FORECASTING EXTREME FINANCIAL RISK

Kay Giesecke* Lisa R. Goldberg†

Cornell University *MSCI Barra, Inc.*

November 29, 2004

Abstract

Extreme value statistics provides a practical, flexible, mathematically elegant framework in which to develop financial risk management tools that are consistent with empirical data. In this introductory survey, we discuss some of the basic tools including power law distributions, the peaks over thresholds estimation procedure and point processes.

Key words: extreme events, normal distribution, extreme value distribution, power law, Pareto distribution, peaks over thresholds, tail index, shortfall risk, Hurst exponent, clustering, contagion, point process, Poisson process, fitness testing

*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA, Phone (607) 255 9140, Fax (607) 255 9129, email: giesecke@orie.cornell.edu, web: www.orie.cornell.edu/∼giesecke.

†MSCI Barra, Inc., 2100 Milvia Street, Berkeley, CA 94704-1113, USA, Phone (510) 649 4601, Fax (510) 848 0954, email: lrg@barra.com.

1 Introduction

The turbulence that dominated financial markets throughout the late 1990's and the early part of the twenty-first century has undermined the effectiveness of standard risk management tools such as the Black-Scholes-Merton option pricing formula and value-at-risk. Because these tools are predicated on the normal distribution, they do not account for extreme events such as the 1998 ruble crisis, the bursting of the internet bubble in 2000, and the ongoing string of defaults resulting from corporate malfeasance.

Consequently, the financial industry is turning to *extreme values statistics* as the basis of a new generation of risk models that account for the turbulence in time series of asset returns. These models incorporate:

- Extreme events: The frequency with which extreme events occur in financial markets is not consistent with forecasts based on the normal distribution.
- Temporal clustering: Extreme events in financial markets are not statistically independent.

Extreme value models take a descriptive, statistical viewpoint of the turbulence observed in financial market data. A more fundamental modeling approach is based on *point processes.* It considers the drivers of turbulence in a probabilistic model. These models incorporate:

- Cyclical dependence: The dependence of counterparties on common macroeconomic factors generates clusters in returns.
- Contagion: The intricate web of relationships among governments, businesses, financial institutions and investors creates intense, unpredictable chain reactions.

Below, we give a brief introduction to some basic techniques from extreme value statistics and point process theory and discuss their application to the modeling of financial risks.

2 A brief, recent history of financial risk modeling

Financial risk is the distribution of future portfolio returns. This definition is valuable since it is both comprehensive and terse. However, it is not immediately useful to someone who wants to measure or manage financial risk. Consider that the future consists of infinitely many possible time horizons and a general distribution is described by an infinite number of quantiles. It is not clear how the distribution at one horizon relates to the distribution at another and portfolios change over time.

These difficulties were addressed in the 1950s by Harry Markowitz, who is widely viewed as the father of modern financial risk management. Markowitz fixes a rebalancing horizon and defines the risk of a portfolio as the standard deviation of its returns, which is the average dispersion of returns around their expected value. Intuitively, this is very appealing. The greater the range of past returns, the less certain we are about what to expect in the future. Markowitz's approach is very practical. It allows portfolio risk to be expressed in terms of the standard deviation of individual securities and the correlations between them. Thus, the abstract concept of diversifying a portfolio is translated into the concrete practice of lowering correlations. By reducing risk to a single number, the Markowitz approach imparts a useful rank ordering on assets and portfolios. It also leads quickly to mean-variance portfolio optimization and the capital asset pricing model, both of which implicitly rely on the normal distribution.

2.1 The normal distribution

The probability $\Phi_{\mu,\sigma}(x)$ of observing a value no greater than x for a normal variable is given by the formula

$$\Phi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy. \tag{1}$$

The probability (1) depends on two parameters: the mean $\mu \in \mathbb{R}$ and the standard deviation $\sigma > 0$.

The normal distribution, also known as the Gaussian distribution, dates back to the eighteenth century. Its role as the error distribution in linear regression was rigorously demonstrated by Carl Friedrich Gauss in 1809. The normal is the most widely used distribution in the physical and social sciences since it arises naturally as the distribution of sums. For example, if you attach a running sum to a sequence of coin flips by adding 1 for each heads and -1 for each tails, the distribution of

$$\frac{\text{sum}_N}{\sqrt{N}} \tag{2}$$

becomes normal as $N \to \infty$.

This is a special case of the *central limit theorem,* which states that as long as the sum in formula (2) is composed of independent draws from a distribution with finite standard deviation, the conclusion of asymptotic normality holds. In other words, the normal distribution arises in connection with sums over large pools of identical, independent observations of a finite standard deviation variable.[1]

It is easy to see that a normal distribution is a poor fit to financial data. Consider the 12% drop in the stock market that occurred on Black Tuesday, 29 October 1929. The normal distribution predicts a daily loss of this magnitude once in every 10^{30} years. This is certainly suspect in light of the 20% drop in the stock market that occurred on Black Monday, only 58 years later.

The frequency of extreme events in financial markets is grossly underestimated by the normal distribution. In other words, the *tails* of the normal distribution are too light relative to the data. A better statistical fit is provided by distributions that assign a higher probability to relatively large values.

