Predicting tail-related risk measures: The Consequences of using GARCH filters for non GARCH data

Amine Jalal   Michael Rockinger

First version: April 2006
Current version: April 2006

This research has been carried out within the NCCR FINRISK project on “Financial Econometrics for Risk Management”
Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data

Amine Jalal*  Michael Rockinger†

April 2006

Abstract

We investigate the consequences for value-at-risk and expected shortfall purposes of using a GARCH filter on various mis-specified processes. In general, we find that the McNeil and Frey (2000) two step procedure has very good forecasting properties. Using an unconditional non filtered tail estimate also appears to perform satisfactorily for expected shortfall measurements but less so for VaR computations. Methods assuming specific densities such as the Gaussian or Student-t may yield wrong predictions. Thus, the use of an adequacy test for filtered data to given densities appears relevant. The paper builds on recent techniques to obtain thresholds for extreme value computations. Statistical tests for the expected shortfall, based on the circular bootstrap, are developed.

Keywords: Extreme value theory, Value at Risk (VaR), Expected Shortfall, GARCH, Markov Switching, Jump Diffusion, Backtesting.

JEL classification: G12, C32.

*HEC Lausanne and FAME. Institute of Banking and Finance, Route de Chavannes 33, CH-1007 Lausanne - Vidy, Switzerland, e-mail: Amine.Jalal@unil.ch.
†HEC Lausanne, Swiss:Finance:Institute, and CEPR. Corresponding author: Michael Rockinger, HEC Lausanne, Institute of Banking and Finance, Route de Chavannes 33, CH-1007 Lausanne - Vidy, Switzerland, Tel: +41 21 692 33 48, e-mail: Michael.Rockinger@unil.ch. Both authors are grateful to Oliver Linton for his comments. The usual disclaimer applies.
Predicting tail-related risk measures: The consequences of using GARCH filters for non GARCH data

April 2006

Abstract

We investigate the consequences for value-at-risk and expected shortfall purposes of using a GARCH filter on various mis-specified processes. In general, we find that the McNeil and Frey (2000) two step procedure has very good forecasting properties. Using an unconditional non filtered tail estimate also appears to perform satisfactorily for expected shortfall measurements but less so for VaR computations. Methods assuming specific densities such as the Gaussian or Student-t may yield wrong predictions. Thus, the use of an adequacy test for filtered data to given densities appears relevant. The paper builds on recent techniques to obtain thresholds for extreme value computations. Statistical tests for the expected shortfall, based on the circular bootstrap, are developed.

Keywords: Extreme value theory, Value at Risk (VaR), Expected Shortfall, GARCH, Markov Switching, Jump Diffusion, Backtesting.

JEL classification: G12, C32.
1 Introduction

Extreme value theory has become a precious tool to assess the likelihood of rare but large events in stock markets. In the finance literature, such estimations have become very popular. In this strand of literature the estimations are typically performed under the assumption that the return generating process is i.i.d.. Actual returns do not obey this assumption since returns’ variability clusters. As pointed out by Mandelbrot (1963), large events tend to be followed by other large events. Such phenomena are typically modeled as ARCH or GARCH processes, see Engel (1982) and Bollerslev (1986). There exist, in the statistics literature related to extreme value theory, elements on how to deal with certain types of dependency, especially to correct standard errors, see Leadbetter, Lindgren and Rootzén (1983), and Hsing (1991b). There also exist links between the ARCH literature and extreme value theory. For instance, de Haan et al. (1989) establish the extremal index for a simple ARCH model. They hint at how to obtain the extremal index of the general GARCH model, and an actual derivation thereof may be found in Mikosch and Starića (2000). Further bridges between the two literatures may be found in Quintos, Fan and Phillips (2001). In these contributions it is shown how to test for the stability of the estimates of the tail index as well as how to adjust the standard errors under ARCH or GARCH specifications.

As an alternative to adjusting standard errors, within a VaR context, McNeil and Frey (2000) propose an interesting technique consisting in first filtering the data, then applying extreme value techniques to the tails of the innovations while bootstrapping the central part of the distribution. From there on, using simulation techniques, it is possible to obtain realistically behaved returns that may be useful for VaR purposes. Our contribution is inspired by this work in that we investigate the consequences of following such a methodology when the GARCH process is mis-specified. To do so, we consider various return generating processes in the spirit of GARCH or EGARCH, a

switching-regime model inspired by the work of Hamilton (1994), as well as a stochastic volatility model such as described by Pan (2000). Last, we consider a pure jump model as is often assumed in the finance literature.

The structure of this paper is as follows. In the next section we very briefly recall the working of the GARCH model mainly to introduce notation, and explain the extreme value method used to describe the tail behavior. In section 3 we describe the possible non GARCH specifications used as hypothetical true data generating processes in the simulations. Section 4 expose the different obtained results when the approach is applied to our different simulated processes. In section 5 we show how expected shortfall is affected in this setting. Section 6 concludes and provides recommendations.

2 Approach

Consider \((X_t, t \in \mathbb{Z})\) a strictly stationary time series representing daily observation of the negative log-return computed for a financial asset price. The dynamic of \(X_t\) is assumed to be governed by

\[ X_t = \mu_t + \sigma_t Z_t, \]

where the innovations \(Z_t\) are a strict white noise process, independent and identically distributed, with zero mean, unit variance and marginal distribution function \(F_{Z}(z)\).\(^2\)

The possibly time varying parameters \(\mu_t\) and \(\sigma_t\) are measurable with respect to \(I_{t-1}\), the information available up to time \(t-1\). Let \(F_{X_t}(x)\) denote the marginal distribution of \(X_t\) and for a horizon \(h \in \mathbb{N}\), let \(F_{X_{t+1}+\ldots+X_{t+h}|I_t}(x)\) denote the predictive distribution of the return over the next \(h\) days, given the knowledge of returns up to and including day \(t\).

