Stockholms universitet

Estimating value at risk - an extreme value approach

Alexandra Hellman

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2015:28
Matematisk statistik
December 2015

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2015:28**
http://www.math.su.se

Estimating value at risk - an extreme value approach

Alexandra Hellman*

December 2015

Abstract

Assessing the probability of extreme and rare events is an important issue in financial risk management, in order to account for potential losses. To model and estimate tail risk adequately, such as value at risk (VaR), is of typical interest. Recent financial disasters has made the common distribution assumption of normality for asset returns questionable. This assumption makes modeling easy but inefficient when the return distribution exhibits heavy tails. An interesting solution to this problem is the extreme value approach, estimation of extreme quantiles. In this thesis we demonstrate how the use of univariate extreme value theory (EVT) can be combined with a GARCH model in order to estimate daily VaR properly. Using backtesting based on historical daily log-returns for OMXS30 and Ericsson the results indicate that the GARCH-EVT approach outperforms other well-known techniques.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: alhe7913@student.su.se. Supervisor: Mathias Lindholm.

Contents

1 Introduction

1.1 Background

Along with a more volatile financial market and the development of more complex financial instruments the interest of measuring risk has increased. Recent financial disasters, such as the bankruptcy of Lehman Brothers in 2008, has triggered and made it even more important for banks, financial institutions and treasury organisations within large corporations to account and cover for extreme outcomes. Extreme outcomes are rare and the estimation is consequently not problem-free.

Value at risk (VaR) is a popular and above all an essential risk measure within market risk (see Jorion 2001), whose origin date back to the late 1980'. Formally, VaR is a quantile-based measure estimating the maximal loss during a given time horizon for a given confidence level. Todays nearly limitless computing power has contributed to existence of multiple techniques, methods and approaches to estimate VaR, both parametric and non-parametric. One popular non-parametric method is the historical simulation (HS) where VaR is estimated from historical price movements, simply using the empirical distribution for the returns. The advantage of HS is that few assumptions are required and the method is easy to implement. However, in order to get accurate estimates from HS, the history has to repeat itself, leading to the assumption of constant volatility. This is a contradiction to the well-known phenomenon of volatility clustering in financial markets, meaning that periods of high volatility are followed by periods of less.

The autoregressive conditional heteroscedastic (ARCH) model and the generalized autoregressive conditional heteroscedastic (GARCH) model are introduced to model such volatility clustering. VaR estimates based on these models are reflecting the current volatility background but tend to underestimate the magnitude of risk. This underestimation results from the assumption of conditional normality, which does not seem to hold for real data. Models based on conditional normality are not that well-suited for analyzing extreme scenarios. An important solution to this problem is the extreme value approach, the estimation of extreme quantiles. The advantages of using extreme value theory (EVT) is that it is based on solid statistical theory for the asymptotic behavior in the tails, and it allow for extrapolation beyond the tail of the distribution. However, EVT applies estimators for independent and identically distributed (iid) variables and do not reflect the current volatility background which is a contradiction to two important features of daily asset returns.

McNeil and Frey (2000) proposes one approach, which will be adapted in this thesis, combining the ideas of all three methods mentioned above. This approach is based on GARCH-modelling and pseudo-maximum likelihood estimation for the conditional volatility, HS for the central part of the distribution and EVT methods for quantile estimation. The residuals (also called innovations) from the GARCH model is said to be at least approximately an iid series, which makes EVT plausible. The distribution of the residuals, estimates of the conditional mean and volatility and EVT based quantile are used for estimating one day ahead VaR. The forecasting of daily VaR will be based on historical data using a moving window of the 1000 most recent log-returns.

1.2 Aim and purpose of the thesis

The purpose of this thesis is to give an introduction on how extreme value theory (EVT), together with time-series models, can be applied in financial risk management when estimating VaR. The aim is to provide a theoretical framework for EVT and present how it can be used together with an AR-GARCH model to forecast the daily VaR. The empirical analysis is based on price-movements for the Swedish OMXS30 index, short for OMX Stockholm 30, as well as the price-movements for Ericsson.

2 Theoretical framework

This section will provide the theoretical tools and formulas that will be useful for the analysis. Since the main interest involves the future possible *losses* the focus throughout the thesis will be on the negative return series, i.e. a loss is a positive value.

2.1 Return series

Let P_t be the daily closing price of an asset at time t. The simple gross return R_t, between time-period $t-1$ and t, is defined as

$$1 + R_t = \frac{P_t}{P_{t-1}} \tag{1}$$

The continuously compounded return or log return, r_t, is defined as the natural logarithm of (1), and will be the formula used in this thesis

$$r_t = \log(1 + R_t) = \log\Big(\frac{P_t}{P_{t-1}}\Big) \tag{2}$$

The loss returns can then be expressed as

$$X_t = -r_t \tag{3}$$

2.2 Value at risk (VaR)

VaR is defined as the maximal loss during a given time period for a given probability and can be applied in both analysis with one asset as several - a portfolio. In a mathematical point of view [9], VaR is defined as the p-th quantile of the distribution F_X

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : F(x) \geq p\} = F_x^{-1}(1-p) \qquad 0 < p < 1 \tag{4}$$

For example, if a portfolio of stocks has a daily 5% VaR of \$1 million, there is a 0.05 probability that the portfolio will fall in value by at least \$1 million over one day. A loss of \$1 million or more on this portfolio is expected on 1 day out of 20 days, because of 0.05 probability.

2.3 Time-series models

In the same way as [8], assume that X_t is a stationary time series representing the daily loss returns. At time t the time-series is said to follow the dynamics of a first order stochastic volatility model

$$X_t = \mu_t + \sigma_t Z_t \tag{5}$$

where μ_t is the expected return at time t, σ_t the volatility at time t and Z_t are a white noise series, i.e. independent and identically distributed with zero mean, unit variance and marginal distribution function $F_Z(z)$. Z_t is commonly assumed to be standard normal distributed. The parameters μ_t and σ_t are assumed to be measurable with respect to the information available up until time $t-1$, denoted I_{t-1}. Based on the model in (5) the VaR measure of interest is

$$\mathrm{VaR}_p^t = \mu_{t+1} + \sigma_{t+1}\mathrm{VaR}^t(Z_p) \tag{6}$$

where μ_{t+1} is the one day ahead predicted expected return, σ_{t+1} is the one day ahead predicted volatility and $\mathrm{VaR}(Z_p)$ is the estimated quantile based on some properly assumed distribution.

