Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach

Alexander J. McNeil & Rüdiger Frey*
Departement Mathematik
ETH Zentrum
CH-8092 Zürich
Tel: +41 1 632 61 62
Fax: +41 1 632 15 23
mcneil@math.ethz.ch
frey@math.ethz.ch

April 27, 2000

Abstract

We propose a method for estimating VaR and related risk measures describing the tail of the conditional distribution of a heteroscedastic financial return series. Our approach combines quasi maximum likelihood fitting of GARCH models to estimate the current volatility and extreme value theory (EVT) for estimating the tail of the innovation distribution of the GARCH model. We use our method to estimate conditional quantiles (VaR) and conditional expected shortfalls (the expected size of a return exceeding VaR), this being an alternative measure of tail risk with better theoretical properties than the quantile. Using backtesting we show that our procedure gives better estimates than methods which ignore the heavy tails of the innovations or the stochastic nature of the volatility. With the help of our fitted models and a simulation approach we estimate the conditional quantiles of returns over multiple day horizons and find evidence of a power scaling law, where the power depends in a natural way on the current volatility level.

J.E.L. Subject Classification: C.22, G.10, G.21

Keywords: Risk Measures, Value at Risk, Financial Time Series, GARCH models, Extreme Value Theory, Backtesting

1 Introduction

The large increase in the number of traded assets in the portfolio of most financial institutions has made the measurement of market risk (the risk that a financial institution incurs losses on its trading book due to adverse market movements) a primary concern for regulators and for internal risk control. In particular, banks are now required to hold a certain amount of capital as a cushion against adverse market movements. According to the Capital Adequacy Directive by the Bank of International Settlement (BIS) in Basle, (Basle Comittee 1996) the risk capital of a bank must be sufficient to cover losses on the bank's trading portfolio over a ten-day holding period in 99% of occasions. This value

is usually referred to as *Value at Risk* (VaR). For purposes of internal risk control most financial firms use a holding period of one day and a confidence level of 95%. From a mathematical viewpoint VaR is of course simply a quantile of the Profit-and-Loss (P&L) distribution of a given portfolio over a prescribed holding period.

In two recent papers, Artzner et al. (1997, 1998) have critisized VaR as a measure of market risk on two grounds. First they show that VaR is not necessarily subadditive. They explain that this may cause problems, if one bases a risk-management system of a financial institution on VaR-limits for individual books. Moreover, VaR gives only an upper bound on the losses that occur with a given frequency; VaR tells us nothing about the potential size of the loss given that a loss exceeding this upper bound has occurred. Artzner et al. propose the use of the so-called *expected shortfall* or *tail conditional expectation* instead of VaR. The tail conditional expectation measures the expected loss given that the loss L exceeds VaR; in mathematical terms it is given by $E[L|L > \text{VaR}]$.

From a statistical viewpoint the main challenge in implementing one of these risk-measures is to come up with a good estimate for the tails of the underlying P&L distribution; given such an estimate both VaR and expected shortfall are fairly easy to compute.

In this paper we are concerned with tail estimation for financial return series. Our basic idealisation is that returns follow a stationary time series model with stochastic volatility structure. There is strong empirical support for stochastic volatility in financial time series; see for instance Pagan (1996) or Frey (1997). The presence of stochastic volatility implies that returns are not necessarily independent over time. Hence with such models there are two types of return distribution to be considered – the conditional return distribution where the conditioning is on the current volatility and the marginal or stationary distribution of the process.

Both distributions are of relevance to risk managers. The tails of the conditional return distribution are essentially the object of interest in computing measures of market risk and will therefore be the focus of this paper. The key concern of a risk manager is the possible extent of a loss caused by an adverse market movement over the next day (or next few days) given the current volatility background. The estimation of unconditional tails provides different, but complementary information about risk. Here we take the long-term view and attempt to assign a magnitude to a specified rare adverse event, such as a 5-year or 10-year loss. This kind of information may be of interest to the risk manager who wishes to perform a scenario analysis and get a feeling for the scale of worst case or stress losses.

Schematically the existing approaches for estimating the P&L distribution of a portfolio of securities can be divided into three groups: the nonparametric historical simulation (HS) method; fully parametric methods based on an econometric model for volatility dynamics and the assumption of conditional normality (e.g. J.P. Morgan's Riskmetrics and most models from the ARCH/GARCH family); and finally methods based on extreme value theory (EVT).

In the HS-approach the estimated P&L distribution of a portfolio is simply given by the empirical distribution of past gains and losses on this portfolio. The method is therefore easy to implement and avoids "ad-hoc-assumptions" on the form of the P&L distribution. However, the method suffers from some serious drawbacks. Extreme quantiles are notoriously difficult to estimate, as extrapolation beyond past observations is impossible. Moreover, quantile estimates obtained by HS tend to be very volatile whenever a large observation enters the sample. Finally, the method is unable to distinguish between periods of high and low volatility, in particular if a long data sample is used to mitigate the influence of the first two problems on the quality of the tail-estimate.

The more refined models within the conditional normality approach such as GARCH-models, which model the dynamics of the conditional variance of asset returns, do yield VaR estimates which reflect the current volatility background. The main weakness of

this approach is that the assumption of conditional normality does not seem to hold for real data. As shown for instance in Danielsson and de Vries (1997b), models based on conditional normality are therefore not well-suited for estimating the distribution of large quantiles of the P&L-distribution.[1]

The estimation of return distributions of financial time series via EVT is a topical issue which has given rise to some recent work (Embrechts, Resnick, and Samorodnitsky 1998, Embrechts, Resnick, and Samorodnitsky 1999, Longin 1997b, Longin 1997a, McNeil 1997, McNeil 1998, Danielsson and de Vries 1997a, Danielsson and de Vries 1997b, Danielsson, Hartmann, and de Vries 1998). In all these papers the focus is on estimating the unconditional (stationary) distribution of asset returns. Longin (1997b) and McNeil (1998) use estimation techniques based on limit theorems for block maxima. Longin ignores the stochastic volatility exhibited by most financial return series and simply applies estimators for the iid-case. McNeil uses a more sophisticated estimation technique which corrects for the clustering of extremal events caused by stochastic volatility. Danielsson and de Vries (1997a,b) use a semiparametric approach based on the Hill-estimator for the tail index. Embrechts, Resnick, and Samorodnitsky (1999) advocate the use of a parametric estimation technique which is based on a limit result for the excess-distribution over high thresholds. This approach will be explained in detail in Section 2.2.

EVT-based methods have two features which make them attractive for tail estimation: They are based on a sound statistical theory, and they offer a parametric form for the tail of a distribution. Hence these methods allow for some extrapolation beyond the range of the data, even if care is required at this point. However, none of the previous EVT-based methods for quantile estimation yields VaR-estimates which reflect the current volatility background. Given the conditional heteroscedasticity of most financial data, which is well-documented by the considerable success of the models from the ARCH/GARCH family, we believe this to be a major drawback of any kind of VaR-estimator.

In order to overcome the drawbacks of each of the above methods we combine ideas from all three approaches. We use GARCH-modelling and pseudo-maximum-likelihood estimation to obtain estimates for the conditional volatility. Statistical tests and exploratory data analysis confirm that the error terms or residuals do form at least approximately an iid series which exhibits heavy tails. We use historical simulation (for the central part of the distribution) and threshold methods from EVT (for the tails) to estimate the distribution of the error terms. The application of these methods is facilitated by the (approximate) independence over time of the residuals. An estimate of the conditional return distribution is now easily constructed from the distribution of the residuals and our estimates of the conditional mean and volatility. This approach reflects two stylized facts exhibited by most financial return series, namely stochastic volatility and the fat-tailedness of conditional return distributions over short time horizons.

In a very recent paper Barone-Adesi, Bourgoin, and Giannopoulos (1998) have independently proposed an approach with some similarities to our own. They fit a GARCH-model to a financial return series and use historical simulation to infer the distribution of the residuals. They do not use EVT-based methods to estimate the tails of the distribution of the residuals. Their approach may work well in large data sets — they use 13 years of daily data — where the empirical quantile provides a reasonable quantile estimator in the tails. With smaller data sets threshold methods from EVT will give better estimates of the tails of the residuals.

We test our approach on various return series. Backtesting shows that it yields better estimates of VaR and expected shortfall than unconditional EVT or GARCH-modelling with normally distributed error terms. In particular, our analysis contradicts Danielsson

---

[1]Note that the *marginal* distribution of a GARCH-model with normally distributed errors is usually fat-tailed as it is a a mixture of normal distributions. However, this matters only for quantile estimation over longer time-horizons; see e.g. Duffie and Pan (1997).

and de Vries (1997b), who state that "an unconditional approach is better suited for VaR estimation than conditional volatility forecasts" (page 3 of their paper). On the other hand, we see that models with normally distributed conditional return distribution yield very bad estimates of the expected shortfall, so that there is a real need for working with leptokurtic error distributions. We also study quantile estimation over longer time-horizons using simulation. This is of interest, if one wants to obtain an estimate of the 10-day VaR (as required by the BIS-rule) from a model fitted to daily data. We find that according to our models the return over k days (from day t to day $t + k$ say) can be obtained by multiplying the quantiles of the one day return by k^{λ_t} with scaling exponent λ_t depending on the value of the volatility at t. As explained in Section 5 this casts some doubts on the usefulness in a VaR context of a scaling-law postulated by Danielsson and de Vries (1997b).

2 Methods

Let $(X_t, t \in \mathbb{Z})$ be a strictly stationary time series representing daily observations of the negative log return on a financial asset price.[2] We assume that the dynamics of X are given by

$$X_t = \mu_t + \sigma_t Z_t, \tag{1}$$

where the innovations Z_t are a strict white noise process (i.e. independent, identically distributed) with zero mean, unit variance and marginal distribution function $F_Z(z)$. We assume that μ_t and σ_t are measurable with respect to $\mathcal{G}_{t-1}$, the information about the return process available up to time $t - 1$.

Let $F_X(x)$ denote the marginal distribution of (X_t) and let $F_{X_{t+1}+...+X_{t+k}|\mathcal{G}_t}(x)$ denote the predictive distribution of the return over the next k days, given knowledge of returns up to and including day t. We are interested in estimating quantiles in the tails of these distributions. For $0 < q < 1$, an unconditional quantile is a quantile of the marginal distribution denoted by

$$x_q = \inf \left\{ x \in \mathbb{R} : F_X(x) \geq q \right\},$$

and a conditional quantile is a quantile of the predictive distribution for the return over the next k days denoted by

$$x_q^t(k) = \inf \left\{ x \in \mathbb{R} : F_{X_{t+1}+...+X_{t+k}|\mathcal{G}_t}(x) \geq q \right\}.$$

We also consider an alternative measure of risk for the tail of a distribution known as the expected shortfall. The unconditional expected shortfall is defined to be

$$S_q = E\left[X \mid X > x_q\right],$$

and the conditional expected shortfall to be

$$S_q^t(k) = E\left[\sum_{j=1}^{k} X_{t+j} \mid \sum_{j=1}^{k} X_{t+j} > x_q^t(k), \mathcal{G}_t \right].$$

We are principally interested in quantiles and expected shortfalls for the 1–step predictive distribution, which we denote respectively by x_q^t and S_q^t. Since

$$\begin{aligned} F_{X_{t+1}|\mathcal{G}_t}(x) &= P\left\{ \sigma_{t+1} Z_{t+1} + \mu_{t+1} \leq x \mid \mathcal{G}_t \right\} \\ &= F_Z((x - \mu_{t+1})/\sigma_{t+1}), \end{aligned}$$

---

[2]In the present paper we test our approach on return series generated by single assets only. However, the method obviously also applies to the time series of profits and losses generated by portfolios of financial instruments and can therefore by used for the estimation of market risk measures in a portfolio context.

these measures simplify to

$$x_q^t = \mu_{t+1} + \sigma_{t+1} z_q, \tag{2}$$
$$S_q^t = \mu_{t+1} + \sigma_{t+1} E[Z \mid Z > z_q], \tag{3}$$

where z_q is the upper qth quantile of the marginal distribution of Z_t which by assumption does not depend on t.

