A comparison of some univariate models for Value-at-Risk and expected shortfall

C. Marinelli
Institut für Angewandte Mathematik, Universität Bonn
Wegelerstr. 6, D-53115 Bonn, Germany

S. d'Addona
Department of International Studies, University of Rome III
Via G. Chiabrera 199, 00145 Rome, Italy

S. T. Rachev
School of Economics and Business Engineering, University of Karlsruhe
Kollegium am Schloss, Bau II, 20.12, R210 D-76128 Karlsruhe, Germany

and

Department of Statistics and Applied Probability
University of California, Santa Barbara, CA 93106, USA

March 15, 2006

Abstract

We compare in a backtesting study the performance of univariate models for Value-at-Risk (VaR) and expected shortfall based on stable laws and on extreme value theory (EVT). Analyzing these different approaches, we are able to test if the sum–stability assumption or the max–stability assumption, that respectively imply α–stable laws and Generalized Extreme Value (GEV) distributions, is more suitable for risk management based on VaR and expected shortfall. Our numerical results indicate that α–stable laws outperform EVT-based risk measures, especially those obtained by the so-called block maxima method.

1 Introduction

This work focuses on the investigation of the predictive power of Value-at-Risk and expected shortfall based on the assumption of Paretian stable returns, comparing their performances with corresponding measures based on the assumption of Gaussian returns as well as on the Extreme Value Theory (EVT). In particular we study the empirical performances of two fully parametric approaches, assuming that returns follow a Gaussian law or an α-stable law, and of some semi-parametric approaches based on limit theorems for maxima of sequences of independent random variables. We also consider, mainly as a benchmark case, a fully non-parametric approach based on empirical processes, which corresponds to the so called historical simulation method.

In the literature, Value-at-Risk (VaR) is commonly accepted as the standard measure of market risk and indicates, in percentage terms, the maximum probable loss on a given portfolio, referring to a specific confidence interval and time horizon. Historically the VaR literature has evolved following both the parametric and the non-parametric (see [17] for a complete historical account and list of references). While in the latter case the probability distribution of future returns is "simulated from the past" in order to estimate the relevant quantile (i.e. the VaR), the parametric approach is based on fitting a certain family of probability laws to observed historical returns.

In the parametric approach the most widely adopted hypothesis is the conditional or unconditional normality of returns (see e.g. [6] for a comprehensive overview). This assumption is motivated by the conception that returns are the outcome of a large number of "microscopic" effects. Hence, the central limit theorem (CLT) provides a theoretically sound argument in favor of Gaussian distribution. The normality assumption, along with the hypothesis of linearity of portfolio returns with respect to the considered risk factors, implies a normal distribution for portfolio returns. Consequently, it is possible to describe the returns' distribution simply with the first two moments, hence VaR can be calculated using the corresponding quantile of a standard Gaussian law.

Even if the normality of returns is intuitively very appealing, its drawbacks are well known in literature. In fact, several empirical studies have shown that financial returns exhibit features like high kurtosis and skewness that are incompatible with the normality assumption (see [11], [10] and [2] among others).

A natural approach to overcome these inconsistencies is to assume that returns follow a stable law, thus saving the CLT argument and explaining heavy tails and asymmetries (a complete account of stable distributions in finance is given in [26]). In particular, stable laws arise as the only possible weak limits of properly normalized sums of i.i.d. random variables, they are heavy tailed (except in the Gaussian subcase), and can exhibit skewness (see e.g. [27]). Moreover, univariate stable models have the potential to provide more realistic estimates of the frequency of large price movements, and therefore they seem preferable to classical models based on the assumption of normally distributed returns (for related discussions see e.g. [16], [13] and [18]).

In the last 10 years there has been intense activity in the application of ideas of extreme value theory to risk management. Roughly speaking, this method is an application of another stability scheme: as α-stable laws are the only laws appearing as (weak) limits of sum of i.i.d. random variables and are stable (better said, closed) with respect to summation, GEV laws are the only weak limits with respect to the operation of pairwise maximum, and they are closed with respect to this operation. In other words, denoting by $\circ$ a binary operation, and writing

$$aX_1 \circ bX_2 \stackrel{d}{=} cX, \tag{1}$$

where X_1, X_2 are i.i.d. copies of X, then (1) defines, respectively, α stable laws when $\circ = +$, and max-stable laws (or equivalently GEV laws) when $x \circ y = \max(x, y)$. One could say that EVT-based methods are semi-parametric, as they do not require a precise parametric assumption on the returns distribution, but they still need fitting procedures for quantities such as block maxima or exceedances over a threshold.

Our contribution is a rather extensive comparison in terms of a backtesting procedure of the two alternative stability scheme described above. Our work is closely related with [22] where an Extreme Value VaR is proposed and tested, and with [14] where the properties of an Extreme Value risk estimator are assessed. However, while the former focuses only on a EVT method, the latter does not provide any information about the out-of-sample (backtesting) performance of the analyzed model.

We also contribute some results about the estimators of VaR and expected shortfall in the stable and EVT framework, computing their (asymptotic) confidence intervals.

Let us introduce some notation and conventions used throughout the paper: vectors will always be column vectors, and v^* denotes the transpose of the vector or matrix v. We shall write $X \sim \eta$ to mean that the law of the random variable X is the (probability) measure η, and $X_n \Rightarrow X$ to mean that the sequence of random variables $(X_n)_{n\in\mathbb{N}}$ converges weakly to X. $N(\mu, \sigma^2)$ denotes the law of a Gaussian random variable with mean μ and variance σ^2. The law of a χ^2 random variable with n degrees of freedom will be denoted by χ^2_n. For $r \in [0, 1]$ we denote by z_r and $\nu_{n,r}$ the r-quantiles of the laws $N(0, 1)$ and χ^2_n, respectively. We shall always denote by $X \sim F$ the random variable of *negative* returns of a financial position. Then Value-at-Risk at confidence level p for our financial position is defined as the p quantile of the distribution F, i.e.

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : F(x) \geq p\}. \tag{2}$$

Since in all cases of interest (in this paper) we deal with random variables X with continuous distribution F, (2) reduces to $\mathrm{VaR}_p(X) = F^{-1}(p)$. Typical choices of p are $p \in \{0.9, 0.95, 0.99\}$.
We shall also assume throughout that the observed (negative) returns X_i, $i = 1, \dots, n$ form an i.i.d. sample from the law F.

The remainder of the paper is organized as follows: section 2 recalls how to compute VaR in a standard univariate Gaussian setting and using only past observation (historical simulation). Asymptotic confidence intervals are obtained in both cases. Sections 3 and 4 derive stable and EVT VaR measures, respectively, together with their asymptotic confidence intervals. Section 5 is devoted to the study of expected shortfall, a risk measure that enjoys better properties than VaR (in particular it is subadditive). All models are empirically tested in section 6. Section 7 concludes.

2 Benchmark VaR

In this section we find point estimates and confidence intervals (sometimes asymptotic, i.e. for n large) for $\mathrm{VaR}_p(X)$ that will be used as benchmark measures for the estimators introduced in the following sections. In particular, we study estimators of VaR based on the Gaussian assumptions and on empirical quantiles.

2.1 Normal VaR

If $X \sim N(0, \sigma^2)$ (we assume, as is commonly done for purposes of VaR estimation, $\mu = 0$), then one has

$$\mathrm{VaR}_p(X) = \sigma z_p,$$

as it immediately follows by well known scaling properties of Gaussian measures. The problem is thus reduced to estimating σ, which can be done as

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2,$$

where, as usual, X_i, $i = 1, \ldots, n$ are i.i.d. random variables with law $F = N(0, \sigma^2)$. It is well known that

$$V := (n-1)\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1},$$

hence the $1-r$ confidence interval for σ^2 is given by

$$\left[\frac{(n-1)\hat{\sigma}_n^2}{\nu_{n-1,1-r/2}}, \frac{(n-1)\hat{\sigma}_n^2}{\nu_{n-1,r/2}}\right]. \tag{3}$$

It is now immediate to obtain confidence intervals for σ, and hence for VaR. However, it is well known that confidence intervals obtained through χ^2 distributions are very sensitive with respect to the normality assumption. A more robust alternative is given by the asymptotic confidence interval that can be obtained by the limiting relation

$$\sqrt{n}(S_n^2 - \sigma^2) \Rightarrow N(0, \mu_4 - \sigma^4), \tag{4}$$

where $S_n^2 := n^{-1}\sum_{i=1}^n X_i^2$ is the sample variance and $\mu_k := \mathbb{E}X^k$. In order to apply (4), which can be easily proved by a direct calculation based on the central limit theorem, one needs to assume $\mathbb{E}X_i^4 < \infty$. An asymptotic confidence interval for σ^2 can now be obtained from (4) as

$$\sigma^2 = S_n^2 \pm \sqrt{\frac{\mu_4 - \sigma^4}{n}}\, z_{r/2}. \tag{5}$$

In order to make this confidence interval operational, we need to replace in (5) σ^4 and μ_4 with consistent estimators. Then, in view of Slutsky's theorem, (5) will still yield asymptotic confidence intervals at level $1-r$. Assuming $\mathbb{E}X^4 < \infty$, μ_4 and σ^4 are consistently estimated by $n^{-1}\sum_{i=1}^n X_i^4$ and $(S_n^2)^2$, respectively.

We shall use confidence intervals for Gaussian VaR derived from both (3) and (5).

