CentER

MANDELBROT'S EXTREMISM

By J. Beirlant, W. Schoutens, JJJ. Segers

December 2004

Discussion Paper

TILBURG ◆ ◆ UNIVERSITY

Mandelbrot's Extremism

Jan Beirlant[*]   Wim Schoutens[†]   Johan Segers[‡]

December 6, 2004

Abstract

In the sixties Mandelbrot already showed that extreme price swings are more likely than some of us think or incorporate in our models. A modern toolbox for analyzing such rare events can be found in the field of *extreme value theory*. At the core of extreme value theory lies the modelling of maxima over large blocks of observations and of excesses over high thresholds. The general validity of these models makes them suitable for out-of-sample extrapolation. By way of illustration we assess the likeliness of the crash of the Dow Jones on October 19, 1987, a loss that was more than twice as large as on any other single day from 1954 until 2004.

JEL: C 13; C 14

Key Words: exceedances; extreme value theory; heavy tails; maxima

1 Introduction

The stock market doesn't like Central Limit Theory: in finance, the well-mannered bell curve of the Gaussian distribution isn't so normal at all. It likes it hotter, spicier, more extreme. Indeed, extreme price swings are more likely than some of us think or incorporate in our models. This insight, which Mandelbrot already had in the sixties, long before the Black-Scholes model was ruling Wall Street, is one of many messages that he has been trying to convey to the financial community for decades now (see e.g. [8]).

The problem with the Normal (Gaussian) distribution lies in the tails. Financial data more often than not exhibit power-law tails, to be treated further on. First evidence of this phenomenon was written down by Mandelbrot in a research report [9] in 1962 (see also [10]), published one year later in [11] and extended in [12] (reprints appeared later several times). Power-law tail behavior was found in the movements of interest rates and prices of cotton, wheat, and railroad stocks. In comparison, Normal tails decay to zero much too fast.

The shortcomings of the Normal distribution are dramatically illustrated by e.g. looking at the Dow Jones Industrial Average. Table 1 lists the ten largest relative down moves of the Dow over the last fifty years (1954–2004).

---

[*]Catholic University Leuven, University Centre for Statistics, W. de Croylaan 54, B-3001 Leuven, Belgium. E-mail: Jan.Beirlant@wis.kuleuven.ac.be

[†]Catholic University Leuven, University Centre for Statistics, W. de Croylaan 54, B-3001 Leuven, Belgium. E-mail: Wim.Schoutens@wis.kuleuven.ac.be

[‡]Department of Econometrics and Operations Research, Tilburg University, P.O.Box 90153, NL-5000 LE Tilburg, the Netherlands. E-mail: jsegers@uvt.nl

Date	Closing	logreturn
19-Oct-87	1738.74	-0.2563
26-Oct-87	1793.93	-0.0838
27-Oct-97	7161.15	-0.0745
17-Sep-01	8920.70	-0.0740
13-Oct-89	2569.26	-0.0716
08-Jan-88	1911.31	-0.0710
26-Sep-55	455.56	-0.0677
31-Aug-98	7539.07	-0.0658
28-May-62	576.93	-0.0588
14-Apr-00	10305.77	-0.0582

Table 1: *Ten largest down moves of the Dow (1954–2004)*

Under the Black-Scholes regime, what is the probability that the Dow will suffer a big loss tomorrow? Everything depends of course on the volatility that you plug in. Figure 1 shows the annualized historical volatility estimated on the basis of, say, a three-year window. Clearly, volatility is not constant and behaves stochastically – another point where Mandelbrot's was a pioneer [13]. In the figure, volatility is typically below 25%. Let us calculate for a 25% vol the frequency of a negative logreturn of -0.0582 or even worse. Under the assumption of Normality, it happens just once every 35 years. In reality, we have witnessed ten in the last 50 years! If the mathematician Thales (c.624–c.546 BC) – one of the ancient derivatives traders – would have been granted eternal live, he would according to the Normal distribution have seen only one down move of -0.0716 or worse up to now. In the last fifty years we had five! A Homo Sapiens would likely have witnessed only one down move of -0.0838 or worse up to now. In a particularly bad month, October 1987, there were two! What is the probability of a down move of -0.25 or worse: It is of the order once in the 10^{53} years (in US language: 100 sexdecillion years, UK language: 100000 octillion years). In contrast, the Big Bang *only* happened around 15×10^9 years ago. The present generation must be really exceptional that God allowed the Dow to crash in October 1987.

What are the main problems with the Black-Scholes model? First of all, it postulates a Normal distribution for logreturns, completely missing the observed tail behavior. Secondly, the environment is changing and the volatility is behaving stochastically over time. Thirdly, the sample paths of the driving Brownian motion are continuous, whereas there is clear evidence that in reality stock prices like to *jump*.