To implement an estimation procedure for these measures we must choose a specific process in the class (1), i.e. a particular model for the dynamics of the conditional mean and volatility. Many different models for volatility dynamics have been proposed in the econometric literature including models from the ARCH/GARCH family (Bollerslev, Chou, and Kroner 1992), HARCH processes (Müller, Dacarogna, Davé, Olsen, Pictet, and von Weizsäcker 1997) and stochastic volatility models (Shephard 1996). In this paper we use the parsimonious but effective GARCH(1,1) process for the volatility and an AR(1) model for the dynamics of the conditional mean; the approach we propose extends easily to more complex models.

In estimating x_q^t with GARCH-type models it is commonly assumed that the innovation distribution is standard normal so that a quantile of the innovation distribution is simply $z_q = \Phi^{-1}(q)$, where $\Phi(z)$ is the standard normal df. A GARCH-type model with normal innovations can be fitted by maximum likelihood (ML) and μ_{t+1} and σ_{t+1} can be estimated using standard 1–step forecasts, so that an estimate of x_q^t is easily constructed using (3). This is close in spirit to the approach advocated in Risk Metrics, but our empirical finding, which we will later show, is that this approach often underestimates the conditional quantile for $q > 0.95$; the distribution of the innovations seems generally to be heavier-tailed or more leptokurtic than the normal.

Another standard approach is to assume that the innovations have a leptokurtic distribution such as Student's t–distribution (scaled to have variance 1). Suppose $Z = \sqrt{(\nu-2)/\nu} T$ where T has a t–distribution on $\nu > 2$ degrees of freedom with df $F_T(t)$. Then $z_q = \sqrt{(\nu-2)/\nu} F_T^{-1}(q)$. GARCH-type models with t–innovations can also be fitted with maximum likelihood and the additional parameter ν can be estimated. We will see in Section 2.2 that this method can be viewed as a special case of our approach; it yields quite satisfactory results as long as the positive and the negative tail of the return distribution are (roughly) equal.

The method proposed in this paper makes minimal assumptions about the underlying innovation distribution and concentrates on modelling its tail using extreme value theory (EVT). We use a two stage approach which can be summarised as follows.

1. Fit a GARCH-type model to the return data making no assumption about $F_Z(z)$ and using a pseudo maximum likelihood approach (PML). Estimate μ_{t+1} and σ_{t+1} using the fitted model and calculate the implied model residuals.

2. Consider the residuals to be a realisation of a strict white noise process and use extreme value theory (EVT) to model the tail of $F_Z(z)$. Use this EVT model to estimate z_q for $q > 0.95$.

We go into these stages in more detail in the next two sections and illustrate them by means of an example using daily negative log returns on the Standard & Poors index.

2.1 Estimating σ_{t+1} and μ_{t+1} using PML

For predictive purposes we fix a constant memory n so that on day t our data consist of the last n negative log returns $(x_{t-n+1}, \ldots, x_{t-1}, x_t)$. We consider these to be a realisation

from a AR(1)–GARCH(1,1) process. Hence the conditional variance of the mean-adjusted series $\epsilon_t = X_t - \mu_t$ is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{4}$$

where $\alpha_0 > 0$, $\alpha_1 > 0$ and $\beta > 0$. The conditional mean is given by

$$\mu_t = \phi X_{t-1}. \tag{5}$$

This model is a special case of the general first order stochastic volatility process considered by Duan (1997), who uses a result by Brandt (1986) to give conditions for strict stationarity. The mean-adjusted series (ϵ_t) is strictly stationary if

$$E\left[\log\left(\beta + \alpha_1 Z_{t-1}^2\right)\right] < 0. \tag{6}$$

By using Jensen's inequality and the convexity of $-\log(x)$ it is seen that a sufficient condition for (6) is that $\beta + \alpha_1 < 1$, which moreover ensures that the marginal distribution $F_X(x)$ has a finite second moment.

Three model is fitted using the pseudo-maximum-likelihood (PML) method. This means that the likelihood for a GARCH(1,1) model with normal innovations is maximized to obtain parameter estimates $\hat{\theta} = (\hat{\phi}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta})^T$. Whilst this amounts to fitting a model using a distributional assumption we do not necessarily believe, the PML method delivers reasonable parameter estimates. In fact, it can be shown that the PML method yields a consistent and asymptotically normal estimator; see for instance Chapter 4 of Gouriéroux (1997).

Estimates of the conditional mean and standard deviation series $(\hat{\mu}_{t-n+1}, \dots, \hat{\mu}_t)$ and $(\hat{\sigma}_{t-n+1}, \dots, \hat{\sigma}_t)$ can be calculated recursively from (4) and (5) after substitution of sensible starting values. In Figure 1 we show an arbitrary thousand day excerpt from our dataset containing the stock market crash of October 1987; the estimated conditional standard deviation derived from the GARCH fit is shown below the series.

Residuals are calculated both to check the adequacy of the GARCH modelling and to use in Stage 2 of the method. They are calculated as

$$(z_{t-n+1}, \dots, z_t) = \left(\frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \dots, \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}\right),$$

and should be iid if the fitted model is tenable. In Figure 2 we plot correlograms for the raw data and their squared values as well as for the residuals and squared residuals. While the raw data are clearly not iid, this assumption may be tenable for the residuals.[3] The stationarity of the fitted model can be checked by verifying that $\hat{\beta} + \hat{\alpha}_1 < 1$.

If we are satisfied with the fitted model, we end stage 1 by calculating estimates of the conditional mean and variance for day $t + 1$, which are the obvious 1–step forecasts

$$\begin{aligned}
\hat{\mu}_{t+1} &= \hat{\phi} x_t, \\
\hat{\sigma}_{t+1}^2 &= \widehat{\alpha_0} + \widehat{\alpha_1} \hat{\epsilon}_t^2 + \hat{\beta} \hat{\sigma}_t^2,
\end{aligned}$$

where $\hat{\epsilon}_t = x_t - \hat{\mu}_t$.

2.2 Estimating z_q using EVT

We begin stage 2 by forming a QQ–Plot of the residuals against the normal distribution to confirm that an assumption of conditional normality is unrealistic, and that the innovation process has fat tails or is leptokurtic – see Figure 3.

---

[3]We also ran some Ljung-Box tests in selected time periods and found no evidence against the iid–hypothesis for the residuals.

We then fix a high threshold u and we assume that excess residuals over this threshold have a generalized Pareto distribution (GPD) with df

$$G_{\xi,\beta}(y) = \begin{cases} 1 - (1 + \xi y/\beta)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-y/\beta) & \text{if } \xi = 0, \end{cases}$$

where $\beta > 0$, and the support is $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\beta/\xi$ when $\xi < 0$.

This particular distributional choice is motivated by a limit result in EVT. Consider a general df F and the corresponding excess distribution above the threshold u given by

$$F_u(y) = P\left\{X - u \leq y \mid X > u\right\} = \frac{F(y + u) - F(u)}{1 - F(u)},$$

for $0 \leq y < x_0 - u$, where x_0 is the (finite or infinite) right endpoint of F. Balkema and de Haan (1974) and Pickands (1975) showed for a large class of distributions F that it is possible to find a positive measurable function $\beta(u)$ such that

$$\lim_{u \to x_0} \sup_{0 \leq y < x_0 - u} |F_u(y) - G_{\xi,\beta(u)}(y)| = 0. \tag{7}$$

For more details consult Theorem 3.4.13 on page 165 of Embrechts, Klüppelberg, and Mikosch (1997).

In the class of distributions for which this result holds are essentially all the common continuous distributions of statistics, and these may be further subdivided into three groups according to the value of the parameter ξ in the limiting GPD approximation to the excess distribution. The case $\xi > 0$ corresponds to the heavy-tailed distributions whose tails decay like power functions such as the Pareto, Student's t, Cauchy, Burr, loggamma and Fréchet distributions. The case $\xi = 0$ corresponds to distributions like the normal, exponential, gamma and lognormal, whose tails decay exponentially; we call such distributions thin-tailed. The final group of distributions are short-tailed distributions ($\xi < 0$) with a finite right endpoint like the uniform and beta distributions.

We assume the the tail of the underlying distribution begins at the threshold u. From our sample of n points a random number $N = N_u > 0$ will exceed this threshold. If we assume that the N excesses over the threshold are iid with exact GPD distribution, Smith (1987) has shown that maximum likelihood estimates $\hat{\xi}_N$ and $\hat{\beta}_N$ of the GPD parameters ξ and β are consistent and asymptotically normal as $N \to \infty$, provided $\xi > -1/2$. Under the weaker assumption that the excesses are iid from $F_u(y)$ which is only approximately GPD he also obtains asymptotic normality results for $\hat{\xi}_N$ and $\hat{\beta}_N$. By letting $u = u_n \to x_0$ and $N = N_u \to \infty$ as $n \to \infty$ he shows essentially that the procedure is asymptotically unbiased provided that $u \to x_0$ sufficiently fast. The necessary speed depends on the rate of convergence in (7). In practical terms this means that our best GPD estimator of the excess distribution is obtained by trading bias off against variance. We choose u high to reduce the chance of bias whilst keeping N large (i.e. u low) to control the variance of the parameter estimates.

Consider now the following equality for points $x > u$ in the tail of F

$$1 - F(x) = (1 - F(u))(1 - F_u(x - u)). \tag{8}$$

If we estimate the first term on the right hand side of (8) using the random proportion of the data in the tail N/n, and if we estimate the second term by approximating the excess distribution with a generalized Pareto distribution fitted by maximum likelihood, we get the tail estimator

$$\hat{F}(x) = 1 - \frac{N}{n}\left(1 + \hat{\xi}_N \frac{x - u}{\hat{\beta}_N}\right)^{-1/\hat{\xi}_N},$$

7

for $x > u$. Smith (1987) also investigates the asymptotic relative error of this estimator and gets a result of the form

$$N^{1/2} \left(\frac{1 - \hat{F}(x)}{1 - F(x)} - 1 \right) \overset{d}{\to} N(0, v^2),$$

as $u = u_n \to x_0$ and $N = N_u \to \infty$, where the asymptotic unbiasedness again requires that $u \to x_0$ sufficiently fast.