2.2 VaR and empirical processes

Let $\mathbb{F}_n$ denote the empirical process of the observed negative returns $X_1, \ldots, X_n$, that is

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \le t),$$

where the X_i are i.i.d. with (unknown) distribution F, and $\mathbb{I}(A)$ stands for the indicator function of the event A. The Glivenko-Cantelli theorem ensures that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F(x)| = 0 \quad \text{a.s.}$$

This suggests that the p quantile $F^{-1}(p)$ can be estimated by

$$\mathbb{F}_n^{-1}(p) = X_{n(i)}, \quad p \in \left(\frac{i-1}{n}, \frac{i}{n}\right],$$

where $X_{n(1)} \le X_{n(2)} \le \ldots \le X_{n(n)}$ are the order statistics.

The asymptotic properties of this estimator are collected in the following proposition, whose proof can be found, e.g., in [31]. The derivative of F, whenever it exists, will be denoted by f.

Proposition 1 *Let $p \in]0,1[$, and assume that F is continuously differentiable at $F^{-1}(p)$, with $f(F^{-1}(p)) > 0$. Then*

$$\sqrt{n}\Big(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\Big) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\mathbb{I}(X_i \le F^{-1}(p)) - p}{f(F^{-1}(p))} + o_P(1),$$

and

$$\sqrt{n}\Big(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\Big) \Rightarrow N\Big(0, \frac{p(1-p)}{f^2(F^{-1}(p))}\Big). \tag{6}$$

Moreover, if $F \in C^1([a,b])$, with $a := F^{-1}(p_1) - \varepsilon$, $b := F^{-1}(p_2) + \varepsilon$ for some $\varepsilon > 0$, and $F'(x) > 0$ for all $x \in [a,b]$, then

$$\sqrt{n}\Big(\mathbb{F}_n^{-1} - F^{-1}\Big) \Rightarrow \frac{B_0}{f(F^{-1}(p))}$$

in $\ell^\infty([a,b])$, where B_0 is a standard Brownian bridge.

If $f^2(F^{-1}(p))$ is known explicitly, or at least can be approximated with a good level of accuracy, then one can obtain confidence intervals from (6). If that is not possible, then the following alternative procedure can be used: let $X_1, \ldots, X_n$ be a random sample from F, and define $U_i = F(X_i)$, so that U_i are independent uniform random variables. Then one has

$$\mathbb{P}\Big(X_{n(k)} < F^{-1}(p) \le X_{n(\ell)}\Big) = \mathbb{P}\Big(U_{n(k)} < p \le U_{n(\ell)}\Big).$$

Choosing k and ℓ so that

$$\frac{k}{n} = p - z_{r/2}\sqrt{\frac{p(1-p)}{n}}$$

and

$$\frac{\ell}{n} = p + z_{r/2}\sqrt{\frac{p(1-p)}{n}},$$

since the events $\{U_{n(k)} < p \le U_{n(\ell)}\}$ and $\{\sqrt{n}\,|G_n^{-1}(p) - p| \le z_{r/2}\sqrt{p(1-p)}\}$ are asymptotically equivalent, then

$$\lim_{n\to\infty} \mathbb{P}\Big(U_{n(k)} < p \le U_{n(\ell)}\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\sqrt{n}\,|G_n^{-1}(p) - p| \le z_{r/2}\sqrt{p(1-p)}\Big) = 1 - r,$$

where G_n^{-1} is the quantile process of the uniform distribution.

3 Stable modelling of VaR

Let us recall that the law of a one-dimensional stable random variable X is explicitly characterized through its characteristic function $\psi(t) = \mathbb{E}e^{itX}$, which can be written as

$$\log \psi(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \Big(1 - i\beta \operatorname{sgn}(t) \tan \frac{\pi\alpha}{2}\Big) + i\mu t & \text{if } \alpha \neq 1 \\ -\sigma |t| \Big(1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \log |t|\Big) + i\mu t & \text{if } \alpha = 1. \end{cases}$$

The parameter $\alpha \in]0, 2]$ is an index of tail thickness, $\beta \in [-1, 1]$ measures skeweness, $\sigma > 0$ and $\mu \in \mathbb{R}$ are scale and location parameters, respectively. The law of a stable random variable will be denoted by $S_\alpha(\sigma, \beta, \mu)$, with obvious meaning of the notation. Note that the characteristic function of a centered (i.e. with $\mu = 0$) symmetric stable law takes the particularly simple form $e^{-\sigma^\alpha |t|^\alpha}$. Moreover, the following scaling and shift property holds: $(X - \mu)/\sigma \sim S_\alpha(1, \beta, 0)$. Although not known in closed form for general parameters, stable laws admit C^∞ density functions (see [27]), which we shall denote by $p(\cdot; \alpha, \beta, \sigma, \mu)$. From a computational point of view, they can be efficiently approximated by numerically inverting the characteristic function, e.g. by numerical integration or by Fast Fourier Transform (see e.g. [24], [23]).

The parameters of a stable law can be fitted to data by maximum likelihood. In particular, setting $\theta = (\alpha, \beta, \sigma, \mu)$, and

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \prod_{k=1}^{n} p(x_k; \alpha, \beta, \sigma, \mu),$$

one has that $\hat{\theta}_n$ is a consistent and asymptotically normal estimator of θ, with

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, J_\theta^{-1}), \tag{7}$$

where $\Theta =]1, 2] \times [-1, -1] \times \mathbb{R}_+ \times \mathbb{R}$ and J_θ is the Fisher information matrix, i.e.

$$J_\theta = \mathbb{E}\left[\nabla_\theta \ell(X; \theta)(\nabla_\theta \ell(X; \theta))^*\right],$$

where $\ell(x; \theta) = \log p(x; \theta)$. For proofs of the above statements we refer to [7]. Computationally, one obtains an initial estimate of θ, using e.g. the quantile-based method of [21], and uses it as starting point for a constrained numerical optimization of the (log) likelihood function.

An interesting alternative is the characteristic function-based method used in [19], where the fit in the tails is particularly emphasized with a very fine sampling of the characteristic function in a neighborhood of the origin (for theoretical properties of this class of estimators see e.g. [25]).

In order to derive (asymptotic) confidence intervals for stable VaR, let us denote by g the following function:

$$\begin{aligned} g_p : \operatorname{int}(\Theta) &\to \mathbb{R} \\ \theta &\mapsto F^{-1}(p; \theta), \end{aligned}$$

where F stands for the distribution function of a $S_\alpha(\sigma, \beta, \mu)$ random variable, $\theta = (\alpha, \beta, \sigma, \mu)$, and p is the (fixed) quantile of interest, e.g. $p = 0.95$ or $p = 0.99$. Since $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, J_\theta^{-1})$, an application of the delta method leads to

$$\sqrt{n}(\widehat{\mathrm{VaR}}_n - \mathrm{VaR}) \Rightarrow N(0, (\nabla g_p(\theta))^* J_\theta^{-1} \nabla g_p(\theta)),$$

where $\mathrm{VaR} := g(\theta)$, and $\widehat{\mathrm{VaR}}_n := g_p(\hat{\theta})$. Applying Slutsky's lemma, one obtains the following asymptotic confidence interval at level $1 - r$:

$$\mathrm{VaR} = \widehat{\mathrm{VaR}}_n \pm z_{r/2} \frac{\sqrt{(\nabla g_p(\hat{\theta}_n))^* J_{\hat{\theta}_n}^{-1} \nabla g_p(\hat{\theta}_n)}}{\sqrt{n}} \tag{8}$$

The argument leading to (8) is of course only formal, but it becomes rigorous if we can prove that g_p is differentiable at θ.

Proposition 2 *Assume that $\theta_0 = (\alpha_0, \beta_0, \sigma_0, \mu_0)$ is such that $1 < \alpha_0 < 2$ and $-1 < \beta_0 < 1$. Then g_p is continuously differentiable at θ_0.*

Proof. Let us assume for now that $\sigma = 1$ and $\mu = 0$, and let $X \sim S_\alpha(1, \beta, 0)$. Then one has

$$\psi(t; \alpha, \beta) = \exp\Big(- |t|^\alpha (1 - i\beta(\operatorname{sgn} t) \tan \frac{\pi\alpha}{2})\Big), \tag{9}$$

and

$$p(x; \alpha, \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} \psi(t; \alpha, \beta) e^{-itx}\, dt.$$

Differentiating with respect to α and β, respectively, in the last expression, and interchanging the order of integration and differentiation, one has

$$\partial_\alpha p(x; \alpha, \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} \partial_\alpha \psi(t; \alpha, \beta) e^{-itx}\, dt,$$

and similarly

$$\partial_\beta p(x; \alpha, \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} \partial_\beta \psi(t; \alpha, \beta) e^{-itx}\, dt.$$

Using the explicit expression for the characteristic function (9), and recalling that stable density functions are C^∞ with respect to x, one has that $p \in C^{1,1}(\mathbb{R} \times G)$, where $G = (1, 2) \times (-1, 1)$. This in turns implies that $F \in C^{1,1}(\mathbb{R} \times G)$, since $F(x; \alpha, \beta) = \int_{-\infty}^x p(y; \alpha, \beta)\, dy$. Recalling that one has, by well-known scaling properties of stable laws,

$$F(x; \alpha, \beta, \sigma, \mu) = \sigma F(x; \alpha, \beta) + \mu,$$

we also get $F \in C^{1,1}(\mathbb{R} \times H)$, where $H = (1, 2) \times (-1, 1) \times \mathbb{R}_+ \times \mathbb{R}$.
Let us now define the function $\Phi : \mathbb{R} \times H \to \mathbb{R}^5$, $\Phi : (x, \theta) \mapsto (F(x; \theta), \theta)$. It is immediately seen that the Jacobian of Φ in a neighborhood U of (x, θ), with $F(x; \theta) = p$, for a given fixed p, is of the form

$$D\Phi(x, \theta) = \left[\begin{array}{c|cccc} p(x, \theta) & 0 & 0 & 0 & 0 \\ \hline * & 1 & & & \\ * & & 1 & & \\ * & & & 1 & \\ * & & & & 1 \end{array} \right],$$

hence $\det D\Phi(x,\theta) \neq 0$: in fact, density functions of stable laws are positive on the whole real line whenever $\alpha > 1$. Therefore Φ is a C^1 diffeomorphism on U, in particular $\theta \mapsto F^{-1}(p,\theta)$ is of class C^1 for any fixed finite p. This is equivalent to the claim that g_p is continuously differentiable at θ_0. □

4 VaR estimates based on Extreme Value Theory

The rationale behind the extreme value theory approach is essentially contained in two theorems, due in their present form to Gnedenko [15] and to Balkema and de Haan [1]. Here we recall only the statements of the two theorems, and we describe what consequences are usually derived from them for the purposes of estimating VaR.