Further insight into the mismatch between the normal distribution and the empirical nature of financial returns is illustrated in Figure 1, which shows daily log-returns to the Standard and Poor's 500 Index between October 1982 and November 2004. In Figure 2, we show a normal simulation with matching length, mean and standard deviation. Profound differences between the Standard and Poor's return series and normal simulation are evident at a glance.

Prescient articles by Benoit Mandelbrot written in the 1960s argue that the normal distribution is a poor tool for financial risk modeling. The claims are based both on relatively high frequency of extreme events as well as the long term statistical dependence present in financial data.

2.2 Standard financial risk management tools implicitly rely on the normal distribution

The normal distribution continues to be the basis of financial risk modeling and management. We mention three of many examples.

[1]The central limit theorem is one of the most important results in probability theory. There is an enormous literature devoted to extensions and generalizations.

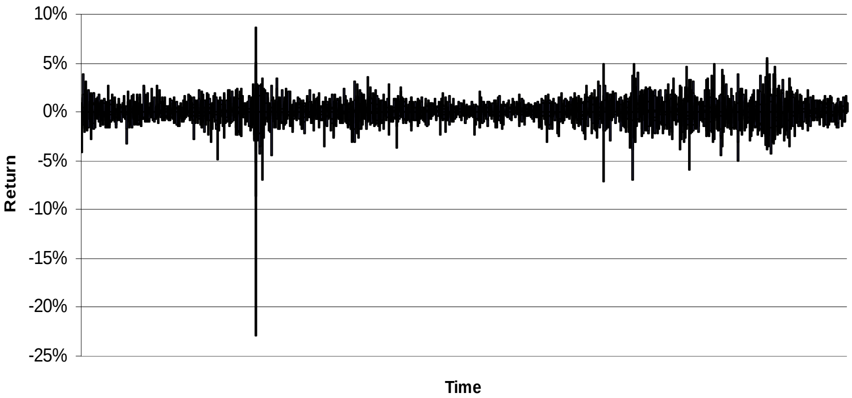

Figure 1: Daily log-returns to the Standard and Poor's 500 Index between October 1982 and November 2004.

2.2.1 The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM), developed by William Sharpe, John Lintner and others in late 1960s, is an equilibrium model of security prices. As prescribed by Markowitz, investors hold mean-variance optimal security portfolios. The CAPM uses a linear regression to decompose portfolio risk into a market component and an idiosyncratic component. The main thrust of the CAPM is that investors should be compensated only for the undiversifiable risk, which is due to the market. As remarked in Section 2.1, the regression incorporates the assumption of normally distributed errors, which represent idiosyncratic returns in the CAPM.

The main difficulty with the CAPM is that it assumes investors care only about mean and variance. It does not provide a way to compensate investors for the risk of extreme events.

2.2.2 The Black-Scholes-Merton formula

The option pricing theory developed by Fischer Black, Myron Scholes and Robert Merton in the early 1970s is one of the most widely used tools in financial management. Together with the CAPM, it is taught in business schools all over the world. The theory is based on the assumption of normally distributed asset returns that are free of autocorrelation. Therefore forecasts option prices are based on a model that underestimates the frequency of extreme returns in

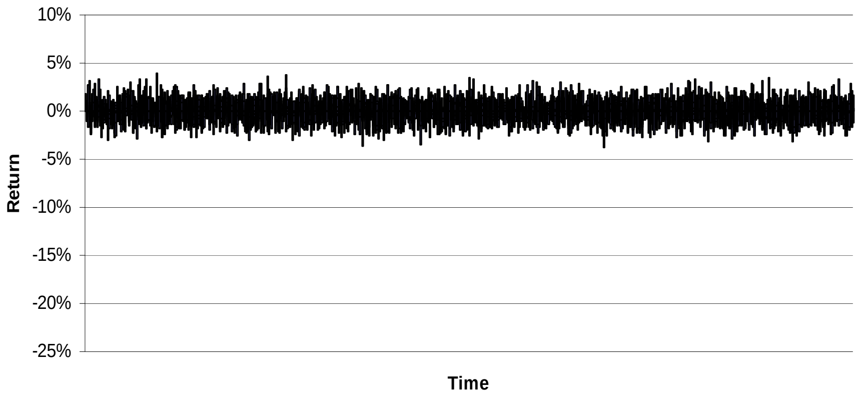

Figure 2: Simulated daily log-returns to the Standard and Poor's 500 Index between October 1982 and November 2004.

the data. Consequently, traders generally do not use the Black-Scholes-Merton to price options. Instead, they use it to convert prices quotes into quotes based on implied volatility

2.2.3 Value-at-risk

Value-at-risk is an industry standard measure of portfolio risk. It is simply a quantile of the portfolio's return distribution. Value-at-risk is quoted in terms of a fixed time horizon h and a percentage α. If the (α, h)-value-at-risk is V, then the loss over a horizon h should be less than V in $\alpha\%$ of cases. For example, if the 99%, one day value-at-risk is \$1 million, then the portfolio loss should be less than 1 million on 99% of days.