2.3.1 Autoregressive model (AR)

An autoregressive model (AR) is a representation of a random process used to describe the time-varying dependence between the variable of interest and its past values. Since daily returns tend to exhibit some serial autocorrelation, an AR model is used to model the dependence and to forecast the conditional mean. An AR model that depends on its r past observations is called an AR model of order r, AR(r), and is mathematically defined as [9]

$$X_t = \phi_0 + \sum_{r=1}^{r} \phi_{t-r} X_{t-r} + \epsilon_t \tag{7}$$

where ϕ_0 is a constant, ϕ_i are coefficients and $\epsilon_t \sim$ iid $(0, \sigma^2)$. The AR(1) model applied in this thesis is expressed as

$$X_t = \phi_0 + \phi_1 X_{t-1} + \epsilon_t \tag{8}$$

where the loss returns at time t depend on the constant ϕ_0, the previous observation X_{t-1} and the innovation ϵ_t, describing the variability or randomness in the model.

2.3.2 ARCH / GARCH models

The autoregressive conditional heteroscedastic (**ARCH**) model was proposed by Engle in 1982 in order to model financial time-series that exhibit time-varying volatility clustering. The conditional variance is modeled as a linear function of past squared residuals, ϵ^2, also known as the ARCH-terms. The general ARCH(p) is defined as [9]

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 \tag{9}$$

where $\alpha_0 > 0$, $\alpha_i \geq$ to keep the conditional volatility postitive. However, financial time-series tend to need too many p in order to fit data well. For this purpose, the generalized autoregressive conditional heteroscedastic (**GARCH**) model was introduced by Bollerslev in 1986. GARCH is also a weighted average of past squared residuals, but adding the GARCH terms, past σ^2. According to [4], it gives parsimonious models that are easy to estimate and has proven to be successful in predicting conditional variances. The general GARCH (p,q), is defined as [9]

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{10}$$

where $\epsilon_t \sim$ iid $(0, \sigma^2)$, $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ are the coefficients of the model such that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. The most commonly used form of the GARCH model is the GARCH (1,1), adapted in this thesis

$$\begin{aligned} X_t = &\quad \mu_t + \epsilon_t \\ \sigma_t^2 = &\quad \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \tag{11}$$

where the volatility at time t depends on the previous squared residual, ϵ_{t-1}^2, the previous conditional variance σ_{t-1}^2 and the sum $\alpha_1 + \beta_1$ indicate how fast the variance tend to go back to its long-run weighted average.

2.4 Extreme Value Theory (EVT)

Extreme value theory (EVT) has since the middle of 1990s become a commonly adaptable approach for estimating financial and insurance risk, see Embrechts. et al (1997) [2]. McNeil (1997) [7] is using the statistical methods of EVT to estimate the tail losses for Danish fire insurance data. Embrecths et. al (1999) [3] use EVT as a risk management tool for several purposes within finance and insurance. The concept behind EVT is to consider the distribution of maxima (or minima) so that the focus is on the tails rather the center of the distribution. EVT plays the same fundamental role for the asymptotic behavior of the extremes as the central limit theorem plays when modelling sums of random variables.

EVT can be divided into two main parametric approaches, *block maxima* which is based on the Fisher-Tippett-Gnedenko theorem and *peaks over threshold* (POT-model) resulting from the Pickands-Balkema-de Haan theorem. For this specific analysis the POT method uses data more efficiently and is therefore considered as the most useful. Thus the focus will be on the POT method.

2.4.1 Generalized extreme value distribution (GEV)

Let $X_1, ..., X_n$ be a sequence of iid random variables with common distribution function $F_X(x)$. The distribution for the maximum $M_n = \max\{X_1, X_2, ..., X_n\}$ is, due to the assumption of independence for the variables, expressed as

$$P(M_n \leq x) = P(X_1 \leq x, ..., X_n \leq x) = F^n(x), \qquad x \in R \tag{12}$$

The Fisher-Tippett-Gnedenko theorem states that if there exists constants $c_n > 0$ and $d_n \in R$ and some non-degenerate distribution function H such that [2]

$$P((M_n - d_n)/c_n) \leq x) = F^n(c_n x + d_n) \to H(x), \qquad \text{as } n \to \infty \tag{13}$$

then H must be one of the three distributions (using terms related to financial markets, μ and $\sigma > 0$)

The Fréchet family: $\xi > 0$ and with CDF

$$H(x) = \left\{ \begin{array}{ll} \exp\Big[-(1 + \xi\Big(\frac{x-\mu}{\sigma}\Big)^{-1/\xi}\Big] & \text{if } x > -1/\xi \\ 0 & \text{otherwise} \end{array} \right\}$$

The Gumbel family: $\xi = 0$ and with CDF

$$H(x) = \exp\Big[-\exp(-\Big(\frac{x-\mu}{\sigma}\Big))\Big] \qquad -\infty < x < \infty$$

The Weibull family: $\xi < 0$ and with CDF

$$H(x) = \left\{ \begin{array}{ll} \exp[-(1+\xi\left(\frac{x-\mu}{\sigma}\right))^{-1/\xi}] & \text{if } x < -1/\xi \\ 0 & \text{otherwise} \end{array} \right\}$$

where μ is the location parameter, σ the scale parameter and ξ is the shape parameter controlling the tail behavior of the limiting distribution indicating the thickness of the tail. The generalized extreme value distribution is a three-parameter combination of these three types of distributions with CDF

$$G_{\mu,\sigma,\xi}(x) = \exp(-[1+\xi(\frac{x-\mu}{\sigma})]^{-1/\xi}), \qquad \xi \neq 0 \tag{14}$$

The parameters μ, σ and ξ are estimated using the maximum-likelihood method (see Appendix). For the block maxima method the time series is divided into equally sized blocks, where the maxima for each block is determined and GEV parameters are fitted for these. The estimated VaR-measure for a financial position with negative log-returns X_t is [9]

$$\widehat{\text{VaR}}_p = \left\{ \begin{array}{ll} \hat{\mu}_n - \frac{\hat{\sigma}_n}{\hat{\xi}_n}\left\{1 - [-n\ \ln(1-p)]^{-\hat{\xi}_n}\right. & \text{if } \xi \neq 0 \\ \hat{\mu}_n - \hat{\sigma}_n\ \ln[-n\ \ln(1-p)] & \text{if } \xi = 0 \end{array} \right\} \tag{15}$$

where n is the length of each block and $(\hat{\xi}_n, \hat{\mu}_n, \hat{\sigma}_n)$ are the estimates obtain using maximum-likelihood. The subscripts are used to indicate that the estimates depend on the size of n.