The first and third problem could be overcome by letting the price dynamics be driven by more flexible processes, allowing for jumps and fatter tails. Moreover, stochastic volatility models are nowadays finding their ways into the business. We especially mention the introduction of stochastic volatility by letting stocks operate in their own "trading time" or "business time", an idea dating back to [13] and nowadays being implemented in e.g. a Lévy-process driven market. For details and an overview on Lévy process models, their stochastic volatility extensions and their applications in derivatives pricing we refer to [15].

In this paper we focus on the modelling of extreme values and the corresponding tail behavior. Extreme value theory is by now a well-developed area of statistics and finds applications in many areas of research: besides finance, it is/can be used in hydrology, cosmology, insurance, pollution and climatology, geology, etc. A basic
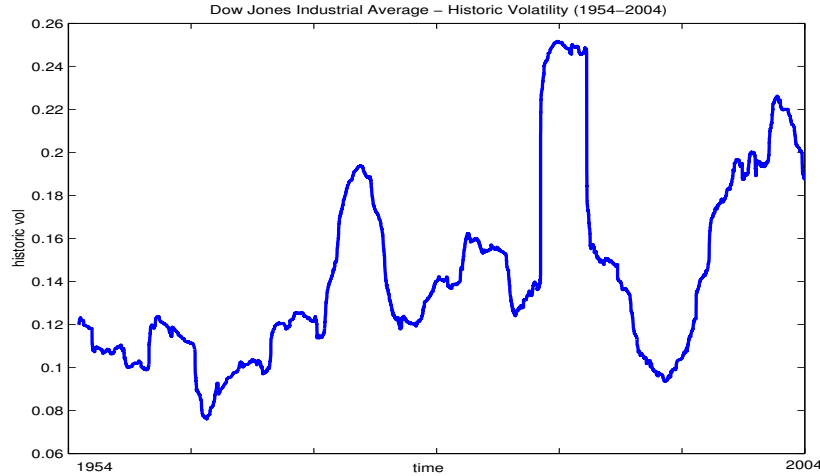
Figure 1: *Dow Jones Industrial Average – Historic Volatility (1954–2004)*

reference text is [2].

2 The tale a tail has to tell

The Normal distribution with mean μ and variance σ^2 exhibits a quadratic decay near infinity of the logarithm of its probability density function:

$$\log f_{\text{Normal}}(x; \mu, \sigma^2) = -\frac{(x-\mu)^2}{2\sigma^2} - \log\left(\sigma\sqrt{2\pi}\right) \sim -\frac{1}{2\sigma^2}x^2 \tag{1}$$

as $x \to \pm\infty$.

A power-law distribution, suggested by Mandelbrot to model the tails of financial returns, has a completely different decay. The distribution goes back to Vilfredo Pareto (1848–1923, economist and very unfortunate speculator on the LME), who suggested it in 1896 as a model for personal income. The distribution of a random variable X is said to follow a power-law or Pareto distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ if

$$\Pr[X > x] = (x/\beta)^{-\alpha}, \qquad \text{for } x \geq \beta. \tag{2}$$

The distribution lives on (β, ∞), but it is its tail decay that is important to arrive at improved tail extrapolation.

The logarithm of its density function is given by

$$\log f_{\text{Pareto}}(x; \alpha, \beta) = \log\left(\alpha\beta^{-\alpha}x^{-\alpha-1}\right) \sim -(\alpha+1)\log(x) \tag{3}$$

as $x \to +\infty$.

The difference between the *light-tailed* Normal distribution (1) and the *heavy-tailed* Pareto distribution (3) is huge. We rely on the power of the image to imprint this message.

Consider the sample of negative daily logreturns of the Dow Jones Industrial Average for the 50-year period from November 9, 1954, until November 8, 2004. We
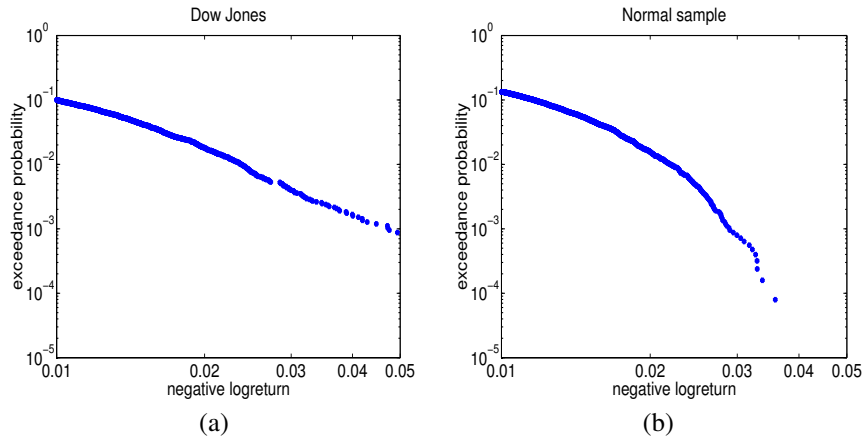
3