Value-at-risk is a convenient measure since it summarizes portfolio risk by a single number. Typically, value-at-risk is based on a normal approximation to the return distribution, in which case it it can be easily calculated as a multiple of the portfolio's standard deviation. However, value-at-risk estimates based on the normal distribution tend to be too low. Further, these estimates say nothing about what to expect on the bad days when value-at-risk is exceeded. We return to this point in Section 3.1.

3 Extreme Value Statistics

A change in focus from the average to the maxiumum takes us beyond the normal to distributions that are a better fit to financial data. Consider a time series (X_i) of observations and set

$$M_n = \max_{j \le n} X_j$$

If the X_i's are independent and identically distributed with distribution function F, then the distribution of M_n is the n-fold product F^n. Unfortunately, this observation is not useful on large samples since

$$\lim_{n \to \infty} F^n = 0.$$

However, the Fisher-Tippett theorem asserts that under very mild assumptions on F, a carefully chosen sequence of affine rescalings $c_n M_n + d_n$ converge in distribution to one of three extreme value distributions. This means that asymptotically, the variable $c_n M_n + d_n$ has one out of three possible distributions. The Fisher-Tippett theorem is the extreme value analog to the central limit theorem. It provides the basis for classification of distributions by the limits of their normalized maxima, which leads to the applications described below.

3.1 Expected shortfall

The (α, h)-value-at-risk V is a probability $\alpha\%$ upper bound for losses over a horizon h. On the $(100 - \alpha)\%$ of bad times when losses reach V over h, how much more do we expect to lose? The answer is *expected shortfall*, also known as *conditional value at risk*. It is defined as the average excess loss over a threshold. Expected shortfall is a function of the threshold, and the shape of this function depends on the loss distribution.

3.1.1 Expected shortfall highlights the differences among distributions

In the figures and analysis below, we reverse signs so that a loss is positive. Therefore, we focus the right tail of the distribution.

Figure 3 displays the expected shortfall for three different return distributions. If the underlying distribution is normal, expected shortfall diminishes

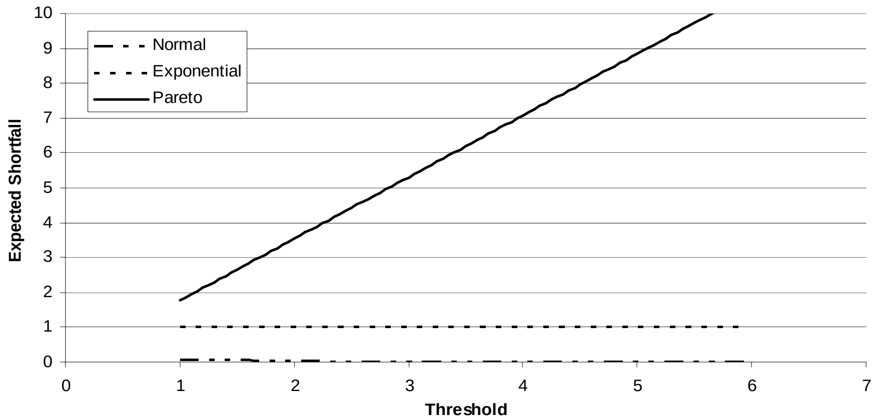

Figure 3: Theoretical expected shortfall as a function of the threshold for the Normal distribution, the exponential distribution and the Pareto distribution.

rapidly to zero as the threshold increases. In other words, once losses have hit a certain threshold, there isn't much more to be concerned about. If the underlying distribution is exponential, expected shortfall is independent of the threshold thanks to the memoryless property of the exponential distribution. Finally we consider the expected shortfall for the Pareto distribution, which is discussed in Section 3.4. The Pareto distribution arises in connection with extreme value statistics in much the same way as the normal distribution arises in connection with sums of independent random variables. Since Pareto distributions take account of extreme events, expected shortfall increases linearly with the threshold. The slope of the line is $\alpha/(\alpha - 1)$ where α is the *tail index*. Thus, a steeper slope, which comes from a greater density of extreme events, corresponds to a smaller tail index.

3.1.2 Expected shortfall is consonant with diversification

It may be tempting to combine value-at-risk measurements based on non-normal distributions with expected shortfall to manage portfolio risk. However, there is a subtle difficulty with this strategy. Value-at-risk estimates may be inconsistent with diversification: the sum of the value-at-risks of the components of a portfolio may be less than the portfolio value-at-risk. In this

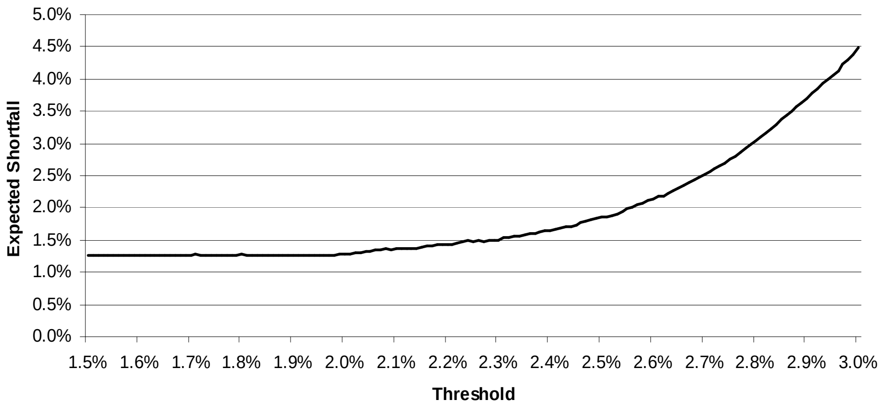

Figure 4: Empirical expected shortfall for daily returns to the S & P 500 index as a function of the threshold.

situation, diversification is penalized and concentration is encouraged.[2] Expected shortfall is always consonant with diversification. This facilitates the consistent allocation of capital to departments and trading desks.