Theorem 3 (Gnedenko) *Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution function F. If there exist a positive sequence $\{a_n\}_{n\in\mathbb{N}}$ and a real sequence $\{b_n\}_{n\in\mathbb{N}}$ such that*

$$\frac{\max(X_1,\ldots,X_n) - b_n}{a_n} \Rightarrow Y \tag{10}$$

as $n \to \infty$ and Y is nondegenerate, then the law of Y is of the generalized extreme value (GEV) type, i.e. its distribution function H is given by

$$H(x) = \exp\Big(- \Big(1 + \xi \frac{x-\mu}{\sigma}\Big)_+^{-1/\xi}\Big). \tag{11}$$

In (11) μ and σ are location and scale parameters, and ξ determines the shape of the distribution: the GEV laws with $\xi > 0$ and $\xi < 0$ correspond to the Fréchet and Weibull distributions respectively, while the case $\xi = 0$ has to be interpreted in the limit $\xi \to 0$ and corresponds to the Gumbel law, i.e $H(x) = \exp\big(- \exp\big(\frac{x-\mu}{\sigma}\big)_+\big)$.

We say that a distribution F is in the max-domain of attraction a GEV law H (in symbols, $F \in D_m(H)$) if it satisfies the hypotheses of theorem 3.

Appealing to theorem 3, at least two ways have been proposed in the literature to estimate high quantiles of probability distributions. In particular, one divides a sample $X_1, X_2, \ldots$ in "blocks" of a given size, say k, and sets

$$\begin{aligned} Y_1 &= \max(X_1, X_2, \ldots, X_k) \\ Y_2 &= \max(X_{k+1}, X_{k+2}, \ldots, X_{2k}) \\ \vdots\ & \quad \vdots \end{aligned}$$

Then, by assuming that the distribution of the block maxima (Y_i) is approximately GEV, one fits a law like (11) to the (Y_i) and computes "block VaR". The procedure is described in detail in subsection 4.1. A refinement of the above procedure consists in relaxing the assumption that block maxima are GEV distributed, assuming instead that block maxima are only in the max-domain of attraction of a GEV law. Such potentially more general algorithms are described in subsection 4.2.

As already mentioned in the introduction, α-stable and GEV laws can be seen as the fixed points of two alternative stability schemes, namely they are the only sum-stable and max-stable laws, respectively (or equivalently, they are the only laws arising as weak limits of normalized sums and maxima, respectively, of i.i.d. random variables). It appears therefore natural to try to compare the performance of the two stability schemes. Since the max-stability scheme does not seem to be useful to estimate daily VaR, we shall backtest "block VaR" models based on the three following assumptions:

- $Y_1, Y_2, \ldots$ are max-stable, i.e. they are GEV distributed;
- $Y_1, Y_2, \ldots$ are in the domain of attraction of a max-stable law;
- $Z_1, Z_2, \ldots$ are sum-stable, i.e. α-stable,

where $Z_1 = X_1 + X_2 + \cdots + X_k$, etc.

Another procedure to estimate VaR (that does not need to divide observations into blocks, and therefore works for daily VaR as well) is based on the following theorem, characterizing the limit distribution of excesses over a threshold of a sequence of i.i.d. random variables.

Theorem 4 (Balkema and de Haan) *Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution function F. Assume that the support of F is $\mathbb{R}$ and that $F \in D_m(H)$, with H max-stable. Then there exists a function $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\lim_{u \uparrow \infty} \sup_{0 \le x \le \infty} \Big| F_u(x) - G_{\xi,\sigma(u)}(x) \Big| = 0,$$

where $F_u(x) = \mathbb{P}(X - u \le x | X > u)$ and $G_{\xi,\sigma}$ is the generalized Pareto distribution:

$$G_{\xi,\sigma}(x) = 1 - \Big(1 + \xi \frac{x}{\sigma}\Big)_+^{-1/\xi}.$$

The method relying on this theorem, sometimes called Peaks over Thresholds (POT) method, is described in subsection 4.4. A natural term of comparison for this method will be the plain assumption of α-stable distributed (daily) returns.

4.1 VaR with max-stable block maxima

Let us define block maxima as follows:

$$Y_k = \max(X_{km}, X_{km+1}, \ldots, X_{k(m+1)-1}),$$

where m is the block size (m could correspond, for instance, to the typical number of trading days in a month or a year). The aim is to obtain estimates of quantiles y_p of Y such that $\mathbb{P}(Y > y_p) = p$, together with their confidence intervals.

Recall, however, that in this case one obtains a VaR estimate over the period covered by the block size. For instance, if the block size corresponds to one year, then we obtain a VaR estimate for annual returns.

Let $\theta = (\xi, \mu, \sigma)$ and $h(\cdot;\theta)$ the density of $H(\cdot;\theta)$. The maximum likelihood estimate $\hat{\theta}_n$ based on the observations $Y_1, \ldots, Y_n$ is given by

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L(\theta),$$

where

$$L(\theta) = \prod_{i=1}^{n} h(Y_i;\theta)\mathbb{I}(1+\xi(Y_i-\mu)/\sigma > 0),$$

$$h(x;\theta) = \frac{1}{\sigma}H(x;\theta)\Big(1+\xi\frac{x-\mu}{\sigma}\Big)^{-1-\frac{1}{\xi}},$$

and $\Theta = \mathbb{R}\times\mathbb{R}\times\mathbb{R}_+$. There are no closed-form expressions for $\hat{\theta}_n$, but the availability of numerical optimization routines renders the task quite simple.
The following result guarantees that in most interesting cases this estimator has good properties (for the proof see [28]).

Proposition 5 *If $\xi > -1/2$ then $\hat{\theta}_n$ is a consistent, asymptotically normal and efficient estimator of θ.*

As it follows from (11), VaR at p level can be estimated as

$$\widehat{\mathrm{VaR}}_n = g_p(\hat{\theta}_n) := \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\Big(1-(-\log p)^{-\hat{\xi}}\Big).$$

In order to obtain confidence intervals for VaR we apply again the delta method, in complete similarity to section 3. In particular one has

$$\mathrm{VaR} = \widehat{\mathrm{VaR}}_n \pm z_{r/2}\frac{\sqrt{(\nabla g(\hat{\theta}_n))^* J_{\hat{\theta}_n}^{-1}\nabla g(\hat{\theta}_n)}}{\sqrt{n}}, \tag{12}$$

where $J_{\hat{\theta}_n}$ is the empirical Fisher information matrix relative to the maximum likelihood estimate $\hat{\theta}_n$. Note that in this case we have an explicit expression for g_p, hence the situation is simpler than in the stable case. The limiting case $\xi = 0$, as observed before, has to be treated separately.

Let us also mention that some estimation procedures for tails and quantiles have been proposed under the assumption that observations are only in the domain of attraction of a max-stable law. Here we limit ourselves to report two algorithms described in [9], to which we refer for further details. The empirical tests will not use the following procedures.

The first method works as follows:

- Estimate the tail parameter ξ through the Hill estimator, i.e.

$$\hat{\xi} = \frac{1}{k}\sum_{j=1}^{k}\log Y_{j,n} - \log Y_{k,n},$$

  where $Y_{n,n} \le Y_{n-1,n} \le \cdots \le Y_{1,n}$ are the order statistics and k is a number to be chosen.

- The tail of the distribution is estimated as

$$\mathbb{P}(Y > x) = \frac{k}{n}\Big(\frac{x}{X_{k+1,n}}\Big)^{-1/\hat{\xi}},$$

and the quantile x_p is estimated as

$$\hat{x}_p = \Big(\frac{n}{k}(1-p)\Big)^{-\hat{\xi}} X_{k+1,n}.$$

The following theorem, due to Dekkers and de Haan ([3]), describes the asymptotic behavior of the quantile estimator and can be used to obtain asymptotic confidence intervals for block VaR.

Theorem 6 *Let $Y_1, \ldots, Y_n$ be i.i.d. with distribution function $F \in D_m(H_\xi)$ with $\xi > 0$. Assume moreover that F has a positive density f of regular variation of order $-1-1/\xi$. Set $p = p_n$, $k = k_n = [n(1-p_n)]$ where $[\cdot]$ denotes integer part. Assume $p_n \to 1$ and $n(1-p_n) \to \infty$. Then*

$$\sqrt{2k}\frac{x_p - Y_{k,n}}{Y_{k,n} - Y_{2k,n}} \Rightarrow N\Big(0, \frac{2^{2\xi+1}\xi^2}{(2^\xi - 1)^2}\Big).$$

Another estimator proposed in [9] (and introduced in [3]) is

$$\hat{x}_p = Y_{k,n} + (Y_{k,n} - Y_{2k,n})\frac{\Big(\frac{k}{n(1-p)}\Big)^{-\hat{\xi}} - 1}{1 - 2^{-\hat{\xi}}}.$$

The following theorem on asymptotic normality of the estimator (see [3]) allows one to construct asymptotic confidence intervals for VaR.

