))$ with the risk mapping approximated by a linear function so that, for a known vector $\boldsymbol{a}(t)$ and a matrix $\boldsymbol{B}(t)$ at time t, we have

$$\boldsymbol{L}(t+1) \approx -(\boldsymbol{a}(t) + \boldsymbol{B}(t)\boldsymbol{z}(t+1)).$$

Then, for the loss $\boldsymbol{L}(t+1)$, we can write

$$\boldsymbol{L}(t+1) \sim \mathcal{N}(-\boldsymbol{a}(t) - \boldsymbol{B}(t)\boldsymbol{\mu}(t), \boldsymbol{B}(t)\boldsymbol{\Sigma}(t)\boldsymbol{B}(t)^\top).$$

In the case where the risk-factor changes are unconditionally normal with zero mean and constant covariance, $\boldsymbol{z}(t+1) = \boldsymbol{C}\,\boldsymbol{w}(t+1)$ with $\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}^\top$ and $\boldsymbol{w}(t+1) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, then $\boldsymbol{a}(t) = \boldsymbol{0}$ and $\boldsymbol{B}(t)$ is the risk-loading matrix used in (3) and defined by (4).

In the conditional setting, $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$ are the conditional mean and covariance, and $\boldsymbol{a}(t)$ and $\boldsymbol{B}(t)$ will depend on the model. Of course, we still have the problem of estimating the parameters of the distribution ($\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$ in this case), as well as the model, and usually maximum likelihood (ML) estimates are employed. A popular conditional estimate for $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$ is to use exponentially weighted moving averages. Given the inherent uncertainty associated with such estimates, however, a Bayesian approach, much like what has been widely used in SP [30], would seem more suitable.

This method, termed the variance-covariance method when a normal distribution is assumed, has some drawbacks. First, linearization of the risk mapping may not be a good approximation. This is particularly true for derivative portfolios that include options with possibly significant nonlinearities. If there are nonlinearities in the risk-mapping function, the linear approximation and resulting risk assessment will be invalid for large moves in the market. The other reservation with this methodology, as we have already stated, is that the normal distribution is too thin tailed to be useful in defining the tail of the observed data. This can be overcome by making alternate distributional assumptions and, as we will see in the sequel, there are natural limiting distributions for extreme values that are quite useful and theoretically sound. When linearization is not valid, a Monte-Carlo approach can be used to simulate future returns based on a presumed model. In fact, this can be extremely powerful when combined with Bayesian methods and computational techniques such as particle filtering [6]. Of course, a choice of models is still required and multivariate generalized autoregressive conditional heteroscedastic models (discussed in the article by David S. Matteson and David Ruppert, also in this issue) with Student-t innovations have found purchase.

One of the issues that makes RM a difficult science is the potential for nonstationary risk factors. For example, in the simple case introduced earlier, where the risk factor is the log price, the modeling of the historical time series of returns can be problematic. Consider the stock returns of MSFT. Today MSFT is a much different company than in the late 1990s during the dot-com frenzy. Back then it was viewed as an Internet-style growth company, while today it looks more like a traditional large-cap value stock. Is it reasonable to use the historical record of MSFT returns to estimate future losses, even with a conditional model imposed? In fact, the search for stationary risk factors that can empirically explain returns is a major goal of RM. For example, we may model a stock's return on a number of company specific, time-varying variables such as its P/E ratio or dividend yield. Even so, it may not be reasonable to assume that, for example, high dividend stocks will always behave in the future like they did in the past. Clearly, in 2008 they did not.

RISK METRICS

The ultimate goal of RM is to summarize the loss distribution by a small set of metrics that can be used to constrain a portfolio, or trading unit, from taking on too much risk. An offshoot of this is to utilize the risk metrics to optimize or sculpt a portfolio to maximize potential profit subject to risk controls. As such, a risk metric should have certain desirable properties and define everyday losses—those observed say once a month—and extreme losses, which could potentially put the financial institution at risk of bankruptcy. The dilemma, of course, is that we can never be sure that the future will behave anything like the past and the act of RM itself can actually induce risk. For example, an RM system may suggest that a portfolio's exposure to a risk factor can be increased but this act, if widely accepted across the financial system, can actually increase what is now known as systemic risk. This negative feedback was apparent during 2008 as the outcome of systemic risk that built up in the financial system based on the then ex-ante observation that home prices never went down.