3.1.3 Empirical expected shortfall

Given a series of observed returns r_i and a threshold u, the empirical expected shortfall function is given by

$$s(u) = \frac{1}{k} \sum_i (r_i - u)^+ \tag{3}$$

where $k = k(u)$ is the number of observations that exceed u.

In Figure 4, we show the empirical expected shortfall for daily returns to the S & P 500 index. Note the resemblance of the empirical shortfall risk functions to the Pareto shortfall risk function. This is not a coincidence.

3.2 Power laws

Vilfredo Pareto was an Italian industrialist who lived during the last half of the nineteenth century and the early part of the twentieth. He was interested

[2]Elliptical value-at-risk estimates, which include normal estimates, are consonant with diversification: an elliptical value-at-risk estimate for a portfolio is an upper bound for the sum of the values-at-risk of the individual positions.

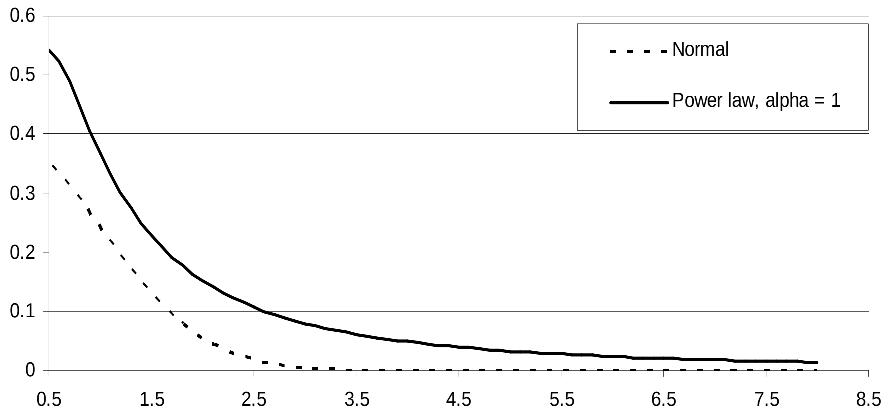

Figure 5: Densities of the Normal distribution and the Power law distribution with exponent 1.

in how wealth is distributed in a society. To study the issue, he gathered data from numerous countries and cultures and he found a remarkably consistent pattern. Wealth distribution follows what we now call a *power law.*

Definition 3.1. *A random variable X with distribution function F follows a power law with exponent α if*

$$(1 - F(x)) \sim x^{-\alpha}$$

Variables with power law distributions scale according to their exponents. For example, consider a society in which wealth follows a power law distribution with exponent 2. Then the fraction of the population that has at least 10 times your net worth is roughly $1/10^2$ times the fraction of the population that has at least your net worth. A power law distribution is easy to identify in a log-log coordinate system since it follows a straight line with slope equal to the negative of its exponent.

What is the relationship between the normal distribution and power law distributions? The latter dominate the former. In other words, the tails of the normal distribution, which measure the probability of extreme events, are much, much smaller than the tails of any power law distribution. We illustrate this in Figure 5.

3.3 Rely on quantiles, not moments

It is standard practice to describe distributions by moments. Often, the first steps of data analysis are to estimate the mean as a measure of the center of the distribution, the standard deviation as a measure of dispersion, the skew as a measure of asymmetry and kurtosis as a measure of the relative frequency of large events. However, when dealing with extreme risks, these measures can fail. By definition, the n^{th} moment of a random variable X is given by the expectation $E[X^n]$. Higher moments are infinite for many distributions that are important for describing extreme events.

Theorem 3.2. *The nth moment of a power law distribution with exponent α is infinite for $n > \alpha$.*

As we discuss below, many extreme value distributions fit to financial data have a tail index of less than 2. Thus, in the presence of extreme events, estimates of kurtosis, skewness, standard deviation and even the mean may be unreliable.

An alternative is to describe a distribution in terms of its quantiles. These exist for every distribution, and they play an important role in the peaks over thresholds estimation described in Section 3.5.

3.4 The generalized Pareto distribution

The probability $G_{\alpha,\beta}(x)$ of observing a value greater than 0 and no greater than x for a generalized Pareto variable is

$$G_{\alpha,\beta} = \begin{cases} 1 - (1 + \frac{x}{\alpha\beta})^{-\alpha} & \alpha < \infty \\ e^{-x/\beta} & \alpha = \infty \end{cases} \tag{4}$$

The probability (4) depends on two parameters: the tail index $\alpha \in (0, \infty]$ and the height $\beta > 0$.