Theorem 7 *Under the same hypotheses of the previous theorem on $X_1, \ldots, X_n$ and on F, assume $n(1-p) \to c$, $c > 0$ fixed. Let $\hat{\xi}$ be the Pickands estimator. Then for every fixed $k > c$ one has*

$$\frac{\hat{x}_p - x_p}{Y_{k,n} - Y_{2k,n}} \Rightarrow \eta,$$

where

$$\eta = \frac{(k/c)^\xi - 2^{-\xi}}{1 - 2^{-\xi}} + \frac{1 - (Q_k/c)^\xi}{e^{\xi H_k} - 1},$$

and the random variables H_k, Q_k are independent, Q_k are exponentially distributed with parameter $2k+1$, and $H_k = \sum_{j=k+1}^{2k} j^{-1}E_j$, with E_i, $i = 1, 2, \ldots$ i.i.d. standard exponentials.

4.2 Exceedences over a threshold

Let u be a fixed threshold and define the conditional distribution of excesses

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u).$$

Then one has

$$F_u(x) = \frac{\mathbb{P}(\{X \leq u + x\} \cap \{X > u\})}{\mathbb{P}(X > u)},$$

hence

$$F(x) = (1 - F(u))F_u(u + x) + F(u).$$

Appealing to theorem 4, one approximates in the previous expression $F_u(u + x)$ by a generalized Pareto distribution $G(x)$ and $F(u)$ by the empirical distribution function at u, i.e. by $1 - n_u/n$, where n_u is the number of observation above the threshold u, getting

$$F(x) \approx 1 - \frac{n_u}{n}\Big(1 + \frac{\xi}{\sigma}(x-u)\Big)^{-1/\xi},$$

from which VaR can be estimated as

$$\widehat{\mathrm{VaR}}_n = g_p(\hat{\theta}_n) := u + \frac{\hat{\sigma}_n}{\hat{\xi}_n}\Big((n(1-p)/n_u)^{-\hat{\xi}_n} - 1\Big).$$

The estimates $\hat{\theta}_n = (\hat{\xi}_n, \hat{\sigma}_n)$ of the parameter vector appearing in the previous formula can be obtained by fitting a generalized Pareto distribution (GPD) to the portion of the data that exceeds the threshold u. Once u has been chosen, then we use maximum likelihood estimation, which is straightforward as the density of GPD is known in closed form.

Let us briefly remark that there is no general rule to optimally select the threshold u. This choice is nonetheless very important, as for u too high the estimator has high variance, and for u too small the estimator becomes biased. In our empirical tests we follow [22] in choosing a random threshold that pick the top 10% of the analyzed data.

Asymptotic approximate confidence intervals for VaR can again be obtained by an argument based on the delta method. In fact, assuming that all negative returns over the threshold u are drawn from a generalized Pareto law, we have (see [29])

$$\sqrt{n_u}(\hat{\theta}_{n_u} - \theta) \Rightarrow N(0, J_\theta^{-1}), \tag{13}$$

provided $\xi > -1/2$, with

$$J_\theta^{-1} = \begin{bmatrix} 2\sigma^2(1+\xi) & \sigma(1+\xi) \\ \sigma(1+\xi) & 1+\xi \end{bmatrix}.$$

We can now write

$$\mathrm{VaR} = \widehat{\mathrm{VaR}}_n \pm z_{r/2}\frac{\sqrt{\big(\nabla g_p(\hat{\theta}_{n_u})\big)^* J_{\hat{\theta}_{n_u}}^{-1}\nabla g_p(\hat{\theta}_{n_u})}}{\sqrt{n_u}}. \tag{14}$$

In the above expression we compute ∇g_p by considering u a constant, even though in practice this is not true. In this sense the confidence intervals obtained in this way are only approximate. Let us mention, however, that there are more refined asymptotic normality results similar to (13) when u is a random threshold – see e.g. [5] and [4]. The asymptotic covariance matrices obtained by these authors seem unfortunately quite difficult to implement.

5 Expected shortfall

Denoting by X the negative return of our financial position, we define as expected shortfall at level p the quantity

$$\mathrm{ES}_p = \mathbb{E}[X | X > \mathrm{VaR}_p(X)].$$

We shall use the shorthand notation $y_p := \mathrm{ES}_p(X)$. Recall that expected shortfall is, under very mild assumptions, the smallest convex measure of risk that dominates Value-at-Risk (see e.g. [12]). Although it is well known that VaR is not a coherent measure of risk, it is subadditive when restricted to elliptic distributions (among which Gaussian and stable laws).

5.1 Empirical shortfall

The following approximation is straightforward:

$$\hat{y}_p = \frac{1}{|I|} \sum_{i \in I} X_i,$$

where I is the set of i such that $X_i > \widehat{\mathrm{VaR}}_p(X)$, and $|I|$ its cardinality. Consistency of this estimator is guaranteed by the law of large numbers.

5.2 Gaussian shortfall

When X is a Gaussian random variable, a simple closed form expression has been obtained in [30]. In particular, if $X \sim N(0, 2)$, the expected shortfall at level p is given by

$$\mathrm{ES}_p(X) = \frac{1}{(1-p)\sqrt{\pi}} \exp\Big(\frac{-(\mathrm{VaR}_p(X))^2}{4}\Big).$$

In the general case $X' \sim N(\mu, \sigma)$ one has

$$\mathrm{ES}_p(X') = \frac{\sigma}{\sqrt{2}} \mathrm{ES}_p(X) + \mu,$$

as follows from well known scaling properties of Gaussian laws.

Assuming $\mu = 0$, recalling that $\mathrm{VaR}_p(X) = \sqrt{2} z_p$ for $X \sim N(0, 2)$, the confidence interval for $\mathrm{ES}_p(X)$ for general $X \sim N(0, \sigma)$ is given by

$$\left[\frac{e^{-z_p^2/2}}{(1-p)\sqrt{2\pi}} \sigma_-, \frac{e^{-z_p^2/2}}{(1-p)\sqrt{2\pi}} \sigma_+ \right],$$

where $[\sigma_-, \sigma_+]$ is the confidence interval for σ (see section 2).

5.3 Stable expected shortfall

For X is α-stable there exists an integral representation of expected shortfall obtained in [30]. In particular, if $X \sim S_\alpha(1, \beta, 0)$, one has

$$\mathrm{ES}_p(X) = \frac{\alpha}{1-\alpha} \frac{|\mathrm{VaR}_p(X)|}{p\pi} \int_{-c}^{\pi/2} \phi(x) \exp(-|\mathrm{VaR}_p(X)|^{\frac{\alpha}{\alpha-1}} v(x))\, dx,$$

where

$$\phi(x) = \frac{\sin(\alpha(c+x)-2x)}{\sin(\alpha(c+x))} - \frac{\alpha\cos^2 x}{\sin^2(\alpha(c+x))},$$

$$v(x) = \cos^{\frac{1}{\alpha-1}}(\alpha c)\Big(\frac{\cos x}{\sin(\alpha(c+x))}\Big)^{\frac{\alpha}{\alpha-1}}\frac{\cos(\alpha c+(\alpha-1)x)}{\cos x},$$

$$c = \frac{1}{\alpha}\arctan\Big(-\operatorname{sgn}(\mathrm{VaR}_p(X))\beta\tan\frac{\pi\alpha}{2}\Big).$$

For general $X' \sim S_\alpha(\sigma,\beta,\mu)$, recall that $\sigma X+\mu \sim X'$, hence

$$\mathrm{ES}_p(X') = \sigma\,\mathrm{ES}_p(X) + \mu.$$

Asymptotic confidence intervals can be obtained again using the delta method. In particular, proposition 2 combined with some other tedious verifications show that the map $g_p^0 :]1,2[\times]-1,1[\to \mathbb{R}$, $g_p^0(\alpha,\beta) := \mathrm{ES}_p(X)$, $X \sim S_\alpha(1,\beta,0)$ is continuously differentiable. Therefore the map $g_p : \mathrm{int}\Theta \to \mathbb{R}$, $g_p(\alpha,\beta,\sigma,\mu) := \sigma g_p^0(\alpha,\beta)+\mu = \mathrm{ES}_p(X)$, $X \sim S_\alpha(\sigma,\beta,\mu)$, is also continuously differentiable. Finally, the delta method yields

$$\sqrt{n}(\widehat{\mathrm{ES}}_n - \mathrm{ES}) \Rightarrow N(0,(\nabla g_p(\theta))^* J_\theta^{-1}\nabla g_p(\theta)),$$

hence

$$\mathrm{ES} = \widehat{\mathrm{ES}}_n \pm z_{r/2}\frac{\sqrt{(\nabla g_p(\hat\theta_n))^* J_{\hat\theta_n}^{-1}\nabla g_p(\hat\theta_n)}}{\sqrt{n}},$$

where J_θ is the Fisher information matrix of (7).