Figure 2: (a) *Negative daily logreturns in the range from 0.01 up to 0.05 for the Dow (1954–2004) versus sample exceedance probabilities (1954–2004). (b) Similarly for a Normal random sample with the same mean and variance.*

compare it with a sample of the same size from a Normal distribution with the same mean and variance. In particular, since the focus in this sample is on the larger losses, we focus on the larger values in the two samples.

Figure 2(a) shows the negative daily logreturns of the Dow in the range from 0.01 up to 0.05, corresponding to losses from 1% up to 5% approximately, versus their sample exceedance probabilities. The most striking feature of the plot is its linearity in loglog-scale. In contrast, panel (b) of the same figure shows the analogous plot for the Normal sample. The plot is distinctly concave, suggesting much smaller exceedance probabilities for the larger losses.

A risk measure currently gaining in popularity is the *expected shortfall*, defined as the expected excess over a given (high) level, conditionally on this level being exceeded. The sample version of the expected shortfall over a certain level is simply the average of the excesses over that level. The expected shortfall over the 1000 largest negative daily logreturns of the Dow are plotted in Figure 3(a). Note that the expected shortfall is increasing with the level: the higher the level being exceeded, the higher the excess by which it will be exceeded! Once more, this is in sharp contrast with panel (b) of the same figure: for a Normal sample with the same mean and variance, the expected shortfall decreases rapidly (note the different axes). In a light-tailed world, given that you exceed a high level, you hardly exceed it at all. But in a heavy-tailed world, once you know you'll get hit, you may get hit much harder than expected!

3 Extreme value theory

3.1 Block maxima

Let S_t be the closing price of the Dow at the end of day t. The negative logreturn at day t is given by
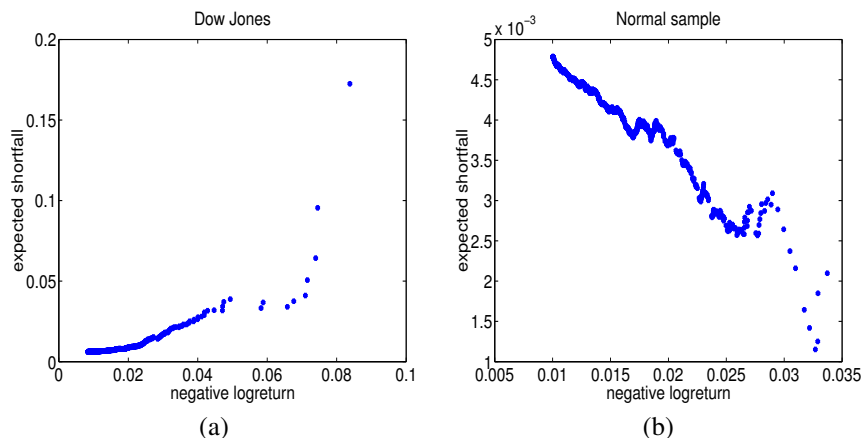
$$X_t = \log S_{t-1} - \log S_t.$$

4

Figure 3: (a) *Expected shortfall over 1000 largest negative daily logreturns of the Dow (1954–2004).* (b) *Similarly for a Normal random sample with the same mean and variance.*

Large down movements of the Dow correspond to high X_t. In particular, the highest X_t in a certain period of time corresponds to the biggest loss of the Dow in that period. Correctly quantifying the possible magnitude of this worst one-period event should be an important ingredient in a prudent risk assessment. Whereas the Normal model is totally inappropriate for this purpose, much more satisfactory tools are available in the field of extreme value theory.

Divide the sample of negative daily logreturns $X_1, \ldots, X_n$ into blocks of (approximately) equal size, say m. For instance, m equal to 20, 60 or 250, corresponds approximately to months, quarters or years, respectively. For each such block, consider the largest observation. That is, we record the largest one-day loss on the Dow for each month, quarter or year. In this way, we obtain for every block size m a sample of block maxima.

For different block sizes m we get different samples. Figure 4(a) shows non-parametric density estimates for the samples of monthly, quarterly and yearly maxima. The features of these densities depend on the block size m: for larger m, the location moves to the right and the dispersion increases. This corresponds to intuition: the larger the block size, the larger we expect the block maxima to be.