Just as the normal distribution arises naturally in conjunction with sums, the Pareto distribution arises naturally in conjunction with excesses over a threshold. To understand this, let X be a random variable with cumulative distribution function F.

Theorem 3.3. *Under very mild hypotheses on F, there is a positive function $\beta(u)$ such that*

$$\lim_{u\to\infty} \frac{1 - F(u + \beta(u)x)}{1 - F(u)} = \left(1 + \frac{x}{\alpha}\right)^{-\alpha}$$

The function β satisfies

$$\lim_{u \to \infty} \beta(u) = \frac{1}{\alpha}$$

It arises in connection with the Karamata representation theorem, which is the mathematical basis of the classification of distributions by the limits of their normalized maxima.

Theorem 3.3 says that if the threshold u is sufficiently high and there are enough independent observations, excesses over a threshold follow a generalized Pareto distribution. This leads immediately to a procedure for estimating distributions of data that include extreme events.

3.5 Peaks over thresholds

Typically, there are abundant observations from the center of a distribution and scarce information about the tail. The semi-parametric *peaks over thresholds* method for estimating distributions exploits this situation. Empirical observations provide estimates of middle and lower quantiles while Theorem 3.3 provides estimates of upper quantiles.

As above, let X be a random variable with distribution function F. The method is predicated on the simple observation that for an u and any value $x > u$, the probability of observing a value of X no greater than x is given by

$$F(x) = F(u) + F(x \,|\, X > u)(1 - F(u)) \tag{5}$$

where $F(x \,|\, X > u)$ is the conditional probability of an observation less than x given that X is at least u.

Suppose we have n identical, independent observations X_i of a real valued variable X with unknown distribution F. A peaks over thresholds estimation of F proceeds as follows:

(1) Fix a threshold u.

(2) Let $Y_1, Y_2, \ldots, Y_k$ denote the subsequence of positive values in the sequence $(X_1 - u)^+, (X_2 - u)^+, \ldots, (X_n - u)^+$.

(3) According to Theorem 3.3, the Y_is approximately follow a generalized Pareto distribution $G_{\alpha,\beta}$. Fit the parameters α and $\beta(u)$ with a maximum likelihood estimation.

(4) The peaks over thresholds estimate of $\hat{F}(x)$ is given by

$$\hat{F}(x) = \begin{cases} \text{card}\,\{i : X_i \leq x\}/n & x \leq u \\ \hat{F}(u) + (1 - \hat{F}(u))G_{\alpha,\beta}(x-u) & x > u. \end{cases} \tag{6}$$

This method can be applied to virtually any real valued variable, no matter how it is distributed.[3] The most delicate issue is choosing the threshold u. The choice is ultimately an empirical question and, in the absence of abundant data, it raises the issue of the variance-bias tradeoff. As u increases, the bias in the estimate of high quantiles diminishes while the variance of the estimates increases.

A plot of the empirical shortfall risk formula (3) sometimes provides an indication of how to choose the threshold u. To minimize variance, u should be as small as possible. In many plots, such as the one in Figure 4, the shortfall index is roughly linear over a range of thresholds. Then, it is reasonable to select u as the smallest value in this range.

A second method for estimating u, known as the *Hill estimator* of the power exponent, can be applied if the distribution F is known to have a positive tail index. Consider the *order statistics*

$$X_{1,n} \geq X_{2,n} \geq \ldots \geq X_{n,n} \tag{7}$$

of the sample. Fix a threshold u and suppose there are $k = k(u)$ observations X_i that are greater than u. The Hill estimator $\hat{\alpha}$ of α is given by

$$\frac{1}{\hat{\alpha}} = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{X_{j,n}}{X_{k,n}}\right) \tag{8}$$

As in the case of shortfall risk, the Hill estimator is a function of the threshold u. A good choice of u is the smallest value that gives a consistent estimate of α. This is illustrated in Figure 6.

3.6 Dependence: the Hurst exponent

Most of the results presented above rely on the assumption of independence. However, this assumption is not always satisfied by financial time series. Below, we describe a well-known method for quantifying the dependence.

[3] There are exceptions, but you have to work to find them.

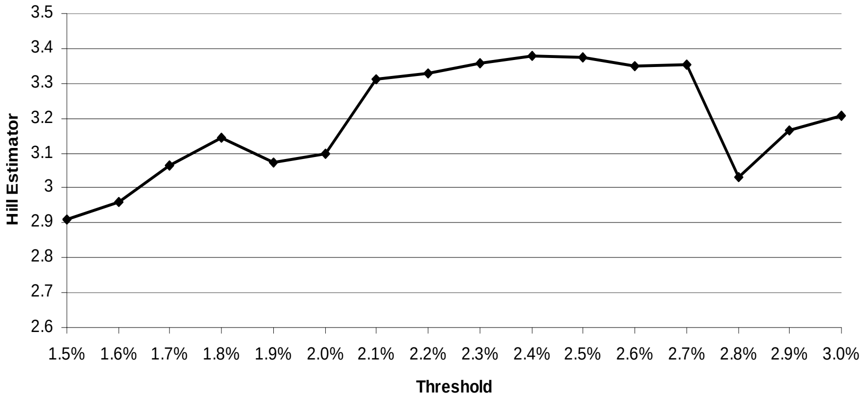

Figure 6: The Hill estimator of the power exponent for daily returns to the S & P 500 index as a function of the threshold.