5.4 EVT-based expected shortfall

Using the POT method one can easily derive a close form expression for the expected shortfall. In fact, if $Y \sim G_{\xi,\sigma}$, then one can verify that, for $\xi < 1$, $\sigma + \xi x > 0$,

$$\mathbb{E}[Y|Y>x] = \frac{x+\sigma}{1-\xi}. \tag{15}$$

Assuming that the distribution of $X-u$, conditional on $X>u$, is GPD, we obtain that the distribution of $X-x_p$, for $x_p>u$, conditional on $X>x_p$, is GPD with parameters ξ and $\sigma+\xi(x_p-u)$. Hence, using (15), one has

$$\mathrm{ES}_p(X) = \mathbb{E}[X|X>\mathrm{VaR}_p(X)] = \frac{\mathrm{VaR}_p(X)}{1-\xi} + \frac{\sigma-\xi u}{1-\xi}.$$

An estimator for $\mathrm{ES}_p(X)$ is therefore obtained by replacing in the previous expressions $\mathrm{VaR}_p(X)$, ξ, and σ with their respective estimators, which were all derived in subsection 4.4.

Asymptotic approximate confidence intervals for expected shortfall can again be obtained by the delta method. Details are omitted, as the relevant issues have already been discussed in previous sections. In particular, the main approximation is to consider the threshold u constant, while in practice it is random.

6 Empirical tests

In this section we present and describe the main empirical results obtained by testing the models introduced in the previous sections. For the empirical test we chose two stock indices, SP500 and NASDAQ, and two stocks, Amazon and Microsoft. All the raw prices are freely available on the web, and the returns are calculated as log-differences on daily data series. The sample periods span from 2-Jan-1990 to 31-Dec-2004 for the SP500 and from 2-Jan-1998 to 31-Dec-2004 for the other series.

In order to better understand the empirical exercise, it is worth looking briefly at the basic characteristics of the analyzed financial series. Table 1 presents, for each of the analyzed series, the first four moments of their distributions. From a preliminary analysis the leptokurtic nature of the returns' series is clearly revealed. In particular the SP500 index, with a kurtosis of 6.67 and a skewness of -0.105, strongly differs from a normal distribution especially in the thickness of the tails. In the same fashion, NASDAQ, Microsoft and Amazon all display clear evidence of fat tails in their distributions.

[Table 1 about here.]

This claim is confirmed by a more detailed analysis: in figure 1 we plot the third and the fourth moment, calculated on a rolling window of 250 data points. It is clear how the behavior of the kurtosis of all the series is far from the one expected for a Gaussian distribution (plotted as a straight line in the graph). In particular both Microsoft and Amazon display a long time span where the kurtosis is well above 6, with peaks of values above 8 for the first one. The same behavior is shown by the SP500, with a kurtosis well above 3 during the period 1990-1997, and peaks of values above 8 during the period 1997-1999. Interestingly enough the deviations from the normality by the kurtosis correspond to a comparable deviation by the skewness parameter (cf. Panel C-D of figure 1).

[Figure 1 about here.]

Having investigated the characteristics of the financial series, we can now turn to a comparative analysis of the VaR models proposed in the previous sections. In particular, we are interested in out-of-sample performances of the different VaR measures proposed. In order to assess them, we calculate for each specification two series of VaRs, with confidence interval of 95% and 99% respectively. All the risk measures are computed on a rolling window of 250 data-points. Subsequently a simple out-of-sample comparison is performed, by testing the VaR measures versus the next day returns. Some preliminary analysis on the estimations can be done by analyzing the time series behavior for the three different VaR measures. Generally speaking the estimations are in line with the empirical returns and present a remarkable level of accuracy in term of their estimation error. In order to make this analysis clearer we plot the last 12 months of estimations[1], along with the confidence intervals, for the Amazon time series, that has the higher historical volatility coupled with a high kurtosis (cf. table 1). The graphs in figures 2 and 3 show estimations that are comparable in magnitude for the three specifications,

[1] We choose to plot only the last year of data for a better readability of the graphs, after having investigated that the analysis in the text can be applied to the whole sample period.

both at a 95% and 99% confidence interval. Moreover, the VaRs based on the Stable assumption seems to have a greater accuracy, given their confidence intervals' tightness. In fact both the Extreme Value estimation based on the Peak over a Threshold approach (GPD), proposed in subsection 4.2, and the Gauss specifications, display a confidence interval of the order of $0.5\% - 1.5\%$ points for the 95% VaR, and $1\% - 2\%$ to peaks of 4% points for the99% VaR. On the contrary, the Stable confidence intervals are below the 0.6% point in both cases.

[Figure 2 about here.]

[Figure 3 about here.]

To further assess the accuracy of the calculated VaR, we perform a simple Proportion of Failure (POF) test, as e.g. in [20]. In particular we calculate:

$$LR = -\,2\log\left(\frac{p_0^x(1-p_0)^{(n-x)}}{p^x(1-p)^{(n-x)}}\right), \tag{16}$$

where p_0 is the probability of an exception implied by the chosen confidence interval, n is the sample size, x is the actual number of exceptions and p is the maximum-likelihood estimator x/n of p_0. Basically this test performs a likelihood-ratio with 5% level, based on the number of exceedences in any given sample, where the null hypothesis is that the estimated value for the exceedences matches its exact value. Given its definition, the test is asymptotically χ^2 distributed with one degree of freedom; thus if the value of the test statistic exceeds the critical value of 3.84, the VaR model can be considered as not reliable with a 95% confidence level.

Table 2 reports the results on the VaR backtesting exercise. Overall, the performance of the three models is good on all the analyzed series, nevertheless some differences can be noted. First the Stable VaR is relatively more accurate than the VaRs based on the Gauss and the GPD assumptions. In fact, while the former never present a LR statistics that exceed the critical value, the Gauss–VaR and the GPD–VaR both are rejected in two out of eight cases respectively. Second, the highest number of failures by GPD and Gauss estimations occurs with the Microsoft data series. This can be ascribed to the high kurtosis of the series, which is probably better captured by fitting a stable law. To further investigate this point, we plot in figure 4 the negative returns of the Microsoft series along with the 95% lower bound for the three models. It is clear that the worse performances pf the Gauss and GPD estimations are due to a more conservative VaR bound in both cases, clearly displayed in the 2000-2002 period for the Gauss estimation and in the 1999-2000 and 2003-2004 periods for the GPD.

[Table 2 about here.]

[Figure 4 about here.]

In the same fashion as the VaR backtesting procedures, we analyze the performance of the Expected Shortfall (ES) proposed is section 5. We calculate for each specification two series of ESs, with confidence interval of 95% and 99% respectively. All the risk measures are computed on a rolling window of 250 data-points. Subsequently a simple comparison test is performed.

Not surprisingly, the preliminary analysis on the estimations and their confidence intervals lead us to basically the same conclusions as in the VaR case. Figures 5 and 6 show estimations that are comparable in magnitude for the three specifications, both at a 95% and 99% confidence interval. Also, the ESs based on the stable assumption seem to have a greater accuracy, given their confidence intervals' tightness. Again the GPD and the Gauss specifications display a confidence interval of the order of $1\% - 3.5\%$ points for the 95% ES, and $1\% - 2.5\%$ for the 99% ES, with a peak of 11% for the GPD. On the contrary, the Stable confidence intervals are below the 0.6% point in both cases[2].

[Figure 5 about here.]

[Figure 6 about here.]

To backtest the ES forecasts, we follow [8] in calculating a measure evaluating the ES performance when returns are violating the corresponding VaR measure. In particular, we calculate the average difference between the realized returns and the forecasted ESs, conditional on having a return below the lower VaR bound[3].

Following [8]'s notation we have:

$$V_1^{ES} = \frac{\sum_{t=t_0}^{t_1} \left(R_t - (ES_t^p)\right) 1_{R_t > VaR_t^p}}{\sum_{t=t_0}^{t_1} 1_{R_t > VaR_t^p}}. \tag{17}$$

Given its specification, the lower the value of the test in absolute term is, the better is the ES estimate.

[Table 3 about here.]

Table 3 displays the result of the V_1 test. Clearly the Expected Shortfall measures estimated on the stock indices perform equally well in all the specified models. The main differences arise in the single stock evaluations; in particular the Stable specifications, both at a 95% and 99% confidence interval, seem to present less accuracy than the other two specifications. This difference is clearer in the Amazon returns' series, and can be ascribed to the less conservative nature of the Stable estimations.

6.1 Block maxima backtesting

Finally we also perform a comparative test on a VaR calculated with the block maxima method (BMM) introduced in subsection 4.1 versus the Gaussian and stable approaches. In practice we calculate the VaR, based on the BMM approach, for all the data series, at three different block sizes: 10, 15, and 20 trading days. In the same fashion we calculate the Gauss-based and the stable-based VaRs on 10, 15 and 20 days and we compare them with the corresponding realized returns.

[2] It is worth noting that these results are ascribable to the whole period analyzed both for the VaR and the ES.

[3] [8] also propose a measure based on the evaluation of values below a threshold calculated on the confidence interval. Given its intuitive definition, we prefer the measure presented in the text.

Results, reported in tables 4 to 7, are quite striking: in all analyzed series the BMM approach is largely “over-conservative”, producing VaR bounds that are difficult to interpret[4]. This lead the log-likelihood ratio test introduced above in rejecting strongly the model in all the data series and at all the confidence intervals. Interestingly enough, even if the results are strongly influenced by performing an estimation on not more than 20 data points, both the Gauss and the stable approaches produce a VaR bound more in line with the observed returns, with several cases where the model cannot be rejected with a 95% probability (cf. tables 4 to 7).