The shapes of these densities may be compared by bringing the different samples of block maxima to a common location and scale. Figure 4(b) shows non-parametric density estimates for the studentized[1] samples of monthly, quarterly and yearly maxima. Up to differences in location and scale, the distributions of the three samples seem to have grossly the same shape.

Now the following postulate does not sound too unreasonable:

> **Block-maximum postulate.** *Up to location and scale, the distribution of block maxima stabilizes as the block size grows indefinitely large.*

In mathematical terms, we postulate the existence of scaling constants $a_m > 0$, centering constants b_m, and a non-degenerate[2] cumulative distribution function (cdf) G such

---

[1] Studentizing a sample means subtracting its mean and dividing by its standard deviation.
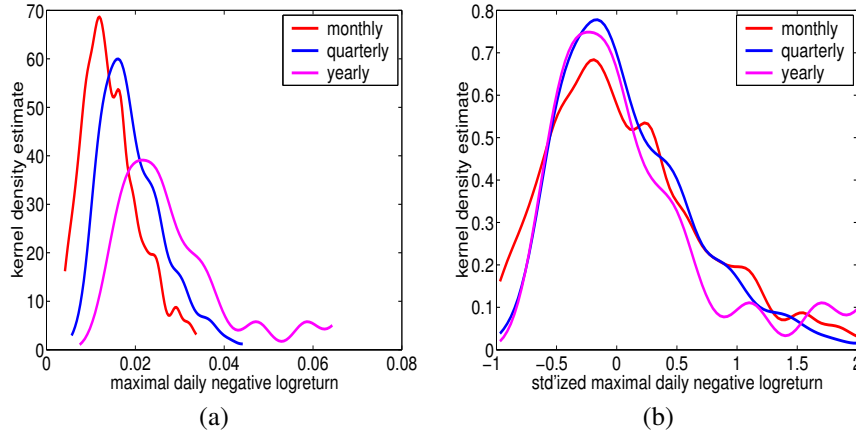[2] A cdf G is non-degenerate if there exists x such that $0 < G(x) < 1$.

5

Figure 4: (a) *Non-parametric estimates of the densities of monthly, quarterly and yearly maxima of negative daily logreturns of the Dow (1954—2004).* (b) *Similarly for studentized samples of maxima.*

that for all large m the distribution of $a_m^{-1}(\max\{X_1, \ldots, X_m\} - b_m)$ can be approximated well by G; or technically

$$\lim_{m \to \infty} \Pr[a_m^{-1}(\max\{X_1, \ldots, X_m\} - b_m) \le x] = G(x). \tag{4}$$

For the sake of exposition, let us assume that the variables X_t are independent and identically distributed with common cdf $F(x) = \Pr[X_t \le x]$. These assumptions are of course unreasonable for a time series of daily logreturns stretching over a fifty-year period. However, the conclusions we will arrive at remain valid under much weaker assumptions involving for instance periodicity and temporal dependence decaying sufficiently fast over time.

There are two basic questions related to equation (4):

(i) Which cdfs G can arise in the limit?

(ii) Given the limit cdf G, how does the cdf F of the variables X_t look like?

The answers to these questions will lead us to the very essence of heavy tails.

The answer to question (i) is surprisingly simple and dates back to a result by Fisher and Tippett from 1928, see [4]: the limit cdf G in (4) must be a member of the three-parameter family of *extreme value distributions*. In a parametrization due to Jenkinson and von Mises, the general form of such a cdf is

$$G(x; \gamma, \mu, \beta) = \begin{cases} \exp\left(-\left(1 + \gamma \dfrac{x - \mu}{\beta}\right)^{-1/\gamma}\right) & \text{if } \gamma \ne 0, \\ \exp\left(-\exp\left(-\dfrac{x - \mu}{\beta}\right)\right) & \text{if } \gamma = 0. \end{cases} \tag{5}$$