2.4.2 Generalized Pareto distribution (GPD)

The more modern approach of EVT is the Peaks over threshold, also known as the POT-model. Instead of fitting a distribution to the maxima as with the GEV distribution, the POT-model involves estimating the conditional distribution of exceedances beyond some threshold u, where the exceedances $X_k - u$ is said to belong to the generalised Pareto distribution (GPD).

Again, consider observations of iid random variables $X_1, ..., X_n$ with common unknown distribution function F_X. According to the Pickands-Balkema-de Haan theorem, $X_k - u$ is well approximated by the GPD. The conditional excess distribution function for X over the given threshold u is defined as [1]

$$F_u(x) = \text{P}(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \qquad \text{for } x \geq 0. \tag{16}$$

where $F_u(x)$ is interpreted as the probability that a loss exceeds u by no more than x given that the threshold is exceeded. The approximation is

$$F_u(x) \approx G_{\xi,\beta}(x) \qquad \text{as } u \to \infty \tag{17}$$

The cumulative distribution function for the GPD is given by

$$G_{\xi,\beta}(x) = \left\{ \begin{array}{ll} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\beta) & \text{if } \xi = 0 \end{array} \right\}$$

where the scale parameter $\beta > 0$, the support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$. The distribution for $F(x)$ can be expressed as

$$F(x) = F(u) + [(1 - F(u))G_{\xi,\beta}(x - u)] \qquad x > u \tag{18}$$

We then require an estimate of $F(u)$ which according to [1] can be approximated by the empirical distribution function $\bar{F}(u) = \frac{n-N_u}{n}$ where n is the total number of observations and N_u the number of observations above the threshold. In this way the tail estimator can be written as

$$F(x) = 1 - \frac{N_u}{n}(1 + \hat{\xi}\Big(\frac{x-u}{\hat{\beta}}\Big))^{-1/\hat{\xi}} \tag{19}$$

and inverting this formula give us the estimate for VaR

$$\widehat{\mathrm{VaR}}_p = u + \frac{\hat{\beta}}{\hat{\xi}} \left\{ \Big[\frac{n}{N_u}(1-p)\Big]^{-\hat{\xi}} - 1 \right\} \tag{20}$$

3 Methodology

This section will present the methods for the analysis, starting with an overview of the data sets. For OMXS30 the set consist of price-movements starting 1991-08-01 until 15-04-30, resulting in 5961 daily log-returns. For Ericsson it is covered from 1989-09-28 until 15-04-30, resulting in 6448 daily log-returns.

3.1 Data analysis

As outlined, there are some features for daily asset returns that need to be considered before modeling, such as the assumption of iid series, the presence of volatility clustering and heavy tails. For illustrative purposes several graphs are used to get a rough sense of this. Starting with the time series for the loss returns.

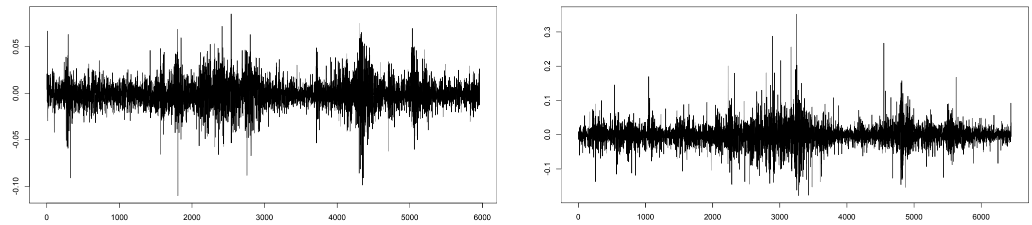

Figure 1: Loss returns for OMXS30 (left) and for Ericsson (right)

As seen in Figure 1 the phenomenon of volatility clustering for the extreme values can be noticed for both assets. There are several periods displaying high volatility followed by periods of less. In Table 1 some descriptive statistics for the time series are summarized. The largest loss for Ericsson is nearly four times the size of OMXS30 indicating a heavier right tail.

Data	Mean	Std.Dev	Skewness	Kurtosis	Min	Max
OMXS30	-0.00035	0.01482	-0.10938	3.91428	-0.11022	0.08527
Ericsson	-0.00026	0.03169	0.65419	9.84829	-0.17768	0.35198

Table 1: Basic statistics for the loss return series.

The qq-plot is a useful graphical tool for studying the tails of a distribution. Consider independent and identically distributed random variables $X_1, ..., X_n$ from which we have observations $x_1, ..., x_n$. The ordered sample for the data is $x_{1,n} \geq$

... $\geq x_{n,n}$. The points in the q-q plot are the pairs

$$\Big(F^{-1}\Big(\frac{n-k+1}{n-1}\Big), x_{k,n}\Big)$$

where $F^{-1}(.)$ are the quantiles of the reference distribution, $x_{k,n}$ the empirical quantiles and $k = 1, ..., n$. Using a qq-plot one should expect that the line is approximately linear if the empirical distribution belong to the reference distribution.