In practice we will actually modify the procedure slightly and fix the number of data in the tail to be $N = k$ where $k \ll n$. This effectively gives us a random threshold at the $(k + 1)$th order statistic. Let $z_{(1)} \geq z_{(2)} \geq \ldots \geq z_{(n)}$ represent the ordered residuals. The generalized Pareto distribution with parameters ξ and β is fitted to the data $(z_{(1)} - z_{(k+1)}, \ldots, z_{(k)} - z_{(k+1)})$, the excess amounts over the threshold for all residuals exceeding the threshold. The form of the tail estimator for $F_Z(z)$ is then

$$\hat{F}_Z(z) = 1 - \frac{k}{n} \left(1 + \hat{\xi}_k \frac{z - z_{(k+1)}}{\hat{\beta}_k} \right)^{-1/\hat{\xi}_k}. \tag{9}$$

For $q > 1 - k/n$ we can invert this tail formula to get

$$\widehat{z_q} = z_{(k+1)} + \frac{\hat{\beta}_k}{\hat{\xi}_k} \left(\left(\frac{1 - q}{k/n} \right)^{-\hat{\xi}_k} - 1 \right). \tag{10}$$

In Table 1 we give threshold values and GPD parameter estimates for both tails of the innovation distribution of the test data in the case that $k = 100$. In Figure 4 we show the corresponding tail estimators (9). We are principally interested in the left picture marked *Losses* which corresponds to large positive residuals. The solid lines in both pictures correspond to the GPD tail estimates and can be seen to model the residuals well. Also shown is a dashed line which corresponds to the standard normal distribution and a dotted line which corresponds to the estimated conditional t distribution in a GARCH model with t innovations. The normal distribution clearly underestimates the extent of large losses and also of the largest gains, which we would already expect from the QQ–plot. The t distribution, on the other hand, underestimates the losses and overestimates the gains. This illustrates the drawbacks of using a symmetric distribution with data which are asymmetric in the tails.

Tail	$z_{(k+1)}$	$\hat{\xi}$	s.e.	$\hat{\beta}$	s.e.
Losses	1.215	0.224	(0.122)	0.568	(0.089)
Gains	1.120	-0.096	(0.090)	0.589	(0.079)

Table 1: Threshold values and maximum likelihood GPD parameter estimates used in the construction of tail estimators for both tails of the innovation distribution of the test data. Note that $k = 100$ in both cases. Standard errors (s.e.s) are calculated using a standard likelihood approach based on the observed Fisher information matrix.

With more symmetric data the conditional t distribution often works quite well and it can, in fact, be viewed as a special case of our method. As already mentioned, it is an example of a heavy-tailed distribution, i.e. a distribution whose limiting excess distribution is GPD with $\xi > 0$. Gnedenko (1943) characterized all such distributions as having tails of the form

$$1 - F(x) = x^{-1/\xi} L(x), \tag{11}$$

where $L(x)$ is a slowly varying function and ξ is the positive parameter of the limiting GPD. $1/\xi$ is often referred to as the *tail index* of F. For the t distribution with ν degrees of freedom the tail can be shown to satisfy

$$1 - F(x) \sim \nu^{(\nu-1)/2} x^{-\nu},$$

so that this provides a very simple example of a symmetric distribution in this class, and the value of ξ in the limiting GPD is the reciprocal of the degrees of freedom (McNeil and Saladin 1997).

Fitting a GARCH model with t innovations can be thought of as estimating the ξ in our GPD tail estimator by simpler means. Inspection of the form of the likelihood of the t–distribution shows that the estimate of ν will be sensitive mainly to large observations so that it is not surprising that the method gives a reasonable fit in the tails although all data are used in the estimation. Our method has, however, the advantage that we have an explicit model for each tail. We estimate two parameters in each case, which gives a better fit in general.

We will also use the GPD tail estimator (9) to estimate the right tail of the negative return distribution $F_X(x)$ and to calculate the unconditional quantile estimate $\hat{x}_q$, an approach that we will call unconditional EVT. We investigate whether this estimate also provides a reasonable estimate of x_q^t. We note however that the assumption of independent excesses over threshold is much less satisfactory for the raw return data. The asymptotics of the procedure are much more poorly understood if applied directly to the raw return data. Even if the procedure can be shown to be theoretically justified, in practice it is likely to give much more unstable results when applied to non–iid data (see Figure 5.5.4. on page 270 of Embrechts, Klüppelberg, and Mikosch (1997) for a related example).

3 Backtesting

We backtest the method on five historical series of log returns: the Standard & Poors index from January 1960 to June 1993, the DAX index from January 1973 to July 1996, the BMW share price over the same period, the US dollar British pound exchange rate from January 1980 to May 1996 and the price of gold from January 1980 to December 1997.

To backtest the method on a historical series $x_1, \ldots x_m$, where $m \gg n$, we calculate $\hat{x}_q^t$ on days t in the set $T = \{n, \ldots, m-1\}$ using a time window of n days each time. In our implementation we have set $n = 1000$ so that we use somewhat less than the last four years of data for each prediction. We always set the constant $k = 100$ so that the largest 100 residuals are considered to come from the tail of the innovation distribution. This means effectively that the 90th percentile of the innovation distribution is estimated by historical simulation, but that higher percentiles are estimated using the GPD tail estimator. On each day $t \in T$ we fit a new AR(1)-GARCH(1,1) model and determine a new GPD tail estimate. Figure 5 shows part of the backtest for the DAX index. We have plotted the negative log returns for a three year period commencing on the first of October 1987; superimposed on this plot is the EVT conditional quantile estimate $\hat{x}_{0.99}^t$ (dashed line) and the EVT unconditional quantile estimate $\hat{x}_{0.99}$ (dotted line).

We compare $\hat{x}_q^t$ with x_{t+1} for $q \in \{0.95, 0.99, 0.995\}$. A violation is said to occur whenever $x_{t+1} > \hat{x}_q^t$. The violations corresponding to the backtest in Figure 5 are shown in Figure 6. We use different plotting symbols to show violations of the conditional EVT, conditional normal and unconditional EVT quantile estimates. In Figure 7 the portion of Figure 6 relating to the crash of October 1987 has been enlarged.

It is possible to develop a binomial test of the success of these quantile estimation methods based on the number of violations. If we assume the dynamics described in (1),

	S&P	DAX	BMW	USDGBP	Gold
Length of Test	7414	5146	5146	3274	3413
0.95 Quantile					
Expected	371	257	257	164	171
Conditional EVT	366 (0.41)	258 (0.49)	261 (0.42)	151 (0.16)	155 (0.12)
Conditional Normal	384 (0.25)	238 (0.11)	210 (0.00)	169 (0.35)	122 (0.00)
Conditional t	404 (0.04)	253 (0.41)	245 (0.23)	186 (0.04)	168 (0.44)
Unonditional EVT	402 (0.05)	266 (0.30)	251 (0.36)	156 (0.29)	131 (0.00)
0.99 Quantile					
Expected	74	51	51	33	34
Conditional EVT	73 (0.48)	55 (0.33)	48 (0.35)	35 (0.37)	25 (0.06)
Conditional Normal	104 (0.00)	74 (0.00)	86 (0.00)	56 (0.00)	43 (0.08)
Conditional t	78 (0.34)	61 (0.11)	52 (0.49)	40 (0.12)	29 (0.22)
Unonditional EVT	86 (0.10)	59 (0.16)	55 (0.33)	35 (0.37)	25 (0.06)
0.995 Quantile					
Expected	37	26	26	16	17
Conditional EVT	43 (0.18)	24 (0.42)	29 (0.28)	21 (0.15)	18 (0.44)
Conditional Normal	63 (0.00)	44 (0.00)	57 (0.00)	41 (0.00)	33 (0.00)
Conditional t	45 (0.11)	32 (0.13)	18 (0.07)	21 (0.15)	20 (0.27)
Unonditional EVT	50 (0.02)	36 (0.03)	31 (0.17)	21 (0.15)	11 (0.08)

Table 2: Backtesting Results: Theoretically expected number of violations and number of violations obtained using our approach (conditional EVT), a GARCH-model with normally distributed innovations, a GARCH-model with Student t innovations, and quantile estimates obtained from unconditional EVT for various return series. p-values for a binomial test are given in brackets.

the indicator for a violation at time $t \in T$ is Bernoulli

$$I_t := 1_{\{X_{t+1} > x_q^t\}} = 1_{\{Z_{t+1} > z_q\}} \sim Be(1-q).$$

Moreover, I_t and I_s are independent for $t, s \in T$ and $t \neq s$, since Z_{t+1} and Z_{s+1} are independent. Therefore

$$\sum_{t \in T} I_t \sim B\left(\text{card}(T), 1-q\right),$$

i.e. the total number of violations is binomially distributed under the model.

Under the null hypothesis that a method correctly estimates the conditional quantiles, the empirical version of this statistic $\sum_{t \in T} 1_{\{x_{t+1} > \hat{x}_q^t\}}$ is from the binomial distribution $B\left(\text{card}(T), 1-q\right)$. If we count more violations than the expected number $(1-q)\text{card}(T)$ we perform a one-sided binomial test of the null hypothesis against the alternative that the method systematically underestimates the conditional quantile. If we count less violations than expected we perform a one-sided binomial test of the null hypothesis against the alternative that the method systematically overestimates the conditional quantile. The corresponding binomial probabilities are given in Table 1 alongside the numbers of violations for each method. A p-value less than 0.05 is interpreted as evidence against the null hypothesis.

In 11 out of 15 cases our approach is closest to the mark. On two occasions GARCH with conditional t innovations is best and on one occasion GARCH with conditional normal innovations is best. In one further case our approach and the conditional t approach are joint best. On no occasion does our approach fail (lead to rejection of the null hypothesis),

whereas the conditional t approach fails twice and the conditional normal approach 11 times. Unconditional EVT fails three times. Figures 6 and 7 give some idea of how the latter two methods fail. The conditional normal estimate of $x_{0.99}^t$ like the conditional EVT estimate responds to changing volatility but tends to be violated rather more often, because it does not take into account the leptokurtosis of the residuals. The unconditional EVT estimate cannot respond quickly to changing volatility and tends to be violated several times in a row in stress periods.

4   Expected Shortfall

The expected shortfall, as defined in Section 2, is an alternative risk measure to the quantile which overcomes some of the theoretical deficiencies of the latter; see Artzner, Delbaen, Eber, and Heath (1999). In particular this risk measure gives some information about the size of the potential losses given that a loss bigger than VaR has occurred. We therefore expect this risk measure to be particularly sensitive with respect to the choice of the model for the tail of the return distribution.

4.1   Estimation

We recall from (3) that the conditional (1–step) expected shortfall is given by

$$S_q^t = \mu_{t+1} + \sigma_{t+1} E\left[Z \mid Z > z_q\right].$$

To estimate this risk measure we require an estimate of the expected shortfall for the innovation distribution $E\left[Z \mid Z > z_q\right]$. For a random variable W with an exact GPD distribution with parameters $\xi < 1$ and β it can be verified that

$$E\left[W \mid W > w\right] = \frac{w + \beta}{1 - \xi}, \tag{12}$$

where $\beta + w\xi > 0$. Suppose that excesses over the threshold u have exactly this distribution, i.e. $Z - u \mid Z > u \sim \mathrm{GPD}(\xi, \beta)$. By noting that for $z_q > u$ we can write

$$Z - z_q \mid Z > z_q = (Z - u) - (z_q - u) \mid (Z - u) > (z_q - u),$$

it can be easily shown that

$$Z - z_q \mid Z > z_q \sim \mathrm{GPD}(\xi, \beta + \xi(z_q - u)), \tag{13}$$

so that excesses over the higher threshold z_q also have a GPD distribution with the same shape parameter ξ but a different scaling parameter. We can use (12) to get

$$E\left[Z \mid Z > z_q\right] = z_q \left(\frac{1}{1 - \xi} + \frac{\beta - \xi u}{(1 - \xi) z_q}\right). \tag{14}$$

This is estimated in the obvious way by using the quantile estimator in (10) and replacing ξ and β by GPD parameter estimates and u by $z_{(k+1)}$. This gives us the conditional expected shortfall estimate

$$\hat{S}_q^t = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{z}_q \left(\frac{1}{1 - \hat{\xi}_k} + \frac{\hat{\beta}_k - \hat{\xi}_k z_{(k+1)}}{(1 - \hat{\xi}_k)\hat{z}_q}\right). \tag{15}$$

4.2 Expected shortfall to quantile ratio for various distributions

From (3) we see that, for μ_{t+1} small, the conditional one–step quantiles and shortfalls of the return process are related by

$$\frac{S_q^t}{x_q^t} \approx \frac{S_q^t - \mu_{t+1}}{x_q^t - \mu_{t+1}} = \frac{E\left[Z \mid Z > z_q\right]}{z_q}.$$

Thus the relationship is essentially determined by the ratio of shortfall to quantile for the noise distribution.