Harold Edwin Hurst was a British civil servant who lived in the late nineteenth and early twentieth century. He worked on the problem of designing dams to regulate the Nile. Hurst realized that it was not enough to know averages and variation in rainfall and river height data. The dams had to be high enough to contain the result of the inevitable clusters of wet years.

To quantify the degree of clustering, Hurst looked at the length of consistent runs of wet years in his data and developed a measure of dependence that is still in common use today. Consider the running total sum_n generated from daily observations X_i of a random quantity X with finite second moment:

$$\mathrm{sum}_n = \sum_{i \leq n} X_i. \tag{9}$$

Let R_n denote the gap between the maximum $M_n = \max_{i \leq n} \mathrm{sum}_i$ and the minimum $m_n = \min_{i \leq n} \mathrm{sum}_i$. The greater the positive dependence in the X_is, the more likely are consistent runs of similar observations and the more rapidly the gap R_n is expected to grow as a function of n. Conversely, the greater the anti-dependence in the X_is, the more slowly the gap R_n is expected to grow as a function of n. To make an even handed comparison, the range R_n must be scaled by the sample standard deviation S_n of the X_is.

Definition 3.4. *The rescaled range statistic of length n is the quotient* R_n/S_n.

If the X_is are independent, the expected value $E[R_n/S_n]$ becomes asymptotically proportional to $n^{.5}$.

Definition 3.5. *Suppose that*

$$E\left[\frac{R_n}{S_n}\right] \sim cn^H. \tag{10}$$

The constant H *is the Hurst exponent.*[4]

Positive dependence in the sample corresponds to $H > .5$ and negative dependence to $H < .5$. Hurst's analysis of rainfall data led to values of H as high as .7. In financial return series, values of H typically fall between .5 and .6, indicating weak positive dependence. An estimate of the Hurst exponent for daily S & P returns in the period October 1982–November 2004 is .69.

4 Point processes

A chaotic market often appears in the wake of a crisis. For example, news of the ruble-denominated sovereign default or fictional earnings at Parmalat generated a market-wide reassessment of the creditworthiness of all issuers. In turn, this caused a dramatic, almost instantaneous flight to quality. This *market contagion* is facilitated by the complex network of relationships among firms, governments and financial institutions. When combined with a recessive economy, contagion can create a chain reaction that leads, for example, to the bursting of a market bubble.

The statistical models considered in Section 3 are descriptive in nature. They do not address the mechanism that generates clusters of extreme events and the ensuing market chaos. Below, we describe a very general class of probabilistic models that provide a natural framework in which to analyze the drivers of turbulence.

A probabilistic model of a chaotic system begins with a mathematical representation of an uncertain future event. The relevant concept is a *random time*, which is a positive random variable T on a set Ω, whose elements represent all possible states of the world. It specifies the future time of an uncertain event in different states of the world. By definition, a *point process* on the positive real numbers is an increasing sequence (T^i) of random times.

Point processes have a wide range of applications. They can be used to model defaults, large drops in the Standard and Poor's 500, heart attacks or

[4] Note the parallel between the rescaled range statistic and a Hölder continuous function f of a real variable. Recall that f is *Hölder continuous with exponent* h if $f(x)-f(y) \sim (x-y)^h$. The modulus of continuity h is analogous to the Hurst exponent.

earthquakes. The concept could hardly be simpler and it seems remarkable at first that such a simplistic construction could be useful. The power of point processes is achieved through overlays of mathematical models of probability and information.

4.1 A sketch of the mathematical basics

Implicit in a point process is a model of observable information and how it evolves over time. Mathematically, this is a *filtration*, which is a family $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$ of collections of subsets of Ω indexed by time t. Each subset of Ω is an event, and the collection $\mathcal{G}_t$ contains all the events that can be distinguished at time t, or the observable information at time t. We assume information is never lost, so $s \leq t$ implies $\mathcal{G}_s \subseteq \mathcal{G}_t$.

It is advantageous to think of the filtration as part of the point process and we require that events that have occurred by time t be "observable at time t." Suppose, for example, that T^i is the random time of the ith default. Then the event that "the ith default is on or before time t," must be observable at time t. This amounts to the mathematical statement:

$$\{\omega \in \Omega \,|\, T^i(\omega) \leq t\} \in \mathcal{G}_t. \tag{11}$$

In the language of probability theory, property (11) makes T^i a *stopping time* with respect to the filtration $\mathbb{G}$.

Associated to the sequence (T^i) of random times is a *counting process* N whose value $N_t(\omega)$ counts the number of events that have occurred by time t if the world is in state ω. The process N contains precisely the same information that is in the sequence (T^i), and it is also called a point process.