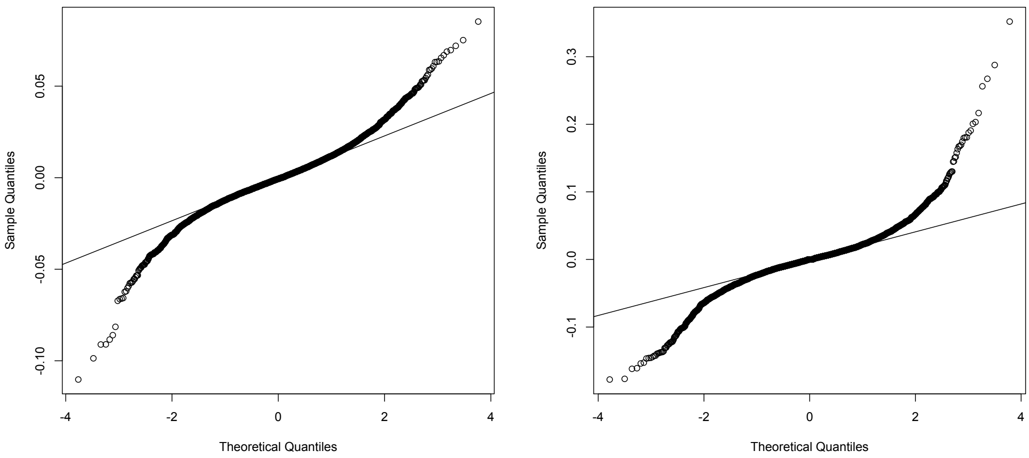

Figure 2: Qq-plots for OMXS30 (left) and Ericsson (right) with the normal quantiles as reference-line indicating that both the returns series are leptokurtic. (No obvious visible comparison between them due to different y-axis)

Figure 2 indicates that the use of EVT for the tails is adequate since the tails are heavier than for the normal distribution. But in order to use EVT the data should be iid, which can be investigated by autocorrelation plots.

An autocorrelation plot is based on the autocorrelation function (ACF) and is used to investigate the assumption of independence and if heteroscedasticity occurs by testing the squared returns. The dotted lines in Figure 3 are a 95% confidence interval for the estimators. If more than 5% exceeds the confidence bound it implies the presence of autocorrelation or heteroscedasticity. It might be hard to see this straight from the plot, so a Ljung-Box test is performed. (For a mathematical framework of the ACF and Ljung-Box test, see Appendix). The p-values from the Ljung-Box test can be found in Table 2.

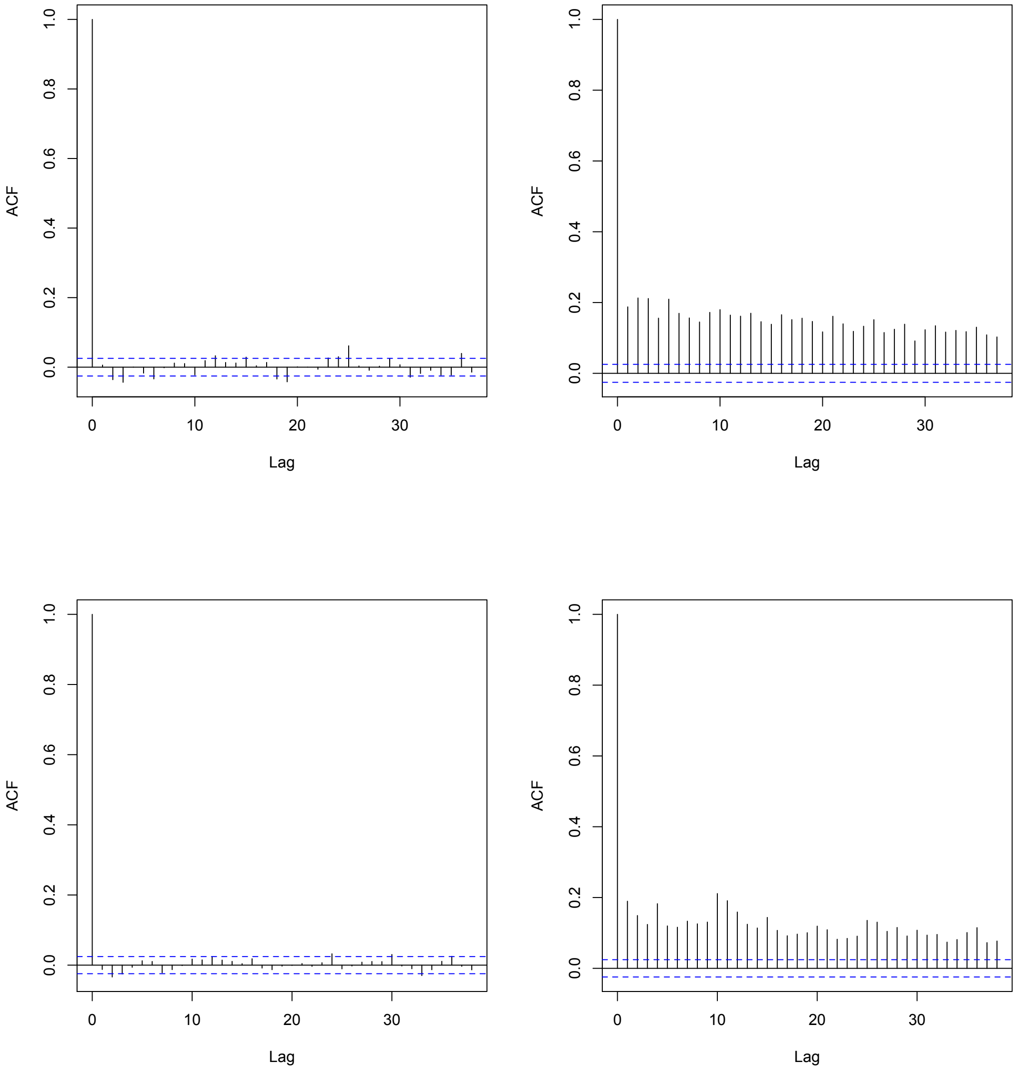

Figure 3: **Upper:** ACF for OMXS30 loss series (left) and for the squared observations (right). **Lower:** ACF for Ericsson loss series (left) and for the squared observations (right).

Based on the ACF plots in Figure 3 and the corresponding Ljung-Box test in Table 2, an AR-GARCH model seems appropriate and necessarily in order to filter the loss returns for autocorrelation and heteroscedasticity.

	data	p-value
Loss returns	OMXS30	0.6423
Squared returns	OMXS30	$< 2e - 16$
Loss returns	Ericsson	0.3224
Squared returns	Ericsson	$< 2e - 16$

Table 2: P-values from the Ljung-Box test for both OMXS30 and Ericsson. No rejection for the null hypothesis of 'no autocorrelation' but a significant result for the presence of heteroscedasticity.