for all x such that $1 + \gamma(x - \mu)/\beta > 0$. The shape parameter γ (sometimes also denoted by ξ) is commonly referred to as the *extreme value index*, while μ and $\beta > 0$ are just parameters for location and scale, respectively.
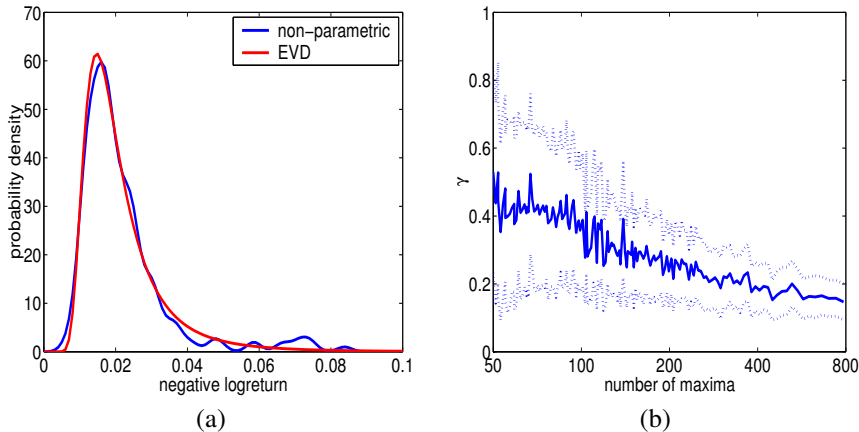
Figure 5: (a) *Non-parametric and parametric density estimates for quarterly maxima of negative daily logreturns of the Dow (1954—2004).* (b) *Maximum likelihood estimates and 95% confidence intervals for* γ *as function of the number of block maxima.*

By a combination of the block-maximum postulate with the Fisher-Tippett result, the distribution of maxima over sufficiently large blocks may be modelled by an extreme value distribution (5); the scaling and centering constants a_m and b_m for block maxima are incorporated into the location and scale parameters μ and β of the extreme value distribution G. The validity of this approach is confirmed by Figure 5(a), comparing a non-parametric density estimate for quarterly maxima with the density of the fitted extreme value distribution, the parameters of the latter being estimated by maximum likelihood.

We wish to stress the point that we did not *a priori* impose the extreme value distribution as a model for maxima over large blocks. Rather than that, the model arises as an inevitable consequence from the block-maximum postulate. This yields some extra justification of the model apart from its apparent goodness of fit, which could also be achieved by other sufficiently flexible parametric families of distributions.

If we repeat the exercise of fitting an extreme value distribution to samples of block maxima for different block sizes, then we expect the estimated location and scale parameters to change with the block size while the shape parameter or extreme value index γ should remain roughly constant. Figure 5(b) shows the various maximum likelihood estimates of γ together with 95% confidence intervals. The estimates are plotted against the number of block maxima, which is equal to the total number of observations in the sample (12 587) divided by the block size.

Although it is hard from the plot to draw any precise inference on the value of the extreme value index γ, there seems little doubt that it is positive. Now one can also prove that for positive γ the tail of an extreme value distribution looks like the tail of a Pareto distribution with shape parameter $\alpha = 1/\gamma$:

$$1 - G(x; \gamma, \mu, \beta) \sim \text{constant} \times x^{-1/\gamma}, \qquad \text{as } x \to \infty.$$

In contrast, according to a Normal model for the daily logreturns, the limiting model for maxima of large blocks would be an extreme value distribution with extreme value index equal to zero, the so-called Gumbel distribution. The tail of the Gumbel dis-
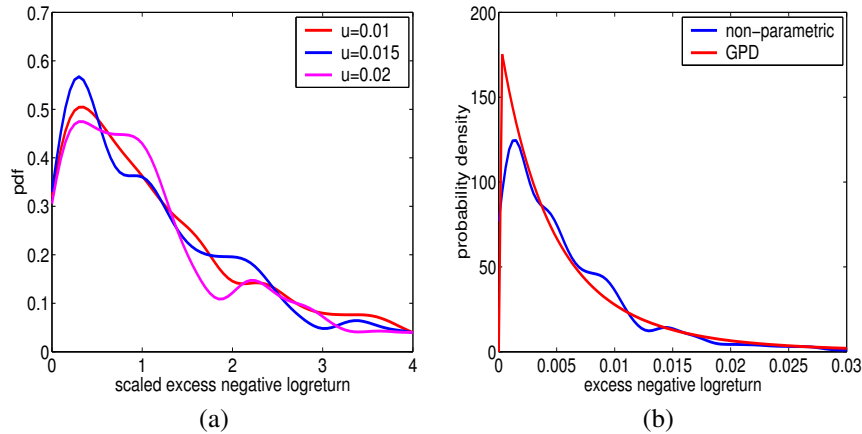
7

Figure 6: (a) *Non-parametric density estimates for scaled excesses over thresholds u of negative daily logreturns of the Dow (1954–2004). (b) Non-parametric and generalized Pareto density estimate for excesses over 0.015 of negative daily logreturns of the Dow (1954–2004).*

tribution, however, decays exponentially fast to zero for large x, much faster than the power-law tail above.