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

7 Conclusions

We have compared the properties of some univariate VaR models, in particular of α–stable and EVT–based models. We argue that comparing Stable and EVT VaR corresponds to testing which one of two stability assumptions performs better for VaR modeling: namely, we implicitly compare sum–stability and max–stability. The two stability schemes give rise, respectively, to α–stable laws and GEV distributions. Even though the EVT approach is quite appealing for its theoretical justification in terms of the theorems of Gnedenko and Balkema and de Haan, and because it applies to a large class of returns distributions, it suffers of several problems when applied in practice. For instance, using the POT approach it is necessary to choose a specific threshold. As noted above, there is no general rule to optimally select this threshold, but this choice is nonetheless very important. In particular, if the chosen threshold is “too high” the estimator has high variance, and if the chosen threshold is “too low” the estimator becomes biased. On the other hand, it seems that the stable procedure requires much less external input (hence it is significantly easier to implement in automated form). A second important issue is that EVT-based methods discard a lot of data, while stable-based ones use all of the data points in the time series. In essence, one could say that good fit of the law where there is more mass contributes to good fit in the tail, even though the EVT approach requires less distributional assumptions.

Our empirical analysis clearly show that α–stable laws outperform GEV distributions for estimating VaR. While GEV VaR gives conservative estimates at 5% level which are close to normal VaR, at 1% level the estimates becomes strongly “over-conservative”, with peaks that are somehow difficult to interpret.

Let us also remark, however, that empirical tests at ”extreme” quantiles (e.g. 99.5% or 99.9%) could be performed in order to asses the models’ behavior ”far out” in the tails of the distribution. Then we would expect EVT models to have a better performance,

[4]In particular, in all the performed estimations, there are several VaR points where the value reaches 150%, producing a bound that is not useful for an economic interpretation.

at least in the case of abundant data. However, we decided to focus on testing quantiles that are commonly used in financial risk management, both to compare our results with the existing literature and to assess the performance of models possibly used by practitioners. Nevertheless, such an analysis may be an interesting topic for future research.

Acknowledgements

S. T. Rachev gratefully acknowledges research support by grants from the Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the German Research Foundation (DFG) and the German Academic Exchange Service (DAAD).

References

[1] A. A. Balkema and L. de Haan. Residual life time at great age. *Ann. Probability*, 2:792–804, 1974.

[2] R. Blattberg and N. Gonedes. A comparison of the stable and student distributions as statistical models of stock prices. *J. of Business*, 47:244–280, 1974.

[3] A. L. M. Dekkers and L. de Haan. On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.*, 17(4):1795–1832, 1989.

[4] J. Diebolt, A. Guillou, and P. Ribereau. Asymptotic normality of the extreme quantile estimator based on the POT method. *C. R. Math. Acad. Sci. Paris*, 341(5):307–312, 2005.

[5] H. Drees, A. Ferreira, and L. de Haan. On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.*, 14(3):1179–1201, 2004.

[6] D. Duffie and J. Pan. An overview of value at risk. *J. of Deriv.*, 4:7–72, 1997.

[7] W. H. DuMouchel. On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Ann. Statist.*, 1:948–957, 1973.

[8] P. Embrechts, R. Kaufmann, and P. Patie. Strategic long-term financial risks: Single risk factors. *Computational Optim. Appl.*, 32:61–90, 2005.

[9] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.

[10] E. Fama. The behaviour of stock market prices. *J. of Business*, 38:34–105, 1965.

[11] E. Fama. Portfolio analysis in a stable paretian market. *Management Sci.*, 3:404–419, 1965.

[12] H. Föllmer and A. Schied. *Stochastic finance: An introduction in discrete time.* Walter de Gruyter & Co., Berlin, 2002.

[13] B. Gamrowski and S. T. Rachev. Financial Models Using Stable Laws. In Y. V. Prohorov, editor, *Applied and Industrial Mathematics*, pages 556–604. 1995.

[14] M. Gilli and E. Kellezi. An application of extreme value theory for measuring financial risk. *Working paper*, 2005.

[15] B. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44:423–453, 1943.

[16] D. W. Jansen and C. G. de Vries. On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *Review of Economics and Statistics*, **73**:18–24, 1991.

[17] P. Jorion. *Value at Risk: the New Benchmark for Controlling Market Risk*. Irwin, 2000.

[18] I. Khindanova, S. Rachev, and E. Schwartz. Stable modeling of value at risk. *Math. Comput. Modelling*, 34(9-11):1223–1259, 2001. Stable non-Gaussian models in finance and econometrics.

[19] I. Khindanova, S. T. Rachev, and E. Schwartz. Stable modeling of value at risk. *Math. Comput. Modelling*, 34(9-11):1223–1259, 2001.

[20] P. H. Kupiec. Techniques for verifying the accuracy of risk management models. *J. Deriv.*, 3:73–84, 1995.

[21] J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Commun. Statist. – Simula.*, **15**:1109–1136, 1986.

[22] A. J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empirical Fin.*, 7:271–300, 2000.

[23] S. Mittnik, S. T. Rachev, T. Doganoglu, and D. Chenyao. Maximum likelihood estimation of stable Paretian models. *Math. Comput. Modelling*, 29(10-12):275–293, 1999.

[24] J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Comm. Statist. Stochastic Models*, 13(4):759–774, 1997. Heavy tails and highly volatile phenomena.

[25] B. L. S. Prakasa Rao. *Asymptotic theory of statistical inference*. John Wiley & Sons Inc., New York, 1987.

[26] S. T. Rachev and S. Mittnik. *Stable Paretian Models in Finance*. John Wiley and Sons, NY, 2000.

[27] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.

[28] R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.

[29] Richard L. Smith. Estimating tails of probability distributions. *Ann. Statist.*, 15(3):1174–1207, 1987.

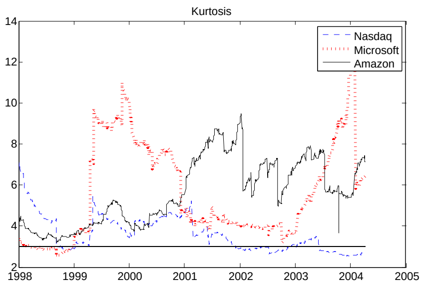
[30] S. Stoyanov, G. Samorodnitsky, S. Ortobelli, and S. T. Rachev. Computing the portfolio Conditional Value-at-Risk in the alpha-stable case. Technical report, Universität Karlsruhe, 2004.

[31] A. W. van der Vaart. *Asymptotic statistics.* Cambridge UP, Cambridge, 1998.

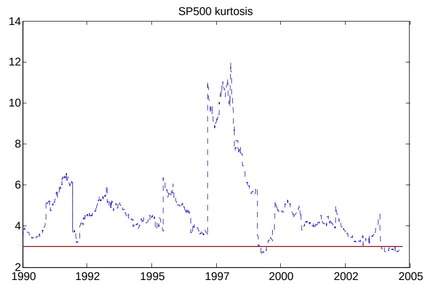
Figure 1: Time series of the kurtosis

This figure plots the skewness and the kurtosis of the analyzed series The moments of NASDAQ, Microsoft and Amazon are shown in Panel A and C respectively, while the ones of SP500 in Panel B and D. The moments are calculated on a rolling window of 250 daily data points. For the sake of comparison straight lines corresponding with a kurtosis of 3 and a skewness of 0 (i.e. for a normal distribution) are provided.

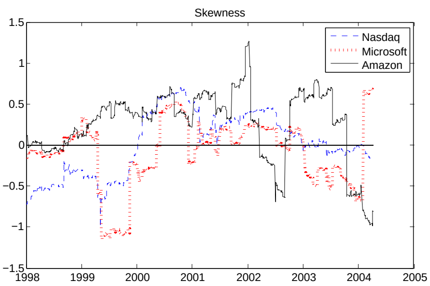
Panel A



Panel B



Panel C



Panel D

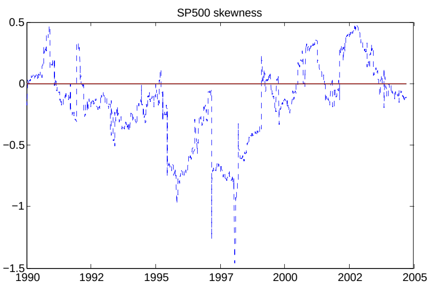
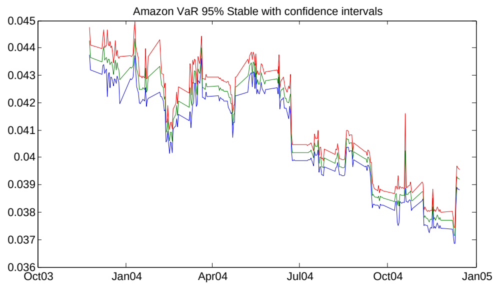


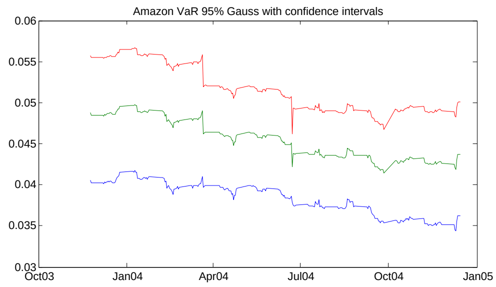
Figure 2: VaR 95% with confidence intervals

This figure plots the 95% VaR estimation for the Stable, Gauss and GPD assumption respectively. The chosen data series is Amazon and it spans from December 2003 to December 2004. The risk measures are calculated on a rolling window of 250 daily data points.