3.2 Estimating VaR

The assumption that μ_t and σ_t are measurable with respect to the information available until $t-1$, results in application also for predictive purposes. The one step ahead predictive distribution is [8]

$$F_{X_{t+1}|I_t}(x) = P(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leq x | I_t) = F_Z\Big(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\Big) \tag{21}$$

which lead us to the VaR estimate

$$\widehat{\text{VaR}}^t{}_p = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \widehat{\text{VaR}}^t(Z_p) \tag{22}$$

where we will follow the framework of [8], using an AR(1)-GARCH(1,1) model to predict the conditional mean and volatility and instead of the commonly assumed standard normal distribution for Z_t we use the quantile estimation based on GPD for the residuals $z_t = \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}$. The estimation will contain a fixed number of $n = 1000$ previous negative log-returns.

The VaR estimation will contain two steps:

1. An AR(1) - GARCH(1,1) model is fitted to the historical loss returns using pseudo maximum-likelihood estimation (see Appendix). From the fitted model, the one day ahead predictions of μ_{t+1} and σ_{t+1} are estimated and the residuals are extracted for using in step 2 and for model validation.

2. Based on the extracted residuals, EVT is applied for quantile estimation using GPD. The parameters for GPD are estimated using ordinary maximum likelihood estimation.

To obtain the parameter estimates $\hat{\theta} = (\hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta})$, the likelihood for the fitted AR(1)-GARCH(1,1) model is maximized in a conventional way using the assumption of normality (see Appendix). At each time t, a collection of 1000 $(\hat{\mu}_{t-n+1}, .., \hat{\mu}_t)$

and $(\hat{\sigma}_{t-n+1}, .., \hat{\sigma}_t)$ are used for calculating the corresponding set of residuals

$$(z_{t-n+1}, ..., z_t) = (\frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, ..., \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}) \tag{23}$$

As seen in Figure 4, the AR(1)-GARCH(1,1) model seems appropriate and there is no evidence against the assumption of iid residuals.

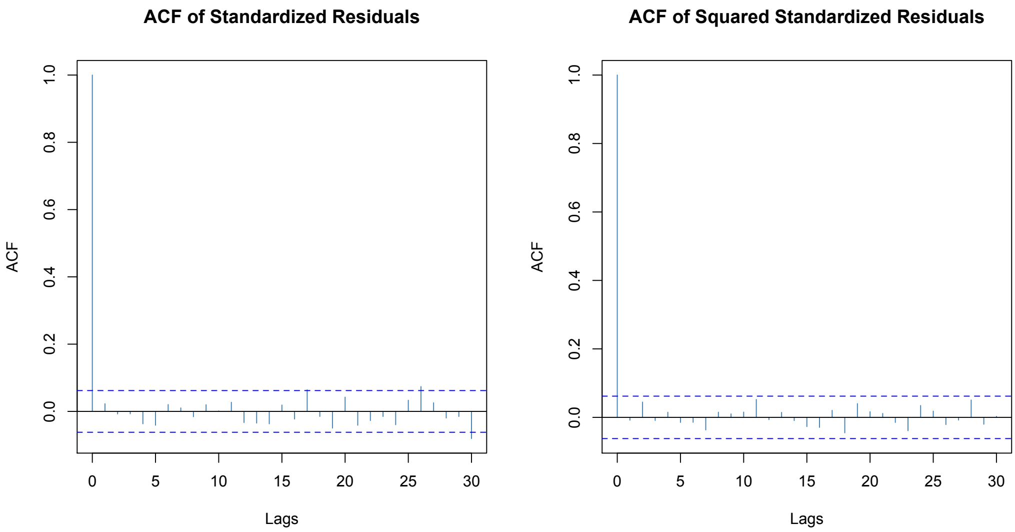

Figure 4: ACF plots for OMXS30 from an arbitrary 1000 days period where the residuals (left) and the squared residuals (right) are displayed, indicating that there are no autocorrelation or heteroscedasticity. P-values from Ljung-Box test (see Appendix), 0.4718 and 0.7939 respectively, confirm this since there is no rejection of the null hypothesis of no autocorrelation and homoscedasticity.

The one day ahead forecast for the conditional mean and standard deviation can then be calculated as

$$\hat{\mu}_{t+1} = \hat{\phi}_1 x_t \tag{24}$$

$$\hat{\sigma}_{t+1} = \sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 \hat{\epsilon}_t^2 + \hat{\beta}\sigma_t^2} \tag{25}$$

3.3 Threshold choice

Moving on to step 2, the challenge in the quantile estimation is the choice of threshold, u. It is a trade-off between observations belonging to the center of the distribution and those belonging to the tail. In this analysis we will continue to follow the same procedure as [8], by fixing $N_u = k$ to be 100 for each estimation period.

In this way the threshold is random within each set but the tail will always contain 10% of the observations such that

$$\widehat{\mathrm{VaR}}(Z_p) = z_{[k+1]} + \frac{\hat{\beta}}{\hat{\xi}} \left\{ \left[\frac{n}{k}(1-p) \right]^{-\hat{\xi}} - 1 \right\} \tag{26}$$

where $z_{[k+1]}$ is the (k+1):th ordered residual.

When the quantile based on GPD is calculated, the combined estimate for VaR in (22) is complete and can be used for backtesting, which will be desribed in the next section. The backtesting will be based on a moving window, such that at each time t, a new set of AR(1)-GARCH(1,1) parameters, residuals and GPD based quantile are estimated. In this way we will have 4961 predictions for OMXS30 and 5448 predictions for Ericsson.

3.4 Backtesting VaR

To verify the accuracy of the predicted VaR, [6] present one well-known method called *backtesting*. It is an essential tool in model validation to observe whether or not the chosen model is adequate. It involves systematically comparing the actual losses with the estimated VaR. Meaning that the estimated VaR at time t can be compared with the actual loss on day $t+1$. For the model to be adequate the expected result should be

$$P(X_{t+1} > \widehat{\mathrm{VaR}}_p^t) = 1 - p \tag{27}$$

If the confidence level is chosen to be 99% (p = 0.99) and the number of observations are 1000, as in this analysis, one should expect 10 violations. So for the model to be satisfying, the expected number of violations for a n days period is approximately $n \cdot (1-p)$.