While the usual statistics, such as the sample moments, are important, as they are in SP, RM is more focused on the right tail of the loss distribution. The standard deviation of the distribution is widely used in traditional portfolio construction but, given that we are concerned with losses, a natural alternative is to examine the (upper) partial moments defined as

$$\mu_{k,\theta} = \int_{\theta}^{\infty} (x-\theta)^k p(x)dx,$$

where $p(x)$ is the pdf of the loss. In particular, with $\theta = \mu$, and $k = 2$, we have the semivariance that is often used in place of variance in portfolio optimizations and financial ratios. By increasing k, we get a more conservative risk metric as extreme values get more weight.

Risk metrics have been developed that provide a better picture of the right tail [12]. Probably the most widely used, for better or for worse, is VaR, which is the quantile of the loss distribution $F^{-1}(\alpha)$ with $F(\cdot)$ the cumulative distribution function (cdf) and $\alpha \in [0, 1]$ a parameter of choice. The formal definition for VaR is

$$\mathrm{VaR}_{\alpha} = \inf[l \in \mathbb{R} : F(l) \geq \alpha], \tag{5}$$

which is the generalized inverse of the cdf. If $F(\cdot)$ is continuous and strictly increasing, then VaR_{α} is the ordinary inverse of the cdf evaluated at α. An example would be a "95, one-day VaR," denoted by $\mathrm{VaR}_{\alpha,\Delta}$ with $\alpha = 95\%$ and $\Delta = 1$, which is the 95th percentile of the one-day loss distribution. Technically, we should specify any conditioning information, such as the size of the historical window used, and often we drop Δ when the period is understood. Unfortunately, many people use VaR incorrectly, statistically speaking. For example, if our $\mathrm{VaR}_{95\%,1} =$ US$1 million, one often hears "we should expect to see a loss of US$1 million or more about one day a month." In reality, the statement should be "conditioned on our model being correct, we can be 95% confident that losses will not exceed US$1 million on any given day." Semantics aside, one reason VaR is so popular is because of the ease of translation into tangible events. It has even been accepted as the standard RM metric for global banking regulators in the Basel accords [2].

The problems with VaR, however, extend beyond its mere interpretation [18], [31]. Consider the following two loss distributions. With portfolio one, I have a 50/50 chance of making or losing US$1. The 75% VaR is simply US$1 (note that losses are positive numbers and the formal VaR definition in (5) is used). Now consider portfolio two that has a 50% chance of making US$1, a 25% chance of losing US$1, and a 25% chance of losing US$1 million. The $\mathrm{VaR}_{75\%}$ is the same, namely US$1, but clearly there is a preference for portfolio one as it is certainly less risky. While this example is oversimplified, it illustrates the main issue with VaR in that it tells us nothing about what might occur upon exceedance. While it is nice to know that we have high confidence in losses being less than a certain amount, we certainly want to know what to expect if losses are larger than the VaR level. The above example, while simple, is actually quite realistic in that it portrays how a VaR-based risk system can be circumvented with nonlinear losses. Since VaR does not contain enough information about the tail of the loss distribution, strategies that potentially have extreme losses can not be easily distinguished from more benign ones.

VaR is also problematic due to its lack of coherency. The principles of a coherent risk metric were first introduced axiomatically in [1] based on economic rationale. Most importantly, a risk metric should be subadditive in that the risk metric of the sum of two loss distributions should be less than the sum of the individual risk metrics. In other words, it should encourage diversification and the merging of two portfolios into one should not create additional risk. Using a risk metric that is nonsubadditive in portfolio optimizations can lead to highly concentrated portfolios that would be deemed quite risky by standard economic arguments. To illustrate the noncoherency of VaR, consider two i.i.d. securities that have a 90% chance of making US$1 and a 10% chance of losing US$5 (a Bernoulli trial). They both have a 90% VaR of −US$1. However, the combined portfolio of 1/2 of each security leads to a higher 90% VaR of +US$2 so that the combination is deemed riskier even though there is clearly a diversification benefit. Note that the formal definition of VaR in (5)—the generalized inverse of the cdf—implies that $\mathrm{VaR}_\alpha = +\$2, \forall \alpha \in (.81, .99]$. This might seem like an unrealistic example, but it illustrates that, for a set of securities that have embedded options or skewed loss profiles, VaR can lead to concentration risk.