It is instructive to compare (14) with the expected shortfall to quantile ratio in the case when the innovation distribution $F_Z(z)$ is standard normal. In this case

$$E\left[Z \mid Z > z_q\right] = \kappa(z_q), \tag{16}$$

where $\kappa(x) = \phi(x)/(1 - \Phi(x))$ is the reciprocal of Mill's ratio and $\phi(x)$ and $\Phi(x)$ are the density and df of the standard normal distribution. Mill's ratio is available to high accuracy in most statistics packages. To get a feeling for the ratio we can use the asymptotic expression

$$\kappa(x) = x\left(1 + x^{-2} + o(x^{-2})\right),$$

as $x \to \infty$, from which it is clear that the expected shortfall to quantile ratio converges to one as $q \to 1$. This can be compared with the limit in the GPD cases; for $\xi > 0$ the ratio under the GPD assumption converges to $(1 - \xi)^{-1} > 1$ as $q \to 1$. For the kind of values of q which interest us we note in passing that a good approximation to the reciprocal of Mill's ratio is $\kappa(x) \approx x\left(1 + (\sqrt{1 + 8/x^2} - 1)/4\right)$; see Johnson and Kotz (1970) for details.

In Table 3 we give values for $E\left[Z \mid Z > z_q\right]/z_q$, the expected shortfall to quantile ratio for the innovation distribution, in both the GPD and normal cases. For the value of the threshold u and the GPD parameters ξ and β we have taken the values obtained from our analysis of the positive residuals from our test data (see Table 1). The table shows that when the innovation distribution is heavy–tailed the expected shortfall to quantile ratio is considerably larger than would be expected under an assumption of normality.

q	0.95	0.99	0.995	$q \to 1$
GPD	1.52	1.42	1.39	1.29
Normal	1.25	1.15	1.12	1.00

Table 3: Values of the expected shortfall to quantile ratio for various quantiles of the noise distribution under two different distributional assumptions. In the first row we assume that excesses over the threshold $u = 1.215$ have an exact GPD distribution with parameters $\xi = 0.224$ and $\beta = 0.568$ (see Table 1). In the second row we assume that the innovation distribution is standard normal.

In the case of the scaled t–distribution, where $F_Z(z) = F_T(z\sqrt{\nu/(\nu - 2)})$, and $T \sim t_\nu$, we can again derive an asymptotic formula

$$E\left[Z \mid Z > z_q\right] = z_q \left(\frac{\nu}{\nu - 1}\left(1 + o(\nu^{-1})\right)\right), \tag{17}$$

where $z_q = \sqrt{(\nu - 2)/\nu}\, F_T^{-1}(q)$ as $q \to 1$. Thus the expected shortfall to quantile ratio converges to $\frac{\nu}{\nu-1}$ as $q \to 1$. As we have already remarked, the value of ξ in the limiting GPD approximation to the excess distribution for a t–distribution with ν degrees of freedom is $1/\nu$, so that the asymptotic result (17) is clearly in line with (14). The results for the t–distribution being similar to those for the GPD-approximation we concentrate on GPD and normal tails in our analyses.

4.3 Backtesting

In Figure 8 we have estimated the expected shortfalls S_q^t, $q \in \{0.95, 0.99\}$, for the BMW series under both the GPD tail assumption and the normal assumption for the innovation distribution. We show the ratios of the GPD–based estimate to the normal estimate. In all days in the backtest the calculated ratio was greater than one. For $q = 0.95$ the ratio reaches values of around 1.2; for $q = 0.99$ the ratio reaches values of around 1.7. This ratio is mainly driven by the estimated value of ξ in the GPD tail estimate, which is clear from the form of (14). Although the conditional 0.95 quantile estimates derived under the GPD and normal assumptions typically do not differ greatly, we see that the same is not true of estimates of the expected shortfall at this quantile. It is thus much more problematic to base estimates of the conditional expected shortfall at even the 0.95 quantile on an assumption of conditional normality when there is evidence that the residuals are heavy–tailed.

We develop a test along similar lines to the binomial test of quantile violation to verify that the GPD–based method gives much better estimates of the conditional expected shortfall than the normal method for our datasets. This time we are interested in the size of the discrepancy between X_{t+1} and S_q^t in the event of quantile violation. We define residuals

$$R_{t+1} = \frac{X_{t+1} - S_q^t}{\sigma_{t+1}} = Z_{t+1} - E\left[Z \mid Z > z_q\right].$$

It is clear that under our model (1) these residuals are iid and that, conditional on $\left\{X_{t+1} > x_q^t\right\}$ or equivalently $\{Z_{t+1} > z_q\}$, they have expected value zero.

It is also possible to standardize these residuals. If we take

$$R_{t+1}^* = R_{t+1} / \sqrt{\text{var}(Z \mid Z > z_q)},$$

these residuals are iid and, conditional on quantile violation, have mean zero and variance one. If we assume that $Z - u \mid Z > u \sim \text{GPD}(\xi, \beta)$, then these residuals have a shifted GPD distribution. With the help of (13) we can show that

$$\text{var}(Z \mid Z > z_q) = \frac{(\beta + \xi(z_q - u))^2}{(1 - \xi)^2(1 - 2\xi)}, \tag{18}$$

for $z_q > u$ and $\xi < 0.5$. We also require an expression for this conditional variance when Z has a normal distribution. In this case

$$\text{var}(Z \mid Z > z_q) = 1 + z_q \kappa(z_q) - \kappa(z_q)^2, \tag{19}$$

where $\kappa(x)$ is again Mill's ratio.

Suppose we again backtest on days in the set T. We can form empirical versions of these residuals on days on which $x_{t+1} > x_q^t$. We will call these residuals exceedance residuals and denote them by

$$\left\{r_{t+1}^* : \ t \in T, \ x_{t+1} > \hat{x}_q^t\right\}, \ \text{where} \ r_{t+1}^* = \frac{x_{t+1} - \hat{S}_q^t}{\hat{\sigma}_{t+1}\sqrt{\widehat{\text{var}}(Z \mid Z > z_q)}}.$$

Here $\hat{S}_q^t$ is an estimate of the shortfall and $\widehat{\text{var}}(Z \mid Z > z_q)$ is an estimate of the variance of the noise distribution truncated at z_q. Under the GPD assumption we reestimate $\hat{S}_q^t$ and $\widehat{\text{var}}(Z \mid Z > z_q)$ every day with the help of (15) and (18). Under the normal assumption we use (16) and (19).

Under the null hypothesis that the dynamics in (1) and our distributional assumption for the tail of the noise distribution are correct, the exceedance residuals should behave

like an iid sample with mean zero and variance one. In Figure 9 we see these standardized exceedance residuals for the BMW series and $q = 0.95$. Clearly for the normal residuals the null hypothesis seems doubtful.

We are particularly interested in the hypothesis of mean zero and we use a bootstrap test which makes no assumption about the underlying distribution of the residuals (see page 224 of Efron and Tibshirani (1993)). This can be applied to either the standardised or unstandardised residuals with similar results. The residuals derived from an assumption of normality always fail the test with p–values in all cases much less than 0.01; the GPD–based residuals are much more plausibly mean zero. In the following Table 4 we give p-values for the test applied to the standardized GPD residuals for all five test series and various values of q. The most problematic series is the S&P series and the null hypothesis is rejected here for all q values (at the 5% level). The null hypothesis is also rejected for the DAX series and $q = 0.995$, but in all other cases it is not rejected. Clearly the expected shortfall is much better estimated under the GPD assumption but, particularly for the indices, there is still some tendency to underestimate. We can conclude that an assumption of conditional normality is useless for the purposes of calculating expected shortfall.

q	0.95	0.99	0.995
S & P	0.04	0.01	0.01
DAX	0.05	0.05	0.03
BMW	0.22	0.05	0.08
USD.GBP	0.23	0.17	0.29
Gold	0.20	0.08	0.13

Table 4: p-values for a one-sided bootstrap test of the hypothesis that the standardised exceedance residuals in the GPD case have mean zero against the alternative that the mean is greater than zero.

5   Multiple Day Returns

In this section we consider estimates of $x_q^t(k)$ for $k > 1$. Among other reasons, this is of interest, if we want to obtain an estimate of the 10-day VaR (as required by the BIS-rule) from a model fitted to daily data. For GARCH-models $F_{X_{t+1}+...+X_{t+k}|\mathcal{G}_t}(x)$ is not known analytically even for a known distribution of the innovation series, so we adopt a simulation approach to obtaining these estimates as follows. Working with the last n negative log returns we fit as before the AR(1)–GARCH(1,1) model and this time we estimate both tails of the innovation distribution $F_Z(z)$.

We simulate noise from this distribution by a combination of bootstrap and simulation from the GPD according to the following algorithm proposed independently by Danielsson and de Vries (1997b).

1. Randomly select a residual from the sample of n residuals.

2. If the residual exceeds $z_{(k+1)}$ sample a GPD($\xi_k^{(1)}$,$\beta_k^{(1)}$) distributed excess y_1 from the right tail and return $z_{(k+1)} + y_1$.

3. If the residual is less than $z_{(n-k)}$ sample a GPD($\xi_k^{(2)}$,$\beta_k^{(2)}$) distributed excess y_2 from the left tail and return $z_{(n-k)} - y_2$.

4. Otherwise return the residual itself.

5. Replace residual in sample and repeat.

This gives points from the distribution

$$
\hat{F}_Z(z) = \begin{cases} \frac{k}{n}\left(1 + \xi_k^{(2)} \frac{|z - z_{(n-k)}|}{\beta_k^{(2)}}\right)^{-1/\xi_k^{(2)}} & \text{if } z < z_{(n-k)} \\ \frac{1}{n}\sum_{i=1}^n 1_{\{z_i \le z\}} & \text{if } z_{(n-k)} \le z \le z_{(k+1)} \\ 1 - \frac{k}{n}\left(1 + \xi_k^{(1)} \frac{z - z_{(k+1)}}{\beta_k^{(1)}}\right)^{-1/\xi_k^{(1)}} & \text{if } z > z_{(k+1)}, \end{cases}
$$

which approximates $F_Z(z)$ for n large.