The reason for recasting the information in the random times (T^i) as a counting process N is that the latter admits a Doob-Meyer decomposition, which is a fundamental construction in the theory of stochastic processes. To minimize technicalities, we make the economically reasonable assumption that the events (T^i) are *totally inaccessible* with respect to the filtration $\mathbb{G}$. This means that the events are surprises: they come unannounced. Mathematically, it means that no T^i can be expressed as the increasing limit of stopping times on any set of positive measure. Under this assumption, the Doob-Meyer decomposition is as follows.

Theorem 4.1. *Suppose the event times (T^i) are totally inaccessible with respect to the filtration $\mathbb{G}$ and increase to infinity along almost every path. Under* *mild growth restrictions, the associated counting process N can be decomposed into a sum*

$$N = A + M \tag{12}$$

where A is a continuous, nondecreasing process and M is a martingale with respect to $\mathbb{G}$.

Recall that a $\mathbb{G}$-*martingale* is a process that is fair, in the sense that its expected change is 0 at every time in almost every state of the world. Thus, the martingale M in formula (12) does not contribute to the upward tendency of the counting process N. This tendency is completely captured by the process A, which is called the $\mathbb{G}$-compensator of N. It is a continuous representation of the upward tendency in the point process.

In many applications, such as forecasting default probabilities and valuing default dependent securities, the compensator is the centerpiece of the model. This is very useful, since the compensator is a a very flexible concept that can be modeled, for example, in terms of conditional probability or information available to investors.

The simplest and most familiar point process is a Poisson process N with intensity λ. In this example, events arrive independently and the waiting time between events is exponential with parameter λ. Thus, the expected number of events in the interval $(s, t]$ given all observable information at time $s < t$ is equal to $\lambda(t - s)$. It follows that $N_t - \lambda t$ defines a $\mathbb{G}$-martingale and the compensator to N_t, is given by λt.

Of course, real world events do not generally follow a Poisson process. They are dependent on each other and also have different waiting time distributions. As we discuss in Section 4.3, an important application of the compensator is to rescale real world data to a standard Poisson process.

4.2 An application: modeling default contagion

Default events are dependent on each other since firms are subject to common macro-economic shocks, which can be propagated and aggravated through the complex network of firm relationships. Extreme market fluctuations can result.

We describe a simple point process model of default events that incorporates this market contagion. Our model is an example of a *self-exciting* process.

Many but not all point processes can be specified by a conditional default rate λ. This is a nonnegative stochastic process whose relationship to the compensator of the point process is given by

$$A_t = \int_0^t \lambda_s ds. \tag{13}$$

In other words, λ is the time derivative of A. The product $dA_t = \lambda_t dt$ gives the conditional probability of having a default in the infinitesimal interval $(t, t+dt]$, given observable information at time t. If $\lambda_s = \lambda$ is constant for all s, then $A_t = \lambda t$ and N follows a Poisson process. In a self-exciting model, the process λ is updated with information about events as they occur.

Consider a fixed collection of firms. The collection can range from the population of all firms in the economy to the counterparties in a security portfolio. Suppose that all firms start out financially healthy and that the first default arrives according to a Poisson process with intensity $\bar{\lambda}$. In other words, the waiting time until the first default is exponential with parameter $\bar{\lambda}$.

At the first default time T^1, we update the intensity $\bar{\lambda}$ according to a rule given by a nonnegative function Λ^1. This rule can take account of the identify of the defaulter at time T_1 or the state of the economy, or any other information that is available at T^1. We assume the second default arrives with intensity given by $\bar{\lambda} + \Lambda^1(t - T^1)$ at time $t \geq T^1$. At default time T^2, we update according to the function Λ^2 and suppose the third default arrives with intensity $\bar{\lambda} + \Lambda^1(t - T^1) + \Lambda^1(t - T^2)$ at time $t \geq T^2$. Continuing, the intensity λ of our model is

$$\lambda_t = \bar{\lambda} + \sum_{i:T^i \leq t} \Lambda^i(t - T^i). \tag{14}$$

The updating functions Λ^i incorporate new information into the intensity λ. A convenient form is

$$\Lambda^i(t) = \alpha_i e^{\beta_i t}, \tag{15}$$

where parameters α_i and β_i specify the influence of the ith default at time T^i on the intensity of defaults after T^i. For examples, the parameter α may reflect the time of the default, the identity and size of the defaulter and the degree of its connectivity to surviving firms. If the ith defaulter is relatively large and has strong connections to surviving firms, future defaults become more likely and there is a large jump in λ. The parameter β_i specifies the rate of the exponential decay of the influence of the ith default on the surviving firms. It can be used to calibrate the model to market prices.

The self-exciting process described in (14) was proposed by Hawkes (1971). It is a natural choice for modeling contagion, since the information about extreme events is immediately incorporated into the model intensity.

4.3 Testing: rescale to the Poisson process

How well to point processes fit empirical data? A powerful result of Meyer provides useful in-sample fitness tests for models of future random events. Under the assumption that the events in a point process are totally inaccessible, the process can be rescaled to Poisson process with unit intensity by a stochastic change of time. The time change is given by the inverse of the compensator. Hence most point processes of practical use in financial risk modeling are Poisson up to a change of time.