3.2 Threshold exceedances

Now let us turn to question (ii): if maxima over increasingly large blocks can be modelled by an extreme value distribution with extreme value index γ, what can be said about the cdf F of the individual negative logreturns X_t? Since the maximum of a large block with high probability will be a value with only a small probability of being exceeded, the only knowledge to be inferred will be on the behavior of $F(x) = \Pr[X_t \leq x]$ for increasingly large x.

To gain some feeling for the problem, let us focus on those negative logreturns X_t that exceed a high threshold u, that is, those losses of at least a given size. The threshold u is high in the sense that it is exceeded by only a small percentage of the observations. For each such u, we obtain a sample of threshold excesses $X_t - u$ from those observations X_t that exceed u. Different thresholds u lead to different samples of excesses, the samples getting smaller as the threshold increases. For the negative daily logreturns of the Dow, the thresholds u of 0.01, 0.015 and 0.02 are exceeded by respectively 10%, 4.2% and 1.8% of the observations.

Samples of excesses corresponding to different thresholds can be compared by bringing them to a common scale, for instance by dividing them by their respective standard deviations. Unlike for block maxima, no additional centering is needed. Figure 6(a) shows non-parametric density estimates of the samples of scaled excesses over the thresholds u equal to 0.01, 0.015 and 0.02. Up to differences in scale, the distributions of the samples of excesses over the different thresholds roughly coincide.

This leads us to formulate the following postulate:

> **Excess postulate.** *Up to scale, the distribution of excesses over a high threshold stabilizes as the threshold u grows indefinitely large.*
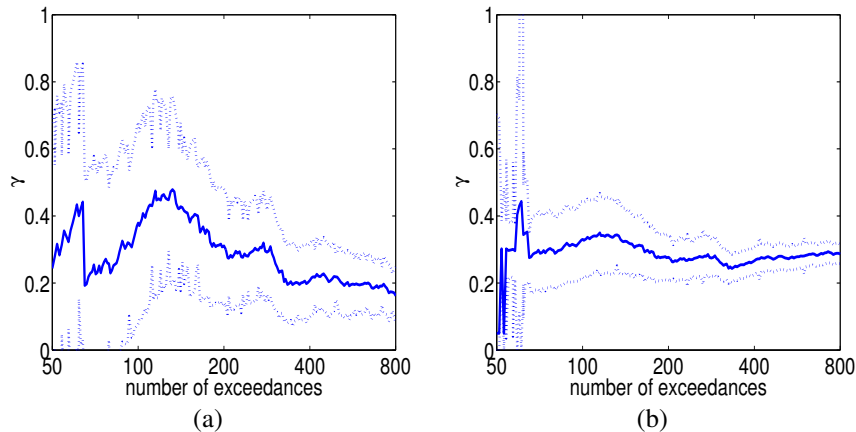
Figure 7: *Maximum likelihood estimates and 95% confidence intervals for extreme value index γ based on negative logreturns of the Dow (1954–2004) for various thresholds as function of the number of threshold exceedances.* (a) *Generalized Pareto model.* (b) *Extended Pareto model [1].*

In mathematical terms, we require the existence of positive scaling constants $a(u)$ and a non-degenerate cdf H such that for all large u the distribution of the scaled excess $(X - u)/a(u)$ can be approximated well by H; or technically,

$$\Pr[(X_t - u)/a(u) \leq x \mid X_t > u] = \frac{1 - F(u + a(u)x)}{1 - F(u)} \rightarrow H(x) \qquad (6)$$

as u increases to the upper end-point of the support of F.

Pickands [14] showed that the excess postulate is equivalent to the block-maximum postulate: convergence in distribution of scaled and centered block maxima is equivalent to convergence in distribution of excesses over high thresholds. Moreover, the limit distributions are related in the following way. The limit distribution G for block maxima is an extreme value distribution with shape parameter γ if and only if the limit distribution H for threshold excesses is a *generalized Pareto distribution* with the same shape parameter γ:

$$H(x; \gamma, \beta) = \begin{cases} 1 - (1 + \gamma x/\beta)^{-1/\gamma} & \text{if } \gamma \neq 0, \\ 1 - \exp(-x/\beta) & \text{if } \gamma = 0, \end{cases} \qquad (7)$$

this for all $x > 0$ such that $1 + \gamma x/\beta > 0$. The parameter $\beta > 0$ is just a scale parameter which can depend on the chosen scaling $a(u)$ of the threshold exceedances.

The excess postulate together with Pickands' result suggests to model excesses over a high threshold by a generalized Pareto distribution (7). For negative daily logreturns of the Dow, Figure 6(b) compares a non-parametric estimate of the density of excesses over the threshold u equal to 0.015 with the fitted generalized Pareto density. The generalized Pareto parameters have been estimated by maximum likelihood.