Using this noise distribution and the fitted GARCH model we simulate 10000 future paths $(x_{t+1}, \dots, x_{t+k})$ and calculate the corresponding cumulative sums which are realisations of $\sum_{j=1}^k X_{t+j} \mid \mathcal{G}_t$. We use these realisations to calculate $x_q^t(k)$. With 10000 simulated paths the sample quantile is a reasonable estimator for $q \le 0.99$.

We are interested in the ratio $x_q^t(k)/x_q^t$ for $k > 1$, i.e. we want to know how we have to scale a conditional quantile estimate for one-day returns in order to obtain an estimate for the same conditional quantile of k-day returns.

If the X_i are iid some theoretical results on the appropriate scaling factor are available. For strictly stable distributions where $X_1 + \dots + X_k \overset{d}{=} k^{1/\alpha} X_1$ for some $\alpha \in (0, 2]$ we get that $x_q(k)/x_q = k^{1/\alpha}$; in the special case of the normal distribution where $\alpha = 2$ we get the famous "square-root of time rule" implemented in RiskMetrics. Next consider the case of iid random variables X_i with heavy-tailed distribution function F_X satisfying (11). Feller (1970), Chapter VIII.8 proved that for $x \to \infty$

$$
(x^{-1/\xi} k L(x))^{-1} P[X_1 + \dots + X_k > x] \to 1 \, .
$$

Hence we obtain the following approximative[4] scaling law for "large" quantiles

$$
x_q(k)/x_q \approx k^\xi \, . \tag{20}
$$

In view of these theoretical results we conjecture that for small k in our setup the scaling factor $x_q^t(k)/x_q^t$ can be approximated by a power function, i.e.

$$
x_q^t(k)/x_q^t \approx k^{\lambda_t} \, ; \tag{21}
$$

however we expect λ_t to depend on the initial volatility σ_t. For $k \to \infty$ a scaling factor of $k^{1/2}$ should be appropriate, as under some technical conditions the central limit theorem holds for a strictly stationary GARCH-model with $\alpha_1 + \beta < 1$. To test this conjecture we fitted a GARCH(1,1) model to the excerpt form the S&P index containing the crash and we used our simulation algorithm to compute $x_q^t(k)$ for $k = 1, 2, \dots, 50$ and three different initial values of σ_t: "high"; "average"; and "low".[5] We found that the power scaling law (21) holds almost perfectly; see Figure 10 for an example. Table 5 gives the estimated values λ_t for the three different values of σ_t. We see that for the higher than average value of σ_t the exponent λ_t is lower than for the average value of σ_t, whereas for the lower than average value of σ_t the exponent λ_t is higher. In view of the stationarity of the volatility process this appears very natural: if the initial value σ_t is high (low) the future volatility will on average be lower (higher) than σ_t, such that the quantile $x_q^t(k)$ increases relatively slowly (relatively fast) in k.

---

[4]Note that for $\alpha < 2$ stable laws are fat-tailed with tail index α; by definition the approximative scaling-law (20) is exact for these distributions.

[5]The average volatility was taken to be the median of $\sigma_{t-n+1}, \sigma_{t-n+2}, \dots \sigma_t$; the high volatility corresponds to the 95% quantile of the observed values of $\sigma_{t-n+1}, \sigma_{t-n+2}, \dots \sigma_t$, the low volatility to the 5% quantile.

	0.95	0.99
low volatility	0.65	0.65
average volatility	0.60	0.59
high volatility	0.48	0.47

Table 5: Values for the exponent λ_t of the scaling law (21) for different initial volatilities.

Based on the scaling law (20) for iid series Danielsson and de Vries (1997b) advocate the use of the scaling factor $k^{1/\alpha}$ when computing the k-day VaR from models fitted to daily data ("$k^{1/\alpha}$-rule"). Here $\alpha = 1/\xi$ is the *tail index* of the marginal distribution F_X of the asset returns introduced in Section 2.2. They do not test this claim empirically. Our results cast some doubts on their "$k^{1/\alpha}$-rule": our example suggests that in stationary models with stochastic volatility structure we should expect a scaling factor that depends on the level of the current volatility relative to the long-term volatility mean. Hence a "universal" scaling law is unlikely to prevail for most financial return series. Moreover, we cannot confirm the conclusion of Danielsson, Hartmann, and de Vries (1998) who claim that "the square root of time formula may lead to an overestimation of VaR when returns are not normally distributed and exhibit fat tails." Our scaling factors are even slightly greater than those obtained by means of the square root of time formula. These findings illustrate again, that approaches to tail estimation which ignore the conditional heteroscedasticity exhibited by most financial return series are not suitable for VaR calculation.

6 Conclusion

The present paper is concerned with tail estimation for financial return series and, in particular, the estimation of measures of market risk such as value at risk (VaR) or the expected shortfall. We fit GARCH–models to return data using pseudo maximum likelihood and use a GPD-approximation suggested by extreme value theory to model the tail of the distribution of the innovations. This approach is compared to various other methods for tail estimation for financial data. Our main findings can be summarized as follows.

- We find that a conditional approach that models the conditional distribution of asset returns against the current volatility background is better suited for VaR estimation than an unconditional approach that tries to estimate the marginal distribution of the process generating the returns. The conditional approach is vindicated by the very satisfying overall performance of our method in various backtesting experiments.

- The distribution of the residuals is found to be often leptokurtic. As an "ad-hoc approach" the innovations can be modeled by a t-distribution where the degree-of-freedom parameter is estimated with Maximum Likelihood. This approach works quite well for return series with symmetric tails but fails when the tails are asymmetric. We find the GPD-approximation to be preferable, because it can deal with asymmetries in the tails. Moreover, this method is based on a sound theoretical theory.

- We advocate the expected shortfall as an alternative risk measure with good theoretical properties. This risk measure is easy to estimate in our model. A comparison of estimates for the expected shortfall using our approach and a standard GARCH-model with normal innovations shows again that the innovation distribution should be modelled by a fat-tailed distribution, preferably using EVT.

In practice, VaR estimation is often concerned with multivariate return series. We are optimistic that our "two-stage-method" can be extended to multivariate series. However, a detailed analysis of this question is left for future research.

References

ARTZNER, P., F. DELBAEN, J. EBER, and D. HEATH (1997): "Thinking Coherently," *RISK*, 10(11), 68–71.

ARTZNER, P., F. DELBAEN, J. EBER, and D. HEATH (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9(3), 203–228.

BALKEMA, A., and L. DE HAAN (1974): "Residual life time at great age," *Annals of Probability*, 2, 792–804.

BARONE-ADESI, G., F. BOURGOIN, and K. GIANNOPOULOS (1998): "Don't look back," *Risk*, 11(8).

Basle Comittee (1996): *Overview of the Amendment of the Capital Accord to Incorporate Market Risk*, Basle Committee on Banking Supervision.

BOLLERSLEV, T., R. CHOU, and K. KRONER (1992): "ARCH modeling in finance," *Journal of Econometrics*, 52, 5–59.

BRANDT, A. (1986): "The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients," *Advances in Applied Probability*, 18, 211–220.

DANIELSSON, J., and C. DE VRIES (1997a): "Tail index and quantile estimation with very high frequency data," *Journal of Empirical Finance*, 4, 241–257.

DANIELSSON, J., and C. DE VRIES (1997b): "Value-at-Risk and extreme returns," FMG-Discussion Paper NO 273, Financial Markets Group, London School of Economics.

DANIELSSON, J., P. HARTMANN, and C. DE VRIES (1998): "The cost of conservatism," *RISK*, 11(1), 101–103.

DUAN, J.-C. (1997): "Augmented GARCH(p,q) process and its diffusion limit," *Journal of Econometrics*, 79, 97–127.

DUFFIE, D., and J. PAN (1997): "An overview of Value at Risk," *The Journal of Derivatives*, (Spring:1997), 7–49.

EFRON, B., and R. TIBSHIRANI (1993): *An introduction to the bootstrap*. Chapman & Hall, New York.

EMBRECHTS, P., C. KLÜPPELBERG, and T. MIKOSCH (1997): *Modelling extremal events for insurance and finance*. Springer, Berlin.

EMBRECHTS, P., S. RESNICK, and G. SAMORODNITSKY (1998): "Living on the Edge," *RISK Magazine*, 11(1), 96–100.

EMBRECHTS, P., S. RESNICK, and G. SAMORODNITSKY (1999): "Extreme Value Theory as a Risk Management Tool," *North American Actuarial Journal, to appear*, 3(2), 30–41.

FELLER, W. (1970): *An Introduction to Probability Theory and Its Applications II*. Wiley, New York.

FREY, R. (1997): "Derivative Asset Analysis in Models with Level-Dependent and Stochastic Volatility," *CWI Quaterly, Amsterdam*, 10, 1–34.

GNEDENKO, B. (1943): "Sur la distribution limite du terme maximum d'une série aléatoire," *Annals of Mathematics*, 44, 423–453.

GOURIÉROUX, C. (1997): *ARCH-Models and Financial Applications*, Springer Series in Statistics. Springer, New York.

JOHNSON, N., and S. KOTZ (1970): *Continuous univariate distributions - 2*. Wiley, New York.

LONGIN, F. (1997a): "Beyond the VaR," Discussion Paper 97-011, CERESSEC.

LONGIN, F. (1997b): "From value at risk to stress testing, the extreme value approach," Discussion Paper 97-004, CERESSEC.

MCNEIL, A. (1997): "Estimating the Tails of Loss Severity Distributions using Extreme Value Theory," *ASTIN Bulletin*, 27, 117–137.

MCNEIL, A. (1998): "Calculating Quantile Risk Measures for Financial Return Series using Extreme Value Theory," preprint, ETH Zürich.

MCNEIL, A., and T. SALADIN (1997): "The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions," in *Proceedings of XXVIIth International ASTIN Colloquium*, pp. 23–43, Cairns, Australia.

MÜLLER, O., M. DACAROGNA, R. DAVÉ, R. OLSEN, O. PICTET, and J. VON WEIZSÄCKER (1997): "Volatilities of different time resolutions – analyzing the dynamics of market components," *Journal of Empirical Finance*, 4(2-3), 213–240.

PAGAN, A. (1996): "The Econometrics of Financial Markets," *Journal of Empirical Finance*, 3, 15–102.

PICKANDS, J. (1975): "Statistical inference using extreme order statistics," *The Annals of Statistics*, 3, 119–131.

RiskMetrics (1995): *RiskMetrics Technical Document, 3rd ed.*,J.P. Morgan.

SHEPHARD, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and other Fields*, ed. by D. Cox, D. Hinkley, and O. Barndorff-Nielsen, pp. 1–55, London. Chapman & Hall.