Another real-world problem is that one cannot sum the individual VaRs of different trading units and be certain of the resulting value bounds the combined VaR from above. Therefore, RM cannot be decentralized using the VaR risk metric creating computational and operational burdens. It can be shown that VaR is subadditive under the ideal conditions of a linear risk mapping and risk factors that have a spherical distribution (e.g., multivariate normal) but this is almost always violated in practice.

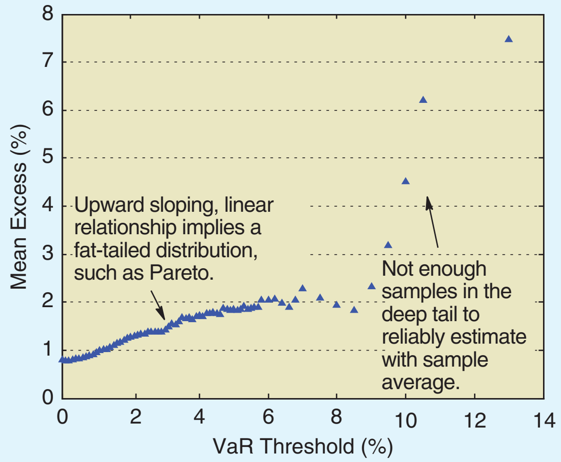
A risk metric that overcomes these difficulties, and that VaR can serve as an adjunct to, is ES and is sometimes called conditional VaR (CVaR) [31]. The ES_α is defined under suitable conditions (an integrable loss, L, and continuous loss cdf, $F(l)$, otherwise a more tedious but nonetheless applicable definition is required) as

$$\mathrm{ES}_\alpha = \mathbb{E}[L|L \geq \mathrm{VaR}_\alpha] = \frac{1}{1 - F(\mathrm{VaR}_\alpha)} \int_{\mathrm{VaR}_\alpha}^{\infty} l \, dF(l).$$

It can be readily seen that ES_α is the conditional mean of the loss distribution given the loss is greater than the α-quantile or VaR_α level. It tells us how much we can expect to lose conditioned on exceeding the VaR_α limit and is a valuable insight that looks deeper into the tail of the loss distribution.

We can estimate ES_α by taking the sample average of the $n(1-\alpha)$ upper-order statistics, from an original sample size n, which converges almost surely. Of course, the problem is that, unless we can rely on Monte Carlo simulations, the derived estimate is poor as $\alpha \to 1$. It can be shown that ES is a coherent risk metric, which is extremely beneficial. Added to the fact that is looks further into the tail, ES gives us a much better picture of risk [18].

In Figure 5, we show the related mean excess function, $e(\mathrm{VaR}_\alpha) = \mathrm{ES}_\alpha - \mathrm{VaR}_\alpha$, versus VaR_α for the S&P 500 (negated) return series, and one can see that the mean excess tends to grow as the VaR_α threshold increases, which is a prominent feature in financial time series. It can be shown that any continuous cdf, F, is uniquely determined by its ES function and for a normal distribution the graph should be downward sloping, inversely proportional to the threshold, which clearly our data reject. For the exponential distribution it is a constant, and the distribution that results in a positive linear relationship, between



[FIG5] Mean excess (conditional mean given threshold exceedance minus threshold) versus VaR (threshold) for S&P 500 negated return data.