We are interested in estimating unconditional and conditional quantiles in the tails of the negative log-return distribution. We remind that for \(0 < q < 1\), the \(q\)th unconditional quantile is a quantile of the marginal distribution denoted by

\[ x_q = \inf\{x \in \mathbb{R} : F_X(x) \geq q\}, \]

\(^2\)Commonly used distributions for \(F_Z(z)\) are the zero mean and unit variance normal or the Student t.
and a conditional quantile is a quantile of the predictive distribution for the return over the next \( h \) days denoted by

\[
x_t^q(h) = \inf\{x \in \mathbb{R} : F_{X_{t+1} + \ldots + X_{t+h}|I_t}(x) \geq q\}.
\]

We also consider the expected shortfall (ES), known to be a measure of risk for the tail of a distribution.\(^3\) The ES is a coherent measure of risk in the sense of Artzner, Delbaen, Elsner, and Heath (2000). The unconditional expected shortfall is defined as

\[
S_q = E[X|X > x_q],
\]

and the conditional expected shortfall is written as

\[
S_q^t(h) = E\left[\sum_{j=1}^{h} X_{t+j} | \sum_{j=1}^{h} X_{t+j} > x_t^q(h), I_t\right].
\]

In this paper we restrict ourselves to the \( h = 1 \) step predictive distribution. Thus, we denote the quantiles respectively by \( x_t^q \) and \( S_q^t \). Since

\[
F_{X_{t+1}|I_t}(x) = P\{\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leq x|I_t\},
\]

trivially it holds that

\[
F_{X_{t+1}|I_t}(x) = F_Z((x - \mu_{t+1})/\sigma_{t+1}).
\]

As a consequence, the quantile and expected shortfall become

\[
x_t^q = \mu_{t+1} + \sigma_{t+1} z_q, \quad (2)
\]

\[
S_q^t = \mu_{t+1} + \sigma_{t+1} E[Z|Z > z_q], \quad (3)
\]

where \( z_q \) is the upper \( q \)th quantile of the marginal distribution of \( Z_t \), which by assumption does not depend on \( t \).

### 2.1 Estimating \( \mu_{t+1} \) and \( \sigma_{t+1} \)

Since the focus of this paper is on volatility filtering, we will deal in our simulations with series of mean 0. In empirical work, the description of data with the GARCH(1,1)

---

\(^3\)The expected shortfall is sometimes called a conditional value at risk or CVaR.
model is a popular way of modelling volatility. We follow this road and model the volatility of the series by
\[ \sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta \sigma_{t-1}^2, \]  
where \( \alpha_0 > 0, \alpha_1 > 0, \) and \( \beta > 0. \)

This model is fitted using the Pseudo Maximum Likelihood (PML) method of Gourieroux, Monfort, and Trognon (1984). If we consider a GARCH(1,1) model with normal innovations, the likelihood is maximized to obtain the parameter estimates \( \hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}) \). It has been shown that the PML method yields a consistent and asymptotically normal estimator. Another approach consists in assuming that the innovations have a leptokurtic distribution such as a Student’s t distribution, scaled to have variance 1, see Bollerslev and Wooldridge (1984). Note that the additional parameter, \( v \) representing the degrees of freedom of the Student t, can be estimated, along with the other parameters by PML.

In order to make predictions, we fix a constant memory \( n \) so that at the end of day \( t \), the data consist of the last \( n \) negative log returns \((x_{t-n+1}, \cdots, x_{t-1}, x_t)\). Estimates of the conditional standard deviation series \((\hat{\sigma}_{t-n+1}, \cdots, \hat{\sigma}_t)\) can be calculated from equation (4), after selection of some sensible starting values. Residuals are calculated both to check the adequacy of the GARCH modelling and as an input for the second stage of the method. The estimate of the conditional variance for day \( t+1 \), is calculated as
\[ \hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_t^2 + \hat{\beta} \hat{\sigma}_t^2. \]

To assess the validity of our estimation, we simulated Gaussian innovations and constructed returns with a conditional variance given by (4). Besides recovering the correct parameters, the two steps of the test developed by Diebold, Gunther, and Tay (1998) did not allow rejection of the fit of the model or rejection of the independency assumption of the residuals.

### 2.2 Estimating \( z_q \) using EVT

In this section, we briefly describe how we obtain the quantile \( z_q \) by applying EVT techniques to the distribution of GARCH filtered innovations. We fix a high threshold
and we assume that excess residuals over this threshold have a generalized Pareto distribution (GPD) with tail index $\xi$,

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

(5)

where $\beta > 0$, and the support is $y \geq 0$ when $\xi \geq 0$, and $0 \leq y \leq -\beta/\xi$ when $\xi < 0$. The choice of this distribution follows from a limit result in EVT. Consider a general distribution function $F$ and the corresponding excess distribution above the threshold $u$ defined by

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

(6)

for $0 \leq y < x_0 - u$, where $x_0$ is the right endpoint of $F$, then it is possible to find, for a large class of distributions $F$, a positive measurable function $\beta(u)$ such that

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

This result was shown by Balkema and de Haan (1974) and Pickands (1975). This result holds for most continuous distributions used in statistics.