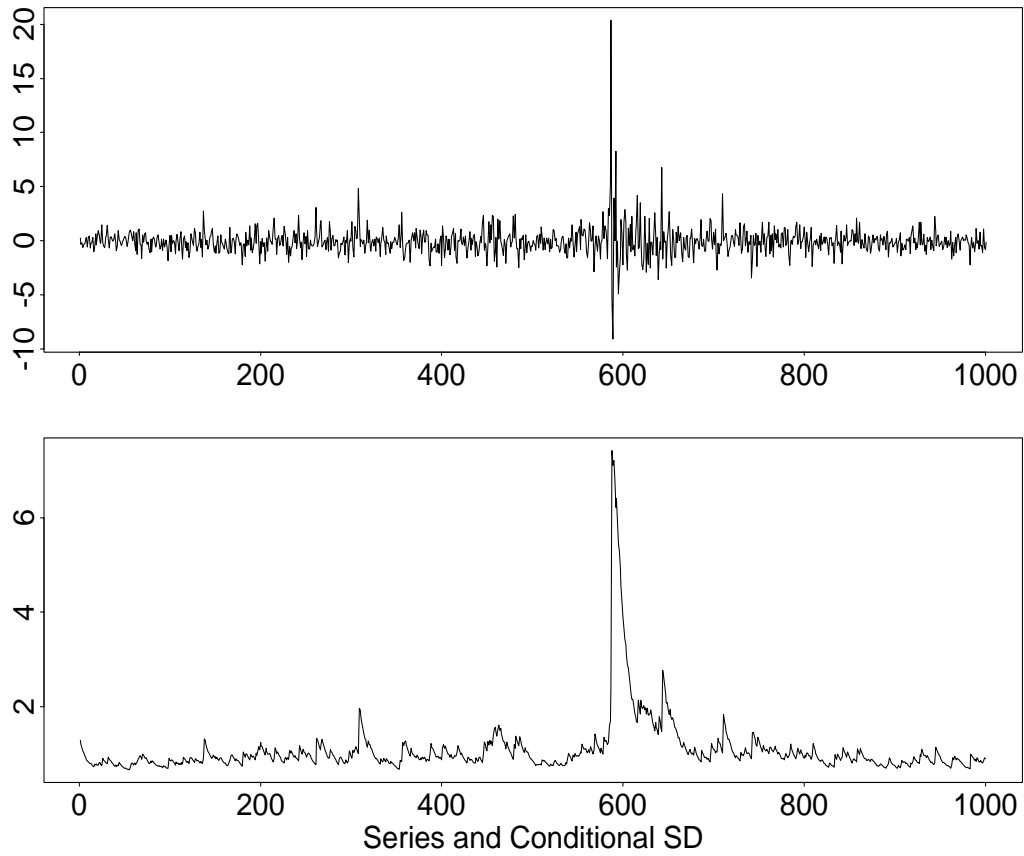SMITH, R. (1987): "Estimating Tails of Probability Distributions," *The Annals of Statistics*, 15, 1174–1207.