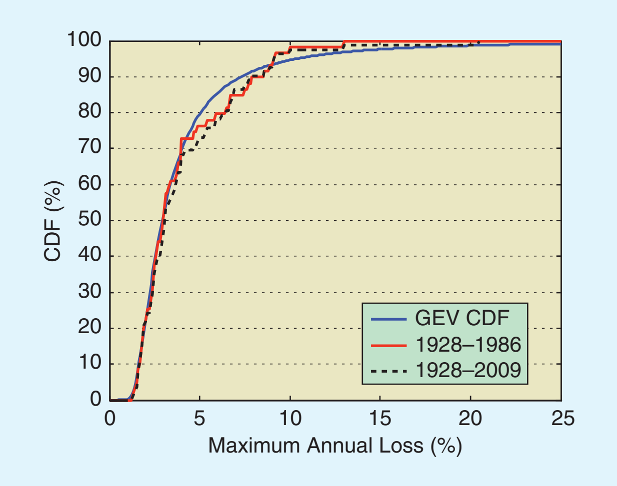
$e(\alpha)$ and VaR_α, is the Pareto distribution, which is a power-law probability distribution that coincides with many types of observable phenomena. The graph in Figure 5 shows a strong linear relationship and the outliers to the far right need to be considered in the context of their large confidence intervals given their extremely small sample size. As it turns out, as we discuss next, there is good reason to consider the Pareto distribution as a choice for estimation based on the asymptotic properties of the loss distribution.

GOING TO THE EXTREME

In RM, we are often put to the task of making probabilistic statements about extreme losses. For example, a financial institution would certainly want to know the chances of its capital being depleted over a given time period. While there is always risk of financial ruin, as in any enterprise, we want to be sure that it is within acceptable bounds and, as scientifically as possible, have a reasonable basis for truth. The extreme value theory (EVT), or the theory of extremes, is a branch of statistics that deals explicitly with this topic providing firm ground for analysis [27].

In many cases, we are interested in the distribution of the maximum of a risk factor over a prescribed time period, such as a year. Specifically, we are interested in the distributional form of a suitable normalization of the maximum, $M_N = \max[X_1, X_2, \cdots, X_N]$, of N i.i.d. RVs, referred to as the block maximum. If the block size N is large enough, the i.i.d. assumption may not be a poor approximation. That said, much of what follows applies to (strictly) stationary processes in general. It turns out that, just as there is a limiting distribution for the normalized sums of RVs, particularly the standard normal, the limiting nondegenerative distribution for the block maximum RV sequence, properly normalized, is the generalized extreme value (GEV) cdf given by

$$H_\epsilon(x) = \begin{cases} \exp(-(1+\epsilon x)^{-1/\epsilon}), & \epsilon \neq 0, \quad 1+\epsilon x > 0 \\ \exp(-e^{-x}), & \epsilon = 0, \quad x > -\infty \end{cases}.$$



[FIG6] Empirical cdf using historical annual maxima of S&P 500 return data (negated) and the GEV from ML estimates.

The variable x can be scaled and translated to obtain a three-parameter cdf, $H_{\epsilon,\mu,\beta}(x) = H_\epsilon((x-\mu)/\beta)$. This result, known as the Fisher-Tippett theorem [26], guarantees the type of limiting distribution, and the parameters $\epsilon \in \mathbb{R}$ and $\beta > 0$ are referred to as its shape and scale parameters, respectively. We should note that $\mu \in \mathbb{R}$ is not the mean of the GEV distribution but rather a location parameter. The GEV comes in three distinct flavors although its continuity with respect to ϵ allows us to reliably use the general form in statistical applications. If $\epsilon > 0$, as is typically the case for financial time series, the GEV is the Fréchet distribution. For $\epsilon = 0$, it is the Gumbel and for $\epsilon < 0$, it is the reversed Weibull distribution. The Fréchet distribution has received most of the attention because it is the limiting form of many underlying distributions used as innovations in financial time-series modeling, including the Pareto and Student-t. More generally, distributions leading to the Fréchet limit have the elegant characterization that their tails are so-called regularly varying [27], which is effectively power-law behavior. An important point to note for the Fréchet is that $\mathbb{E}(X^k) = \infty$ for $k > 1/\epsilon$ so, in particular, for the variance to exist we would need $\epsilon < 1/2$.