In our case we assume that the tail of the underlying distribution begins at the threshold $u$, with $N$, the random number of exceeding observations. In the following section, we will discuss how the threshold may be obtained and the $N$ then results by counting exceedances of the threshold. For a sample with total size $n$ the random proportion of extremes is then $N/n$. Since the aim of the paper is to investigate the effect of using a GARCH filter when the hypothesis that excesses over the threshold are i.i.d, must be ruled out. The parameters $\xi$ and $\beta$ are estimated by maximum likelihood. Smith (1985) has shown that the maximum likelihood estimates $\hat{\xi}$ and $\hat{\beta}$ of the GPD parameters $\xi$ and $\beta$ are consistent and asymptotically normal as $N \to \infty$, provided $\xi > -1/2$, when the $N$ exceedances are i.i.d with exact GPD distribution. Even under an approximate GPD distribution, parameter estimates $\hat{\xi}$ and $\hat{\beta}$ are unbiased and asymptotically normal, provided a sufficient rate of convergence in (14). Under the assumption of dependent data, the GPD-based tail estimator is still asymptotically valid, but provides much less stable results compared to the i.i.d case. Embrechts et al. (1997) provide a related example involving an AR(1) process.
The following equality holds for points \( x > u \) in the tail of \( F \)

\[
1 - F(x) = (1 - F(u))(1 - F_u(x - u)).
\]

If we estimate the first term, \( 1 - F(u) \), using the random proportion of the data in the tail, i.e. \( N/n \), and if we estimate the term \( 1 - F_u(x - u) \), where \( F_u(x) \) is defined in (6), by approximating the excess distribution with a GPD fitted by maximum likelihood, we get the tail estimator

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

for \( x > u \). Let \( z_{(1)} \geq z_{(2)} \geq z_{(3)} \geq \ldots \geq z_{(n)} \) represent the ordered residuals. If we fix the number of data in the tail to be \( N = k \), this give us a random threshold at the \((k+1)\)th order statistic. The GPD with parameters \( \xi \) and \( \beta \) 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

\[
\hat{F}_Z(z) = 1 - \frac{k}{n} \left(1 + \frac{z - z_{(k+1)}}{\beta} \right)^{-1/\xi}
\]

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

\[
\hat{Z}_{q,k} = z_{(k+1)} + \frac{\beta}{\xi} \left( \left(1 - q \frac{k}{n} \right)^{-\xi} - 1 \right),
\]

the \( q^{th} \) quantile of the data distribution.

At this stage, we realize that there exist various possibilities to compute the VaR and ES forecasts. First, we notice that for a set of \( M \) simulations and quantile \( q \) one would expect \( qM \) exceedances of the threshold. Then, it is possible to apply the GPD estimation to unfiltered raw data and using (2) and (7) one may obtain forecasts for VaR thresholds. We will refer to such a technique as Unconditional EVT. One may also apply a GARCH(1,1) with normal innovations as a filter, and then run EVT techniques on the residuals. This is the approach of McNeil and Frey (2000). We will call this approach Conditional EVT. Then, it is also possible to filter the data with a normal or a Student t, and use the quantiles associated to these distributions, combined with (2) to obtain VaR forecasts. These approaches will be named Conditional normal and Conditional t.
2.3 Threshold selection and definition of extreme values

The choice of the threshold \( u \) is an important issue in EVT, since it may have severe consequences on the tail estimates. Danielsson and De Vries (1997) have developed bootstrap methods for optimal threshold selection in the context of the Hill estimator that applies to Fréchet type tails. The bootstrap threshold selection is, however, very time consuming. For this reason, we draw on the recent technique of Gonzalo and Olmo (2004) that builds on Pickands (1975) who proposed a single step approach to threshold selection. This method selects as a threshold, the order statistic that minimizes the distance \( d^\infty \) between the empirical conditional distribution function \( F_{u,n} \) and the fitted GPD \( \xi(u),\theta(u) \). Remind that \( F_{u,n} \) is given by

\[
F_{u,n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{u < x_i < x\}} \frac{1}{\sum_{j=1}^{n} \mathbb{1}_{\{x_j > u\}}}.
\]

Equivalently, after performing a change of variable, \( y = x - u \), we may consider

\[
\hat{F}_u(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{0 < y_i < y\}} \frac{1}{\sum_{j=1}^{n} \mathbb{1}_{\{y_j > 0\}}}.
\]

The distance \( d^\infty \) can be written as a function of \( u \), as

\[
d^\infty \left( \hat{F}_u, \text{GPD}_{\xi(u),\theta(u)} \right) = \sup_{0 \leq y \leq \infty} |\hat{F}_u(y) - \text{GPD}_{\xi(u),\theta(u)}(y)|.
\]

The optimal threshold is then,

\[
u^* = \arg \min_u d^\infty \left( \hat{F}_u, \text{GPD}_{\xi(u),\theta(u)} \right).
\]

Gonzalo and Olmo (2004), propose a weighted version of the distance \( d^\infty \) in order to account for the estimation pitfalls that may result from the lack of observations when \( u \) is close to \( x_0 \), the right endpoint of \( F \). The distance \( d^\infty \) is generalized to

\[
d^{WP} \left( \hat{F}_u, \text{GPD}_{\xi(u),\theta(u)} \right) = k^\epsilon \sup_{0 \leq y \leq \infty} |\hat{F}_u(y) - \text{GPD}_{\xi(u),\theta(u)}(y)|.
\]

with \( 0 \leq \epsilon \leq \frac{1}{2} \) and \( k = \sum_{j=1}^{n} \mathbb{1}_{\{x_j > u\}} \).

When \( \epsilon = 0 \), \( d^{WP} \) reduces to the distance \( d^\infty \), and to the Kolmogorov-Smirnov (KS) statistic when \( \epsilon = 1/2 \). The corresponding threshold choice is defined by

\[
u^{WP*} = \arg \min_u d^{WP} \left( \hat{F}_u, \text{GPD}_{\xi(u),\theta(u)} \right).
\]
Gonzalo and Olmo (2004) show that the tail index based on a weighted Pickands’s distance criteria with $\epsilon = 1/2$, is less biased than the approach based on Pickand’s distance $\epsilon = 0$.¹

3 Various non-GARCH data generating processes

In this section, we describe the various alternative data-generating processes. Then we consider alternatives to the benchmark GARCH (1,1) model that will have a volatility dynamics given by (4) and Gaussian innovations. We select a model that dynamically incorporates volatility asymmetries, both stochastic volatility and jumps, and a model with switching regime volatility. These dynamics are assumed to be potentially true DGPs for actual asset return series. Presently, we turn to describing how we perform the various simulations.