Figure 1: 1000 day excerpt from series of negative log returns on Standard & Poors index containing crash of 1987; lower plot shows estimate of the conditional standard deviation derived from PML fitting of AR(1)–GARCH(1,1) model
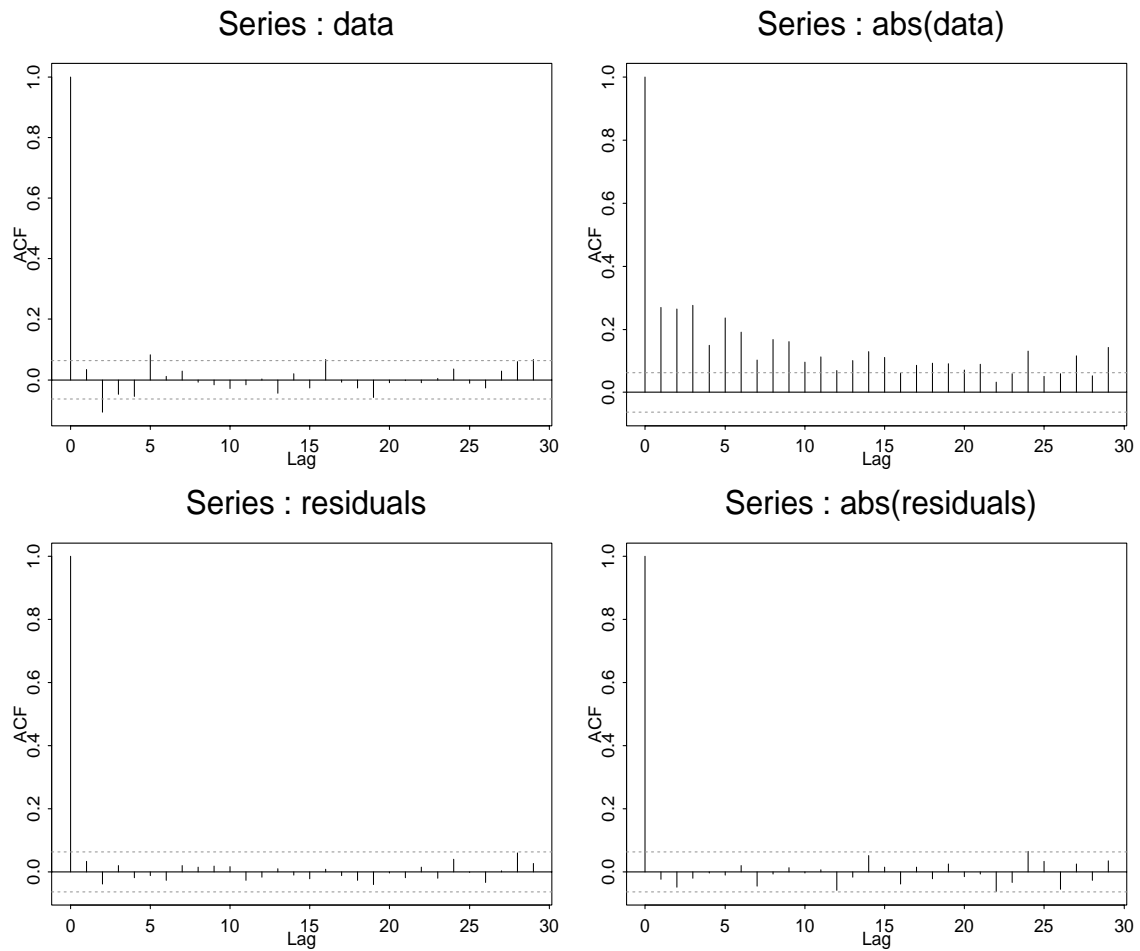
Figure 2: Correlograms for the raw data and their squared values as well as for the residuals and squared residuals. While the raw data are clearly not iid, this assumption may be tenable for the residuals.
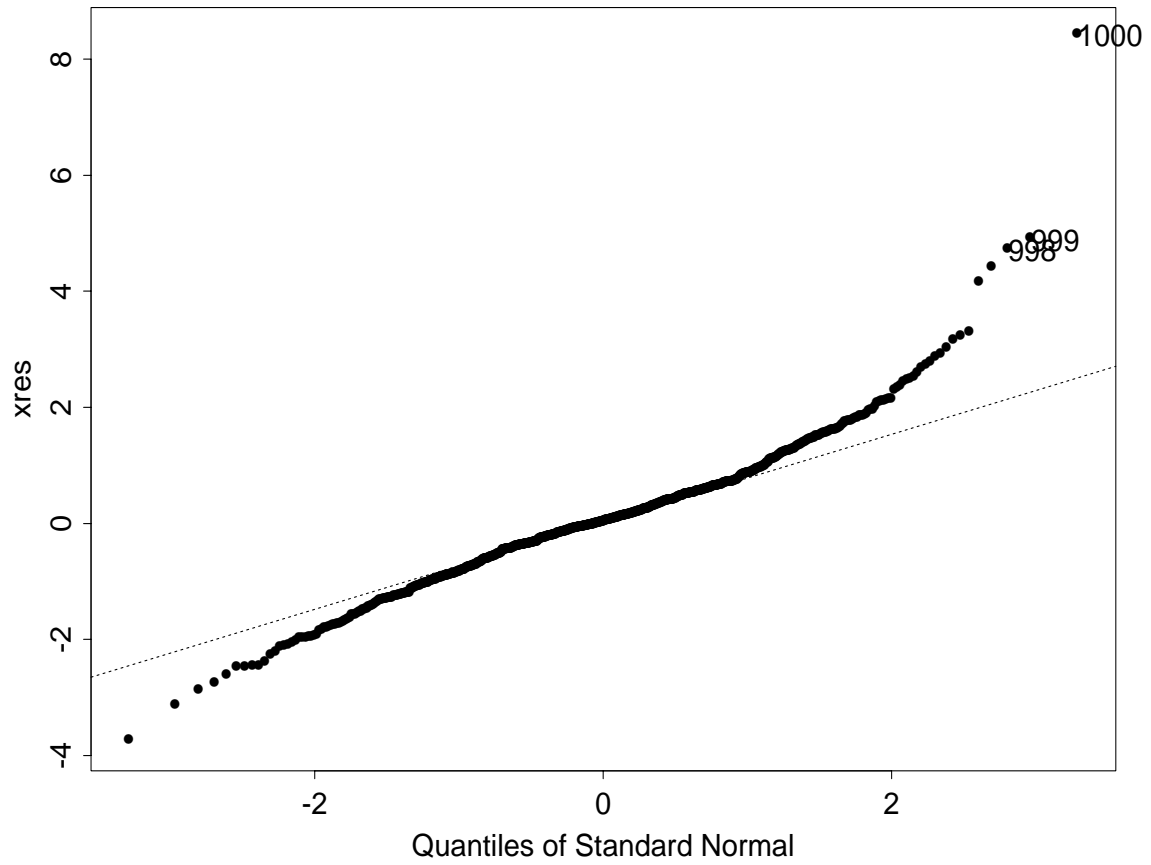
Figure 3: Quantile-quantile plot of residuals against the normal distribution shows residuals to be leptokurtotic.
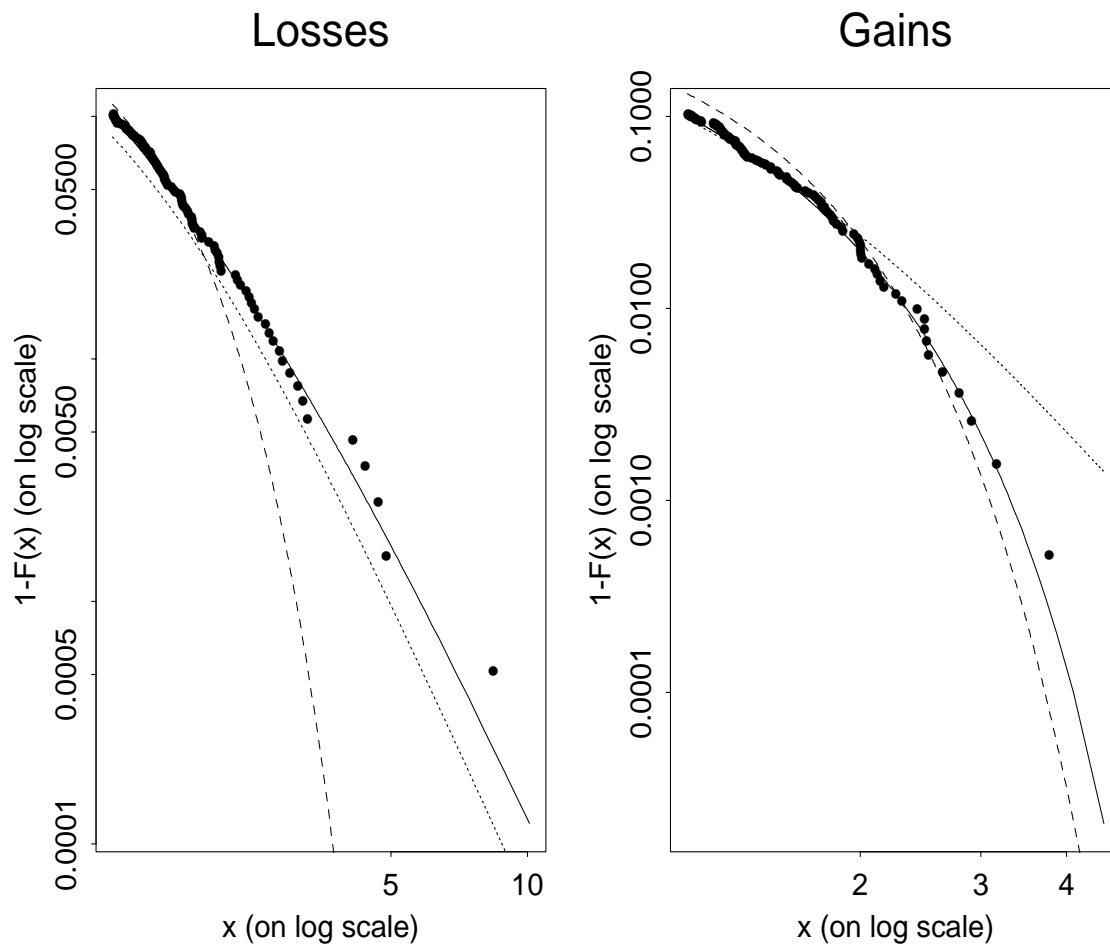
Figure 4: GPD tail estimates for both tails of the innovations distribution. The points show the empirical distribution of the residuals and the solid lines represent the tail estimates. Also shown are the df of the standard normal distribution (dashed) and the df of the t–distribution (dotted) with degrees of freedom as estimated in an AR(1)–GARCH(1,1) model with t–innovations.
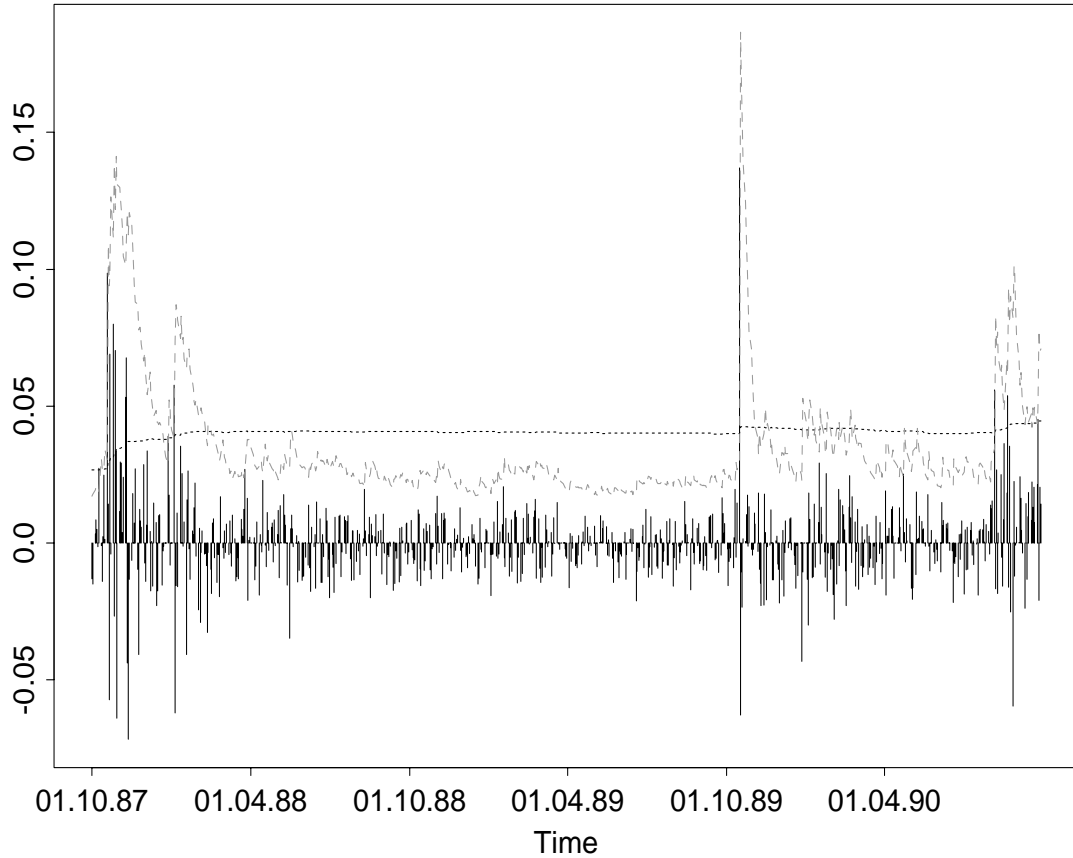
Figure 5: Three years of the DAX backtest beginning in October 1987 and showing the EVT conditional quantile estimate $\hat{x}^t_{0.99}$ (dashed line) and the EVT unconditional quantile estimate $\hat{x}_{0.99}$ (dotted line) superimposed on the negative log returns. The conditional EVT estimate clearly responds quickly to the high violatility around the 1987 stock market crash.
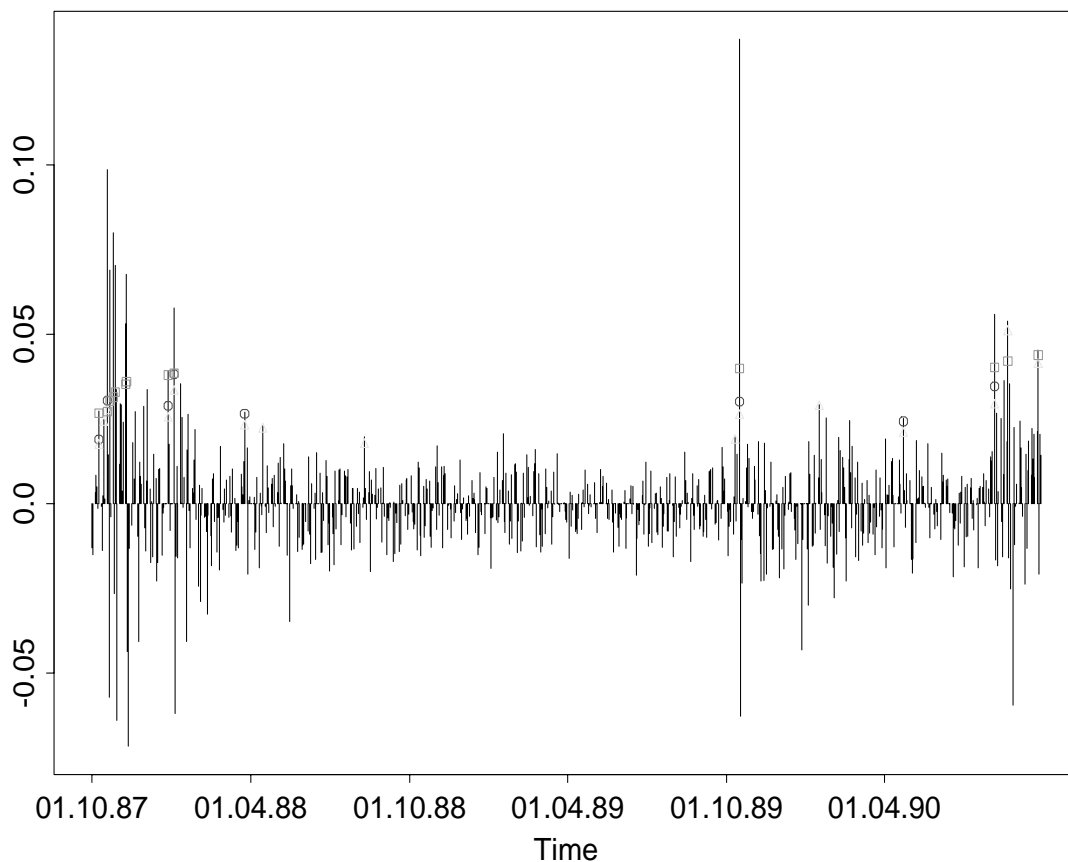
Figure 6: Violations of $\hat{x}^t_{0.99}$ and $\hat{x}_{0.99}$ corresponding to the backtest in Figure 5. Triangles, circles and squares denote violations of the conditional normal, conditional EVT and unconditional EVT estimates respectively. The conditional normal estimate like