With this limiting distribution in hand, we can look at our stock return data and get a first glimpse at how EVT can be applied in practice. In particular, we fit the annual maximum of the (negated) daily stock returns from 1928 to 1986 to see where the extreme event of 1987 fits into a predictive distribution. We use a simple approach, using the ML estimates for the parameters as their true values to construct the predictive distribution for the 1987 maximum. Given we have only 59 annual maxima to work from, a more holistic Bayesian approach would seem better suited and is an area for future research. Our parameter estimates were $\mu = 2.49\%$, $\beta = 1.18\%$, and $\epsilon = .489$ so the derived distribution barely has a finite second moment. Shown in Figure 6 are the empirical cdfs of the data for both prior to 1987 and using the complete 82 years of data (1928–2009). We also plot the estimated cdf conditioned on the pre-1987 data using the ML estimates. One can see the fat right tail of the Fréchet distribution, which fits the historical (in-sample) data fairly well. Based on the pre-1987 data, we would have stated that the probability of the maximum daily loss in 1987 being at least 20% was 1.35% (the actual maximum for 1987 was 20.47%) or, crudely speaking, a once every 75-year type storm. While clearly a low-probability event, our simple model would not have been that surprised with the 1987 extreme market crash and certainly any veritable financial institution would want to weather such a financial storm.

While analyzing the block maximum distribution is useful in RM, it is only a glimpse into the potential risk and is quite wasteful of data. We would much prefer to make probabilistic statements about losses at a higher frequency (e.g., daily) and the block maximum approach requires a significant sample size, to take the maximum of, to apply the asymptotic results and the GEV. Fortunately, one of the major results in EVT is that the family of distributions for RVs whose maxima converge to the GEV will have a conditional excess distribution, $F_u(x) = P[X - u \le x \,|\, X > u]$, that converges to the cdf of the generalized Pareto distribution (GPD),

$$G_{\epsilon,\beta_u}(x) = 1 - \left(1 + \frac{\epsilon x}{\beta_u}\right)^{-1/\epsilon}, \ \epsilon, \ \beta_u > 0; x \ge 0.$$

The shape parameter ϵ is the same as in the limiting GEV distribution, and we have only shown the case for $\epsilon > 0$, typical of the data we investigate and the parameter β_u is, once again, a scale parameter. Typically, in applications, u is fixed and therefore the subscript on β is dropped. This result, the Pickands-Balkena-de Haan theorem [26], is extremely powerful and widely applicable, effectively stating that the Pareto distribution ($\epsilon > 0$) is the canonical form for modeling extreme losses over high thresholds for most underlying distributions of our interest. Since, by definition, we will observe few samples in the historical record above, say, the 99.99% quantile (a feasible survival test level), this result allows us to estimate such quantities with the data at hand. In practice, we choose a "high enough" threshold u to estimate the parameters of the GPD and then use its properties to compute even more extreme conditional distributions and statistics.

The mean of the GPD, assuming it is defined ($\epsilon < 1$), is $\mathbb{E}(x) = \beta/(1-\epsilon)$ and it can be shown with some basic algebra and probability manipulations that, for a higher threshold $v \ge u$, $F_v(x) = G_{\epsilon,\beta+\epsilon(v-u)}(x)$ so that the mean excess function, defined in the previous section, can be written as

$$\mathrm{e}(v) = \frac{\beta + \epsilon(v-u)}{1-\epsilon}.$$

The function $\mathrm{e}(v)$ is linear in v and this fact is often used in a visual test to admit a GPD model as well as to choose the threshold u. Recall, in Figure 5, the linear appearance of the empirical mean excess versus threshold graph that would suggest a Pareto distribution ($\epsilon > 0$) is an appropriate model for deriving a conditional distribution upon exceedance.

To compute the unconditional tail distribution, we may write for $x \ge u$

$$Q(x) = 1 - F(x) =$$
$$= \mathbb{P}(X > u)\mathbb{P}(X > x | X > u)$$
$$= Q(u)\left(1 + \frac{\epsilon(x-u)}{\beta}\right)^{-1/\epsilon}.$$

If we know $Q(u)$, or can reliably estimate it from our data (e.g., N_u/N where N_u is the number of samples that are greater than u, then we can use the above formula for the computation of various unconditional tail probabilities. The trick is that we need a high enough threshold so the asymptotic GPD is appropriate but low enough to get a good estimate of $Q(u)$. An area of future research is using robust Bayesian techniques to provide for a true predictive distribution upon marginalization over the parameter space with appropriate priors. With some minor effort, we can write

$$\mathrm{VaR}_\alpha = u + \frac{\beta}{\epsilon}\left(\left(\frac{1-\alpha}{Q(u)}\right)^{-\epsilon} - 1\right)$$