3.1 Asymmetric GARCH models

In the basic GARCH model, since only squared residuals enter the equation, positive and negative past values have a symmetric effect on the conditional volatility. Empirically, negative returns tend to be followed by larger increases in volatility than equally large positive returns (Black, 1976 and Christie, 1982).

Nelson (1991) proposed the following exponential GARCH (EGARCH) model to allow for leverage effects:

$$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \frac{\alpha_i |\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{q} \beta_j \log(\sigma_{t-j}^2)$$  (8)

The total effect of $\varepsilon_{t-i}$ is $(1 + \gamma_i)|\varepsilon_{t-i}|$ when there is “good news” or $\varepsilon_{t-i}$ is positive. When $\varepsilon_{t-i}$ is negative or there is “bad news”, the total effect of $\varepsilon_{t-i}$ is $(1 - \gamma_i)|\varepsilon_{t-i}|$. Under the leverage effect, bad news have a larger impact on volatility than good news.

Hence, the value of $\gamma_i$ would be expected to be negative.

¹In order to validate the approach, we have simulated samples of 1000 observations following various Student-t distributions. For the Student t, the degree of freedom parameter is known to be equal to the tail index. Use of this distribution, therefore, proves a simple setting where to verify consistency properties. Indeed, for the approach of Gonzalo and Olmo (2004), we verified that the obtained estimates were close to the corresponding theoretical values.
3.2 Jump diffusion models

Following Pan (1997), we present a model for asset returns that incorporates jumps. At each point of time, the occurrence of a jump is dictated by Bernoulli trials whereas the jump-size is assumed to be Gaussian. Under a discrete-time setting, let $\varepsilon = \{\varepsilon_t : t = 1, 2, \ldots\}$ be a sequence of i.i.d. random variables with standard normal distribution, $J = \{J_t : t = 1, 2, \ldots\}$ be Bernoulli trials with success probability $p$, and $Z = \{Z_t : t = 1, 2, \ldots\}$ be a sequence of i.i.d. random variables normally distributed with mean $\mu_Z$ and variance $\sigma_Z^2$. We assume that the various sets $\{\varepsilon\}$, $\{J\}$, and $\{Z\}$ are mutually independent. As a first DGP, we model the return process as follows

$$X_t = \mu + \sigma_t \varepsilon_t + Z_t J_t,$$

$$\sigma_t^2 = c J_{t-1} + a_0 + a_1 (X_{t-1} - \mu)^2 + a_2 \sigma_{t-1}^2,$$  \hspace{1cm} (9)

where $\mu, a_0, a_1, a_2 \in \mathbb{R}$. In that model, the time-$t$ jump arrival is dictated by $J_t$, a Poisson counter with intensity $\lambda$, while the jump size is modelled by $Z_t$. At time $t$, the marginal movement in returns is modelled by $\varepsilon_t$ with a stochastic volatility $\sigma_t$. Both $J$ and $\varepsilon$ contribute to the dynamics of $\sigma_t$. In this paper we simulate the general model (9), and a restriction thereof where volatility is held constant, hence, where only the mean is allowed to jump. In Pan (1997), the fit of the restricted model is far less satisfactory than the general model with stochastic volatility. However, one cannot exclude this restricted model as a potential representation of the dynamics of an asset return.

3.3 Switching regime volatility model

Switching regime models present a further alternative that proved useful in modelling financial time series. See for instance Duecker (1997), Gray (1996), Hamilton (1989), Hamilton and Liu (1996), as well as Hamilton and Susmel (1994). See also van Norden and Schaller (1997) and Timmermann (2000) who have proposed a Markov switching regime volatility model where returns obey a mixture of normal distributions. Following van Norden and Schaller (1997), such a model may be written as

$$X_t = \mu + |\sigma_1 S_t + \sigma_0 (1 - S_t)| \varepsilon_t,$$ \hspace{1cm} (10)
The innovations $\varepsilon_t$ are independent and identically distributed normal innovations with mean 0 and variance 1. The state variable, $S_t$, is a Markov chain taking the values 0 and 1 with transition probabilities $p = [p_{00}, p_{01}, p_{10}, p_{11}]$ such that

\[
Pr[S_t = 1|S_{t-1} = 1] = p_{11}, \quad Pr[S_t = 0|S_{t-1} = 1] = p_{01}, \\
Pr[S_t = 1|S_{t-1} = 0] = p_{10}, \quad Pr[S_t = 0|S_{t-1} = 0] = p_{00},
\]

where $p_{11} + p_{01} = 1$ and $p_{10} + p_{00} = 1$. Such a model may be easily estimated with the EM algorithm, see Kitagawa (1987) or Hamilton (1989), or via PML as in van Norden and Schaller (1997).

4 Implementation and empirical results

Presently, we wish to discuss the way we perform our simulations. Each sample that we use in our simulation involves $n=1'000$ observations. We simulate each of the previously described processes 20'000 times and evaluate for each simulation the corresponding VaR and ES.

As mentioned in the previous sections, to test the method, we simulate six processes. Our benchmark data generating process is a GARCH(1,1) with Gaussian innovations. Our alternative specifications consist in a GARCH(1,1) with Student t innovations, an EGARCH, a switching volatility regime model, a stochastic volatility process with jumps, and a pure jump diffusion. Since our simulations aim at describing actual returns, we use estimates corresponding to actual stock market data. Parameters for both the stochastic volatility with jumps process and the pure jump process have been estimated in Pan (1997) on daily returns of the SP500 composite index using data from 1986 to 1997.\(^5\) Concerning the switching regime volatility model, parameters have been estimated in van Norden and Schaller (1997) based on CRSP value-monthly returns over the period January 1927 to December 1989.\(^6\)

\(^{5}\)The parameter values used for the stochastic volatility with jumps model are the following: $\mu = 0$ (in Pan (1997) $\mu$ was set to 0.1842 ), $\mu_Z = -0.0183$, $\sigma_Z^2 = 0.0024$, $a_0 = 1.1273 \times 10^{-4}$, $a_1 = 0.0363$, $a_2 = 0.9494$, $c = 0.0275$, and $p = 0.0124$. For the pure jumps process the parameters are: $\mu = 0$ (in Pan (1997) $\mu = 0.1938$), $\mu_Z = -0.0043$, $\sigma_Z^2 = 0.007$, $a_0 = 0.0113$.