From these violations a binomial test, also called bernoulli trials, can be constructed in order to verify the significance of the violations. Under the null hypothesis that the model is correctly the number of violations is said to follow a binomial probability distribution [6]

$$f(x) = \binom{T}{x} p^x (1-p)^{T-x} \tag{28}$$

where T is the total number of predictions, x the number of violations and p is the probability from the VaR level. The p-value for this test is in a conventional way set to be 5%, a confidence level of 95%, and not to be confused with the confidence level for the VaR estimates. The confidence levels used for VaR are 95%, 99% and 99,5%.

The VaR estimates from the GARCH-EVT approach will be compared with ordinary GARCH(1,1) estimates with assumption of conditional normality and unconditional EVT where the VaR estimates is simply the quantile estimation based on GPD.

3.5 Software

The software used for this analysis is R. R is frequently used for both statistical and data analysis and can easily be downloaded from the official website `http://www.r-project.org/`.

The packages required for the analysis are *evir*, *fExtremes* and *fGarch* where the first two are used for extreme value theory and the latter for the time series modeling. The functions are implemented by the formulas outlined in section 2 and the methods described in section 3.

4 Results

Table 3 summarizes the number of expected violations and the actual VaR violations that occurred during the test period for the different methods. The p-values indicate the success of the estimation method based on hypothesis tests for the number of violations observed as compared to the expected number of violations.

As seen in Figure 5, the unconditional EVT do not reflect current volatility background and as expected this method express several violations in stress periods. However, one should keep in mind that this method is not really appropriate since EVT applies for iid series. The GARCH(1,1) with normally distributed innovations respond to the changing volatility but fails the binomial test in each set. This indicate the need for models based on heavy tail distributions. (Plots for backtesting of the other quantile-levels and for Ericsson can be found in Appendix, where the same results applies as for Figure 5).

	OMXS30	**ERICSSON**
Total predictions	4961	5448
0.95 Quantile		
Expected	248	272
GARCH(1,1)	296 (0.00)	234 (0.00)
GARCH-EVT	275 (0.08)	286 (0.40)
Unconditional EVT	265 (0.27)	283 (0.51)
0.99 Quantile		
Expected	49	54
GARCH(1,1)	88 (0.00)	85 (0.00)
GARCH-EVT	56 (**0.35**)	57 (**0.73**)
Unconditional EVT	71 (0.00)	69 (0.06)
0.995 Quantile		
Expected	24	27
GARCH(1,1)	53 (0.00)	64 (0.00)
GARCH-EVT	31 (**0.22**)	29 (**0.70**)
Unconditional EVT	37 (0.02)	40 (0.02)

Table 3: Results from the backtesting-procedure, with p-values from the binomial test in the brackets. GARCH(1,1) with normally distributed innovations and the unconditional EVT, only using quantile-estimation based on GPD.

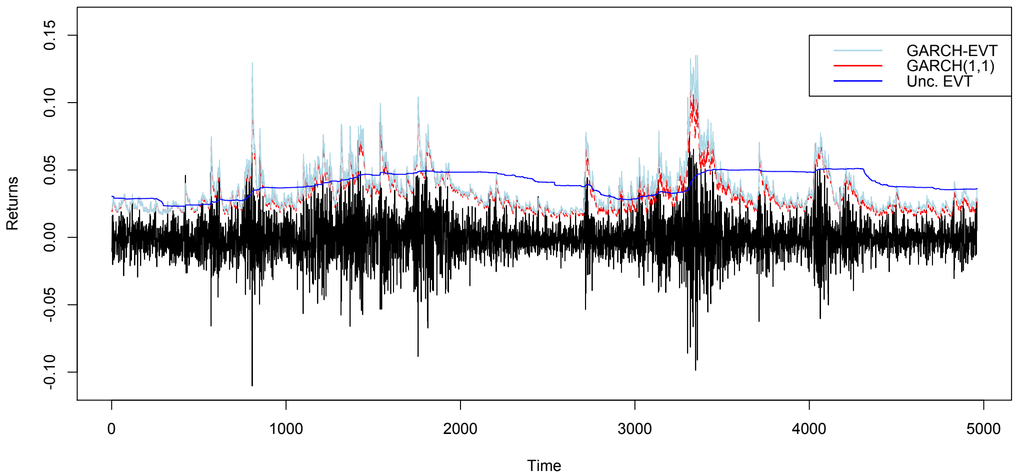

Figure 5: Backtesting for quantile 0.99 for OMXS30

5 Conclusion

The main purpose of this thesis was to demonstrate how EVT along with time-series models can be applied in risk management, in this case when forecasting daily VaR. This analysis was restricted to the case with single assets and one day predictions, in reality this type of analysis usually involves portfolios and multiple days.

As seen from the backtesting result in Table 3, the dynamic GARCH-EVT approach outperforms the other methods in 4 cases out of 6. This states the obvious importance of modeling the volatility in order to get accurate estimates for VaR as well as the clear need for EVT for higher quantiles ($p > 0.95$).

Since the main interest for the analysis was the ability to forecast VaR, the significance for the parameters in the AR-GARCH modeling has been of minor interest. The fact that there might be periods within the 'moving window' where some parameter/s are insignificant is ignored.

As discussed in section 3.3 the choice of threshold is critical. For further analysis there would be interesting to use other thresholds. One could also try different calibration periods, such as less than 1000 days.

Acknowledgement

I wish to express my gratitude to my supervisor Mathias Lindholm, at Stockholm University, for his guidance, time and patience throughout the work of this thesis.

References

[1] Davison, Anthony C., and Richard L. Smith. *Models for exceedances over high thresholds.* Journal of the Royal Statistical Society. Series B (Methodological) (1990): 393-442.

[2] Embrechts P., Klüppelberg C., and Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance.* Spring Verlag, Berlin.

[3] Embrechts, P., Resnick, S.I. and Samorodnitsky, G. (1999) *Extreme value theory as a risk management tool.* North American Actuarial Journal.

[4] Engle. R (2001) *GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics.* Journal of Economic Perspectives—Volume 15, Number 4—Fall 2001—Pages 157–168.