Panel A



Panel B



Panel C

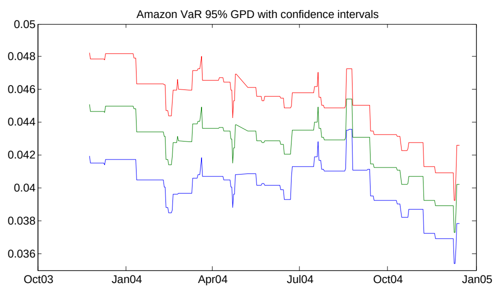
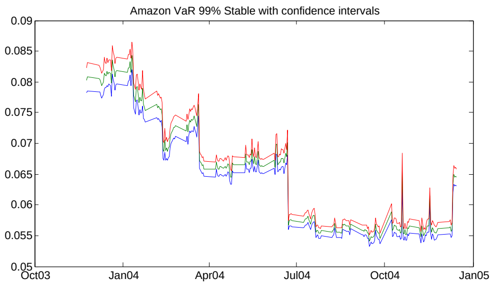


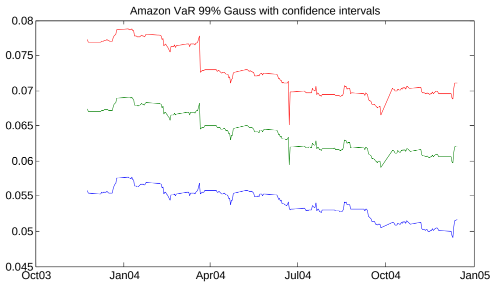
Figure 3: VaR 99% with confidence intervals

This figure plots the 99% VaR estimation for the Stable, Gauss and GPD assumption respectively. The chosen data series is Amazon and it spans from December 2003 to December 2004. The risk measures are calculated on a rolling window of 250 daily data points.

Panel A



Panel B



Panel C

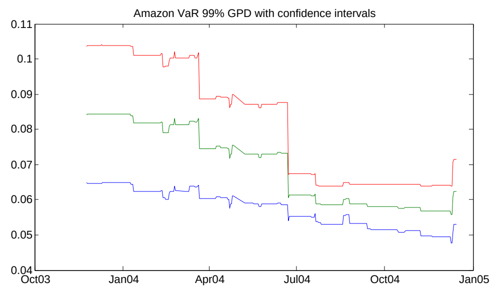


Figure 4: $\mathrm{VaR}_{95\%}$ Lower Bounds for Microsoft

This figure plots the negative returns of the Microsoft series, along with the 95% VaR lower bounds for Stable, Gauss and GPD model respectively. The risk measures are calculated on a rolling window of 250 daily log returns.

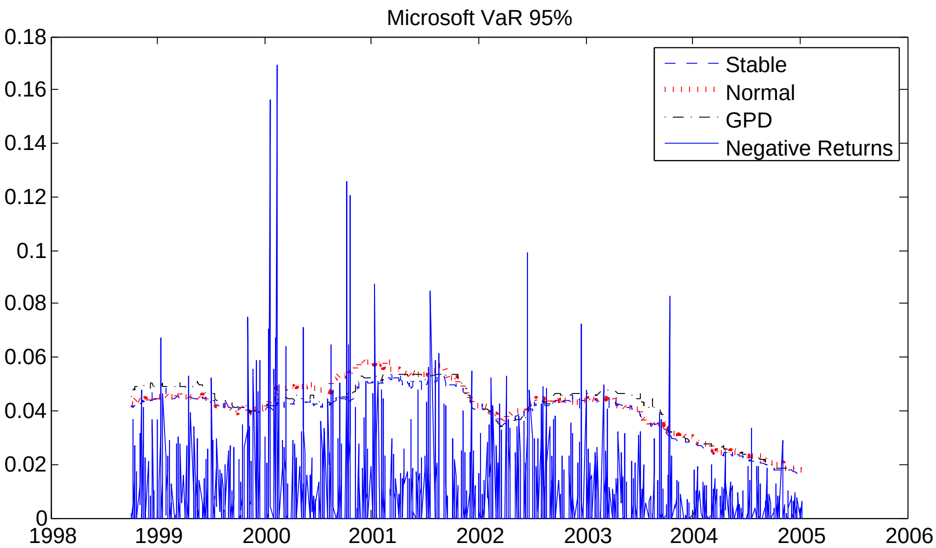
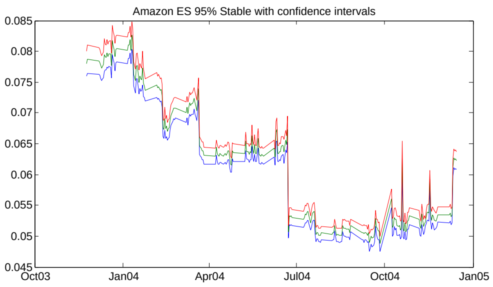


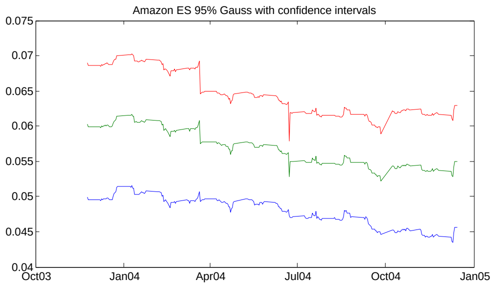
Figure 5: Expected Shortfall 95% with confidence intervals

This figure plots the 95% ES estimation for the Stable, Gauss and GPD assumption respectively. The chosen data series is Amazon and it spans from December 2003 to December 2004. The risk measures are calculated on a rolling window of 250 daily data points.

Panel A



Panel B



Panel C

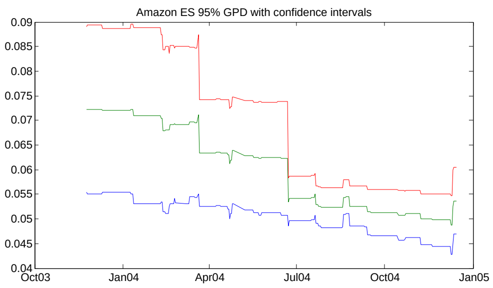
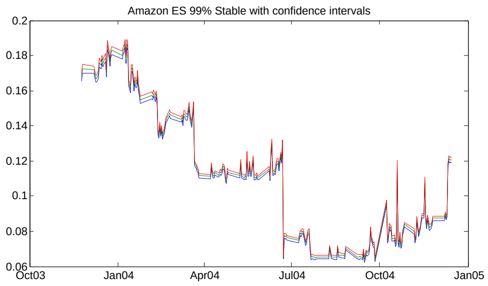


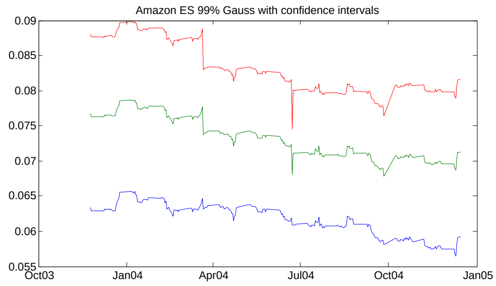
Figure 6: Expected Shortfall 99% with confidence intervals

This figure plots the 99% ES estimation for the Stable, Gauss and GPD assumption respectively. The chosen data series is Amazon and it spans from December 2003 to December 2004. The risk measures are calculated on a rolling window of 250 daily data points.

Panel A



Panel B



Panel C

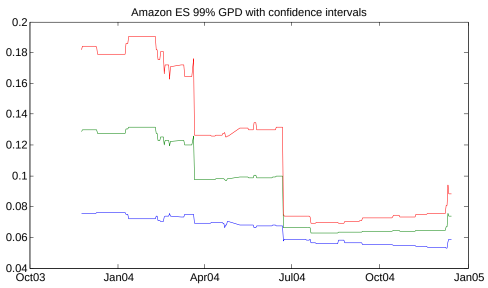


Table 1: Descriptive Statistics of Financial Series

This table reports the first four moments of the analyzed time series. All returns are calculated as log-differences on daily data series. The sample periods span from 2-Jan-1990 to 31-Dec-2004 for the SP500 and from 2-Jan-1998 to 31-Dec-2004 for the other series.

	Descriptive Statistics			
	Mean	Standard deviation	Skewness	Kurtosis
SP500	0.000	0.010	-0.105	6.666
NASDAQ	0.000	0.021	0.071	5.571
MICROSOFT	0.000	0.025	-0.145	7.882
AMAZON	0.001	0.053	0.318	6.498

Table 2: Value at Risk Backtesting

This table reports the results of a Value at Risk backtesting on the proposed models. All returns are calculated as log-differences on daily data series. Panel A results are from a sample period from 2-Jan-1990 to 31-Dec-2004, while Panel B to D results are from 2-Jan-1998 to 31-Dec-2004. The first two columns display the empirical violations and their percentages of the returns with respect to the VaR_p bound. The last column shows the result of the POF test, where $*$ indicates a 95% rejection of the VAR model.