This fitting exercise for various thresholds u, hopefully leads to estimates of the extreme value index γ that do not vary too much. Figure 7(a) shows maximum likelihood estimates and 95% confidence intervals for a range of thresholds as a function of the number of threshold exceedances. The scale on the horizontal axis is chosen so as to facilitate a comparison with the estimates for γ using block maxima in Figure 5(b).

9

A disturbing feature of the plot is that the estimates for γ vary strongly according to the chosen threshold. This is basically due to the model error that arises when treating the limit relation (6) as an exact equality for high enough threshold. Explicitly modelling the approximation error results in an extension of the generalized Pareto model described in [1], see also [2], pp. 117–118. The corresponding estimates of the extreme value index in Figure 7(b) are much more stable over various thresholds and suggest a value for the extreme value index γ around 0.3.

Now fix a high threshold u and fit a generalized Pareto distribution to the excesses over u. Then we arrive at the following model for F in the region $[u, \infty)$: for $x \geq u$,

$$
\begin{aligned}
1 - F(x) &= \{1 - F(u)\}\frac{1 - F(u + (x - u))}{1 - F(u)} \\
&\approx \{1 - F(u)\}\{1 + \gamma(x - u)/\beta\}^{-1/\gamma}.
\end{aligned}
\tag{8}
$$

Here $1 - F(u)$ can be estimated by the sample proportion of exceedances over the threshold u, while γ and β can be estimated by fitting a generalized Pareto distribution to the excesses over u.

In particular, we see that for positive extreme value index γ, the right tail of the cdf F of the negative logreturns behaves like a Pareto distribution with index $\alpha = -1/\gamma$:

$$
1 - F(x) \sim \text{constant} \times x^{-1/\gamma}, \qquad \text{as } x \to \infty.
$$

In contrast, the Normal model for F would entail an exponential rather than a polynomial decay to zero of the tail function $1 - F$. As a consequence, exceedance probabilities over high levels would be grossly underestimated.

4 Return levels and return periods

The return period of a high level x is defined as the number of years $T = T(x)$ such that exceedances over x are observed once every T years on average. If there are 250 observations per year, as is the case for daily returns of the Dow, this means that the probability of exceeding x should be $1/(250T)$. Hence the return period T (in years) for a high level x is formally defined through the relation

$$
1 - F(x) = \frac{1}{250T}.
$$

Conversely, the return level given a return period T is defined as that level $x = x(T)$ solving the previous equation. It can be thought of as that value x that is exceeded once every T years on average.

Now suppose that you are given the sample of negative daily logreturns of the Dow in the fifty-year period 1954–2004 excluding the crash of October 19, 1987. The largest observation in your sample is about 0.084. Then what would be your estimate of the return period of a negative logreturn of 0.25 as observed on that particular day in October 1987? Or conversely, what would be your estimate of return levels with return periods of fifty years and more? As explained in the introduction, to such questions, the Normal model would give completely misleading answers.

Instead, the tail model (8) is based on a very general theory with only a minimum of assumptions. Extrapolating from the generalized Pareto model fitted to exceedances over a high threshold yields estimates of return levels corresponding to periods longer
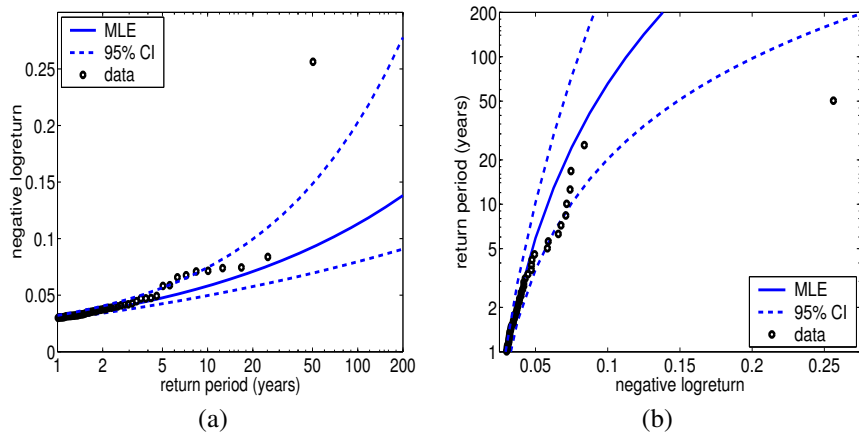
Figure 8: *Maximum likelihood estimates and profile likelihood 95% confidence intervals for* (a) *high return levels and* (b) *high return periods (in years) for negative daily logreturns of the Dow based on generalized Pareto model for excess negative logreturns over 0.02 excluding the crash of October 19, 1987.*

than the observation period and return periods corresponding to levels that are higher than observed so far.