\(^{6}\)The parameters obtained by van Norden and Schaller (1997) are: $\mu = 0.0071$, $\sigma_0 = 0.0392$, $\sigma_1 = 0.1180$, and the transition probabilities are $p_{00} = 0.991$, and $p_{11} = 0.9452$. 

12
To assess the quality of the VaR prediction capability, we compare $x^t_q$ with $x^t_{t+1}$ for $q \in \{0.95, 0.99, 0.995\}$. A violation is said to occur whenever $x^t_{t+1} > x^t_q$.

We develop a binomial test of the success of this quantile estimation method based on the number of violations. Assuming the dynamics initially introduced and described in section 1, the indicator for a violation at time $t$ is Bernoulli

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

For two distinct time periods, $t \neq s$, since $Z_{t+1}$ and $Z_{s+1}$ are independent, it follows that $I_s$ and $I_t$ are independent. Therefore, the total number of violations is binomially distributed under the model,

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

Under the null hypothesis that a method correctly estimates the conditional quantiles, the empirical version of the statistic $\sum_{t \in T} 1\{X_{t+1} > x^t_q\}$ follows the binomial distribution $B(\text{card}(T), 1 - q)$. We perform a two-sided binomial test of the null hypothesis against the alternative that the method has a systematic estimation error and gives too few or too many violations. Since our approach is simulation based, the induced p-values are based on the obtained empirical distribution of the test statistic. A p-value less than or equal to 0.05 will be interpreted as evidence against the null hypothesis. The level of this test statistic with corresponding binomial probabilities are given in Table 1, between parenthesis, alongside the estimates of the number of violations for each process simulation. Table 1 shows that on no occasion the filter approach fails systematically. Then, following these empirical results we note that the approach does not fail for any mis-specification with stochastic volatility, as well as for the constant volatility case.

Inspection of Table 1 reveals that the VaR thresholds obtained on raw data tend to perform remarkably well. The largest deviations are found for the pure jumps model where 51.10 rather than 50 violations occur for a 95% quantile. For higher thresholds, the unconditional model has some difficulties with the pure jumps model where for the 99.5% quantile on average 8 rather than 5 violations take place.
Turning to the two step method proposed by McNeil and Frey (2000), named ‘Conditional EVT’ the rejection rate tends to be in general very close to the theoretical threshold. Hence, this method performs extremely well. For instance, for a stochastic volatility with jumps model, at the 99.5% level, the average rejection rate is 4.98 rather than 5. For lower values of the quantile, say the 95% level, there is some deterioration for the models involving jumps. For instance, for the pure jumps model, the rejection rate becomes 57 rather than 50.

Next, we inspect the parametric models. These are obtained after filtering the data with a GARCH(1,1) and either normal or Student-t distributed innovations. The high quantiles are then obtained using the normal or Student-t distribution. First, we verify that under good specification, the VaR forecasts perform fine. Indeed, if we focus on the GARCH with normal distributions, filtering with the GARCH and normal innovations distribution yields in all cases the best VaR rejection rates. For instance, we find an average rejection rate of 5.01 rather than of 5 for the 99.5 quantile.

We verify the same when filtering a GARCH(1,1) with Student-t innovations by a GARCH(1,1) with Student-t innovations. For instance, at the 99% quantile, we find a rejection of 10.09 rather than of 10 sharp.

All other situations correspond to a mis-specification. We notice, e.g. the first column containing statistics, that filtering the true GARCH(1,1) with normal innovations with a GARCH involving Student t distributed innovations leads to a deterioration of the VaR prediction. For instance, at the 99.5% level, the rejection rate becomes 3.76 rather than 5. This suggests that filtering with a Student-t tends to remove too much of the tails. Symmetrically, see the second column of statistics, filtering the data with a GARCH filter with normal innovations, if the true DGP involves Student-t distributed errors, does poorly. For instance, rather than obtaining 5 rejections at the 99.5% rejection rate, we obtain 11.37, more than twice as much. This finding corroborates the idea that the normal distribution will do poorly for obtaining high thresholds.

Further investigations involving the ‘Conditional t’ filter reveal that for the switching regime model, as well as the SV model with jumps, the rejection rate is generally good, however, not as good as the two step procedures of McNeil and Frey (2000). So far, we may conclude that this two step procedure, consisting first in applying a
GARCH(1,1) filter to the data, then using EVT techniques, appears to perform a remarkable job in combination with a well chosen threshold estimation such as the one of Gonzalo and Olmo (2004).

5 Expected shortfall

The concept of Value-at-Risk, VaR, is a quantile-based risk measure. Thus, VaR condenses all the risk of a portfolio into a single number that describes the magnitude of the likely losses. It has undesirable properties such as lack of sub-additivity, i.e., the VaR of a portfolio with two instruments may be greater than the sum of the individual VaRs of these two instruments, and total absence of information on the size of the loss exceeding the VaR. The expected shortfall, ES, is an alternative risk measure to the quantile-based risk-measures such as VaR, which overcomes the deficiencies of the latter, see Artzner et al. (2000). The ES provides information of the average size of a potential loss given that a loss bigger than VaR has occurred.