	Panel A: SP500		
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	190	5.4%	1.071
$\text{Stable}_{99\%}$	36	1.0%	0.014
$\text{Gaussian}_{95\%}$	156	4.4%	2.616
$\text{Gaussian}_{99\%}$	48	1.4%	4.14*
$\text{GPD}_{95\%}$	163	4.6%	1.122
$\text{GPD}_{99\%}$	39	1.1%	0.377
	Panel B: NASDAQ		
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	69	4.6%	0.597
$\text{Stable}_{99\%}$	14	0.9%	0.081
$\text{Gaussian}_{95\%}$	61	4.0%	3.108
$\text{Gaussian}_{99\%}$	18	1.2%	0.534
$\text{GPD}_{95\%}$	64	4.2%	1.924
$\text{GPD}_{99\%}$	14	0.9%	0.081
	Panel C: MICROSOFT		
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	64	4.2%	1.924
$\text{Stable}_{99\%}$	9	0.6%	2.902
$\text{Gaussian}_{95\%}$	55	3.6%	6.415*
$\text{Gaussian}_{99\%}$	13	0.9%	0.307
$\text{GPD}_{95\%}$	56	3.7%	5.773*
$\text{GPD}_{99\%}$	10	0.7%	1.968
	Panel D: AMAZON		
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	67	4.4%	1.034
$\text{Stable}_{99\%}$	13	0.9%	0.307
$\text{Gaussian}_{95\%}$	62	4.1%	2.680
$\text{Gaussian}_{99\%}$	22	1.5%	2.800
$\text{GPD}_{95\%}$	55	3.6%	6.415*
$\text{GPD}_{99\%}$	14	0.9%	0.081

Table 3: Expected Shortfall Backtesting

This table reports results of the backtesting procedure on the expected shortfall measures. All returns are calculated as log-differences on daily data series. The sample periods span from 2-Jan-1990 to 31-Dec-2004 for the SP500 and from 2-Jan-1998 to 31-Dec-2004 for the other series.

	Stable 95%	Stable 99%	Gauss 95%	Gauss 99%	GPD 95%	GPD 99%
SP500	0.002	0.006	0.002	0.004	0.001	0.004
NASDAQ	0.004	0.005	0.001	0.007	0.000	0.012
MICROSOFT	0.006	0.019	0.008	0.015	0.004	0.015
AMAZON	0.032	0.060	0.011	0.021	0.001	0.004

Table 4: Block Maxima VaR Backtesting on SP500

This table reports the results of a Value at Risk backtesting on the Block Maxima approach for the Microsoft series (2-Jan-1990 to 31-Dec-2004). Returns are calculated as log-differences on daily data series. Panel A-C results are from a block of 10, 15 and 20 days respectively. The first two columns display the empirical violations and their percentages of the returns with respect to the VaR_p bound. The last column shows the result of the POF test, where $*$, $**$ and $***$ indicates a 95%, 99% and 99.9% rejection of the VAR model.

	SP500		
	Panel A: 10 days block		
	Violations	Percentage	POF
$\mathrm{Stable}_{95\%}$	123	3.5%	19.009***
$\mathrm{Stable}_{99\%}$	32	0.9%	0.322
$\mathrm{Gaussian}_{95\%}$	88	2.5%	56.822***
$\mathrm{Gaussian}_{99\%}$	18	0.5%	10.439**
$\mathrm{BMM}_{95\%}$	2	0.1%	340.006***
$\mathrm{BMM}_{99\%}$	1	0.0%	61.808***
	Panel B: 15 days block		
	Violations	Percentage	POF
$\mathrm{Stable}_{95\%}$	142	4.0%	7.585**
$\mathrm{Stable}_{99\%}$	42	1.2%	1.211
$\mathrm{Gaussian}_{95\%}$	78	2.2%	72.474***
$\mathrm{Gaussian}_{99\%}$	28	0.8%	1.641
$\mathrm{BMM}_{95\%}$	17	0.5%	246.905***
$\mathrm{BMM}_{99\%}$	9	0.3%	28.198***
	Panel C: 20 days block		
	Violations	Percentage	POF
$\mathrm{Stable}_{95\%}$	223	6.3%	11.945***
$\mathrm{Stable}_{99\%}$	102	2.9%	84.344***
$\mathrm{Gaussian}_{95\%}$	121	3.4%	20.551***
$\mathrm{Gaussian}_{99\%}$	32	0.9%	0.322
$\mathrm{BMM}_{95\%}$	17	0.5%	246.905***
$\mathrm{BMM}_{99\%}$	9	0.3%	28.198***

Table 5: Block Maxima VaR Backtesting on NASDAQ

This table reports the results of a Value at Risk backtesting on the Block Maxima approach for the NASDAQ series (2-Jan-1998 to 31-Dec-2004). Returns are calculated as log-differences on daily data series. Panel A-C results are from a block of 10, 15 and 20 days respectively. The first two columns display the empirical violations and their percentages of the returns with respect to the VaR_p bound. The last column shows the result of the POF test, where $*$, $**$ and $***$ indicates a 95%, 99% and 99.9% rejection of the VAR model.

	Violations	Percentage	POF
Panel A: 10 days block			
$\text{Stable}_{95\%}$	39	2.6%	22.346***
$\text{Stable}_{99\%}$	11	0.7%	1.236
$\text{Gaussian}_{95\%}$	19	1.3%	62.691***
$\text{Gaussian}_{99\%}$	5	0.3%	9.202**
$\text{BMM}_{95\%}$	9	0.6%	97.661***
$\text{BMM}_{99\%}$	3	0.2%	14.585***
Panel B: 15 days block			
$\text{Stable}_{95\%}$	49	3.2%	11.083***
$\text{Stable}_{99\%}$	24	1.6%	4.506*
$\text{Gaussian}_{95\%}$	32	2.1%	33.309***
$\text{Gaussian}_{99\%}$	12	0.8%	0.687
$\text{BMM}_{95\%}$	3	0.2%	129.152***
$\text{BMM}_{99\%}$	3	0.2%	14.585***
Panel C: 20 days block			
$\text{Stable}_{95\%}$	83	5.5%	0.771
$\text{Stable}_{99\%}$	28	1.9%	8.910**
$\text{Gaussian}_{95\%}$	42	2.8%	18.467***
$\text{Gaussian}_{99\%}$	14	0.9%	0.081
$\text{BMM}_{95\%}$	6	0.4%	111.831***
$\text{BMM}_{99\%}$	5	0.3%	9.202**

Table 6: Block Maxima VaR Backtesting on Amazon

This table reports the results of a Value at Risk backtesting on the Block Maxima approach for the Amazon series (2-Jan-1998 to 31-Dec-2004). Returns are calculated as log-differences on daily data series. Panel A-C results are from a block of 10, 15 and 20 days respectively. The first two columns display the empirical violations and their percentages of the returns with respect to the VaR_p bound. The last column shows the result of the POF test, where $*$, $**$ and $***$ indicates a 95%, 99% and 99.9% rejection of the VAR model.

	Violations	Percentage	POF
Panel A: 10 days block			
$\text{Stable}_{95\%}$	43	2.9%	17.229***
$\text{Stable}_{99\%}$	13	0.9%	0.304
$\text{Gaussian}_{95\%}$	33	2.2%	31.507***
$\text{Gaussian}_{99\%}$	9	0.6%	2.894
$\text{BMM}_{95\%}$	9	0.6%	97.571***
$\text{BMM}_{99\%}$	4	0.3%	11.625***
Panel B: 15 days block			
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	60	4.0%	3.549
$\text{Stable}_{99\%}$	17	1.1%	0.237
$\text{Gaussian}_{95\%}$	38	2.5%	23.691***
$\text{Gaussian}_{99\%}$	14	0.9%	0.080
$\text{BMM}_{95\%}$	4	0.3%	122.809***
$\text{BMM}_{99\%}$	2	0.1%	18.193***
Panel C: 20 days block			
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	63	4.2%	2.268
$\text{Stable}_{99\%}$	29	1.9%	10.218**
$\text{Gaussian}_{95\%}$	37	2.5%	25.140***
$\text{Gaussian}_{99\%}$	16	1.1%	0.056
$\text{BMM}_{95\%}$	6	0.4%	111.736***
$\text{BMM}_{99\%}$	1	0.1%	22.866***

Table 7: Block Maxima VaR Backtesting on Microsoft

This table reports the results of a Value at Risk backtesting on the Block Maxima approach for the Microsoft series (2-Jan-1998 to 31-Dec-2004). Returns are calculated as log-differences on daily data series. Panel A-C results are from a block of 10, 15 and 20 days respectively. The first two columns display the empirical violations and their percentages of the returns with respect to the VaR_p bound. The last column shows the result of the POF test, where $*$, $**$ and $***$ indicates a 95%, 99% and 99.9% rejection of the VAR model.

	Violations	Percentage	POF
	Panel A: 10 days block		
$\text{Stable}_{95\%}$	34	2.3%	29.827 ***
$\text{Stable}_{99\%}$	14	0.9%	0.080
$\text{Gaussian}_{95\%}$	22	1.5%	54.569***
$\text{Gaussian}_{99\%}$	6	0.4%	7.156**
$\text{BMM}_{95\%}$	4	0.3%	122.809***
$\text{BMM}_{99\%}$	1	0.1%	22.866 ***
	Panel B: 15 days block		
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	56	3.7%	5.747*
$\text{Stable}_{99\%}$	20	1.3%	1.471
$\text{Gaussian}_{95\%}$	32	2.1%	33.249***
$\text{Gaussian}_{99\%}$	6	0.4%	7.156**
$\text{BMM}_{95\%}$	3	0.2%	129.054***
$\text{BMM}_{99\%}$	2	0.1%	18.193***
	Panel C: 20 days block		
	Violations	Percentage	POF
$\text{Stable}_{95\%}$	83	5.5%	0.782
$\text{Stable}_{99\%}$	35	2.3%	19.365***
$\text{Gaussian}_{95\%}$	40	2.7%	20.953***
$\text{Gaussian}_{99\%}$	9	0.6%	2.894
$\text{BMM}_{95\%}$	3	0.2%	129.054***
$\text{BMM}_{99\%}$	3	0.2%	14.569***