Figure 8 shows maximum likelihood estimates and profile likelihood 95% confidence intervals for high return periods and high return levels. The estimates are based on the generalized Pareto model fitted to negative daily logreturns of the Dow exceeding 0.02 and excluding the event of October 19, 1987. Also shown on the plots are the largest negative logreturns versus their empirical return periods.

The negative logreturn of October 19, 1987, clearly stands out as the lone circle in the upper right of the two plots. According to panel (a), it falls within the confidence interval for the return level with a return period 200 years. Similarly, according to period (b), the confidence interval for the return period of a negative daily logreturn of 0.25 contains a period of 200 years. So, although unlikely, the event on October 19, 1987, is according to this elementary extreme value analysis far from impossible at all. And, maybe this generation is not that exceptional under an *abnormal* God.

5 Conclusion

The Normal distribution is a particularly inadequate model for the larger values in financial data. In contrast to the Normal's exponentially light tails, time series of returns on interest rates, commodity prices and stock prices often exhibit power-law tail behavior, as already revealed by Mandelbrot's pioneering work in the sixties. In reality, large price movements occur much more often than can be accounted for by the Normal model.

A modern toolbox for analyzing such rare events can be found in the field of *extreme value theory*. A few very simple diagnostic plots already show the huge difference between the occurrence of larger losses for the Dow Jones versus those in a Normal pseudo-random sample with the same mean and variance.

At the core of extreme value theory lies the modelling of maxima over large blocks

of observations by the three-parameter family of extreme value distributions or, equivalently, excesses over high thresholds by the two-parameter family of generalized Pareto distributions. The use of these two families of distributions is an inevitable consequence of some very general assumptions, the block-maximum postulate or the excess postulate.

These models can be used to compute return levels for return periods longer than the observation period and to compute return periods of levels that have not been surpassed thus far. Doing the exercise for the Dow Jones shows that the single-day loss of about 25% on October 19, 1987, is far from impossible given the other data, although the second largest loss in the period 1954–2004 is not larger than 9%. For how to hedge in crash scenarios we refer to e.g. [7].

Not treated in this paper have been the effects of temporal dependence, which may cause large losses to cluster over time, in periods of high volatility for instance. More refined models designed to account for such temporal dependence for extremes include Poisson process models and Markov chain models [2, 3]. Another issue is that of the choice of threshold defining the excesses to which to fit the generalized Pareto distribution. Recently developed techniques offer a solution through stabilizing the estimates over a large range of thresholds [1, 2].

References

[1] Beirlant, J., Joossens, E., and Segers, J. (2004) Discussion of "Generalized Pareto Fit to the Society of Actuaries' Large Claims Database" by A. Cebrián, M. Denuit and Ph. Lambert. *North American Actuarial Journal* **8**, 108–111.

[2] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley.

[3] Ferro, C.A.T. and Segers, J. (2003) Inference for clusters of extreme values. *Journal of the Royal Statistical Society, Series B*, **65**, 545–556.

[4] Fisher, R.A. and Tippett, L.H.C. (1928) On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* **24**, 180–190.

[5] Gnedenko, B.V. (1943) Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44**, 423–453.

[6] Gumbel, E.J. (1958) *Statistics of Extremes*. Columbia University Press.

[7] Hua, P. and Wilmott, P. (1999) *Crash Modelling, Value at Risk and Optimal Hedging*. OFRC Report 1999-MF-06.

[8] Hudson, R.L. and Mandelbrot, Benoit B. (2004) *The (Mis)Behavior of Markets – A Fractal View of Risk, Ruin and Reward*. Perseus Books Group.

[9] Mandelbrot, Benoit B. (1962) The variation of certain speculative prices. *IBM Research Report NC-87*.

[10] Mandelbrot, Benoit B. (1962b) Sur certains prix spéculatifs: faits empiriques et modèle basé sur les processus stables additifs de Paul Lévy. *Comptes rendus (Paris)* **254**, 3968–3970.

[11] Mandelbrot, Benoit B. (1963) The variation of certain speculative prices. *Journal of Business* **36**, 394–419.

[12] Mandelbrot, Benoit B. (1967) The variation of some other speculative prices. *Journal of Business* **40**, 393–413.

[13] Mandelbrot, Benoit B. and Taylor, H.M. (1967) On the distribution of stock price differences. *Operations Research* **15**, 1057–1062.

[14] Pickands III, J. (1975) Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119–131.

[15] Schoutens, W. (2003) *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley.