International Journal of Theoretical and Applied Finance
Vol. 10, No. 6 (2007) 1043–1075

World Scientific
www.worldscientific.com

A COMPARISON OF SOME UNIVARIATE MODELS FOR VALUE-AT-RISK AND EXPECTED SHORTFALL

CARLO MARINELLI

Institut für Angewandte Mathematik, Universität Bonn
Wegelerstr. 6, D-53115 Bonn, Germany
marinelli@wiener.iam.uni-bonn.de

STEFANO D'ADDONA

Department of International Studies, University of Rome III
Via G. Chiabrera 199, 00145 Rome, Italy
daddona@uniroma3.it

SVETLOZAR 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
rachev@pstat.ucsb.edu

Received 15 March 2006
Accepted 18 October 2006

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 test whether 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 models tend to outperform pure EVT-based methods (especially those obtained by the so-called block maxima method) in the estimation of Value-at-Risk, while a peaks-over-threshold method turns out to be preferable for the estimation of expected shortfall. We also find empirical evidence that some simple semiparametric EVT-based methods perform well in the estimation of VaR.

Keywords: Value-at-Risk; expected shortfall, Paretian stable laws; extreme value theory.

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 several 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 the maximum probable loss on a given portfolio, referring to a specific confidence level and time horizon. Historically the literature on VaR has evolved following both the parametric and the non-parametric approach (see, e.g., [9, 22] 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., [9] 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 distribution of returns 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 extensively documented 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 [2, 13, 14], 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 [31]). 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., [32]). Moreover, models based on stable laws 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, 19, 23]).

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 sums of i.i.d. random variables and are stable (better said, closed) with respect to summation, Generalized Extreme Value (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

$$aX_1 \circ bX_2 \stackrel{d}{=} cX + d, \tag{1.1}$$

where X_1, X_2 are i.i.d. copies of X, a, b, $c \in \mathbb{R}_+$, $d \in \mathbb{R}$, then (1.1) 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 robust with respect to the distribution F of returns, as only very mild assumptions are required, in particular no specific parametric assumption on F is necessary — see, e.g., [12]. They still need, however, 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 to [17, 26], where EVT-based estimates for VaR and expected shortfall are proposed and tested. However, both papers focus on EVT methods only, and 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. In particular, we provide *analytic* expressions for asymptotic confidence intervals for estimates of VaR and expected shortfall for a set of models widely used in the industry. These expressions are based on the delta method, and are rather straightforward to implement once the parameters of the corresponding distributions are estimated, together with their confidence intervals. Approximate confidence intervals for VaR and expected shortfall are obtained in [17], mainly using profile likelihood and bootstrap techniques, but only for EVT-based methods. Confidence intervals in the stable case seem to be new, to the best of our knowledge we have not been able to find them in the literature.

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 the random variable of negative returns of a financial position and by F its distribution. Then Value-at-Risk at confidence level p for our financial position is defined as the p quantile of the distribution F, i.e.,

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : F(x) \geq p\}. \tag{1.2}$$

Typical choices of p are $p \in \{0.9, 0.95, 0.99\}$. We shall also assume throughout the paper that the observed (negative) returns X_i, $i = 1, \ldots, n$ form an i.i.d. sample from the distribution 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 Sec. 6. Section 7 concludes.

Some results contained in a preliminary version of the present paper were announced in [20] and published in incomplete form in [21]. This paper significantly improves on [21] in several respects: we compute for most estimators (asymptotic) confidence intervals, while [21] only gives point estimates. In the stable case, while [21] uses plain Monte Carlo methods to estimate VaR and expected shortfall, we use analytic formulas and precise numerical integration, achieving a much higher level of accuracy (especially for the estimation of tail integrals, i.e., expected shortfall, plain Monte Carlo methods are highly inaccurate and converge very slowly). Moreover, while [21] considers only a peaks-over-threshold method in the class of EVT-based methods, we study the block maxima method as well as two semiparametric methods based on estimates of the tail index and on order statistics. Finally, we include in our empirical tests some of the time series used by [21] comparing the corresponding results.

2. Benchmark VaR

In this section we find point estimates and confidence intervals (some of them 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 assumption 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, $i = 1, \dots, n$ 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

$$\left[\frac{(n-1)\hat{\sigma}_n^2}{\nu_{n-1,1-r/2}}, \frac{(n-1)\hat{\sigma}_n^2}{\nu_{n-1,r/2}}\right]. \tag{2.1}$$

Consequently, it is straightforward 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

$$\sqrt{n}(S_n^2 - \sigma^2) \Rightarrow N(0, \mu_4 - \sigma^4), \tag{2.2}$$

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 (2.2), 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 (2.2) as

$$\sigma^2 \in \left[S_n^2 - \sqrt{\frac{\mu_4 - \sigma^4}{n}} z_{r/2}, S_n^2 + \sqrt{\frac{\mu_4 - \sigma^4}{n}} z_{r/2}\right]. \tag{2.3}$$

In order to make this confidence interval operational, we need to replace in (2.3) σ^4 and μ_4 with consistent estimators. Then, in view of Slutsky's theorem, (2.3) 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 (2.1) and (2.3).

2.2. *VaR and empirical processes*

Let $\mathbb{F}_n$ denote the empirical process of the observed negative returns $X_1, \dots, X_n$, 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 $X_{n(1)} \le X_{n(2)} \le \cdots \le X_{n(n)}$ are the order statistics.

The asymptotic properties of this estimator are collected in the following proposition, whose proof can be found, e.g., in [36]. The derivative of F, whenever it exists, will be denoted by f.

Proposition 2.1. *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

$$\sqrt{n}\left(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\right) \Rightarrow N\left(0, \frac{p(1-p)}{f^2(F^{-1}(p))}\right). \tag{2.4}$$

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 (2.4). If that is not possible, then the following alternative procedure can be used: let $X_1, \dots, X_n$ 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

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

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 Modeling 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 skewness, $\sigma > 0$ 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 [32]), 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., [27, 28]).

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

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, J_\theta^{-1}), \tag{3.1}$$

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 [10]. Computationally, one obtains an initial estimate of θ, using e.g. the quantile-based method of [25], 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 [23], 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 also [30]).

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

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

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 $1 - r$:

$$\mathrm{VaR} \in \left[\widehat{\mathrm{VaR}}_n - \zeta z_{r/2}, \widehat{\mathrm{VaR}}_n + \zeta z_{r/2}\right], \tag{3.2}$$

where

$$\zeta = \frac{\sqrt{(\nabla g_p(\hat{\theta}_n))^* J_{\hat{\theta}_n}^{-1} \nabla g_p(\hat{\theta}_n)}}{\sqrt{n}}.$$

The argument leading to (3.2) is of course only formal, but it becomes rigorous if we can prove that g_p is differentiable at θ.

Proposition 3.1. *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.*

See the Appendix for the proof.

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 [18] 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 4.1 [18]. *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*

$$\frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \Rightarrow Y \tag{4.1}$$

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

$$H(x) = \exp\left(-\left(1 + \xi\frac{x - \mu}{\sigma}\right)_+^{-1/\xi}\right). \tag{4.2}$$

In (4.2) 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., $H(x) = \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)_+\right)$.

We say that a distribution F is in the max-domain of attraction of a GEV law H (in symbols, $F \in D_m(H)$) if it satisfies the hypotheses of Theorem 4.1.

Appealing to Theorem 4.1, 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 m, and sets

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

Then, by assuming that the distribution of the block maxima (Y_i) is approximately GEV, one fits a law of the type (4.2) to the (Y_i), and using simple properties of the distribution function of the maximum of a finite family of i.i.d. random variables, obtains an estimate of VaR. The procedure is described in detail in Sec. 4.1.

Another procedure to estimate VaR is based on the following theorem, which characterizes the limit distribution of excesses over a threshold of a sequence of i.i.d. random variables.

Theorem 4.2 [1]. *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*

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

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

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

The method relying on this theorem, sometimes called Peaks over Thresholds (POT) method, is described in Sec. 4.2.

A natural term of comparison for these methods, which, roughly speaking, are based on the assumptions that returns are in the domain of attraction of a max stable law, will be the assumption of α-stable distributed (daily) returns, i.e., that returns are sum-stable.

4.1. *VaR with max-stable block maxima*

Let us define block maxima as follows:

$$Y_{k+1} = \max(X_{km+1}, X_{km+2}, \ldots, X_{(k+1)m}),$$

where m is the block size (m could correspond, for instance, to the typical number of trading days in a week, or two weeks, or a month). Assuming that the random variables Y_k are independent and (approximately) distributed like a GEV law with distribution function $H(x)$, we have $\mathbb{P}(X_1 \le x_p) = p$ if and only if $\mathbb{P}(Y_1 \le x_p) = p^m$, hence $x_p = H^{-1}(p^m)$. This simple observation suggests the following procedure to estimate x_p: let $\theta = (\xi, \mu, \sigma)$ and denote by $h(\cdot;\theta)$ 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 [33]).

Proposition 4.1. *If $\xi > -1/2$ then $\hat{\theta}_n$ is a consistent, asymptotically normal and efficient estimator of θ.*

As it follows from (4.2), VaR at p level can be estimated as

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

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

$$\mathrm{VaR} \in \Big[\widehat{\mathrm{VaR}}_n - \zeta z_{r/2}, \widehat{\mathrm{VaR}}_n + \zeta z_{r/2}\Big], \tag{4.3}$$

with

$$\zeta = \frac{\sqrt{(\nabla g(\hat{\theta}_n))^* J_{\hat{\theta}_n}^{-1} \nabla g(\hat{\theta}_n)}}{\sqrt{n}},$$

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. It is important to remark that in the above expressions n is not the total number of observed returns, but only the total number of block maxima.

4.2. *Exceedances over a threshold*

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

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

Then one has

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

hence

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

Appealing to Theorem 4.2, 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 [26] in choosing a random threshold that selects the top 10% of the observed negative returns.

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 [34])

$$\sqrt{n_u}(\hat{\theta}_{n_u} - \theta) \Rightarrow N(0, J_\theta^{-1}), \tag{4.4}$$

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

$$\mathrm{VaR} \in \Big[\widehat{\mathrm{VaR}}_n - \zeta z_{r/2}, \widehat{\mathrm{VaR}}_n + \zeta z_{r/2}\Big], \tag{4.5}$$

where

$$\zeta = \frac{\sqrt{(\nabla g_p(\hat{\theta}_{n_u}))^* J_{\hat{\theta}_{n_u}}^{-1} \nabla g_p(\hat{\theta}_{n_u})}}{\sqrt{n_u}}.$$

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 (4.4) when u is a random threshold — see, e.g., [7, 8]. The asymptotic covariance matrices obtained by these authors seem unfortunately quite difficult to implement in terms of computational complexity.

4.3. *Two semiparametric approaches*

We shall describe two "semiparametric" approaches that assume only that the distribution of losses is in the domain of attraction of a max-stable law. Both methods need an estimate $\hat{\xi}$ of the tail index and use, apart of $\hat{\xi}$, only the order statistics of the sample. Common to both methods is also the choice of a threshold parameter, in analogy to the POT method described above.

The first method (see, e.g., [12]) assumes that the tail of F can be sufficiently well approximated by a Pareto tail, i.e., that $F(x) = 1 - x^{-1/\xi}$, for x large enough. In particular one proceeds in two steps:

(a) Estimate the tail parameter ξ through the Hill estimator

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

where k is a number to be chosen, and $X_{1,n} \geq X_{2,n} \geq \cdots \geq X_{n,n}$ stands for the order statistics in descending order.

(b) Estimate the tail of the distribution by

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

and the quantile x_p by

$$\hat{x}_p = \Big(\frac{n}{k}(1-p)\Big)^{-\hat{\xi}} X_{k+1,n}. \tag{4.6}$$

The choice of k in the above estimator is a very delicate issue, and there is an extensive literature on the subject (see [12] and references therein). Usually k is chosen by visual inspection of a Hill plot, and as such is not amenable to an automated computer implementation. On the other hand, a two stage bootstrap method has been proposed in [4] to find a k that minimizes the asymptotic mean square error. Unfortunately this result provides little guidance in the finite sample case (see also [3] for a related discussion).

Let us also mention that (approximate) asymptotic confidence intervals for x_p can be constructed assuming n/k and $X_{k+1,n}$ constants in (4.6), and using the corresponding asymptotic confidence interval for the Hill estimator of ξ (which is asymptotically normal, see [12]). This method unfortunately suffers of many drawbacks, mainly due to the problem mentioned in the above paragraph. Therefore we shall limit ourselves to compute confidence intervals for the next method we are going to present, which instead behaves well in numerical experiments.

An alternative estimator for x_p, that again assume only that F is in the domain of attraction of a max-stable law, but uses more observations of the available sample, has been introduced in [5]. In particular, let us set

$$\hat{x}_p = X_{k,n} + (X_{k,n} - X_{2k,n}) \frac{\left(\frac{k}{n(1-p)}\right)^{\hat{\xi}} - 1}{1 - 2^{-\hat{\xi}}},$$

where

$$\hat{\xi} = \frac{1}{\log 2} \log \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}},$$

is the Pickand's estimator and k is a "threshold" to be chosen appropriately.

The following theorem on asymptotic normality of the estimator (see [5]) allows one to construct asymptotic confidence intervals for VaR.

Theorem 4.3. *Let $X_1, \ldots, X_n$ be i.i.d. with distribution function $F \in D_m(H_\xi)$. Assume moreover that F has a positive density f of regular variation of order $-1 - 1/\xi$ and that $n(1 - p_n) \to c$, $c > 0$ fixed. Then for every fixed $k > c$ one has*

$$\frac{\hat{x}_{p_n} - x_{p_n}}{X_{k,n} - X_{2k,n}} \Rightarrow \eta,$$

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 is standard gamma distributed with parameter $2k + 1$, and $H_k = \sum_{j=k+1}^{2k} j^{-1} E_j$ in distribution, with E_i, $i = k + 1, \ldots, 2k$ i.i.d. standard exponentials.

In practice p_n is fixed and the theorem is used assuming $p_n = p$ and $\xi = \xi_n$ (see [5]).

In the empirical section we shall see that these two simple estimators of extreme quantiles perform quite well, even without fine-tuning the choice of the threshold k. This observation is important because, while the classical method of visual inspection is simply infeasible in a back-testing study with over 1,000 samples, an "automated" procedure such as the two-stage bootstrap method mentioned above would be computationally very expensive, and in general far from optimal in the finite sample case.

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., [15]). 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 [35]. In particular, if $X \sim N(0,2)$, the expected shortfall at level p is given by

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

In the general case $X' \sim N(\mu, \sigma^2)$ 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 $X \sim N(0,\sigma^2)$ 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 Sec. 2).

5.3. *Stable expected shortfall*

In the α-stable case there exists an integral representation of expected shortfall obtained in [35]. In particular, if $X \sim S_\alpha(1,\beta,0)$, one has

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

where

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

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

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

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 3.1 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

$$\mathrm{ES} \in \left[\widehat{\mathrm{ES}}_n - \zeta z_{r/2}, \widehat{\mathrm{ES}}_n + \zeta z_{r/2}\right],$$

where

$$\zeta = \frac{\sqrt{(\nabla g_p(\hat{\theta}_n))^* J_{\hat{\theta}_n}^{-1} \nabla g_p(\hat{\theta}_n)}}{\sqrt{n}}$$

and J_θ is the Fisher information matrix of (3.1).

5.4. *EVT-based expected shortfall*

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

$$\mathbb{E}[Y|Y > x] = \frac{x+\sigma}{1-\xi}. \tag{5.1}$$

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 (5.1), 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 Sec. 4.2.

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.

The first semiparametric method presented in Sec. 4.3 also implies an estimator for expected shortfall. In particular, let $\hat{F}(x) = 1 - (n/k)X_{k+1,n}^{1/\xi} x^{-1/\xi}$ the approximation to the tail of F, then we have

$$\widehat{\mathrm{ES}}_p = \int_{\hat{x}_p}^{\infty} x\hat{f}(x)\, dx,$$

where $\hat{f}(x) = \hat{F}'(x)$. A direct calculation shows that

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

On the other hand, the second method of Sec. 4.3 does not yield an estimate of expected shortfall, or at least we have not been able to derive one.

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), two stocks (Amazon and Microsoft), and two exchange rates (USD/GPB and USD/YEN). The exchange rates series are chosen so that we perfectly overlap two series used in [21]. This allows us to give a direct comparison with their results. All the raw prices are freely available on the web, and the returns are calculated as log-differences on daily data series.[1] The sample periods span from 2-Jan-1990 to 31-Dec-2004 for the SP500 and the exchange rates, 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

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
USD/GBP	0.000	0.006	−0.257	5.348
USD/YEN	0.000	0.007	−0.506	7.036

[1] We restrict ourselves to consider daily data for two reasons: the first and most important is that the industry and regulatory standard is to compute VaR and related risk measures on a daily basis. Moreover, studying lower frequencies (such as weekly or monthly) would considerably decrease the size of our samples, possibly invalidating the asymptotic properties of many, if not all, of the estimators proposed.

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, Amazon and the exchange rates all display clear evidence of fat tails in their distributions.

This claim is confirmed by a more detailed analysis: in Fig. 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 of the kurtosis correspond to a comparable deviation of the skewness parameter (cf. Panel C-D of Fig. 1). Same figures can be observed for both the exchange rates. In particular, the USD/GBP exchange rate displays the highest peak in the kurtosis around 1997, even if the series seem to have a stable kurtosis around 4.

The presence of tails heavier than Gaussian is also confirmed by analyzing Fig. 2, where QQ plots of all the series versus a normal distribution are shown: the sample quantiles in the tails strongly deviate from the corresponding normal quantiles.

Extending this analysis to other distributions, in Fig. 3 we collect QQ plots of the series with the highest sample kurtosis for each category (indices, stocks and FX rates) against the exponential distribution (as suggested in [12]) and the stable distribution. While the QQ plots versus the exponential distribution (left hand side panels) clearly indicate "heavy tails" in the series, the plots versus the stable distribution show a good fit in the center of the distribution, and tails "lighter" than stable far out in the tail. One should bear in mind, however, that such QQ plots are typical of stable laws (see, e.g., [29]), and are essentially due to small sample size, and should not be considered as evidence against the stable hypothesis.

Finally, again using the series with the highest sample kurtosis in each category, we compare their empirical density both with a stable and a Gaussian density. These plots, shown in Fig. 4, clearly indicate the good fit of stable laws also in the middle of the density.

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 proposed VaR measures. In order to assess them, we calculate for each specification two series of VaRs, with confidence level 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, comparing the VaR estimates 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

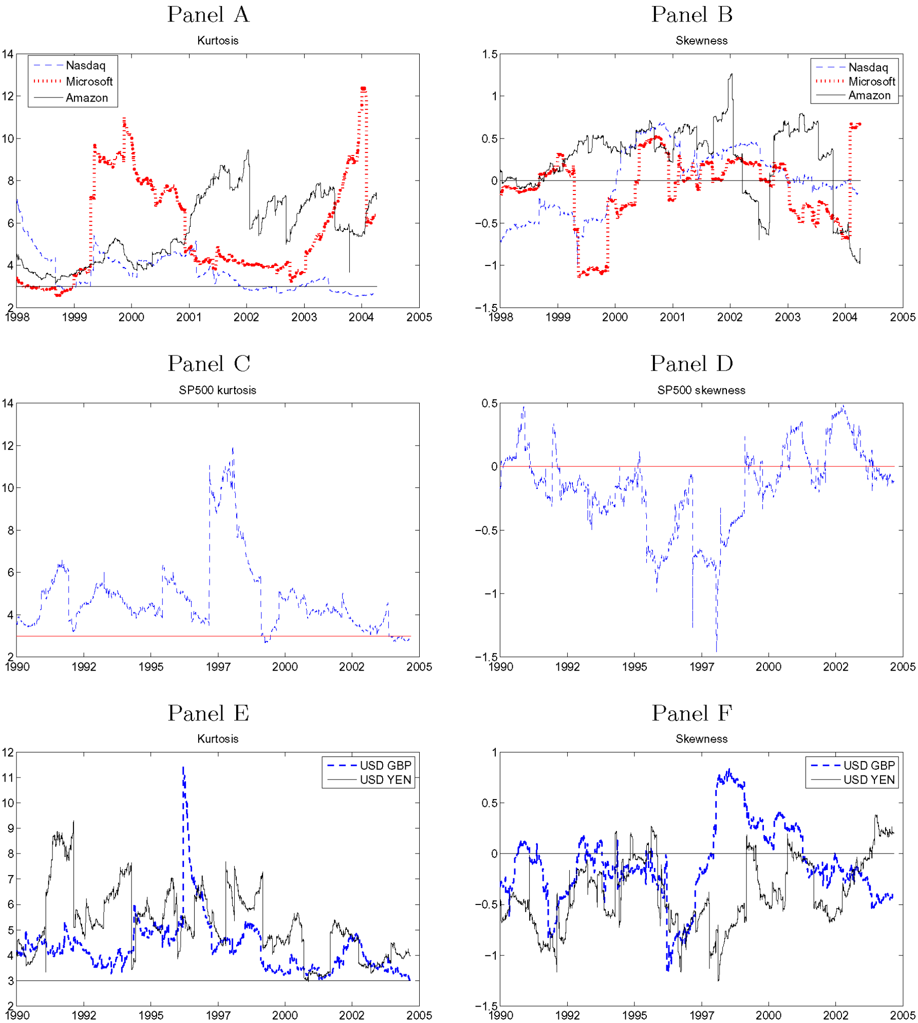


Fig. 1. Time series of the kurtosis.

This figure plots the skewness and the kurtosis of the analyzed series. The moments of NASDAQ, Microsoft and Amazon are shown in Panels A and B respectively, the ones of SP500 in Panels C and D, while the ones on the exchange rates are displayed in Panels E and F. 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.

estimations are in line with the empirical returns and present a remarkable level of accuracy in terms of their estimation error. In order to make this analysis clearer we plot the estimations for the last 12 months,[2] along with the confidence intervals,

[2]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.

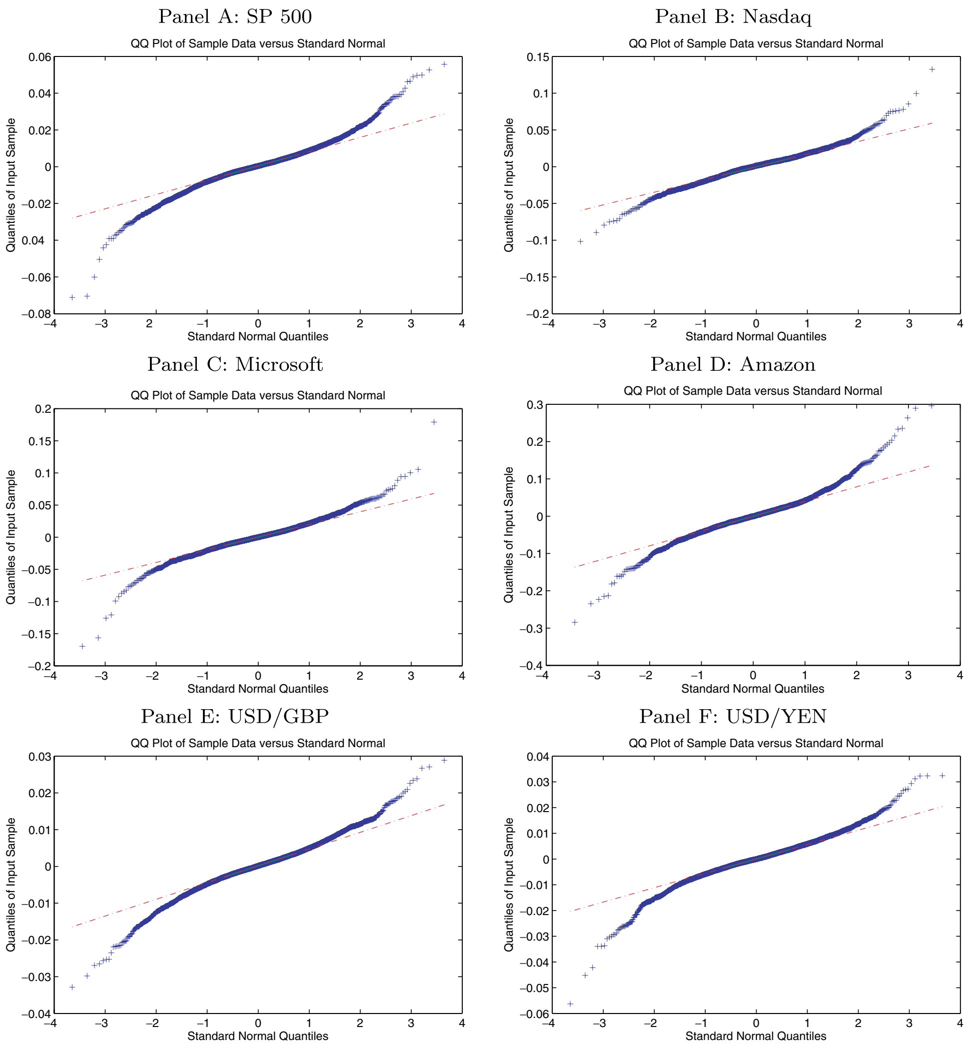
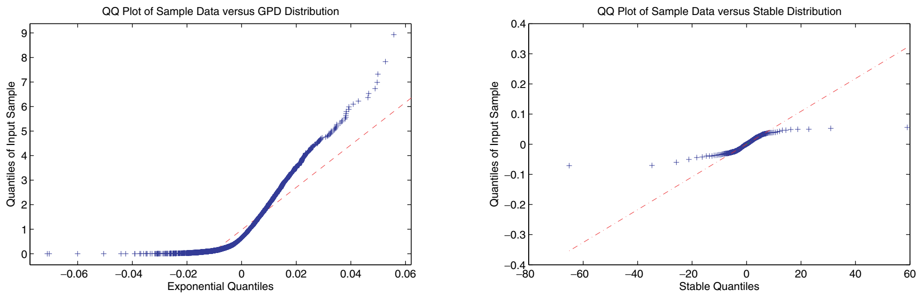


Fig. 2. QQ Plots versus the normal distribution.

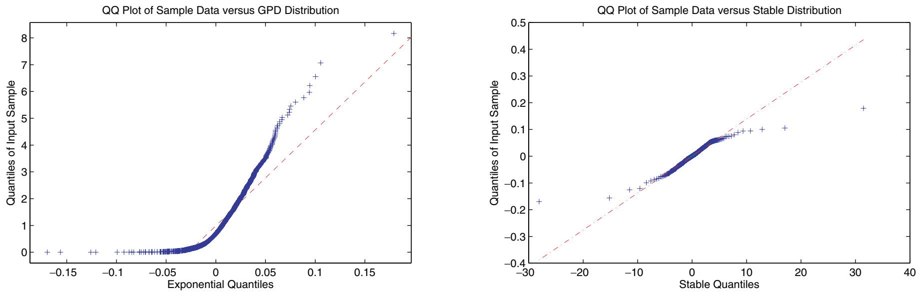
This figure shows the QQ Plots against the Normal distribution of the analyzed series. Indices are shown in Panel A and B, stocks in Panel C and Panel D, and exchange rates are displayed in Paned E and Panel F.

for the Amazon time series, that has the highest historical volatility coupled with a high kurtosis (cf. Table 1). The graphs in Figs. 5 and 6 show estimations that are comparable in magnitude for the three specifications, both at a 95% and 99% confidence level. Moreover, the VaRs based on the stable assumption seem to have a greater accuracy, given the tightness of their confidence intervals. In fact, while the estimation based on a semi-parametric approach (Pickand) displays a confidence interval quite large in absolute values, both the Extreme Value estimation based on

Panel A: SP 500



Panel B: MICROSOFT



Panel C: USD/YEN

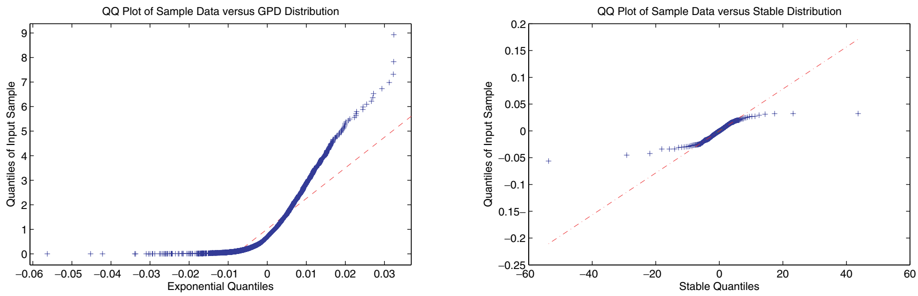
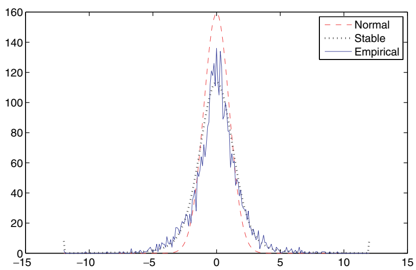


Fig. 3. QQ Plots versus "fat tails" distributions.

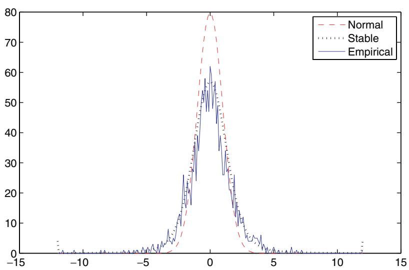
This figure shows the QQ Plots against the standard Stable distribution and the Exponential distribution. For each of the analyzed categories (indices, stocks and exchange rates) the series with the highest kurtosis is plotted. Left hand side panels display the QQ Plots for the Stable distribution, right hand side panels display QQ Plots for the Exponential distribution.

the Peak over a Threshold approach (GPD), proposed in Sec. 4.2, and the Gauss specification, display a confidence interval of the order of 0.5%−1.5% for the 95% VaR, and 1%−2% to peaks of 4% for the 99% VaR. On the contrary, the stable confidence intervals are below the 0.6% in both cases. This phenomenon is simply explained by the fact that the asymptotic confidence interval for stable VaR is

Panel A: SP 500



Panel B: MICROSOFT



Panel C: USD/YEN

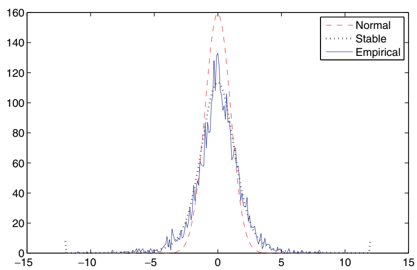


Fig. 4. Distributions' comparison.

This figure compares the empirical distributions with the Standard Stable distribution and the Standard Normal distribution. For each of the analyzed categories (indices, stocks and exchange rates) the series with the highest kurtosis is plotted.

computed on the basis of all observations in the sample, while EVT-based confidence intervals rely only on the observations exceeding a certain threshold.

To further assess the accuracy of the calculated VaR, we perform a simple Proportion of Failure (POF) test, as, e.g., in [24]. In particular we calculate:

$$LR = -2\log\left(\frac{p_0^x(1-p_0)^{(n-x)}}{p^x(1-p)^{(n-x)}}\right), \tag{6.1}$$

where p_0 is the probability of an exception implied by the chosen confidence level, 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 (6.1) is the likelihood ratio statistics based on the number of exceedances in any given sample, where the null hypothesis is that the estimated value for the exceedances 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

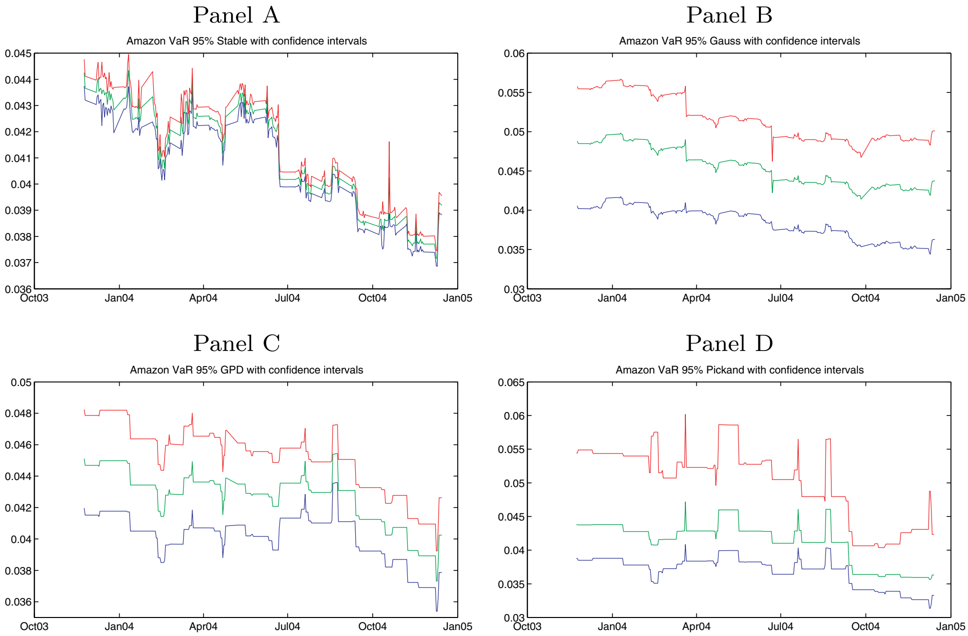


Fig. 5. VaR 95% with confidence intervals.

This figure plots the 95% VaR estimation for the Stable, Gauss, GPD, and Pickand estimator 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.

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 presents a LR statistics that exceed the critical value, the Gauss–VaR and the GPD–VaR are rejected in five and two out of twelve cases respectively. Second, the highest number of failures by GPD and Gauss estimations occurs with the Microsoft and the USD/YEN data series. This can be ascribed to the high kurtosis of the two series, which is probably better captured by fitting a stable law. To further investigate this point, we plot in Fig. 7 the negative returns of the Microsoft series along with the 95% lower bound for the three models. It is clear that the worse performances of 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.

Given the large amount of tests performed, it is difficult, at a first glance, to draw general conclusions on the performances of the different estimation methods. As a partial solution, we consider the overall performance by a pooled analysis. In practice we calculate the same statistics presented for the financial series by pooling together the obtained violation for each method. Results, reported in Panel G of Table 2, clearly show a good performance of the stable hypothesis both at 95% and 99% confidence level. While the same conclusion can be reached for the Hill

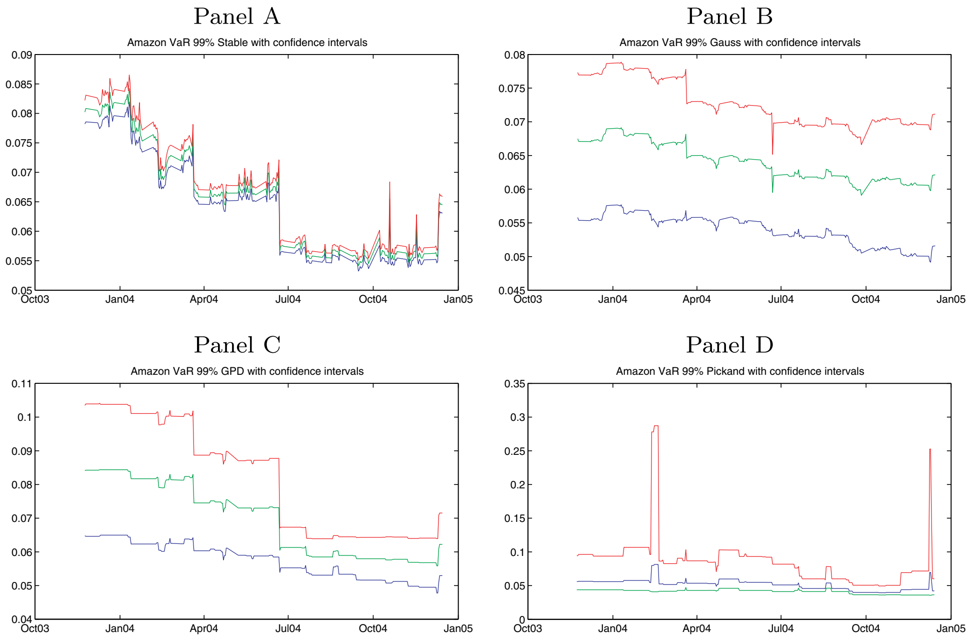


Fig. 6. VaR 99% with confidence intervals.

This figure plots the 99% VaR estimation for the Stable, Gauss, GPD and Pickand estimator 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.

estimator, the other methods are performing poorly either at 95% confidence level (empirical, GPD and Pickand) or in both cases, as in the Gaussian case.

In the same fashion as for the VaR backtesting procedure, we analyze the performance of the estimators of Expected Shortfall (ES) proposed is Sec. 5. We calculate for each specification two series of ES's, at confidence level 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 estimators and their confidence intervals lead us to essentially the same conclusions as in the VaR case. Figures 8 and 9 show estimations that are comparable in magnitude for the three specifications, both at 95% and 99% confidence level. Also, the ES's 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%, with a peak of 11% for the GPD. On the other hand, the stable confidence intervals are below the 0.6% in both cases, similarly as for the estimators of VaR.

To backtest the ES forecasts, we follow [11] 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

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. Panels A, E and F are based on a sample period from 2-Jan-1990 to 31-Dec-2004, while Panel B to D are based on 2-Jan-1998 to 31-Dec-2004. Last panel displays a pool analysis on the joint performance of the series. 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.

	Violations	Percentage	LR
Panel A: SP500			
$\mathrm{Empirical}_{95\%}$	171	4.8%	0.186
$\mathrm{Empirical}_{99\%}$	38	1.1%	0.202
$\mathrm{Stable}_{95\%}$	190	5.4%	1.071
$\mathrm{Stable}_{99\%}$	36	1.0%	0.014
$\mathrm{Gaussian}_{95\%}$	156	4.4%	2.616
$\mathrm{Gaussian}_{99\%}$	48	1.4%	4.14*
$\mathrm{GPD}_{95\%}$	163	4.6%	1.122
$\mathrm{GPD}_{99\%}$	39	1.1%	0.377
$\mathrm{Pickand}_{95\%}$	175	5.0%	0.014
$\mathrm{Pickand}_{99\%}$	31	0.9%	0.554
$\mathrm{Hill}_{95\%}$	191	5.4%	1.214
$\mathrm{Hill}_{99\%}$	38	1.1%	0.202
Panel B: NASDAQ			
$\mathrm{Empirical}_{95\%}$	66	4.4%	1.162
$\mathrm{Empirical}_{99\%}$	12	0.8%	0.687
$\mathrm{Stable}_{95\%}$	69	4.6%	0.597
$\mathrm{Stable}_{99\%}$	14	0.9%	0.081
$\mathrm{Gaussian}_{95\%}$	61	4.0%	3.108
$\mathrm{Gaussian}_{99\%}$	18	1.2%	0.534
$\mathrm{GPD}_{95\%}$	64	4.2%	1.924
$\mathrm{GPD}_{99\%}$	14	0.9%	0.081
$\mathrm{Pickand}_{95\%}$	64	4.2%	1.924
$\mathrm{Pickand}_{99\%}$	9	0.6%	2.902
$\mathrm{Hill}_{95\%}$	72	4.8%	0.168
$\mathrm{Hill}_{99\%}$	14	0.9%	0.081
Panel C: MICROSOFT			
$\mathrm{Empirical}_{95\%}$	58	3.8%	4.600*
$\mathrm{Empirical}_{99\%}$	10	0.7%	1.578
$\mathrm{Stable}_{95\%}$	64	4.2%	1.924
$\mathrm{Stable}_{99\%}$	9	0.6%	2.902
$\mathrm{Gaussian}_{95\%}$	55	3.6%	6.415*
$\mathrm{Gaussian}_{99\%}$	13	0.9%	0.307
$\mathrm{GPD}_{95\%}$	56	3.7%	5.773*
$\mathrm{GPD}_{99\%}$	10	0.7%	1.968
$\mathrm{Pickand}_{95\%}$	58	3.8%	4.600*
$\mathrm{Pickand}_{99\%}$	10	0.7%	1.968
$\mathrm{Hill}_{95\%}$	61	4.0%	3.108
$\mathrm{Hill}_{99\%}$	11	0.7%	1.236
Panel D: AMAZON			
$\mathrm{Empirical}_{95\%}$	55	3.6%	6.415$*$
$\mathrm{Empirical}_{99\%}$	13	0.9%	0.307

Table 2. (*Continued*)

	Violations	Percentage	LR
Panel D: AMAZON			
$\text{Stable}_{95\%}$	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
$\text{Pickand}_{95\%}$	54	3.6%	7.095*
$\text{Pickand}_{99\%}$	13	0.9%	0.307
$\text{Hill}_{95\%}$	60	4.0%	3.571
$\text{Hill}_{99\%}$	12	0.8%	0.687
Panel E: USD/YEN			
$\text{Empirical}_{95\%}$	176	5.0%	0.001
$\text{Empirical}_{99\%}$	37	1.1%	0.089
$\text{Stable}_{95\%}$	192	5.5%	1.470
$\text{Stable}_{99\%}$	42	1.2%	1.242
$\text{Gaussian}_{95\%}$	145	4.1%	6.136*
$\text{Gaussian}_{99\%}$	48	1.4%	4.207*
$\text{GPD}_{95\%}$	162	4.6%	1.220
$\text{GPD}_{99\%}$	45	1.3%	2.522
$\text{Pickand}_{95\%}$	169	4.8%	0.305
$\text{Pickand}_{99\%}$	32	0.9%	0.307
$\text{Hill}_{95\%}$	191	5.4%	1.293
$\text{Hill}_{99\%}$	42	1.2%	1.242
Panel F: USD/GBP			
$\text{Empirical}_{95\%}$	176	5.0%	0.001
$\text{Empirical}_{99\%}$	33	1.0%	0.086
$\text{Stable}_{95\%}$	193	5.5%	1.658
$\text{Stable}_{99\%}$	25	0.7%	3.333
$\text{Gaussian}_{95\%}$	172	4.9%	0.101
$\text{Gaussian}_{99\%}$	54	1.5%	8.697*
$\text{GPD}_{95\%}$	172	4.9%	0.101
$\text{GPD}_{99\%}$	38	1.1%	0.216
$\text{Pickand}_{95\%}$	170	4.8%	0.225
$\text{Pickand}_{99\%}$	23	0.7%	4.881*
$\text{Hill}_{95\%}$	185	5.3%	0.466
$\text{Hill}_{99\%}$	33	0.9%	0.144
Panel G: Pool analysis			
$\text{Empirical}_{95\%}$	702	4.6%	4.021*
$\text{Empirical}_{99\%}$	143	0.9%	0.438
$\text{Stable}_{95\%}$	775	5.1%	0.548
$\text{Stable}_{99\%}$	139	0.9%	0.993
$\text{Gaussian}_{95\%}$	651	4.3%	15.814*
$\text{Gaussian}_{99\%}$	203	1.3%	16.313*
$\text{GPD}_{95\%}$	672	4.4%	9.981*
$\text{GPD}_{99\%}$	160	1.1%	0.530
$\text{Pickand}_{95\%}$	690	4.6%	6.760*
$\text{Pickand}_{99\%}$	118	0.8%	2.263
$\text{Hill}_{95\%}$	760	5.0%	0.033
$\text{Hill}_{99\%}$	150	1.0%	0.007

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
USD/GBP	0.002	0.010	0.001	0.001	0.001	0.000
USD/YEN	0.001	0.004	0.003	0.003	0.002	0.003

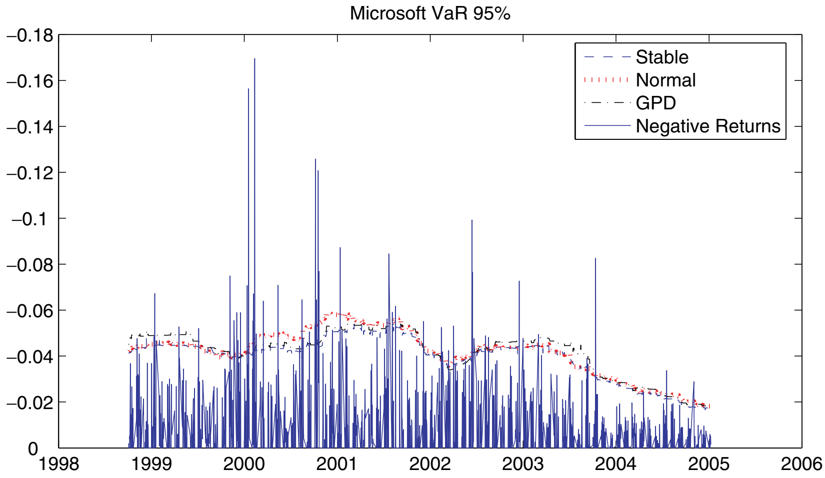


Fig. 7. $\mathrm{VaR}_{95\%}$ lower bounds for microsoft.

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.

the forecasted ES's, conditional on having a (negative) return exceeding the corresponding VaR estimate.[3]

The test statistic is defined as follows:

$$V = \left| \frac{\sum_{k=1}^{n}(X_k - (\mathrm{ES}_{p,k}))\mathbb{I}_{X_k > \mathrm{VaR}_{p,k}}}{\sum_{k=1}^{n}\mathbb{I}_{X_k > \mathrm{VaR}_{p,k}}} \right|. \tag{6.2}$$

Given its definition, the lower the value of the V in absolute term is, the better the ES estimate is.

Table 3 displays the result of the test statistic V. Clearly the ES measures estimated on the stock indices perform equally well in all the specified models.

[3] [11] also proposes 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.

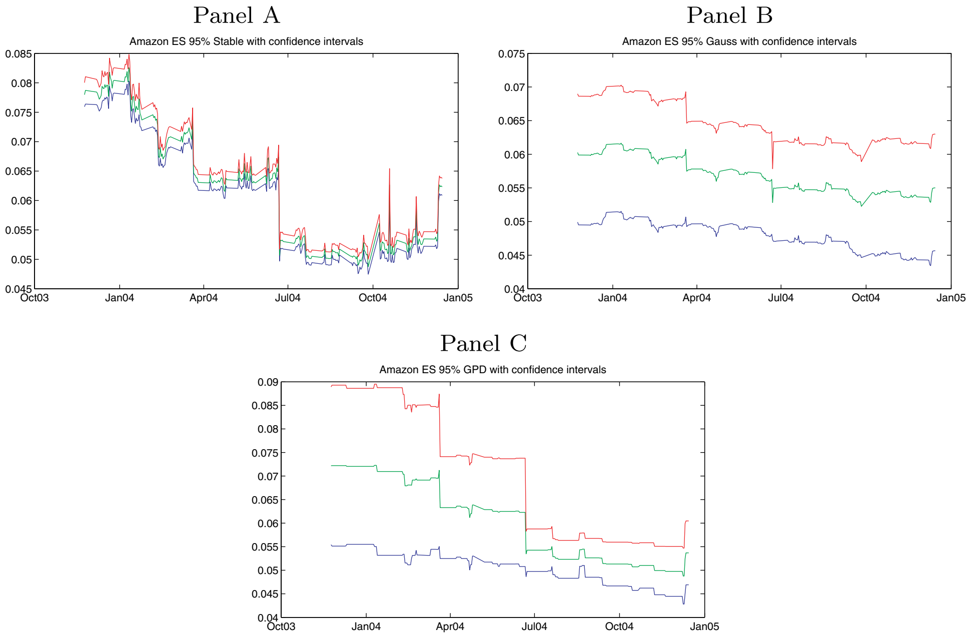


Fig. 8. Expected shortfall 95% with confidence intervals.

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.

The main differences arise in the single stock evaluations; in particular the stable specifications, both at 95% and 99% confidence level, seem to present less accuracy than the other two specifications. This difference is clearer in the Amazon return series, and can be seen as a consequence of the "conservative" nature of the stable estimations (for which the mean realized shortfall is less than the expected shortfall implied by the fitted stable law).

6.1. *Block maxima backtesting*

Finally we also perform a test on a VaR calculated with the block maxima method (BMM) introduced in Sec. 4.1. In practice we calculate the VaR, based on the BMM approach, at two different block sizes: 10 and 25. For each class of assets (indices, stocks and exchange rates), we select the series with the highest estimated kurtosis. As an aid for interpreting the results on BMM, we provide in Fig. 10 both a record development plot and a block maxima plot for the three selected series, with a block size of 25.

Results, reported in Table 4, are quite striking: in all analyzed series the BMM approach is largely "over-conservative", producing VaR bounds that are difficult

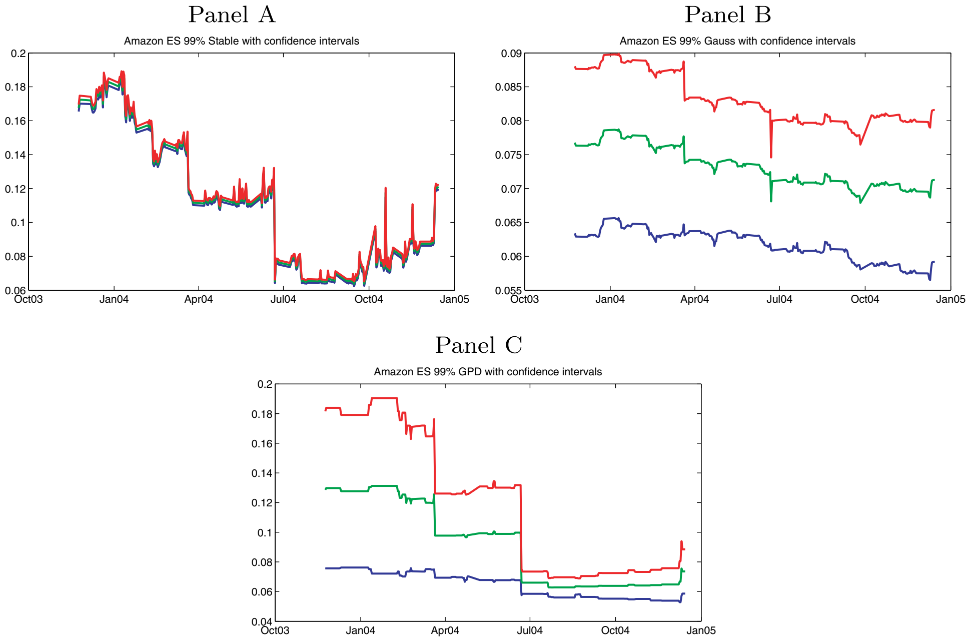


Fig. 9. Expected shortfall 99% with confidence intervals.

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.

to interpret.[4] This leads the log-likelihood ratio test introduced above to strongly reject the model in all the data series and at all the confidence levels.

7. Conclusions

We have compared the properties of some univariate models that are commonly used for purposes of risk management, 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 presents some potentially difficult issues 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

[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.

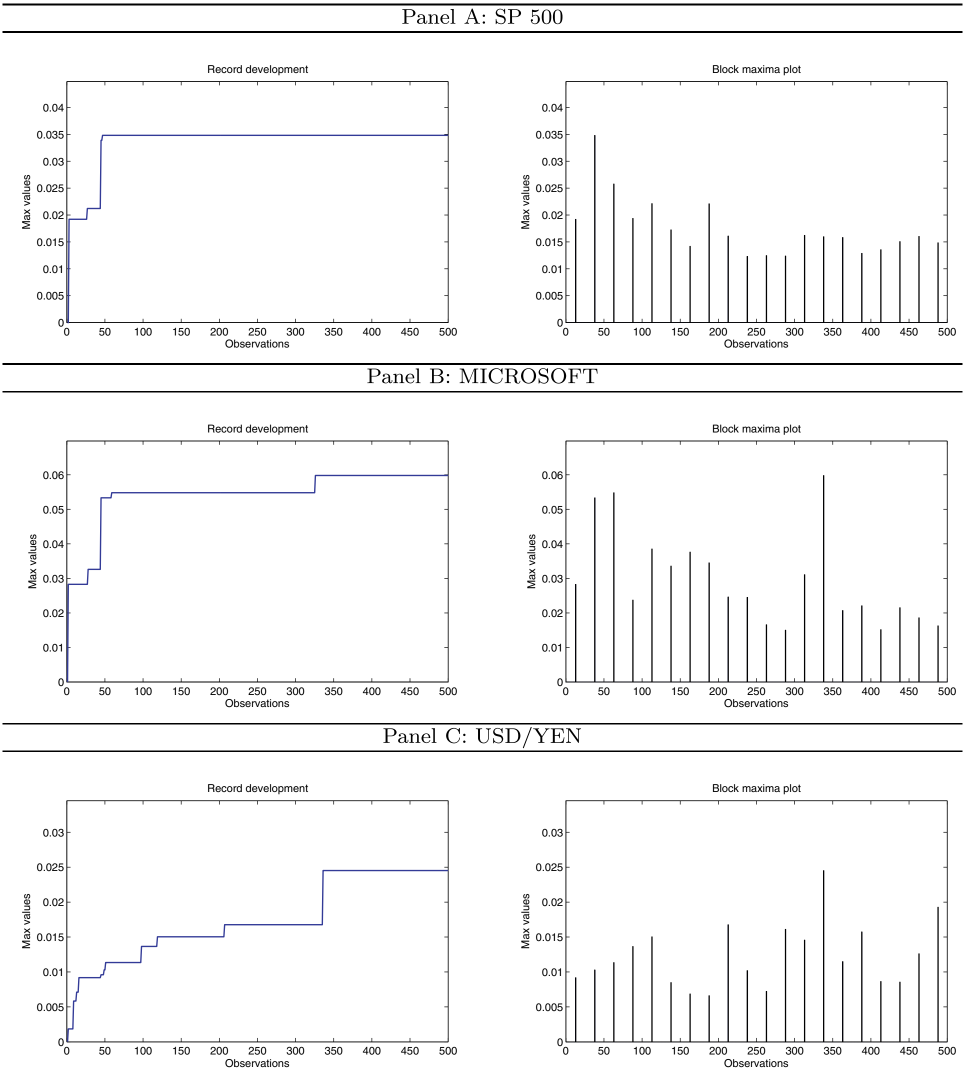


Fig. 10. Block maxima plots.

This figure shows both the record development plots (left hand side figures) and the block maxima plots (right hans side figures). For each of the analyzed categories (indices, stocks and exchange rates) the series with the highest kurtosis is plotted. The size of the blocks is 25 and the last 500 data points of the series re considered, corresponding to roughly 2 years of data.

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 α-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 large amount of observed data, while stable-based ones use all the data points in the time series. In essence, one could say that good fit of the law

Table 4. Block maxima VaR backtesting.

This table reports the results of a Value-at-Risk backtesting on the Block Maxima approach for SP500, Microsoft and USD/YEN series (2-Jan-1990 to 31-Dec-2004). Returns are calculated as log-differences on daily data series. Panel A results are from a block of 10, while Panel B results are from 25 days. 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% and 99.9% rejection of the VAR model.

	Violations	Percentage	LR
	SP 500		
Panel A: 10 days block			
$\text{BMM}_{95\%}$	11	0.3%	277.977**
$\text{BMM}_{99\%}$	3	0.1%	50.106**
Panel B: 25 days block			
$\text{BMM}_{95\%}$	123	3.5%	19.009**
$\text{BMM}_{99\%}$	6	0.2%	37.579**
	MICROSOFT		
Panel A: 10 days block			
$\text{BMM}_{95\%}$	58	3.8%	4.576*
$\text{BMM}_{99\%}$	7	0.5%	5.459*
Panel B: 25 days block			
$\text{BMM}_{95\%}$	237	15.7%	238.606**
$\text{BMM}_{99\%}$	24	1.6%	4.518*
	USD/YEN		
Panel A: 10 days block			
$\text{BMM}_{95\%}$	3	0.1%	330.469**
$\text{BMM}_{99\%}$	1	0.0%	61.632**
Panel B: 25 days block			
$\text{BMM}_{95\%}$	71	2.0%	84.417**
$\text{BMM}_{99\%}$	2	0.1%	55.263**

where there is more mass contributes to good fit in the tail, even though the EVT approach requires less distributional assumptions.

Our empirical analysis does not uniquely identify the "best" approach. However, it definitely provides evidence that α-stable laws outperform the so-called block maxima method for estimating VaR. Especially at 99% level the estimates becomes strongly "over-conservative", with peaks that are somehow difficult to interpret. On the other hand, the POT method is significantly better than the block maxima method, yielding good VaR approximations at 99% levels that are comparable to the α-stable-based estimates. Two simple alternatives, again based on the theory of extreme value distributions, have the advantages of being particularly simple to implement, and yield good VaR estimates. However, they rely on tail estimators which may in general require a large amount of observations to achieve good relative accuracy, and they need tuning as the POT method.

It is worth noting that some of our empirical results seem to be in conflict with a similar analysis presented in [21]. However, as explained above, our methods (especially those involving stable laws) are definitely more precise, as we make extensive use of analytic results without relying on crude Monte Carlo techniques.

In [21] the authors also fit to data symmetric stable laws, but unfortunately they do it in an incorrect way: namely they simply fit a general (skewed) stable law, and set a posteriori $\beta = 0$. As a consequence, their results on the corresponding estimates of risk measures are rather fortuitous. While a correct procedure would just be to fit symmetric stable laws to data, we chose not to consider this method as it is quite artificial, and it does not look meaningful to fit a symmetric distribution to samples with pronounced skewness.

Let us finally remark 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, 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.

Appendix

Proof of Proposition 3.1. *Let us assume for now that* $\sigma = 1$ *and* $\mu = 0$, *and let* $X \sim S_\alpha(1, \beta, 0)$. *Then one has*

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

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 (A.1), 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. □

Acknowledgments

C. Marinelli acknowledges the hospitality and financial support of IMPAN, Warsaw and IHÉS, Bures-sur-Yvette, through an IPDE fellowship. 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. Probab.* **2** (1974) 792–804.

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

[3] B. Bradley and M. Taqqu, Financial risk and heavy tails, *Handbook of Heavy-Tailed Distributions in Finance*, S. T. Rachev, ed. (Elsevier, Amsterdam, 2003), pp. 35–103.

[4] J. Danielsson, L. de Haan, L. Peng and C. G. de Vries, Using a bootstrap method to choose the sample fraction in tail index estimation, *J. Multiv. Analysis* **76** (2001) 226–248.

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

[6] F. X. Diebold, T. Schuermann and J. D. Stroughair, Pitfalls and opportunities in the use of extreme value theory in risk management, Center for Financial Institutions Working Papers 98-10, Wharton School Center for Financial Institutions, University of Pennsylvania (1998).

[7] 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** (2005) 307–312.

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

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

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

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

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

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

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

[15] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time* (Walter de Gruyter & Co., Berlin, 2002).

[16] B. Gamrowski and S. T. Rachev, Financial models using stable laws, in *Applied and Industrial Mathematics*, Y. V. Prohorov, ed. (1995), pp. 556–604.

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

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

[19] D. W. Jansen and C. G. de Vries, On the frequency of large stock returns: Putting booms and busts into perspective, *Review Econ. Stat.* **73** (1991) 18–24.

[20] F. C. Harmantzis, C. Marinelli and S. T. Rachev, A comparative study of some univariate models for value-at-risk, *INFORMS Applied Probability Conference Beijing*, China (2004).

[21] F. C. Harmantzis, L. Miao and Y. Chien, Empirical study of Value-at-Risk and expected shortfall model with heavy tails, *J. of Risk Finance* **7** (2006) 117–135.

[22] P. Jorion, *Value at Risk: The New Benchmark for Controlling Market Risk* (Irwin, 2000).

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

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

[25] J. H. McCulloch, Simple consistent estimators of stable distribution parameters, *Commun. Statist. — Simula.* **15** (1986) 1109–1136.

[26] 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** (2000) 271–300.

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

[28] J. P. Nolan, Numerical calculation of stable densities and distribution functions, *Comm. Statist. Stochastic Models* **13** (1997) 759–774.

[29] J. P. Nolan, User Manual for STABLE 4.0 — Mathematica Version, Robust Analysis, Inc. (2005).

[30] B. L. S. Prakasa Rao, *Asymptotic Theory of Statistical Inference* (John Wiley & Sons Inc., New York, 1987).

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

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

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

[34] R. L. Smith, Estimating tails of probability distributions, *Ann. Statist.* **15** (1987) 1174–1207.

[35] S. Stoyanov, G. Samorodnitsky, S. T. Rachev and S. Ortobelli, Computing the portfolio conditional Value-at-Risk in the α-stable case, to appear in *Probab. Math. Statist.*

[36] A. W. van der Vaart, *Asymptotic Statistics* (Cambridge Univeristy Press, Cambridge, 1998).