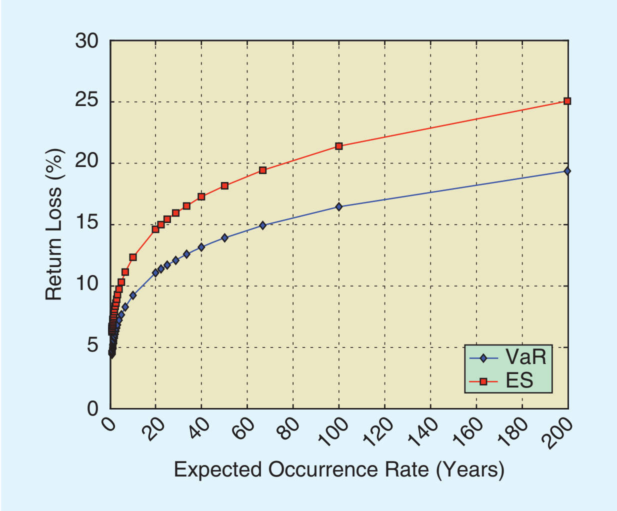
and

$$\mathrm{ES}_\alpha = \frac{\mathrm{VaR}_\alpha}{1-\epsilon} + \frac{\beta - \epsilon u}{1-\epsilon}.$$

To illustrate the use of EVT in practice, let us suppose we wish to estimate the one-day, 99.99% VaR level (a roughly one in 40 trading-year event) and the ES upon exceeding that threshold. We used the 300 or so largest values with a threshold, u, of 3% to estimate $Q(u)$ and the parameters β and ϵ of the tail distribution with the GPD assumption. Using the ML estimates as true values, we compute $\mathrm{VaR}_{99.99\%} = 13.2\%$ and $\mathrm{ES}_{99.99\%} = 17.3\%$, which can be used to ascertain adequate capital levels to manage a modestly severe financial storm.

Figure 7 illustrates how a financial institution might use the GDP extreme-value analysis. Shown on the x -axis is the expected occurrence rate that corresponds to a level of α for which VaR_α and ES_α can be computed. For each occurrence rate, we plot VaR_α and ES_α from the stock market return data. So, for example, if a financial institution desires to weather a once in a 100-year market crash, they should stress test their portfolio using a minimum 17% return loss with the expected loss, in that event, being about 22%. In fact, RM would suggest looking even deeper into the tail of the loss distribution so that a conditional survival probability at some acceptable level can be determined and capital levels maintained to survive such an event. What may be surprising to some is the frequency of occurrence of extreme events based on the empirical data with, for example, a 10% daily decline in the stock market not being that rare of an event at all.

At this point, the reader will have noted that the results presented in this section are unconditional methods for estimating tail probabilities. To deal with a more general stationary process requires more tact such as estimating the model parameters and using EVT on the residuals. More refined approaches need to be explored and an exciting area of research should include Bayesian estimation of tail probabilities, model selection, and model assessment. We have also not covered the case of multivariate risk factors where modeling the tail regions and correlations is a more complex problem and worthy of future research.



[FIG7] Estimated return losses versus expected occurrence rate for the S&P 500 daily returns using the GPD.

RISK MANAGEMENT AND SIGNAL PROCESSING

We hope our article leaves no doubt about the abundance of problems in RM where SP, and more specifically, the theory of estimation, model selection, filtering, system identification, and sequential SP can be applied. Recently, Einhorn compared VaR to "an airbag that works all the time, except when you have a car accident" [8]. This seemingly innocuous statement is actually a significant deficiency and thereby a challenge in RM. Hopefully, the SP community can help in meeting the challenge by identifying, quantifying, and correcting this deficiency.

For RM, we adopt a model and use it to estimate the probabilities of extreme outcomes and the obtained results are conditioned on the validity of the model. How do we quantify the uncertainty of the assumed model and incorporate it into our risk assessment? How should we choose among different models and monitor their performance in real time? In a different situation, we may use more than one model for computing risks and would like to fuse their results in an optimal fashion. How should this be done? All of these questions may be answered with intricate use of the theories of model selection and model assessment.