5.1 Estimation

The conditional one-step expected shortfall, in the context of this paper, i.e., with mean constant and equal to 0, is given by

\[ S^t_q = \sigma_{t+1} E[Z|Z > z_q], \]

where the model governing \( \sigma_{t+1} \) has been already discussed in the previous section. As shown by McNeil and Frey (2000), one then obtains

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

These GPD-based estimates give us the conditional expected shortfall estimate

\[ \hat{S}^t_q = \hat{\sigma}_{t+1} \hat{z}_q \left[ \frac{1}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} \hat{z}_q}{(1 - \hat{\xi})\hat{z}_q} \right]. \]

5.2 A Block-bootstrap test of ES violations

To backtest the method, we develop a one sided test. This test will compare standardized returns with the expected shortfall to establish a criteria to investigate if
expected shortfalls correctly describes actual average exceedances. We are interested in the size of the discrepancy between $X_{t+1}$ and $S_q^t$ in the event of a quantile violation, i.e., $x_{t+1} > \hat{x}_q^t$. We define residuals as the random variables

$$e_{t+1} = \frac{X_{t+1} - S_q^t}{\sigma_{t+1}} = Z_{t+1} - E[Z|Z > z_q].$$

For the particular case where the estimated model is correctly specified, these residuals are i.i.d., and, conditional on $\{X_{t+1} > x_q^t\}$ or equivalently $\{Z_{t+1} > z_q\}$, they have an expected value of zero. To measure ES violations, we then consider residuals that correspond to days when returns actually exceed the VaR threshold. Thus,

$$e^{E}_{t+1} = \frac{x_{t+1} - \hat{S}_q^t}{\hat{\sigma}_{t+1}} \text{ if } x_{t+1} > \hat{x}_q^t.$$

Our test will be based on the observation that if $e^{E}_{t+1}$ is systematically too large, then the conditional shortfall measure does not describe correctly the distribution. Formally, we test the hypothesis that $e^{E}_{t+1}$ has mean zero against the alternative hypothesis that the mean is greater than 0. For a mis-specified model, since the data generating process is non i.i.d., e.g. a GARCH process, the excesses over a certain threshold are also clustered, ruling out the i.i.d. hypothesis for the exceedances. For this reason, we use a dependent bootstrap test that makes no assumption about the residual’s underlying distribution. The test is based on the usual $t$-statistic that we obtain using the algorithm of Efron and Tibshirani (1993, p224) but in a dependent bootstrap setup. The p-values are deduced from the empirical distribution of the $t$-statistic that we obtain from our simulations.

For processes exhibiting weak dependence structure, the concept of moving block bootstrap has been proposed (Künsch 1989; Liu and Singh, 1992; Politis and Romano, 1992a, 1994; Bühlmann 1994). The principle consist in randomly selecting blocks of some length from the original data and in concatenating them together to form a resample. There exist several improvements to this method. The idea behind the circular blocks bootstrap (Politis and Romano, 1992b; Shao and Yu, 1993) consist in wrapping the data around a circle. This ensures that each of the original observations has an equal chance of appearing in a simulated series and that all the blocks have the same length. When the data exhibit a stronger dependence structure, one may
require the approach based on whitening and post-blackening described in Davison and Hinkley (1997), or the block of blocks bootstrap originally proposed by Künsch (1989). The stationary bootstrap is another variation of the moving block bootstrap, where the length of the block is allowed to be random (Politis and Romano, 1994).

The choice of the block length is the main critical issue in block bootstrapping. For the block bootstrap to be effective, the length should be large enough so that it includes most of the dependence structure, but not too large so that the number of blocks becomes insufficient. Bühlmann and Künsch (1999), propose some advises on how to select the block length. In our paper, we use the circular bootstrap and we select the block length following the algorithm based on the spectral density estimation as proposed in Politis and White (2003). Note that the circular bootstrap has the property of higher-order accurate estimation of the distribution of the sample mean (Politis and Romano, 1992b).

Insert Table 2 somewhere here

In Table 2 we display various statistics for the one-sided block-bootstrap test of the hypothesis that the exceedance’s residuals have mean zero. The numbers without parenthesis represent the simulation based p-values for the test. The values between parenthesis represent first the average block length and second the standard deviation thereof, obtained using the spectral density estimation. For instance, the cell corresponding to a GARCH with Student-t innovations as true DGP, and a 99% quantile shows for the NeC Neil and Frey (2000) method an average p-value of 0.53. Inspection of the histogram of p-values shows that in no case, could the forecasted ES value be rejected. The average block length used in the resampling procedure is 3.15 with a standard deviation of 0.95. This indicates that in general relatively small blocks need to be taken while still capturing dependency in a satisfactory manner.

The overall picture that emerges from this table is that on no occasion does the Conditional EVT method of NeC Neil and Frey (2000) fail the test. However, we note that an Unconditional EVT may perform as well as it’s conditional counterpart for this ES exercise. The conditional normal approach perform well only for GARCH and EGARCH with normal innovations and is systematically rejected for all the other
processes which distribution present heavier tails. The conditional t estimator fail the test at the 99% and 99.5% confidence level when the process consist in pure jumps, but has the highest p-values when the others processes are considered.

6 Conclusion

The true temporal dependency of financial returns is a complex issue. As a way to improve relevant measures for risk management, one could consider a two step procedure: First, filtering the returns through a more or less complex GARCH model, and, second, estimating the tail parameters using the assumption of i.i.d. data. This two step procedure has been proposed by McNeil and Frey (2000). With these estimates it is possible to forecast VaR and ES measures.

In this paper we investigate the consequences of using GARCH filtered returns when the data is generated either by some GARCH but with non-Gaussian innovations or some non-GARCH type process such as a switching regime process or a stochastic volatility with jumps model. Our findings may be summarized as follows. We find that filtering returns with a GARCH involving normal or Student-t distributed innovations and using parametric VaR forecasts may yield to bad results. We find that it is important to correctly specify the threshold by using for instance the Gonzales and Olmo (2004) procedure. Focusing on expected shortfall forecasts, obtained for the two step procedure of McNeil and Frey (2000), we find after using a circular bootstrap method to account for possible dependency among the exceedances, that one cannot reject the assumption that the forecasted expected shortfall measure captures actual shortfalls in a satisfactory manner. We also find that ES forecasts obtained on raw data yield, from a statistical point of view, very good results.