[5] de Haan L. and Ferreira A. (2006) *Extreme Value Theory - An introduction.* Springer Science + Business Media LLC, New York.

[6] Jorion, P. (2007) *Value at Risk - The new benchmark for managing financial risk.* Third edition, McGraw-Hill.

[7] McNeil A. J. (1997) *Estimating the tails of loss severity distributions using extreme value theory.* ASTIN Bulletin, 27.

[8] McNeil, A. J., & Frey, R. (2000). *Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach.* Journal of empirical finance, 7(3), 271-300.

[9] Tsay R. S. (2010) *Analysis of Financial Time Series, Third Edition.* John Wiley & Sons Inc, New Jersey.

6 Appendix

6.1 Skewness and kurtosis

Let $x_1, .., x_N$ be a random sample of X with N observations. Then the sample *skewness* $\hat{S}$ and sample *kurtosis* $\hat{K}$ are defined as [9]

$$\hat{S} = \frac{1}{(N-1)\hat{\sigma}^3} \sum_{n=1}^{N} (x_n - \hat{\mu})^3 \tag{29}$$

$$\hat{K} = \frac{1}{(N-1)\hat{\sigma}^4} \sum_{n=1}^{N} (x_n - \hat{\mu})^4 \tag{30}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the sample mean and variance respectively. As a reference the skewness for a normally distributed sample is 0 and the kurtosis 3. (Excess kurtosis would be 0, i.e. $\hat{K}$ - 3 = 0)

6.2 Maximum likelihood estimation

The idea behind this method is to find the values for the parameters such that the observations are as likely as possible to occur. Let $x_1, ..., x_n$ be observed values from n random variables $X_1, ..., X_n$ with the same distribution function. The probability density function f is then

$$f_{X_1,..,X_N}(x_1, ..., x_n) = \prod_{k=1}^{n} f(x_k)$$

and the likelihood-function is defined as

$$L = f_{X_1,..,X_N}(x_1, ..., x_n)$$

The values that maximizes the likelihood-function is then the maximum-likelihood estimates, the ML-estimates. It is equivalent and often more preferable to use the log-likelihood-function

$$\ell = \log(L)$$

The estimates are then found by differentiate the log-likelihood function with respect to the parameter, and set equal to zero.

Since the underlying distribution for the data used in this analysis is unknown, the so called *pseudo-maximum likelihood* method is used instead. The difference between the ordinary maximum likelihood is that the estimates are obtained by maximizing a function that is related to the log-likelihood. For example, a normal distribution can be used for estimating the parameters even though the empirical distribution has heavier tails than a normal distribution. As mentioned above, financial time series often exhibit heavy tails but [8] present that maximum likelihood can still be used since it provides consistent estimators.

6.3 Autocorrelation function (ACF)

According to [9], the correlations between the variable of interest, in this case X_t, and its past values X_{t-1}, X_{t-2} and so on are referred to as serial correlations or autocorrelations. The lag-ℓ autocorrelation coefficient is defined as

$$\rho_\ell = \frac{Cov(X_t, X_{t-\ell})}{\sqrt{Var(X_t)Var(X_{t-\ell})}} = \frac{Cov(X_t, X_{t-\ell})}{Var(X_t)}$$

The lag-1 sample autocorrelation of X_t is given by

$$\hat{\rho_1} = \frac{\sum_{t=2}^{N}(X_t - \bar{x})(X_{t-1} - \bar{x})}{\sum_{t=1}^{N}(X_t - \bar{x})^2}$$

6.4 Ljung-Box test

The Ljung-Box test is one way to test the significance of the autocorrelation coefficient. According to [9] the test-statistic is defined as

$$Q(m) = N(N+2)\sum_{\ell=1}^{m}\frac{\hat{\rho}_\ell}{N-\ell}$$

where N is the length of the time-series, $\hat{\rho}$ is the autocorrelation coefficient. H_0 is rejected if $Q(m) > \chi^2$ for the alternative that dependence is present.

When testing for serial dependence the $a_t = \sigma_t Z_t$ is used. When testing for heteroscedacity, ARCH effect, a_t^2 is used.

6.5 Figures

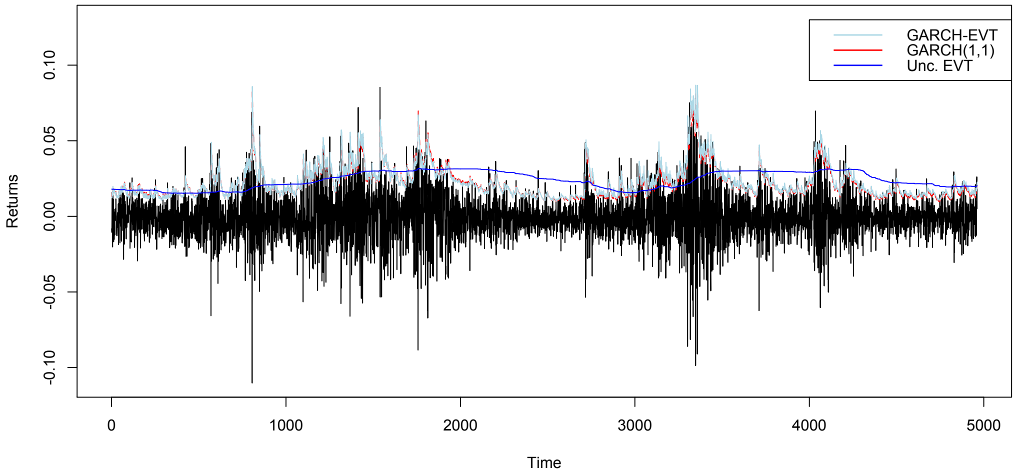

Figure 6: Backtesting for quantile = 0.95 for OMXS30

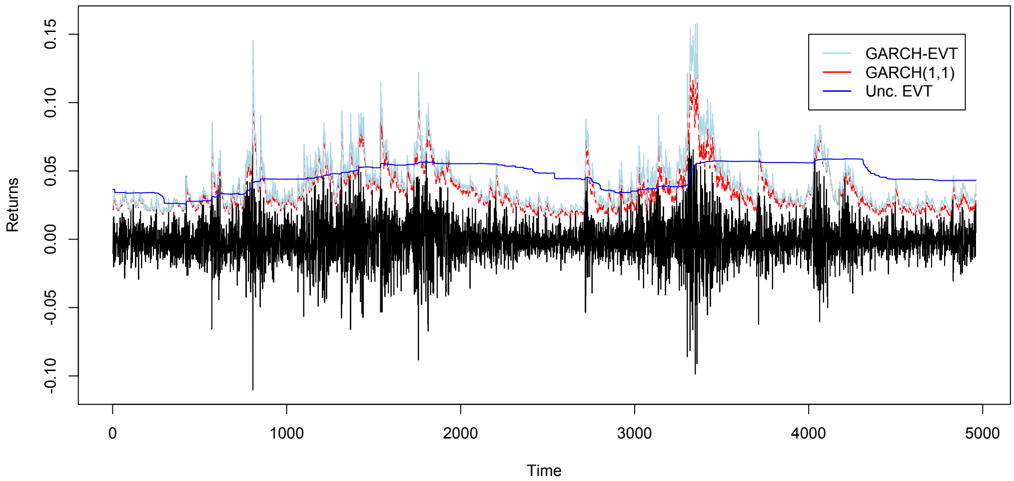

Figure 7: Backtesting for quantile 0.995 for OMXS30

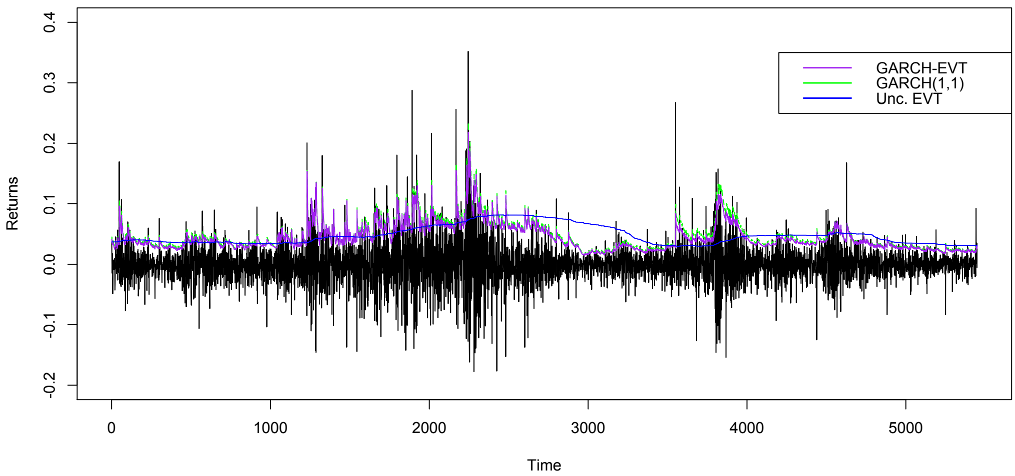

Figure 8: Backtesting for quantile 0.95 for Ericsson.

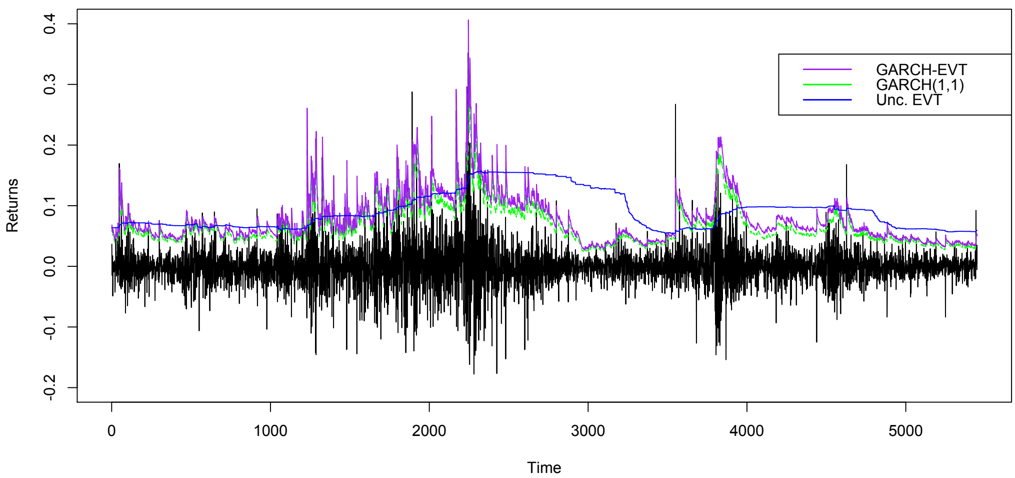

Figure 9: Backtesting for quantile 0.99 for Ericsson

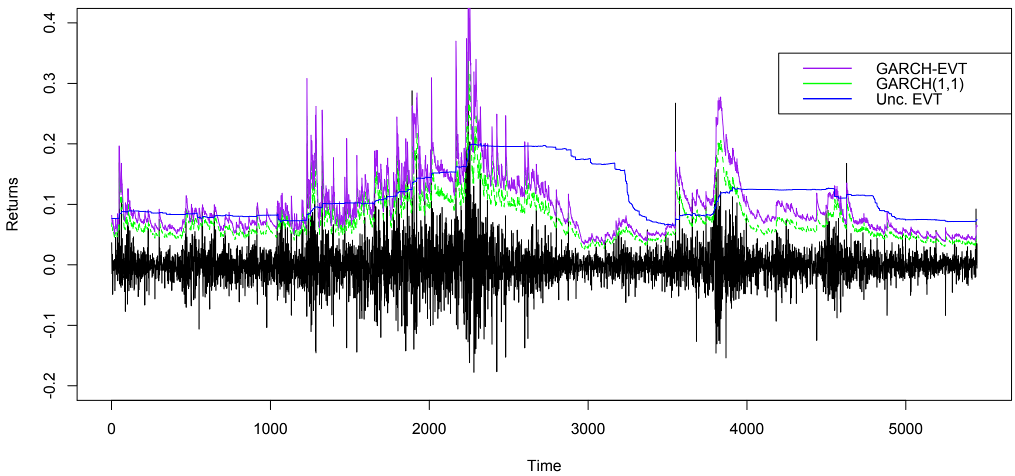

Figure 10: Backtesting for quantile 0.995 for Ericsson