We need to strive to improve current risk models and build new ones that are more sophisticated and more robust, particularly with regard to systemic risk. Here, both parametric and nonparametric models are of interest and especially models that are based on minimal assumptions. We need to consider the long-range dependence of the data both temporally as well as across asset classes. For example, could we have used the lessons learned from the historical stock market crashes to at least have portrayed a potential housing market decline?

With the advances of SP methods, we can afford to study RM with models of high complexities that include nonlinearities and many hidden unknowns. Of particular interest are computational methods based on the Bayesian paradigm and that employ Monte Carlo sampling (including Markov chain Monte Carlo sampling, particle filtering, and population Monte Carlo sampling) and that have found extensive use in signal and image processing. We hope this article leaves the reader with no doubt that the science of SP can add much to that of RM.

ACKNOWLEDGMENTS

The work of Petar M. Djurić was supported by the National Science Foundation under Award CCF-1018323 and by the Office of Naval Research under Award N00014-09-1-1154.

AUTHORS

Douglas E. Johnston (djohnst@verizon.net) received his B.S. degree in computer science and his M.S. and Ph.D. degrees in electrical engineering from Stony Brook University. Most recently, he was at Field Street Capital Management, a hedge fund in New York. He is also an adjunct professor at Stony Brook University in both the engineering and business schools. He has worked on Wall Street since 1994 in both research and proprietary trading and, prior to that, he was a research engineer in the defense industry. He is a member of the IEEE Signal Processing Society. His research interests are in the self-destructive nature of financial time series.

Petar M. Djurić (djuric@ece.sunysb.edu) received his B.S. and M.S. degrees in electrical engineering from the University of Belgrade and his Ph.D. degree in electrical engineering from the University of Rhode Island. He is with Stony Brook University, where he is a professor in the Department of Electrical and Computer Engineering. He works in the area of statistical signal processing, and his primary interests are in the theory of signal modeling, detection, and estimation and application of the theory to a wide range of disciplines. He has been invited to lecture at universities in the United States and overseas and has served on numerous committees for the IEEE. During 2008–2009, he was Distinguished Lecturer of the IEEE Signal Processing Society. He was the area editor for special issues of *IEEE Signal Processing Magazine* and associate editor of *IEEE Transactions on Signal Processing*. He has also been on the editorial boards of many professional journals. In 2007, he received the IEEE Signal Processing Magazine Best Paper Award. He is a Fellow of the IEEE.

REFERENCES

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Math. Finance*, vol. 9, no. 3, pp. 203–228, 1999.

[2] Basel Accords. (2006) Basel II: Revised international capital framework. [Online]. Available: http://www.bis.org/publ/bcbsca.htm

[3] R. Baysal and J. Staum, "Empirical likelihood for value-at-risk and expected shortfall," *J. Risk*, vol. 11, no. 1, pp. 3–32, 2008.

[4] D. A. Clifton, L. Tarassenko, N. McGrogan, D. King, S. King, and P. Anuzis, "Bayesian extreme value statistics for novelty detection in gas-turbine engines," in *Proc. IEEE Aerospace Conf.*, 2008.

[5] G. Connor, L. Goldberg, and R. Korajczyk, *Portfolio Risk Analysis*. Princeton, NJ: Princeton Univ. Press, 2010.

[6] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

[7] N. Dunbar, *Inventing Money*. New York: Wiley, 2000.

[8] D. Einhorn and A. Brown, "Private profits and socialized risk," *Global Assoc. Risk Prof. Rev.*, ch. 1, no. 42, pp. 10–26, 2008.

[9] P. Embrechts, C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*. New York: Springer-Verlag, 2003.

[10] A. N. Gavrilov, "Fluctuations of seafloor backscatter data from multibeam sonar systems," *IEEE J. Oceanic Eng.*, vol. 35, no. 2, pp. 209–219, 2010.

[11] K. Giesecke and L. Goldberg, "Forecasting extreme financial risk," in *Risk Management: A Modern Perspective*, M. Ong, Ed. Hoboken, NJ: Wiley, 2005.

[12] K. Giesecke, T. Schmidt, and S. Weber, "Measuring the risk of large losses," *J. Investment Manage.*, vol. 6, no. 4, pp. 1–15, 2008.

[13] L. Goldberg, M. Hayes, J. Menchero, and I. Mitra, "Extreme risk analysis," *J. Perform. Manage.*, vol. 14, no. 3, pp. 17–30, 2010.

[14] F. Harmantzis and L. Miao, "Empirical study of fat-tails in maximum drawdown: The stable Paretian modeling approach," presented at the Quantitative Methods in Finance Conf., Sydney, Australia, 2005.

[15] H. Hurst, "Long-term storage capacity of reservoirs," *Trans. Amer. Soc. Civil Eng.*, vol. 116, pp. 778–808, 1951.

[16] J. E. Ingersoll, Jr., *Theory of Financial Decision Making*. Lanham, MD: Rowman & Littlefield, 1987.

[17] L. A. Johnston and V. Krishnamurthy, "Performance analysis of a dynamic programming track before detect algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 228–242, 2002.

[18] K. Kuester, S. Mittnik, and M. Paolella, "Value-at-risk prediction: A comparison of alternative strategies," *J. Financial Econometrics*, vol. 4, no. 1, pp. 53–89, 2006.

[19] M. Lefebvre, *Applied Stochastic Processes*. New York: Springer-Verlag, 2007.

[20] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[21] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[22] A. Lo, *Hedge Funds: An Analytic Perspective*. Princeton, NJ: Princeton Univ. Press, 2010.

[23] M. L. de Prado and A. Peijan, "Measuring loss potential of hedge fund strategies," *J. Alternative Investments*, vol. 7, no. 1, pp. 7–31, 2004.

[24] D. G. Luenberger, *Investment Science*. New York: Oxford Univ. Press, 1998.

[25] S. Maslov and Y.-C. Zhang, "Probability distribution of drawdowns in risk investments," *Physica A: Statist. Mech. Applicat.*, vol. 262, no. 1–2, pp. 232–241, 1999.

[26] A. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management*. Princeton, NJ: Princeton Univ. Press, 2005.

[27] S. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer-Verlag, 2007.

[28] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwood, MA: Artech House, 2004.

[29] S. J. Roberts, "Extreme value statistics for novelty detection in biomedical data processing," *IEE Proc. Sci., Meas. Technol.*, vol. 147, no. 6, pp. 363–367, 2000.

[30] J. J. K. Ó. Ruanaidh and W. J. Witzgerald, *Numerical Bayesian Methods Applied to Signal Processing*. New York: Springer-Verlag, 1996.

[31] S. Sarykalin, G. Serraino, and S. Uryasev, "Value-at-risk vs. conditional value-at-risk in risk management and optimization," in *Tutorials in Operations Research,* Z. L. Chang, S. Raghavan, and P. Gray, Eds. Hanover, MD: INFORMS, 2008, pp. 270–294.

[32] H. Thode, Jr., *Testing for Normality*. New York: Marcel Dekker, 2002.

[33] S. Thurner, J. Farmer, and J. Geanakoplos, "Leverage causes fat tails and clustered volatility," *Cowles Foundation Discussion Series*, no. 1745, Yale Univ., 2010.

[34] G. A. Tsihrintzis and C. L. Nikias, "Fast estimation of the parameters of alpha-stable impulsive interference," *IEEE Trans. Signal Processing*, vol. 44, no. 6, pp. 1492–1503, 1996.

[35] B. V. de Melo Mendes and V. R. Brandi, "Modeling drawdowns and drawups in financial markets," *J. Risk*, vol. 6, no. 3, pp. 53–69, 2004.

[36] J. de la Vega, "Confusión de confusiones," in *Extraordinary Popular Delusions and the Madness of Crowds and Confusión de Confusiones*, M. S. Fridson, Ed. New York: Wiley, 1966, pp. 125–211.

[37] W. Willinger, M. S. Taqqu, M. Leland, and D. Wilson, "Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements," *Statist. Sci.*, vol. 10, no. 1, pp. 67–85, 1995.

[SP]