Suppose we want to test a point process N whose events are totally inaccessible. An example is the multi-firm default model described in Section 4.2. Meyer's time change result implies the following test procedure, which is described in detail in Giesecke & Goldberg (2004):

(1) Fit the model compensator A to an observed sequence (T^i) of past defaults and other observable information. Rating agencies, for example, maintain extensive default histories for their universe of rated firms. In the context of the contagion model (14), this means we fit the intensity parameters (α_i, β_i).

(2) Transform the observed sequence (T^i) with the compensator A into a new sequence (S^i) by setting

$$U^i = A_{T^i}. \tag{16}$$

(3) Test the hypothesis that the transformed sequence U^i is a standard Poisson process with respect to the transformed information sets $\mathcal{G}_{A_t}$.

There are a variety of standard tests of the Poisson property available. For example, we can test whether the inter-arrival time $(U^i - U^{i-1})$ is independent and standard exponential conditional on the information observed at time S^{i-1}, represented by $\mathcal{G}_{U^{i-1}}$, for all i. Furthermore, we can test whether the number of arrivals $\sum_i 1_{\{A_s < U^i \leq A_t\}}$ in the transformed interval $(A_s, A_t]$ is Poisson with parameter $(A_t - A_s)$ conditional on the information observed at the transformed time A_s, represented by $\mathcal{G}_{A_s}$.

If the Poisson hypothesis is rejected, then the model compensator does not provide the correct rescaling for the observed arrival times. In this case the model does not provide a good fit to the observed sequence of event times.

5 Further Reading

The framework of risk measures such as value-at-risk and expected shortfall is established in Artzner, Delbaen, Eber & Heath (1999). See Embrechts, McNeil & Straumann (2001) for an analysis of value-at-risk in the context of normal distributions.

An excellent mathematical overview of many of the topics discussed in Section 3 is Embrechts, Klüppelberg & Mikosch (1997). This fine book includes clear exposition of the main topics, countless examples and an invaluable guide to the literature. A careful proof of the the Fisher-Tippett theorem, the Karamata representation theorem and the classification of extreme value distributions is in Resnick (1987).

A practical discussion of how to implement the peaks of thresholds estimation in one variable is Frey & Saladin (1997), and the mathematical foundations underlying its extension to higher dimensions is Balkema & Embrechts (2004). The non-trivial relationship between the Hill estimator (8) and the power exponent involves the analysis of order statistics, which is exposed, for example in Reiss (1989).

A rich, informative source for point processes is Daley & Vere-Jones (2003). This book includes a section on self-exciting processes, which were first proposed by Hawkes (1971). See also Hawkes & Oakes (1974). A theoretical discussion of the role of compensators in default models is in Giesecke (2001). A careful, elegant proof of the Doob-Meyer decomposition is in Meyer (1966). A modern exposition of Meyer's time change result and applications to testing default models is in Giesecke & Goldberg (2004). Meyer's original exposition can be appreciated by those who read French in Meyer (1971).

An enjoyable, non-technical history of mathematical finance with an emphasis on risk due to extreme events is Mandelbrot & Hudson (2004).

References

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber & David Heath (1999), 'Coherent measures of risk', *Mathematical Finance* **9**(3), 203–228.

Balkema, Guus & Paul Embrechts (2004), Multivariate excess distributions. Working Paper, ETH Zurich.

Daley, Daryl & David Vere-Jones (2003), *An Introduction to the Theory of Point Processes*, Springer-Verlag, New York.

Embrechts, Paul, Alexander J. McNeil & Daniel Straumann (2001), Correlation and dependence in risk management, *in* M.Dempster, ed., 'Risk management: value at risk and beyond', Cambridge University Press, Cambridge.

Embrechts, Paul, Claudia Klüppelberg & Thomas Mikosch (1997), *Modelling Extremal Events*, Springer-Verlag, Heidelberg.

Frey, Alexander & Thomas Saladin (1997), The peaks over thresholds method for estimating high quantiles of loss distributions, *in* 'Proceedings of 28th International ASTIN Colloquium'.

Giesecke, Kay (2001), Default and information. Working Paper, Cornell University.

Giesecke, Kay & Lisa Goldberg (2004), Fitness tests for multi-firm default models. Working Paper, Cornell University.

Hawkes, Alan G. (1971), 'Spectra of some self-exciting and mutually exciting point processes', *Biometrika* **58**(1), 83–90.

Hawkes, Alan G. & David Oakes (1974), 'A cluster process representation of a self-exciting process', *Journal of Applied Probability* **11**, 493–503.

Mandelbrot, Benoit & Richard L. Hudson (2004), *The (Mis)behavior of Markets*, Basic Books, New York.

Meyer, Paul-Andre (1966), *Probability and Potentials*, Blaisdell, London.

Meyer, Paul-Andre (1971), Démonstration simplifée d'un théorèm Knight, *in* 'Séminaire de Probabilités V, Lecture Note in Mathematics 191', Springer-Verlag Berlin, pp. 191–195.

Reiss, Rolf-Dieter (1989), *Approximate Distributions of Order Statistics*, Springer, New York.

Resnick, Sidney (1987), *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, Heidelberg.