The overall lesson from this work is that VaR or ES forecasting obtained by parametric models after filtering by a GARCH(1,1) may be dangerous. If this approach is taken, a careful test of the adequacy of the density of the residuals needs to be made. One popular test of adequacy is the Diebold, Gunther and Tay (1998) test. In particular, if changes in regime are expected, or if the data contains lots of jumps, then particular care needs to get exercised.
The McNeil and Frey (2000) two step approach appears to provide very good VaR and ES forecasts. Extreme value techniques applied to raw data may yield satisfactory ES forecasts whereas VaR properties may not be so satisfactory.
References


Captions

Table 1: Investigation of prediction properties of VaR models. The various columns present statistics for various data generating processes. The statistics are obtained for 20'000 simulations, each involving a sample of size 1'000. The various cells, along the horizontal, contain first for different quantile levels the theoretical amount of exceedances. Then, ‘Unconditional EVT’ corresponds to a situation where VaR is computing using no filter, thus by fitting a GPD to raw data. ‘Conditional EVT’ is the McNeil and Frey (2000) approach.

It is also possible to filter with GARCH(1,1) and normal or Student-t innovations and, then, use the quantiles of the normal or the Student-t to obtain VaR forecasts, yielding the ‘Conditional normal’ or ‘Conditional t’ statistics. The statistic without parenthesis represents the average rejection rate. The figures in parenthesis represent the average binomial test statistics of adequacy of the VaR threshold and corresponding average p-value.

Table 2: Investigation of the prediction properties of Expected Shortfalls. The various cells correspond to the ones of Table 1. The statistic without parenthesis is the average p-value for a one-sided block-bootstrap test of the hypothesis that standardized exceedances of residuals have mean zero against the alternative that the mean is greater than zero. The p-values are obtained running 20'000 simulations. Between parenthesis are the mean and standard deviations of the optimal block lengths. The block lengths are obtained using the spectral density estimation algorithm.
<table>
<thead>
<tr>
<th>Data generating process</th>
<th>GARCH with Normal innovations</th>
<th>GARCH with student-t innovations</th>
<th>EGARCH with normal innovations</th>
<th>Switching regime volatility</th>
<th>Stochastic volatility with jumps</th>
<th>Pure jumps</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.95 Quantile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Unconditional EVT</td>
<td>50.89</td>
<td>50.79</td>
<td>51.42</td>
<td>49.74</td>
<td>51.02</td>
<td>51.10</td>
</tr>
<tr>
<td></td>
<td>(2.62; 0.99)</td>
<td>(2.15; 0.99)</td>
<td>(3.11; 0.99)</td>
<td>(2.38; 1)</td>
<td>(2.58; 0.99)</td>
<td></td>
</tr>
<tr>
<td>Conditional EVT</td>
<td>50.66</td>
<td>50.51</td>
<td>50.75</td>
<td>51.03</td>
<td>54.03</td>
<td>57.29</td>
</tr>
<tr>
<td></td>
<td>(5.42; 0.97)</td>
<td>(3.46; 0.99)</td>
<td>(4.69; 0.98)</td>
<td>(4.46; 0.99)</td>
<td>(5.48; 0.93)</td>
<td></td>
</tr>
<tr>
<td>Conditional normal</td>
<td>49.99</td>
<td>41.63</td>
<td>49.85</td>
<td>38.51</td>
<td>50.83</td>
<td>39.82</td>
</tr>
<tr>
<td></td>
<td>(5.81; 0.98)</td>
<td>(6.01; 0.98)</td>
<td>(5.75; 0.98)</td>
<td>(5.80; 0.63)</td>
<td>(6.27; 0.96)</td>
<td></td>
</tr>
<tr>
<td>Conditional t</td>
<td>49.81</td>
<td>49.77</td>
<td>49.95</td>
<td>48.25</td>
<td>50.94</td>
<td>49.18</td>
</tr>
<tr>
<td></td>
<td>(5.77; 0.98)</td>
<td>(5.01; 0.99)</td>
<td>(5.82; 0.94)</td>
<td>(5.79; 0.99)</td>
<td>(5.74; 0.99)</td>
<td></td>
</tr>
<tr>
<td><strong>0.99 Quantile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Unconditional EVT</td>
<td>10.42</td>
<td>10.48</td>
<td>9.96</td>
<td>11.33</td>
<td>10.31</td>
<td>12.15</td>
</tr>
<tr>
<td></td>
<td>(2.01; 0.98)</td>
<td>(1.77; 0.99)</td>
<td>(2.65; 0.95)</td>
<td>(1.94; 0.99)</td>
<td>(2.06; 0.98)</td>
<td>(2.44; 0.95)</td>
</tr>
<tr>
<td></td>
<td>(3.14; 0.93)</td>
<td>(1.84; 0.99)</td>
<td>(2.89; 0.95)</td>
<td>(1.87; 0.99)</td>
<td>(2.01; 0.99)</td>
<td>(2.42; 0.94)</td>
</tr>
<tr>
<td>Conditional normal</td>
<td>9.91</td>
<td>15.69</td>
<td>10.63</td>
<td>12.10</td>
<td>14.45</td>
<td>15.58</td>
</tr>
<tr>
<td></td>
<td>(3.82; 0.97)</td>
<td>(3.42; 0.59)</td>
<td>(2.91; 0.96)</td>
<td>(3.09; 0.92)</td>
<td>(3.21; 0.74)</td>
<td>(3.03; 0.63)</td>
</tr>
<tr>
<td>Conditional t</td>
<td>8.49</td>
<td>10.09</td>
<td>9.17</td>
<td>9.30</td>
<td>10.57</td>
<td>13.82</td>
</tr>
<tr>
<td></td>
<td>(2.63; 0.97)</td>
<td>(2.52; 0.99)</td>
<td>(2.68; 0.98)</td>
<td>(2.55; 0.98)</td>
<td>(2.57; 0.98)</td>
<td>(2.90; 0.82)</td>
</tr>
<tr>
<td><strong>0.995 Quantile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Unconditional EVT</td>
<td>5.41</td>
<td>5.39</td>
<td>5.14</td>
<td>6.05</td>
<td>5.39</td>
<td>8.21</td>
</tr>
<tr>
<td></td>
<td>(1.48; 0.98)</td>
<td>(1.28; 0.99)</td>
<td>(1.73; 0.95)</td>
<td>(1.52; 0.98)</td>
<td>(1.48; 0.98)</td>
<td>(1.58; 0.79)</td>
</tr>
<tr>
<td>Conditional EVT</td>
<td>5.06</td>
<td>5.20</td>
<td>5.13</td>
<td>5.09</td>
<td>4.98</td>
<td>8.06</td>
</tr>
<tr>
<td></td>
<td>(1.96; 0.93)</td>
<td>(1.36; 0.99)</td>
<td>(1.85; 0.94)</td>
<td>(1.34; 0.99)</td>
<td>(1.37; 0.99)</td>
<td>(1.55; 0.83)</td>
</tr>
<tr>
<td>Conditional normal</td>
<td>5.01</td>
<td>11.37</td>
<td>5.56</td>
<td>8.55</td>
<td>9.51</td>
<td>13.38</td>
</tr>
<tr>
<td></td>
<td>(2.06; 0.97)</td>
<td>(2.89; 0.26)</td>
<td>(2.17; 0.95)</td>
<td>(2.65; 0.65)</td>
<td>(2.63; 0.51)</td>
<td>(3.14; 0.09)</td>
</tr>
<tr>
<td>Conditional t</td>
<td>3.76</td>
<td>5.04</td>
<td>4.27</td>
<td>5.06</td>
<td>6.21</td>
<td>11.84</td>
</tr>
<tr>
<td></td>
<td>(1.82; 0.97)</td>
<td>(1.85; 0.98)</td>
<td>(1.86; 0.98)</td>
<td>(1.87; 0.98)</td>
<td>(1.83; 0.95)</td>
<td>(2.92; 0.21)</td>
</tr>
</tbody>
</table>

Table 1
<table>
<thead>
<tr>
<th>Data generating process</th>
<th>GARCH with Normal innovations</th>
<th>GARCH with student t innovations</th>
<th>EGARCH with normal innovations</th>
<th>Switching regime volatility</th>
<th>Stochastic volatility with jumps</th>
<th>Pure jumps</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95 Quantile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unconditional EVT</td>
<td>0.51 (5.53; 2.02)</td>
<td>0.58 (5.44; 2.08)</td>
<td>0.42 (5.68; 2.21)</td>
<td>0.38 (5.32; 2.11)</td>
<td>0.49 (5.50; 2.10)</td>
<td>0.95</td>
</tr>
<tr>
<td>Conditional EVT</td>
<td>0.43 (5.71; 2.19)</td>
<td>0.61 (5.81; 2.21)</td>
<td>0.43 (5.65; 2.19)</td>
<td>0.46 (5.65; 2.20)</td>
<td>0.89 (5.99; 2.24)</td>
<td>0.84</td>
</tr>
<tr>
<td>Conditional normal</td>
<td>0.52 (5.69; 2.13)</td>
<td>0.01 (5.04; 1.97)</td>
<td>0.38 (5.58; 2.14)</td>
<td>0.01 (4.84; 1.89)</td>
<td>0.06 (5.76; 2.17)</td>
<td>0.01</td>
</tr>
<tr>
<td>Conditional t</td>
<td>0.99 (5.70; 1.92)</td>
<td>0.67 (5.78; 2.12)</td>
<td>0.99 (5.64; 2.15)</td>
<td>0.99 (5.50; 2.08)</td>
<td>0.99 (5.31; 2.09)</td>
<td>0.99</td>
</tr>
</tbody>
</table>

| 0.99 Quantile               |                               |                                |                               |                             |                                |            |
| Unconditional EVT           | 0.49 (3.12; 0.98)             | 0.57 (3.18; 0.94)              | 0.40 (3.15; 0.96)             | 0.60 (3.24; 0.97)           | 0.46 (3.13; 0.97)              | 0.87       |
| Conditional EVT             | 0.44 (3.12; 0.96)             | 0.53 (3.15; 0.95)              | 0.44 (3.16; 0.96)             | 0.49 (3.09; 0.94)           | 0.59 (3.05; 0.94)              | 0.53       |
| Conditional normal          | 0.53 (3.01; 0.97)             | 0.01 (3.54; 1.21)              | 0.43 (3.11; 1.03)             | 0.02 (3.24; 1.14)           | 0.04 (3.45; 1.18)              | 0.00       |
| Conditional t               | 0.99 (2.72; 0.99)             | 0.68 (3.01; 1.01)              | 0.99 (5.82; 0.93)             | 0.86 (2.88; 1.11)           | 0.94 (2.93; 0.83)              | 0.01       |

| 0.995 Quantile              |                               |                                |                               |                             |                                |            |
| Unconditional EVT           | 0.54 (2.37; 0.76)             | 0.62 (2.39; 0.75)              | 0.42 (2.32; 0.78)             | 0.71 (2.48; 0.74)           | 0.5 (2.36; 0.76)               | 0.97       |
| Conditional EVT             | 0.41 (2.30; 0.79)             | 0.52 (2.30; 0.78)              | 0.41 (2.31; 0.78)             | 0.49 (2.27; 0.79)           | 0.53 (2.29; 0.77)              | 0.63       |
| Conditional normal          | 0.55 (2.07; 1.01)             | 0.01 (3.16; 0.88)              | 0.47 (2.25; 0.91)             | 0.01 (2.82; 0.93)           | 0.4 (2.98; 0.92)               | 0.00       |
| Conditional t               | 0.99 (3.39; 0.84)             | 0.63 (1.74; 0.86)              | 0.99 (2.06; 0.84)             | 0.73 (1.87; 0.85)           | 0.80 (2.44; 0.79)              | 0.01       |

Table 2