the conditional EVT estimate responds to changing volatility but tends to be violated rather more often, because it does not take into account the leptokurtosis of the residuals. The unconditional EVT estimate cannot respond quickly to changing volatility and tends to be violated several times in a row in stress periods.
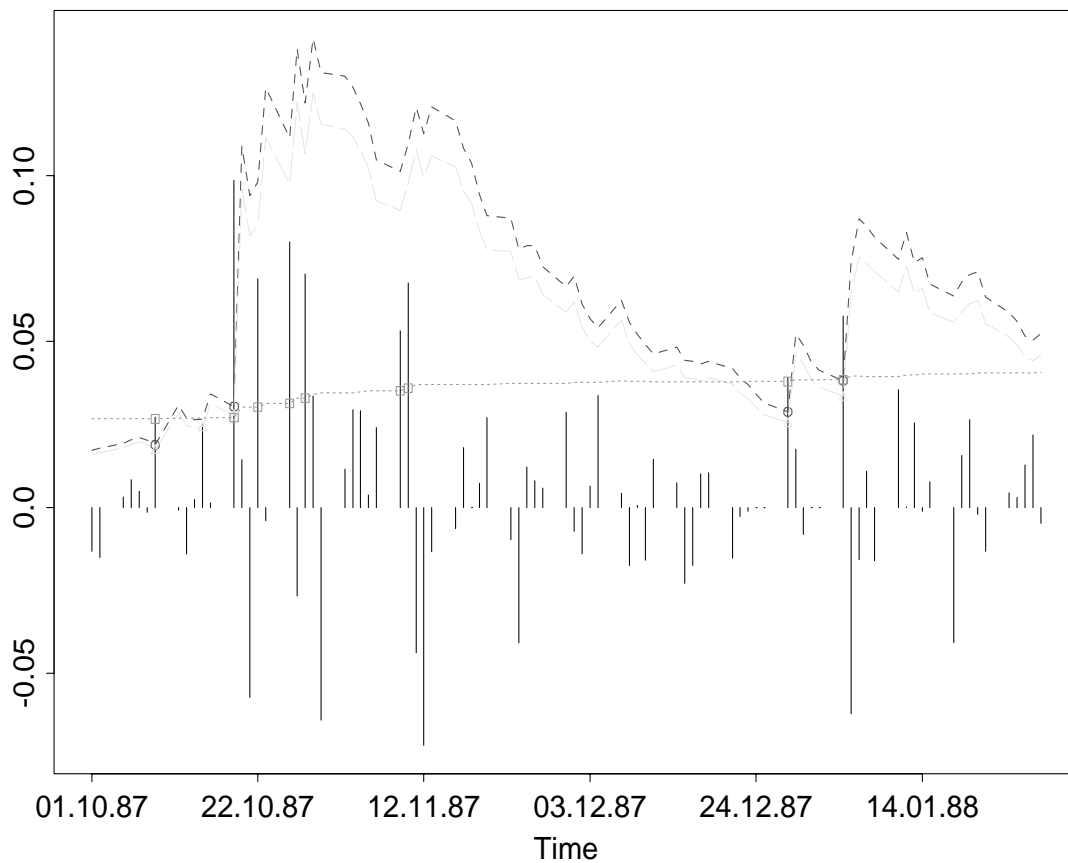
Figure 7: Enlarged section of Figure 6 corresponding to the crash of 1987. Triangles, circles and squares denote violations of the conditional normal, conditional EVT and unconditional EVT estimates respectively. The dotted line shows the path of the unconditional EVT estimate, the dashed line shows the path of the conditional EVT estimate and the long dashed line shows the conditional normal estimate.
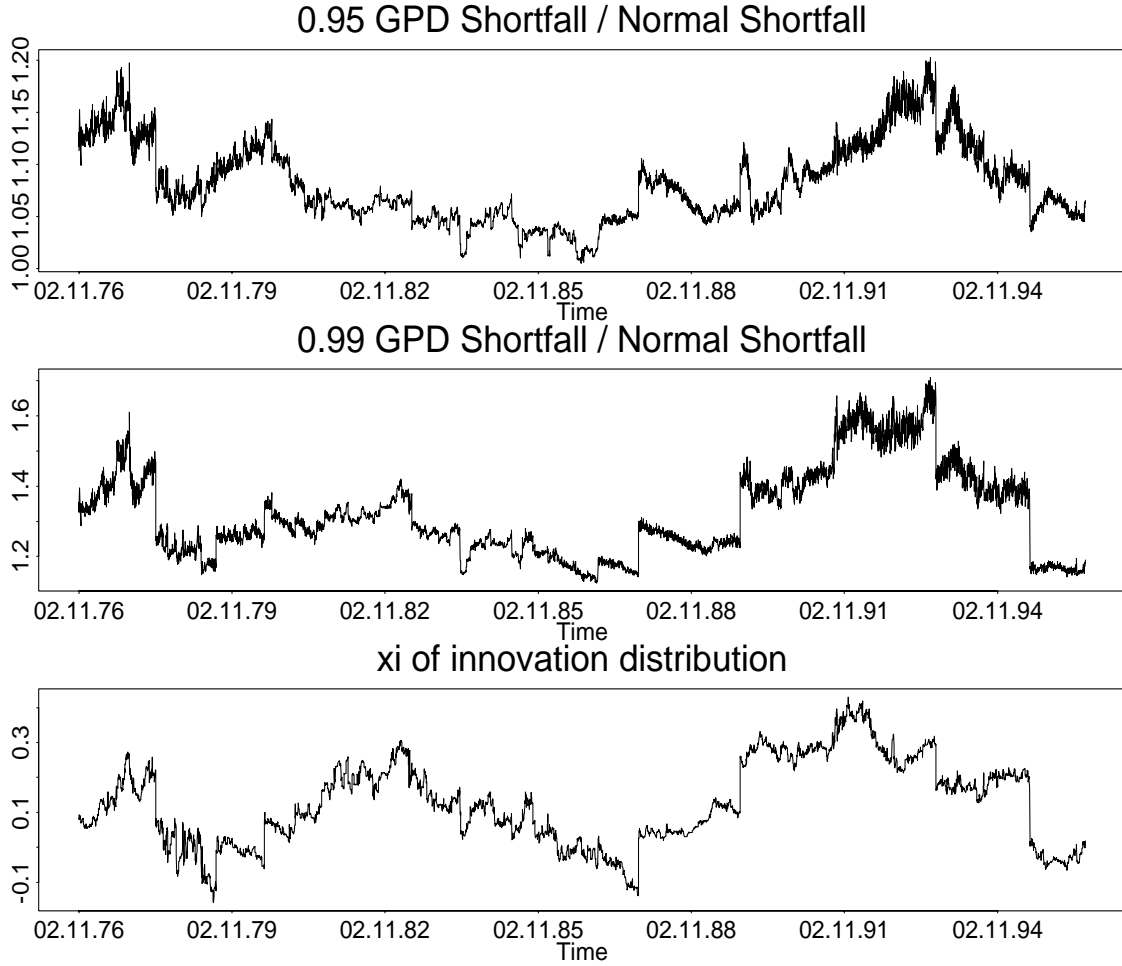
Figure 8: Ratio of expected shortfall calculated with GPD tail estimator to expected shortfall calculated under conditional normality assumption for the BMW data series. Top graph shows calculations for $S_{0.95}^t$ and middle graph those for $S_{0.99}^t$. Lower graph shows ξ value used to construct GPD tail estimate.
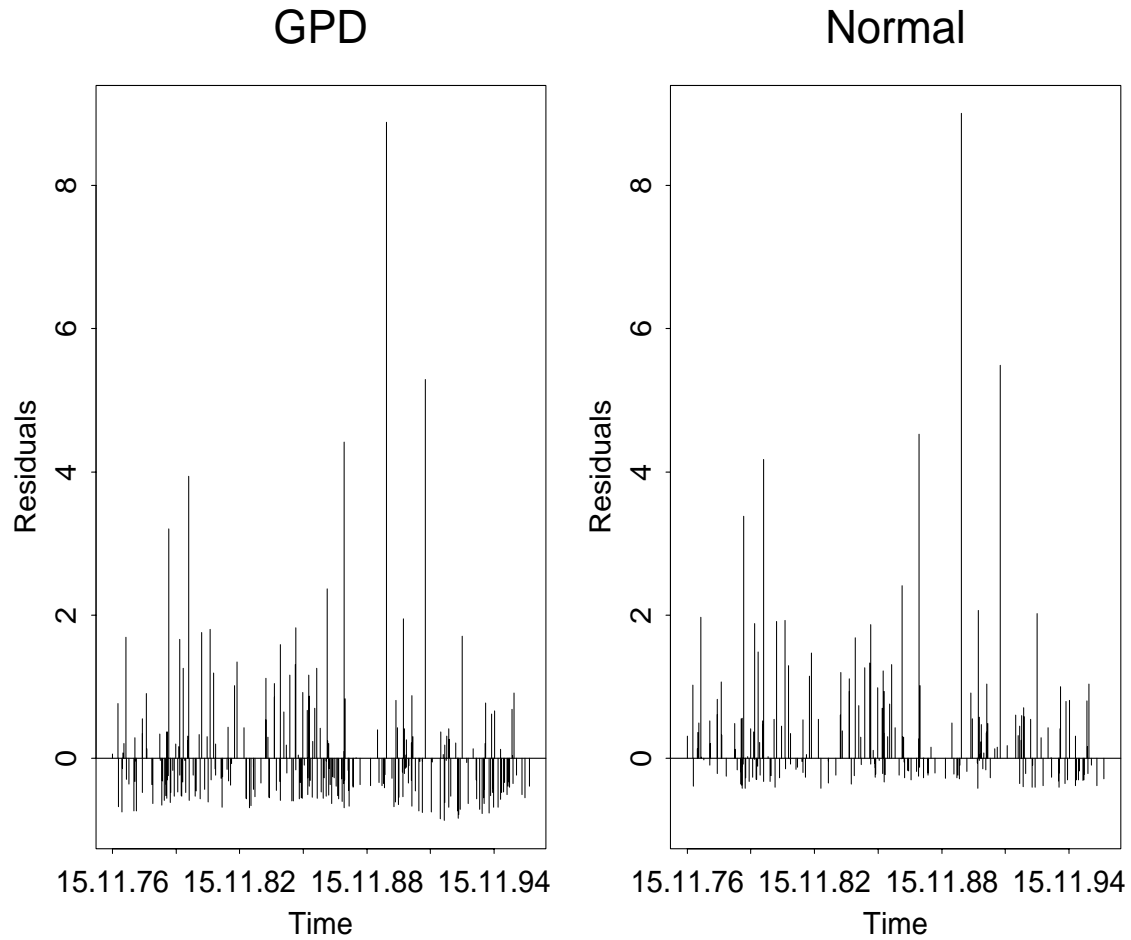
Figure 9: Standardized exceedance residuals for the BMW series and $q = 0.95$. Under the null hypothesis that the dynamics in (1) are correct and that the distributional assumption above the graph is correct, these should have mean zero and variance one. The right graph shows clear evidence against the conditional normality assumption; the left graph shows the assumption of a conditional GPD tail is more reasonable.
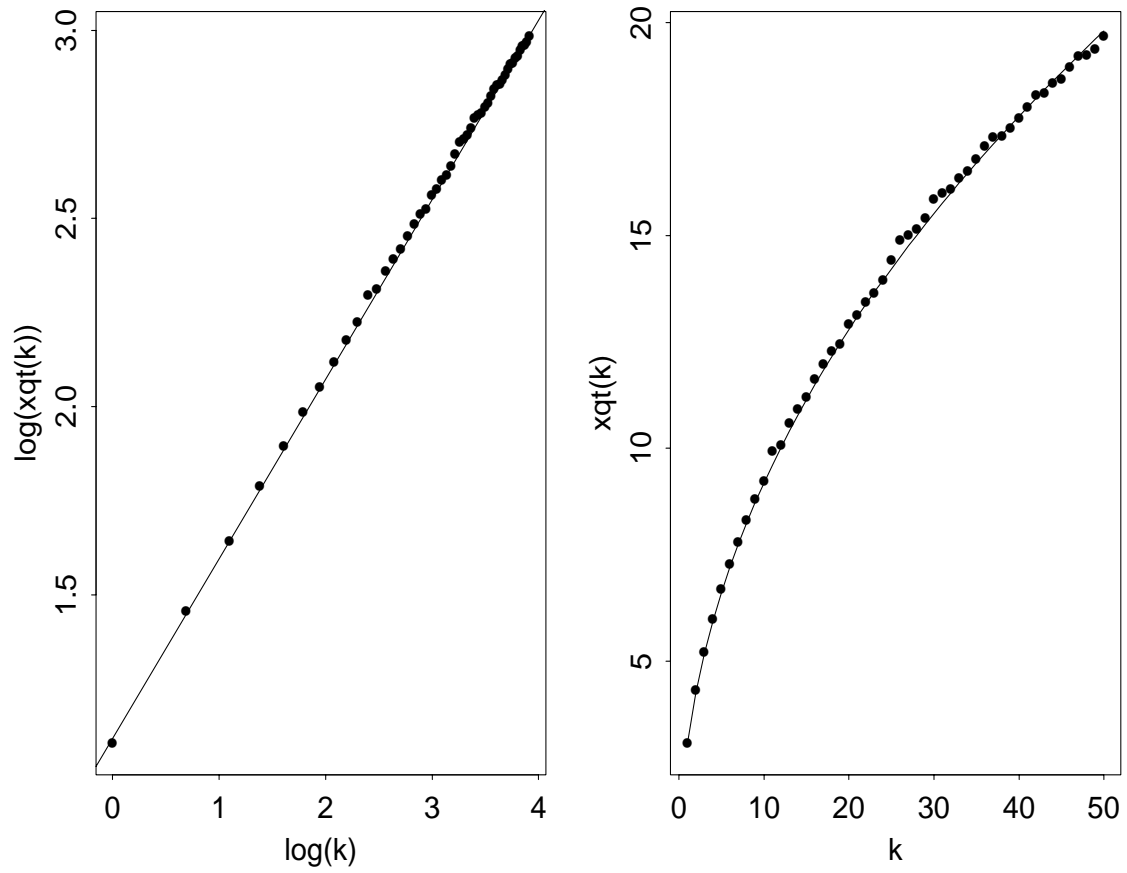
Figure 10: Demonstration of power scaling for the conditional 95% quantile of the k day return for a high value of the current volatility. Slope of the left hand graph is λ_t