# RISK EVALUATION IN FINANCIAL RISK MANAGEMENT: PREDICTION LIMITS AND BACKTESTING

**Ralf Pauly and Jens Fricke**

Working Paper 76
July 2008

INSTITUT FÜR EMPIRISCHE WIRTSCHAFTSFORSCHUNG
University of Osnabrueck
Rolandstrasse 8
49069 Osnabrück
Germany

**Risk evaluation in financial risk management: prediction limits and backtesting***

Ralf Pauly and Jens Fricke

# Contents



* We are thankful to Olena Brustina for her work on bootstrapped prediction intervals and Philip Bußmann for his support in analyzing the financial time series.

# 1 Motivation

Value-at-Risk ($VaR$) and Expected Shortfall (ES) are statistical risk measures of potential losses. Inappropriate forecasting methods can systematically lead to an understatement of market risk which can result in real losses. Therefore, financial risk managers are concerned with the precision of forecasting techniques. As significant real losses may endanger the economic stability, regulations are imposed on financial institutions. They are required to hold a certain amount of capital - a risk-based capital charge - for unexpected losses. In the Basel II regulation, the capital charge is increased for forecasting methods which underestimate the $VaR$. There, an ex-post backtesting procedure translates the degree of risk underestimation into a capital charge increase.

The focus of this paper is the understatement of market risk. We want to assess the degree of underestimating the risk by selected univariate $VaR$ and ES forecasting procedures. This model validation may help to discriminate between them and discard inappropriate methods.

The underestimation can be measured by the rate in which the $VaR$ and the ES exceed an upper confidence limit. This ex-ante analysis, which only holds for Monte Carlo data, is related to the ex-post backtesting which can be based on real data. There, forecasting methods for VaR can be considered as reliable when the relative number of observable losses greater than the forecasted VaR is in line with a certain percentage fixed in the VaR. The relative number is an estimation of the fixed percentage. A systematic overestimation indicates a systematic underestimation of risk. A similar statement holds for the ES.

Our ultimate goal in this paper is to clarify, whether there exists a relationship between the underestimation in the backtesting procedure and in the interval forecasting, i.e., whether the accuracy of prediction limits can be empirically evaluated by backtesting based on real data.

By relating the ex-ante analysis with confidence limits to the ex-post backtesting the validation of underestimating the risk will be of relevance for empirical risk studies. For reliable methods, $VaR$ and ES point forecasts can be supplemented by valid confidence limits, which give additional information for financial risk management.

The paper is organized as follows. The framework of risk evaluation will be briefly discussed in section 2. In section 3, we present the different forecasting procedures. In section 4, we review the resampling technique to generate the $VaR$ and the ES confidence intervals for the GARCH model according to Chistoffersen and Gonçalves (2005). In section 5, we describe the Monte Carlo setup which generates data with volatility and fat tail features similar to real data. We discuss the simulation results and present an analysis of real data supplemented with confidence limits. Section 6 summarizes our results.

# 2 Risk evaluation

In univariate approaches, a series of relative losses $y_t$ ( negative log returns) on a given financial asset or portfolio in period $t$ ($t = 1, \ldots, T$) is typically modelled as follows:

$$(2.1) \qquad y_t = \mu_t + u_t$$

$$(2.2) \qquad u_t = \sqrt{h_t}\varepsilon_t$$

where $\mu_t$ is the conditional mean of $y_t$ and $h_t$ is its conditional variance (conditional on the information at time $t-1$). The sequence $\{\varepsilon_t\}$ is an independently and identically distributed (i.i.d.) process with mean zero, variance one, and distribution function $F$. For evaluating the performance of selected forecast procedures in a broader framework, we consider cases in which $F$ is a standard normal distribution, $\varepsilon_t \sim \mathcal{N}(0,1)$, a standard skew-Student distribution, $\varepsilon_t \sim t(k,s)$ ($k$ is the number of degrees of freedom and $s$ the skewness parameter, compare Bauwens and Laurent (2002) and (2005) and Jondeau, Poon and Rockinger (2007)) and a standardized general error distribution, $\varepsilon_t \sim GED(\lambda)$ proposed by Nelson (1991) for GARCH models, see also Johnson et. al.(1980) (compared with the normal distribution, $\lambda < 2$ leads to a higher kurtosis). In comparison with the normal distribution, with $t(k,s)$ and $GED(\lambda)$ we can specify distributions with thicker tails and greater peakedness which are typical characteristics of financial time series.

The volatility dynamics are modelled by a stationary $GARCH(1,1)$ model for $h_t$:

$$(2.3) \qquad h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \delta h_{t-1}, \quad 0 < \alpha_1 + \delta < 1.$$

For estimating potential losses, we consider two popular risk measures, the Value-at-Risk ($VaR$) and the Expected Shortfall (ES). Many firms assess their market risk using $VaR$ models. The Basel II regulation of market-risk charge is based on $VaR$ methods. $VaR$ has been defined as the minimum loss that occurs with a probability $p$ over a fixed time horizon. We focus on the probability $p = 0.01$ and on the time horizon of one day. The relative $1-$day $VaR_{1-p}(1|T)$ – for \$1 invested – is the conditional quantile of the losses distribution, based on information at time $T$:

$$(2.4) \qquad P\left(y_{T+1} > VaR_{1-p}\left(1|T\right)\right) = p$$

Similarly , the $ES_{1-p}(1|T)$ measures, given the information at time $T$, the expected loss in the $p$% worst cases for one day ahead, i.e., when the loss $y_{T+1}$ exceeds the $VaR_{1-p}\left(1|T\right)$:

$$ES_{1-p}(1|T) = \mathrm{E}[y_{T+1}|y_{T+1} > VaR_{1-p}(1|T)] \tag{2.5}$$

Given (2.1) and (2.2), we can express the $VaR_{1-p}(1|T)$ using the quantile $q_{1-p}$ of the distribution of the i.i.d. error $\varepsilon_t$:

$$VaR_{1-p}(1|T) = \mu_{T+1} + \sqrt{h_{T+1}}q_{1-p} \tag{2.6}$$

where $q_{1-p}$ is defined in $P(\varepsilon \leq q_{1-p}) = 1 - p$. Similarly, given (2.1) and (2.2), we can show that

$$ES_{1-p}(1|T) = \mu_{T+1} + \sqrt{h_{T+1}}ES_{1-p}, \tag{2.7}$$

where $ES_{1-p} = \mathrm{E}[\varepsilon|\varepsilon > q_{1-p}]$.

In empirical analysis, we cannot compute the true values of $VaR_{1-p}(1|T)$ and $ES_{1-p}(1|T)$, their components $\mu_{T+1}$, $h_{T+1}$, $q_{1-p}$ and $ES_{1-p}$ are unknown. In practice, these components have to be estimated to forecast the unknown risk measures $VaR$ and ES. Using resampling techniques, Christoffersen and Gonçalves have quantified the forecast error by constructing a prediction interval for the future $VaR_{1-p}(1|T)$ and $ES_{1-p}(1|T)$. Quantifying the uncertainty by forecast intervals is important, because reliable intervals allows risk managers to make more informed decisions.

However, such an interval can systematically underestimate the true risk measures. In empirical risk studies it is therefore of importance to avoid forecasting procedures with low reliability, i.e. with a strong systematic underestimation of risk. Our goal in this paper is to evaluate the reliability of forecasted intervals. The evaluation relates the ex-ante forecast with the ex-post backtesting, based on information at time $T$. There, we examine whether the loss $y_{T+i}$ exceeds the $VaR_{1-p}(1|T-i), i = 0, 1, \ldots, 249$, or equivalently, whether standardized losses $\tilde{y}_{T-i}$ exceed $q_{1-p}$, i.e. whether

$$\tilde{y}_{T-i} = \frac{y_{T-i} - \mu_{T-i}}{\sqrt{h_{T-i}}} = \varepsilon_{T-i} > q_{1-p}.$$

With i.i.d. errors $\varepsilon$, the number of exceedances $H$ has a Binomial Distribution, $H \sim \mathcal{B}(n, p)$ with $n = 250$.

Similarly with $P(\varepsilon > ES_{1-p}) = p_E$, the number of exceedances $H_E$,

$$\tilde{y}_{T-i} = \frac{y_{T-i} - \mu_{T-i}}{\sqrt{h_{T-i}}} = \varepsilon_{T-i} > ES_{1-p},$$

has a Binomial Distribution, $H_E \sim \mathcal{B}(n, p_E)$ with $n = 250$.

In backtesting, we measure the reliability of forecasting procedures by estimating $p$ by $\hat{p} = H/n$ and $p_E$ by $\hat{p}_E = H_E/n$. A systematic overestimation of $p, \hat{p} > p$, and of $p_E$, $\hat{p}_E > p_E$, indicate a systematic underestimation of risk.

# 3 Forecasting procedures

We will consider eight different approaches to forecast Value-at-Risk and Expected Shortfall.

## 3.1 Historical Simulation (HS) and Historical Volatility (HV)

The $HS$ calculates the $VaR(1|T)_{1-p}$ and the $ES_{1-p}(1|T)$ using the empirical distribution of past losses. The $HS$ forecast is

$$\widehat{VaR}^{HS}_{1-p}(1|T) = \hat{q}^{HS}_{1-p}(\{y_T\}) \tag{3.1}$$

where $\hat{q}_{1-p}(\{y_T\})$ denotes the $(1-p)$-th empirical quantile of the losses data $\{y_T\}_{t=1}^{T}$. The $HS$ forecast of $ES_{1-p}(1|T)$ is given by

$$\widehat{ES}^{HS}_{1-p}(1|T) = \frac{1}{|\{y_t > \widehat{VaR}^{HS}_{1-p}(1|T)\}|} \sum_{y_t > \widehat{VaR}^{HS}_{1-p}(1|T)} y_t, \tag{3.2}$$

where $|\{y_t > \widehat{VaR}^{HS}_{1-p}(1|T)\}|$ denotes the number of losses $\{y_t\}_{t=1}^{T}$ that are above the forecasted $\widehat{VaR}^{HS}_{1-p}(1|T)$.

Assuming constant mean, $\mu_t = \mu$, and volatility, $h_t = h = \sigma^2$, and errors $\varepsilon_t$ with a standard normal distribution, the sequence of losses $\{y_t\}$ is an i.i.d process with mean $\mu$ and variance $\sigma^2$, $y_t \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$

Thus, with $\hat{u} = \bar{y}$ and $\widehat{\sigma^2} = s^2_y = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$ the $VaR_{1-p}(1|T)$ is

$$\widehat{VaR}^{HV}_{1-p}(1|T) = \hat{\mu} + \hat{\sigma} z_{1-p}, \tag{3.3}$$

where $z_{1-p}$ is the $(1-p)$-th quantile of the standard normal distribution, and

$$\widehat{ES}^{HV}_{1-p}(1|T) = \hat{\mu} + \hat{\sigma}\frac{\phi(1-p)}{p}, \tag{3.4}$$

where $\phi$ is the density function of the standard normal distribution.

## 3.2 Exponentially weighted moving average (EWMA)

**EWMA-N and EWMA-FHS**

Following a widely used method proposed by Risk Metrics, which set an industrywide standard, the variances $h_t$ are considered as changing over time and modelled using an exponentially weighted moving average ($EWMA$) approach. Formally, the forecast for time $T+1$ is a weighted average of the previous forecast and the latest loss in form of

$$h_{T+1} = (1-\lambda)(y_T - \mu)^2 + \lambda h_T \qquad 0 < \lambda < 1$$

with starting value $h_1 = \sigma^2$. In substituting $h_T$ by observable losses, the $\lambda$ parameter places geometrically declining weights on past observations. Therefore, $\lambda$ is called the decay factor. In Risk Metrics, the decay factor $\lambda$ has been chosen to be equal to 0.94. With innovation $\varepsilon_t \sim \mathcal{N}(0,1)$, the forecast of $VaR$ is

$$\widehat{VaR}^{E-N}_{1-p}(1|T) = \hat{\mu} + \sqrt{\hat{h}_{T+1}} z_{1-p} \tag{3.5}$$

and of $ES$ is

$$\widehat{ES}^{E-N}_{1-p}(1|T) = \hat{\mu} + \sqrt{\hat{h}_{T+1}} \frac{\phi(1-p)}{p}, \tag{3.6}$$

where $\hat{h}_{T+1} = 0.06(y_T - \bar{y})^2 + 0.94\hat{h}_T$ with $\hat{h}_1 = s_y^2$.

Dropping the assumption of the normal distribution, the distribution free $FHS$ forecast of VaR is

$$\widehat{VaR}^{E-FHS}_{1-p}(1|T) = \hat{\mu} + \sqrt{\hat{h}_{T+1}} \hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\}), \tag{3.7}$$

where $\hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\})$ ist the $(1-p)$-th empirical quantile of EWMA standardized losses

$$\tilde{y}^E_t = \frac{y_t - \hat{\mu}}{\sqrt{\hat{h}_t}}$$

with $\hat{h}_t = 0.06(y_{t-1} - \bar{y})^2 + 0.94\hat{h}_{t-1}$ and $\hat{h}_1 = s_y^2$. The distribution free forecast of ES is

$$\widehat{ES}^{E-FHS}_{1-p}(1|T) = \hat{\mu} + \sqrt{\hat{h}_{T+1}} \widehat{ES}^{HS}_{1-p}(\{\tilde{y}^E_t\}), \tag{3.8}$$

where

$$\widehat{ES}^{HS}_{1-p}(\tilde{y}^E_t) = \frac{1}{|\{\tilde{y}^E_t > \hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\})\}|} \sum_{\tilde{y}^E_t > \hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\})} \tilde{y}^E_t$$

with $|\{\tilde{y}^E_t > \hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\})\}|$ the number of exceedances of $\hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\})$.

## 3.3 GARCH based forecasts

**GARCH-N and -FHS as well as GARCH-GPD and -Hill**

According to (2.3), the volatility $h_{T+1}$ is forecasted by

$$\tilde{h}_{T+1} = \tilde{\alpha}_0 + \tilde{\alpha}_1(y_T - \bar{y})^2 + \tilde{\delta}\tilde{h}_T$$

with the QML estimates $\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\delta}$ and the starting value $\tilde{h}_1 = s_y^2$.

Compared to the EWMA-N approach, the GARCH-N forecasts $\widehat{VaR}_{1-p}^{G-N}(1|T)$ and $\widehat{ES}_{1-p}^{G-N}(1|T)$ only differ by the estimation of the volatility $h_t,\ \ t = 2, \ldots, T+1$.
For the distribution free approach, we build $GARCH$ standardized losses

$$\tilde{y}_t^G = \frac{y_t - \bar{y}}{\sqrt{\tilde{h}_t}} \quad , t = 1, \ldots, T,$$

With these losses we compute the $(1-p)$-th quantile $\hat{q}_{1-p}^{HS}(\{\tilde{y}_t^G\})$ and the Expected Shortfall $\widehat{ES}_{1-p}^{HS}(\{\tilde{y}_t^G\})$ to calculate the forecasts

$$\widehat{VaR}_{1-p}^{G-FHS}(1|T) = \hat{\mu} + \sqrt{\tilde{h}_{T+1}}\hat{q}_{1-p}^{HS}(\{\tilde{y}_t^G\}) \tag{3.9}$$

and

$$\widehat{ES}_{1-p}^{G-FHS}(1|T) = \hat{\mu} + \sqrt{\tilde{h}_{T+1}}\widehat{ES}_{1-p}^{HS}(\{\tilde{y}_t^G\}) \tag{3.10}$$

In Extreme Value Theory (EVT), the focus of interest is not the entire distribution but the relevant tail of the distribution. The tail (or peak-over-threshold) approach considers the exceedances over a high threshold. Following McNeil and Frey (2000), we consider two different $EVT$ estimators of $q_{1-p}$ and of $ES_{1-p}$, respectively. The first one in based is the Generalized Pareto distribution ($GPD$), the so-called $GPD$-based estimation. The second one has been proposed by Hill and is related to the Fréchet extreme value distribution, the so-called Hill-based estimation. As we look for an estimation of $q_{1-p}$ and of $ES_{1-p}$, the $GPD$-based and the Hill-based estimators are based on i.i.d. variables,i.e. on the standardized losses $\tilde{y}_t = (y_t - \mu)/\sqrt{h_t} = \varepsilon_t$ with distribution function $F$. In the GARCH-GPD approach, we fix a high threshold $u$ and assume that the excess residuals $\omega_t = \varepsilon_t - u$ over the threshold $u$ have a $GPD$ with distribution function

$$F_{\xi,\beta}(\omega) = \begin{cases} 1 - (1 + \xi\omega/\beta)^{-1/\xi}, & \xi \neq 0, \\ 1 - exp(-\omega/\beta), & \xi = 0, \end{cases} \tag{3.11}$$

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

The choice of the $GPD$ is motivated by a limit result in $EVT$. According to it, the function $F_{\xi,\beta}(\omega)$ is approximately equal to the corresponding function $F_u(\omega)$ of the excesses $\omega_t$, i.e. $F_{\xi,\beta}(\omega) \approx F_u(\omega)$, where $F_u(\omega)$ is given by

$$
\begin{aligned}
(3.12) \qquad F_u(\omega) &= P(\varepsilon - u < \omega | \varepsilon > u) \\
&= \frac{F(u+\omega) - F(u)}{1 - F(u)}, \quad \omega = \varepsilon - u > 0
\end{aligned}
$$

From the approximation $F_{\xi,\beta} \approx F_u$ and a transformation of (3.12), we get

$$
(3.13) \qquad 1 - F(u - \omega) = [1 - F(u)][1 - F_u(\omega)] \approx [1 - F(u)][1 - F_{\xi,\beta}(\omega)]
$$

Using this result, we compute the $GPD$-based estimators $\hat{q}^G_{1-p}$ and $\widehat{ES}^G_{1-p}$.

Let $\tilde{y}_{(t)} = \varepsilon_{(t)}$ denote the $t$-th order statistics of $\varepsilon_t (i.e.\ \varepsilon_{(t)} \geq \varepsilon_{(t-1)}$ for $t = 2, \ldots, T)$ and let $T_u$ denote the number of standardized losses $\tilde{y}$ that exceed $u$. A natural estimate of $F(u)$ is given by

$$
\hat{F}(u) = \frac{T - T_u}{T},
$$

where $T_u$ is the number of exceedances above the threshold $u$. We specify the threshold $u$ indirectly by fixing a probability value $p^*$ near $p$, say $p^* = 0.02$ or $0.05$. With $p^*$ we calculate the $(1 - p^*)$-th empirical quantile $\hat{u}_{1-p^*}(\{\tilde{y}^G_T\})$ of the standardized losses $\{\tilde{y}^G_T\}^T_{t=1}$. Here, $1 - F(\hat{u}_{1-p^*}) = p^*$. With the estimated threshold $\hat{u}$, we get estimations $\hat{q}^E_{1-p}$ for the GPD quantile and $\widehat{ES}^G_{1-p}$ for the GPD Expected Shortfall, for further details see McNeil and Frey (2000), see also McNeil, Frey and Embrechts (2005).

In the GARCH-Hill approach, we suppose that the tail of the distribution of $\varepsilon$ is well approximated by the distribution function

$$
(3.14) \qquad F(\varepsilon) = 1 - L(\varepsilon)\varepsilon^{-1/\xi} \approx 1 - c\varepsilon^{-1/\xi}, \varepsilon > u, \xi > 0,
$$

where $L(\varepsilon)$ is a slowly varying function. As in the GPD approach, here, we also indirectly fix the threshold. With the estimated threshold, we get estimations $\hat{q}^H_{1-p}$ and $\widehat{ES}^H_{1-p}$ for the Hill method, for details compare Christoffersen and Gonçalves (2005).

# 4 Prediction limits

Here, we briefly describe the resampling technique to compute $VaR$ and $ES$ symmetric confidence intervals introduced by Christoffersen and Gonçalves (2005). The focus is on the GARCH-FHS approach with $\mu_t = \mu$. As our special interest is the evaluation of risk underestimation, we consider the one-sided upper confidence limit, too.

According to equation (2.3), the prediction of volatility

$$h_{T+1} = \alpha_0 + \alpha_1(y_T - \mu)^2 + \delta h_T \tag{4.1}$$

depends on information available at time $T$ and on the unknown parameters $\mu, \alpha_0, \alpha_1$ and $\delta$. For large $T$, we can express $h_T$ as a function of the sequence of past losses $\{y_t\}_{t=1}^T$ in form of

$$h_T = \sigma_y^2 + \alpha_1 \sum_{j=0}^{T-2} \delta^j[(y_{T-j-1} - \mu)^2 - \sigma_y^2] \tag{4.2}$$

where $\sigma_y^2 = \alpha_0/(1 - \alpha_1 - \delta)$, please refer to Christoffersen and Gonçalves (2005).

With the estimator of moments $\hat{\mu} = \bar{y}$ and $\widehat{\sigma^2} = s_y^2$ and the QML estimators $\tilde{\alpha}_0, \tilde{\alpha}_1$ and $\tilde{\delta}$, we will generate bootstrap losses $\{y_t^*\}_{t=1}^T$ by a resampling technique. These losses will be used to reestimate the unknown parameters. With the sequence of past losses $\{y_t\}_{t=1}^T$ and the estimated parameters we can renew forecasts for $h_{T+1}$ according to equation (4.1) and (4.2). The replication of resampling allows us to consider the estimation risk in forecasting $VaR(1|T)$ and $ES(1|T)$.

**Bootstrap Algorithm for GARCH-FHS**

*Step 1.* Compute $\hat{u}, \widehat{\sigma^2}, \tilde{\alpha}_0, \tilde{\alpha}_1$ and $\tilde{\delta}$ with observed losses $\{y_t\}_{t=1}^T$ and standardized losses $\{\tilde{\varepsilon}_t\}_{t=1}^T$, where $\tilde{\varepsilon}_t = \tilde{y}_t^G = (y_t - \hat{\mu})/\sqrt{\tilde{h}_t}$.

*Step 2.* Generate a bootstrap pseudo series of standardized losses $\{\tilde{\varepsilon}_t^*\}_{t=1}^T$ by resampling with replacement from the standardized losses $\{\tilde{\varepsilon}_t\}_{t=1}^T$. Using this bootstrap series $\{\tilde{\varepsilon}_t^*\}_{t=1}^T$, compute a bootstrap pseudo series of losses $\{y_t^*\}_{t=1}^T$ by the recursions

$$\tilde{h}_t^* = \tilde{\alpha}_0 + \tilde{\alpha}_1(y_{t-1}^* - \hat{\mu})^2 + \tilde{\delta}\tilde{h}_{t-1}^*$$

$$y_t^* = \hat{\mu} + \sqrt{\tilde{h}_t^*}\tilde{\varepsilon}_t^* \quad , t = 2, \ldots, T,$$

where $\tilde{h}_1^* = \tilde{h}_1 = \widehat{\sigma^2}$ and $y_1^* = \hat{\mu} + \sqrt{\tilde{h}_1^*}\tilde{\varepsilon}_1^*$. With the bootstrap pseudo-data $\{y_t^*\}_{t=1}^T$, calculate the bootstrap estimates $\hat{u}^*, \widehat{\sigma^{2*}}, \tilde{\alpha}_0^*, \tilde{\alpha}_1^*$ and $\tilde{\delta}^*$.

*Step 3.* Make a bootstrap prediction $\tilde{h}_{T+1}^*$ according to

$$\tilde{h}_{T+1}^* = \tilde{\alpha}_0^* + \tilde{\alpha}_1^*(y_T - \hat{\mu}^*) + \tilde{\delta}^*\tilde{h}_T^*$$

with

$$\tilde{h}_T^* = \widehat{\sigma^{2*}} + \tilde{\alpha}_1^* \sum_{j=0}^{T-2} \tilde{\delta}^{*j}[(y_{T-j-1} - \hat{\mu}^*)^2 - \widehat{\sigma^{2*}}]$$

*Step 4.* Build bootstrap forecasts of $VaR(1|T)$ and $ES(1|T)$ by applying the $HS$ procedure on the series of standardized bootstrap losses $\{\tilde{\varepsilon}_t^*\}_{t=1}^T$. This application yields

$$\widehat{VaR}_{1-p}^*(1|T) = \hat{\mu}^* + \sqrt{\tilde{h}_{T+1}^*}\hat{q}_{1-p}^{HS}(\{\tilde{\varepsilon}_t^*\}),$$

where $\hat{q}_{1-p}(\{\tilde{\varepsilon}_t^*\})$ is the $HS$ quantile computed with standardized bootstrap losses $\tilde{\varepsilon}_t^*$, and

$$\widehat{ES}^*_{1-p}(1|T) = \hat{u}^* + \sqrt{\tilde{h}^*_{T+1}}\widehat{ES}^{HS}_{1-p}(\{\tilde{\varepsilon}_t^*\}),$$

where $\widehat{ES}^{HS}_{1-p}(\{\tilde{\varepsilon}_t^*\})$ is the $HS$ Expected Shortfall computed with $\tilde{\varepsilon}_t^*$.

*Step 5.* Repeat step 2 to 4 a large number of times, $B$ say, and obtain a sequence of bootstrap $\widehat{VaR}^*_{1-p,i}(1|T), i = 1, \ldots, B$, and a sequence of bootstrap $\widehat{ES}^*_{1-p,i}(1|T)$.

*Step 6.* Compute the symmetric $(1-\alpha)$-confidence interval for $VaR_{1-p}(1|T)$ (prediction interval PI)

$$PI_{1-\alpha}(VaR_{1-p}(1|T)) = [\hat{q}^{HS}_{\alpha/2}(\{\widehat{VaR}^*_{1-p,i}(1|T)\}^B_{i=1}), \hat{q}^{HS}_{1-\alpha/2}(\{\widehat{VaR}^*_{1-p,i}(1|T)\}^B_{i=1})]$$

and the one-sided upper confidence limit (upper prediction limit UPL)

$$UPL_{1-\alpha}(VaR_{1-p}(1|T)) = \hat{q}^{HS}_{1-\alpha}(\{\widehat{VaR}^*_{1-p,i}(1|T)\}^B_{i=1}),$$

where $\hat{q}^{HS}_{1-\alpha}(\{\widehat{VaR}^*_{1-p,i}(1|T\}^B_{i=1})$ is the $(1-\alpha)$-quantile of the bootstrap-data $\{\widehat{VaR}^*_{1-p,i}(1|T)\}^B_{i=1}$. Similarly, build the symmetric $(1-\alpha)$-confidence interval for $ES_{1-p}(1|T)$

$$PI_{1-\alpha}(ES_{1-p}(1|T)) = [\hat{q}^{HS}_{\alpha/2}(\{\widehat{ES}^*_{1-p,i}(1|T)\}^B_{i=1}), \hat{q}^{HS}_{1-\alpha/2}(\{\widehat{ES}^*_{1-p,i}(1|T)\}^B_{i=1})]$$

and the one-sided upper confidence limit

$$UPL_{1-\alpha}(ES_{1-p}(1|T)) = \hat{q}^{HS}_{1-\alpha}(\{\widehat{ES}^*_{1-p,i}(1|T)\}^B_{i=1}),$$

where $\hat{q}^{HS}_{1-\alpha}(\{\widehat{ES}^*_{1-p,i}(1|T)\}^B_{i=1})$ is the $(1-\alpha)$-quantile of the bootstrap-data $\{\widehat{ES}^*_{1-p,i}(1|T)\}^B_{i=1}$.

With probability $1-\alpha$ the $PI_{1-\alpha}(VaR_{1-p}(1|T))$ prediction interval should cover the true forecast $VaR_{1-p}(1|T)$ which can be generated in a Monte Carlo study. There, we can examine whether the $PI_{1-\alpha}(VaR_{1-p}(1|T))$ covers the true $VaR_{1-p}(1|T)$ or not. This examination can be repeated with $m$ generated time series of length $T$ ($m$=number of replications within the Monte Carlo study). For a reliable forecast procedure the coverage rate C-R should be close to the promised rate $1-\alpha$.

When the true $VaR_{1-p}(1|T)$ exceeds the upper limit of the prediction interval PI-U $=$ $\hat{q}^{HS}_{1-\alpha/2}$, *i.e.* $VaR_{1-p}(1|T) >$ PI-U, then $PI_{1-\alpha}(VaR_{1-p}(1|T))$ underestimates the risk. The degree of a systematic underestimation of risk can be measured by the relative number of exeedances in the $m$ replications, denoted as rate of exeedance E-R. With higher number of cases where E-R exceeds $\alpha/2$, the risk underestimation gets higher. With the upper prediction limit $UPL_{1-\alpha}$, we can measure the systematic underestimation by the exeedance rate E-R. For E-R$>\alpha$, there is a systematic risk underestimation.

Similarly, we can examine the risk underestimation of forecasting procedures for ES.

# 5 Simulation and real data results

According to equation (2.6) and (2.7), the reliability of forecasts for $VaR_{1-p}(1|T)$ and $ES_{1-p}(1|T)$ depends on the efficiency of estimates for the mean $\mu$ of losses, the volatility $h_{T+1}$ and for the quantile $q_{1-p}$ of standardized losses $\tilde{y}_t = (y_t - \mu)/\sqrt{h_t} = \varepsilon_t$. The considered eight forecast procedures differ in their estimation of $h_{T+1}$ and $q_{1-p}$. Especially, methods, which are based on the standard normal distribution, may underestimate the risk of financial time series with heavy-tailed distributions, since in reality $q_{1-p}$ will be greater than the quantile $z_{1-p}$ of the normal distribution.

In our Monte Carlo design, we follow the Basel II recommendation for a scenario type evaluation of historical range for volatilities and of historical extreme risk situations, compare Basel Committee on Banking Supervision (2006), Market risk - The internal Model approach, 5 Stress testing; p.198. In addition, are analyze the effects of historical differences in the size of heavy tails.

## 5.1 Monte Carlo design

Based on a separate study investigating the effect of selected thresholds $u$ on the coverage rate C-R, we fix $u$ indirectly by its 95% quantile for the GARCH-GPD procedure and by its 98% quantile for the GARCH-Hill method.

For the prediction interval $PI_{1-\alpha}$ and the upper prediction limit $UPL_{1-\alpha}$ we set the confidence level $1 - \alpha = 0.90$. The simulation is based on $m = 25000$ Monte Carlo replications, each with $B = 100$ bootstrap replications. The length of the time series is $T = 1000$. According to the Basel II regulation, we estimate $p = 0.01$ with 250 replications. Thus, the average estimate $\overline{\hat{p}}$ is based on 100 replications.

As indicated in the introduction, the main purpose of our paper is to evaluate the risk underestimation of commonly used methods for risk forecasting. To provide quantitative evidence through a Monte Carlo study, we consider the degree of volatility and heavy tails within real financial time series, i.e. the historical range of variation for volatilities and for heavy tails.

Following the EWMA-approach of Risk Metrics, we estimate the volatility $h_t$ of each considered real financial time series $\{y_t\}_{t=1}^{T}$ by $\hat{h}_t = 0.06(y_{t-1} - \bar{y})^2 + 0.94\hat{h}_{t-1}$ with $\hat{h}_1 = s_y^2$. Thereby, we compute standardized losses $\tilde{y}_t^E = (y_t - \bar{y})/\sqrt{\hat{h}_t}$ and the quantile estimation $\hat{q}_{1-p}^{HS}(\{\tilde{y}_t^E\})$. As a high volatility can lead to a high risk, we measure the degree of **v**olatility by the relative mean **e**xcess ($VE$) over $\overline{\sqrt{\hat{h}}}$ as follows:

$$VE = \frac{1}{|\{\sqrt{\hat{h}_t} > \overline{\sqrt{\hat{h}}}\}|} \sum_{\sqrt{\hat{h}_t} > \overline{\sqrt{\hat{h}}}} \frac{\sqrt{\hat{h}_t} - \overline{\sqrt{\hat{h}}}}{\overline{\sqrt{\hat{h}}}} \times 100,$$

with $|\{\sqrt{\hat{h}_t} > \overline{\sqrt{\hat{h}}}\}|$ the number of exceedances of $\overline{\sqrt{\hat{h}}}$.

The size of heavy tails is measured by the relative **s**urplus of the **q**uantile estimation

$\hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\})$ over the corresponding normal distribution quantile $z_{1-p}$ $(QS)$ as follows

$$QS = \frac{\hat{q}^{HS}_{1-p}(\{\tilde{y}^E_t\}) - z_{1-p}}{z_{1-p}} \times 100$$

As Basel II fixes $p$ at 0.01, the corresponding normal quantile is $z_{0.99} = 2.326$.

Applying the Risk Metrics EWMA-approach to selected financial time series of Dow Jones, Nasdaq, FTSE and Nikkei, we can describe each of them by their characteristics $(QS, VE)$. The scatter diagram in Figure (5.1) informs about the volatility risk $VE$ and the tail risk $QS$. $QS$ and $VE$ are positive correlated and their center is near (12%, 36%).

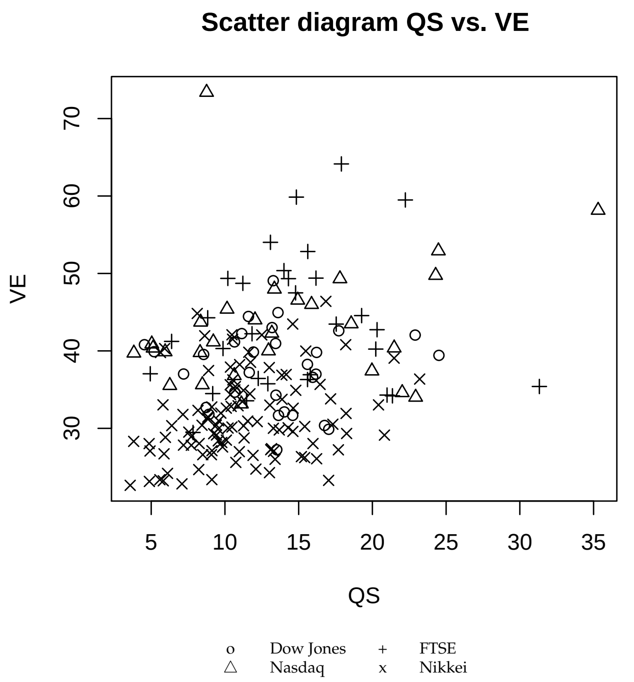



Figure 5.1

Some financial time series have a considerably higher volatility risk than $VE = 36\%$ and tail risk than $QS = 12\%$. They can be regarded as financial time series with a particularly high risk. Especially for such time series a risk manager needs reliable VaR and ES forecasts. To evaluate forecast methods in high risk situations, we generate time series with a higher $VE$ and a higher $QS$.

The basic choice of the parameters in the data generating process (DGP) is based on the $QML$ estimation of a $GARCH(1,1)$ model with daily DAX data. According to this estimation result, we set the parameters in (2.1) and (2.3) for the base line DGP with changing volatility as follows: $\mu = -0.0005, \alpha_0 = 0.000018, \alpha_1 = 0.06$ and $\delta = 0.85$. These parameter values yield a case of medium VE risk. From the volatility parameter values the expected volatility $\sigma^2$ is $\sigma^2 = \alpha_0/(1 - \alpha_1 - \delta) = 0.0002$. This value has been taken for the case of constant risk, say of no VE , i.e. constant volatility. Here, $h_t = h = \sigma^2 = 0.0002$. To increase the volatility compared to the medium VE-case, we double the value $\alpha_1$ from $\alpha_1 = 0.06$ to $\alpha_1 = 0.12$ and we reduce $\alpha_0$ from $\alpha_0 = 0.000018$

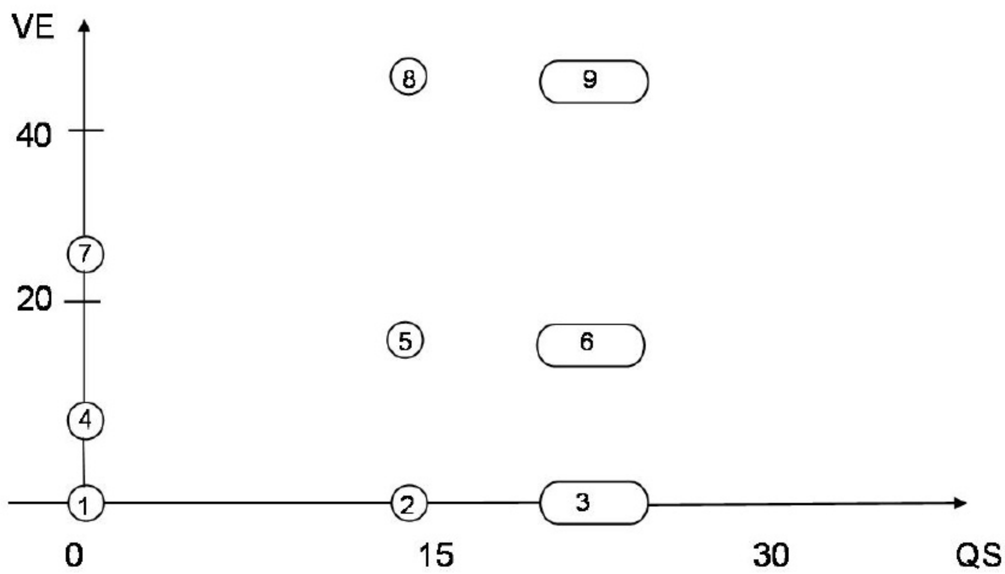
to $\alpha_0 = 0.000006$ for keeping the long run volatility at the level $\sigma^2 = 0.0002$. This change yields the case of high VE risk. Thus, in all three VE-cases the long run risk is identical.

To capture the observed high $QS$ tail risk, we substitute the standard normal distribution – no QS – by the symmetric $t(4)$-distribution – low QS – and by the asymmetric $t(4, 1.1)$- and the symmetric $GED(0.75)$-distribution – high QS. Combining the three VE-cases with the QE-cases yields nine cases of tail and volatility risk. Values of QS and VE of a presample of 100 time series of length 5000 regarding the nine cases are listed in the following Table (5.1) and Figure (5.2) visualizes the differences.

Table 5.1: Characteristic $(\overline{QS}, \overline{VE})$ of $DGP$

| | $\overline{QS}$ | $\overline{VE}$ | | |
|---|---|---|---|---|
| | | no | medium | high |
| | | $h_t = \sigma^2 = 0.002$ | $\alpha_1 = 0.06$ | $\alpha_1 = 0.12$ |
| no: | $\mathcal{N}(0,1)$ | $(0,0)$ | $(0,9)$ | $(0,28)$ |
| medium: | $t(4)$ | $(14,0)$ | $(14,16)$ | $(14,46)$ |
| high: | $t(4,1.1)$ | $(21,0)$ | $(21,17)$ | $(21,43)$ |
| | $GED(0.75)$ | $(27,0)$ | $(27,17)$ | $(26,46)$ |

$\overline{QS}$ and $\overline{VE}$ are averages from a presample of 100 time series of length 5000, respectively. The values of $\overline{QS}$ only slightly differ from their corresponding theoretical values.




Case 1: no $QS$, no $VE$; Case 2: medium $QS$, no $VE$; Case 3: high $QS$, no $VE$
Case 4: no $QS$, medium $VE$; Case 5: medium $QS$, medium $VE$; Case 6: high $QS$, medium $VE$
Case 7: no $QS$, high $VE$; Case 8: medium $QS$, high $VE$; Case 9: high $QS$, high $VE$

Figure 5.2: Representation of nine different $DGP$ cases

These nine cases cover a wide range of volatility and tail risk, which allow us to analyse extensively the prediction ability of the eight selected univariate forecasting procedures. We are interested in the precision of their VaR and ES interval forecasts – especially in the upper prediction limit –, i.e. the underestimation risk of their VaR and ES point forecast. In addition, we analyze the relationship between the risk underestimation measured by the prediction interval – the percentage of point forecasts $VaR(1|T)$ exceeding the $UPL$ – and the measurement of risk by backtesting, i.e. the percentage $\hat{p}$ of losses $y_{T+1}$ exceeding the point forecast $VaR(1|T)$. With backtesting a risk manager can assess the reliability of a forecast method empirically. For reliable methods, VaR and ES point forecasts can be supplemented by valid upper prediction limits $UPL(VaR)$ and $UPL(ES)$, which give additional information about the risk situation.

The precision of the interval forecasts is measured by the width of the prediction interval PI-With as percentage of the true VaR point forecast and by the coverage rate C-R which should not fall short of the 90% confidence level. As our main interest is the underestimation of risk, we focus on the upper prediction limit $UPL_{0.9}$ which measures the precision by its length. The reliability of the $UPL_{0.9}$ can be measured by its exceedance rate E-R. When the exceedance rate E-R considerably surpasses the nominal 10% level, then the prediction limit is not reliable as it underestimates the risk.

The control of underestimation for real financial time series is regulated by Basel II in form of the backtesting procedures. When the percentage $\hat{p}$ of relative losses is significantly higher than the target level $p = 1\%$ then the concerned methods underestimate the market risk.

Of further interest is the relationship between $\hat{p}$ and the exceedance rate E-R. The estimation $\hat{p}$ in the backtesting procedure may indicate not only that the point predictions are precise but also that the upper prediction limits are reliable.

## 5.2 Simulation results

Table (8.1) to (8.6) present the main results of six cases, the three levels of VE risk are combined with the no QS risk case and the high QS risk case of the $t(4, 1.1)$ distribution. These tables contain results of VaR as well as of ES forecasts. The first two columns inform about the relative PI-Width and its coverage rate C-R for VaR and ES, respectively. These columns allow a comparison with results given in Christoffersen and Gonçalves (2005). The next three columns inform about the 90% upper prediction limit UPL, the exceedance rate E-R and the backtesting $\hat{p}$ regarding the underestimation risk, our main interest. For Non-GARCH procedures the results are presented not only for the sample size $T = 1000$, but also for the shorter sample size $T = 500$, as we expect, that the Non-GARCH procedures are able to pick up some of the dynamics in the volatility movement in a shorter sample size, but that they are less able to do so in a greater sample size.

According to Table (8.1) to (8.6), the coverage rate of the HS method often gets worse with sample size (as stated in Christoffersen and Gonçalves). In the case of medium volatility risk VE, the HS method has a coverage rate of around 70% for $T = 500$, 20% points below the nominal rate of 90%. In the case of high volatility risk VE, the coverage rate drastically goes down to 30%.

For the no and medium VE cases, the EWMA-FHS method shows a similar performance. Contrary to the HS method, the coverage rate C-R of EWMA-FHS does not decrease in the high volatility case. Whereas, their exceedance rates E-R only slightly differ, compare Table (8.5) and (8.6), – this discrepancy between the coverage rate C-R and the exeedance rate E-R may result from the asymmetric distribution of volatility. Their PI-Widths are considerably higher than those of the HV and EWMA-N methods. However, these normal distribution based models can have a terribly low effective coverage rate. In the non-normal DGP-cases, their coverage rate can fall clearly below 40%. Their underestimation risk is very high. The rates of exceedance higher than the 90% UPL are often much too high. For the non-normal DGP-cases, these rates can attain more than 80% instead of 10%. These extremely high exceedance rates E-R are related to high backtesting values $\hat{p}$, which are confirming the underestimation. The distribution free methods HS and the EWMA-FHS perform better. However, their underestimation risk is still high. Their exceedance rates are clearly above 20%. For the greater sample size of $T = 1000$, the results do not change significantly.

In the normal case with no volatility, case 1 ”no QS risk and no VE risk”, the HV method is appropriate. There is no model risk for HV. As expected, the HV model performs best. Here, the HV procedure nearly attains the nominal exceedance rate of 10% as well as the nominal coverage rate of 90%.

In case 1, all GARCH procedures nearly attain the good HV performance regarding the C-R, the UPL, the E-R and the backtesting $\overline{\hat{p}}$. The overparameterization does not result in a great loss of accuracy.

As expected, the GARCH-N procedure performs best in its appropriate case 4 ”no QS risk, medium VE risk” and case 5 ”no QS risk, high VE risk” because the underlying data generating process fits to the procedure. Here, the GARCH-N method nearly attains the nominal exceedance rate of 10% and the nominal coverage rate of 90% with a relatively small PI-Width. The good performance is confirmed by the backtesting, $\hat{p}$ only slightly overestimates the nominal value $p = 0.01$. However, the GARCH-N procedure is not robust. In non-normal distribution DGP-cases, the GARCH-N method has terribly low effective coverage – coverage rates below 15% for $GED(0.75)$ and $t(4, 1.1)$ as data generating process – and extremely high underestimation – exceedance rates above 90%.

Contrary to the GARCH-N method, the GARCH-FHS and -GPD as well as GARCH-Hill are relatively robust. In the normal distribution DGP-cases, GARCH-FHS and -GPD methods have nearly the same coverage rate C-R and the same upper prediction limits UPL as the GARCH-N VaR forecast, and regarding C-R and UPL, the GARCH-Hill only slightly differs with its ES forecasts from the GARCH-N ES forecasts.

In the non-normal DGP-cases, the coverage rate C-R of GARCH-FHS and -GPD for VaR forecasts and of GARCH-Hill for ES forecasts only slightly fall below the target level of 90%. But the exceedance rates E-R exceed the target level of 10% and can even reach 20% in the extreme risk case 9 ”high volatility risk VE and high tail risk QS” for GARCH-FHS and -GPD in predicting the VaR and for GARCH-Hill in predicting the ES.

In addition to their high coverage rate C-R and their relative low exceedance rate E-R, prediction limits UPL of the three non-normal GARCH-based models are relatively precise as their length is mostly shorter compared to the UPL of HS and of EWMA-

FHS which are ranking in reliability behind these non-normal GARCH models and clearly before models with normal distributions. In the extreme risk case 9, the UPL of HS and EWMA-FHS are around 20% higher than the UPL of GARCH-FHS, -GPD and GARCH-Hill, respectively.

For the important non-normal DGP-cases, Table (5.2) shows a strong positive correlation between the backtesting $\overline{\hat{p}}$ and the exceedance rate E-R of the VaR as well as between the backtesting $\overline{\hat{p}}_E$ and the exceedance rate E-R of the ES. According to Table (5.2), the correlation between backtesting $\overline{\hat{p}}$ for VaR and backtesting $\overline{\hat{p}}_{ES}$ for ES is very high, especially in the non-normal DGP-cases. As a consequence of this relationship, the backtesting for ES can be based on the backtesting for VaR.

Table 5.2: Correlations between backtesting values
and bootstrap exceedance rates E-R

| | | QS | | | |
|---|---|---|---|---|---|
| | | no | medium | high | |
| | VE | $\mathcal{N}(0,1)$ | $t(4)$ | $t(4,1.1)$ | $GED(0.75)$ |
| $\overline{\hat{p}}$ | no | 0.90 | 0.83 | 0.92 | 0.94 |
| and | medium | 0.50 | 0.95 | 0.97 | 0.99 |
| E-R($VaR$) | high | 0.58 | 0.81 | 0.89 | 0.93 |
| $\overline{\hat{p}}_E$ | no | 0.79 | 0.90 | 0.93 | 0.98 |
| and | medium | 0.79 | 0.98 | 0.97 | 0.99 |
| E-R($ES$) | high | 0.47 | 0.90 | 0.94 | 0.96 |
| $\overline{\hat{p}}$ | no | 0.89 | 0.98 | 0.99 | 0.99 |
| and | medium | 0.63 | 0.97 | 1.00 | 0.99 |
| $\overline{\hat{p}}_E$ | high | 0.93 | 0.98 | 0.99 | 0.99 |

For all cases, Figure (5.3a) and (5.3b) visualize the relationship between backtesting and exceedance rates for VaR and ES. As $p_E$ only slightly varies around the average $\overline{p}_E$ in the considered four distributions, the scatterplot in Figure (5.3c) visualizes the strong positive relationship between $\overline{\hat{p}}$ and $\overline{\hat{p}}_E$.

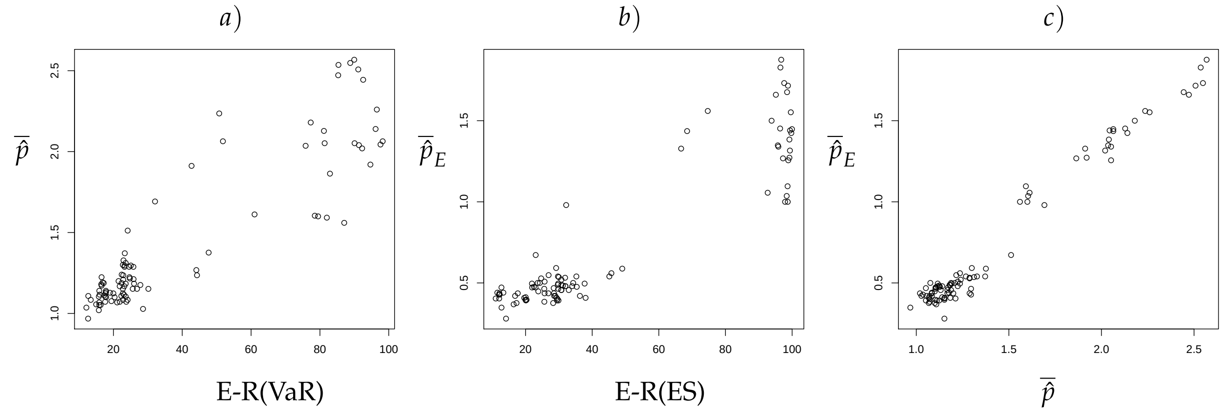



Figure 5.3: Scatterplots of backtesting values and bootstrap exceedance rates E-R

Let us examine the relationship between the ex-post backtesting and the ex-ante UPL as well as the underestimation risk in detail. For the most interesting cases with (medium and high) volatility risk VE, Table (5.3) informs about relevant characteristic values for the non-normal forecasting models – normal distribution based prediction models are discarded here, as they have an unacceptable high underestimation risk. The presented characteristic values regarding the cases with (medium and high) volatility are:

- the average backtesting $\overline{\overline{\hat{p}}}$,
- the average exceedance rate $\overline{\text{E-R}}$ for VaR and ES,
- the average upper prediction limit $\overline{\text{UPL}}$ as percentage compared to the UPL of the reference model GARCH-GPD.

Table 5.3: Characteristic average values in the cases with medium and high volatility risk VE

| | $\overline{\overline{\hat{p}}}^1$ | $\overline{\text{UPL}}$(VaR)$^2$ | $\overline{\text{E-R}}$(VaR) | $\overline{\text{UPL}}$(ES)$^2$ | $\overline{\text{E-R}}$(ES) |
|---|---|---|---|---|---|
| HS | 1.28 | 13.6% | 24.2% | 12.6% | 30.2% |
| EWMA-FHS | 1.21 | 11.8% | 22.9% | 12.9% | 28.1% |
| GARCH-FHS | 1.15 | 1.0% | 17.4% | 0.3% | 28.3% |
| GARCH-GPD | 1.11 | — | 17.3% | — | 27.2% |
| GARCH-Hill | 1.20 [3] | -2.2% | 24.8% | 6.2% | 15.7% |

1) $\overline{\overline{\hat{p}}}$ multiplied by 100

2) average percentage deviation of the UPL of the other models from the UPL of the reference model GARCH-GPD

3) in comparison to the other models the backtesting values $\overline{\hat{p}}_E$ of the GARCH-Hill model are close to the corresponding true values $p_E$ .

The GARCH-GPD has been chosen, because it performs best regarding the underestimation risk captured by the ex-post backtesting $\hat{p}$ and the ex-ante exceedance rate E-R for VaR. Next to the GARCH-GPD comes GARCH-FHS, which has nearly the same performance in predicting the VaR, i.e. nearly the same UPL, E-R and a bit higher backtesting $\hat{p}$, $\overline{\overline{\hat{p}}} = 1.11$ for GPD and $\overline{\overline{\hat{p}}} = 1.15$ for FHS. With $\overline{\overline{\hat{p}}} = 1.20$ and E-R(VaR)= 24.8% the underestimation risk of GARCH-Hill is clearly greater than the corresponding values of GARCH-GPD and -FHS. The high value of 24.8% depends on a relative low UPL estimation of the VaR. On the other hand, its UPL for ES is around 6% higher which lets GARCH-Hill performing best regarding the E-R(ES). Compared to these three GARCH-models, the other two non-normal models HS and EWMA-FHS have a significantly higher underestimation risk.

In the Basel II regulation, a risk underestimation of a certain degree yields an additional capital charge. The Basel II regulation accepts four exceedances in 250 forecasts without augmenting an additional penalty factor on-top of the usual factor $M$ which is fixed at $M = 3$ (four exceptions are within the green zone with no penalty increase, compare Jorion (2007)). Four exceptions yield an estimate $\hat{p} = 0.016$, which clearly overestimates the nominal value of $p = 0.01$. A number of exceedances in that amount or higher is related to a high underestimation of the risk in terms of the $\overline{\text{E-R}}$(VaR), compare Figure (5.3a) and Table (8.1) – (8.6). Especially the normal distribution based models HV, EWMA-N and GARCH-N are concerned. These models drastically underestimate the risk in cases with medium and high tail risk QS. Therefore, they are inappropriate for a sound risk analysis of real financial data. Their underestimation risk can be detected by the Basel II backtesting procedure. The backtesting values $\hat{p}$ are greater than 0.016 – the number of exceptions falls in the yellow or red zone with an additional penalty factor for the capital charge – indicating a considerably high underestimation risk. As we will see in the following data analysis, the increase in the penalty factor may be regarded as not high enough.

The Non-GARCH models HS and EWMA-FHS perform better. However, their underestimation risk is still high. Regarding the prediction of VaR, these two models have an E-R clearly above 20%, although their UPL are more than 10% longer as those of GARCH-GPD and FHS, which have an E-R clearly under 20%. The backtesting values $\overline{\overline{\hat{p}}}$ of these two GARCH models are clearly under $p = 0.012$, whereas the backtesting values $\overline{\overline{\hat{p}}}$ of the HS and the EWMA-FHS model are above $p = 0.012$, i.e. $\overline{\overline{\hat{p}}} = 0.0121$ for EWMA-FHS and $\overline{\overline{\hat{p}}} = 0.0128$ for HS, compare Table (5.3). This means, that in the backtesting HS and EWMA-FHS models often have $\hat{p}$ values in the neighborhood of 0.012 or 3 exceptions – the second highest number of exceptions in the green zone with no penalty factor increase. Thus, the Basel II backtesting procedure will accept models as reliable, which considerably underestimate the risk. To avoid such an underestimation, the Basel II backtesting procedure should be more restrictive.

In general, the simulation results show a systematic risk underestimation, even in the models which perform quite well, which are GARCH-GPD and -FHS for VaR predictions and GARCH-Hill for ES forecasts. Considering their underestimation risk and taking precautions, a conservative risk manager should use the reliable GARCH-GPD or -FHS models to generate upper prediction limits UPL related to VaR forecasts and the reliable GARCH-Hill models to calculate upper prediction limits UPL related to ES forecasts.

A conservative risk manager and a rigorous external risk controller will take the estimation risk into account and consider the upper prediction limit UPL as an additional risk measure. To give an idea, what this additional consideration may mean for the capital charge of market risk, let us analyze the case of medium volatility risk VE and high tail risk QS represented by the $t(4, 1.1)$ distribution. According to Table (5.4), for VaR as well as for ES the risk capital charge would increase more than 10%, when the capital charge is defined on the considered four risk measures VaR, UPL(VaR), ES and UPL(ES).

Table 5.4: A broader risk evaluation in the case
"medium VE and high QS with $t(4, 1.1)$"

| $VaR$ | $UPL$ | $ES$ | $UPL$ |
|---|---|---|---|
| 3.842 | 4.428 | 5.996 | 6.766 |
| | [15.25%][1] | | [12.84%][2] |

VaR and UPL: GARCH-FHS, ES and UPL: GARCH-Hill
1) percentage increase due to the estimation risk of VaR forecast
2) percentage increase due to the estimation risk of ES forecast

## 5.3 Real data results

According to the Basel II rules, four real financial time series have been analyzed with the GARCH-FHS model, these are the indizes of Dow Jones, DAX, FTSE and Nikkei. Therefore, we predicte the VaR on a day-by-day basis over a period of seven years taking into account a rolling estimation sample as requested by the Basel II guidelines. Table (5.5) informs about the backtesting results, the average of the relative 1-day ahead VaR forecasts and its bootstrap 90% upper prediction limit UPL. Figure (5.4) visualizes the development of the relative losses $y$, the 99% 1-day VaR and its 90% UPL for the Dow Jones in the year 2002.

Table 5.5: Basel II analysis with the GARCH-FHS model
and bootstrap UPL

| | $NE$ | $YZ$ | $RZ$ | $M$ | $\overline{\text{VaR}}$ | $\overline{\text{UPL}}$ | $\overline{\hat{p}}$ | $NP$ |
|---|---|---|---|---|---|---|---|---|
| Dow Jones | 23 | 2 | 0 | 3.02 | 2.3 | 3.0 | 1.3 | 1785 |
| Dax | 22 | 0 | 0 | 3.00 | 3.3 | 3.8 | 1.2 | 1807 |
| FTSE | 28 | 5 | 0 | 3.07 | 2.4 | 2.8 | 1.6 | 1796 |
| Nikkei | 24 | 2 | 0 | 3.03 | 3.2 | 3.6 | 1.4 | 1685 |

$NE$ = Number of exceedances
$YZ$ = Number in yellow zone within Basel II regulation
$RZ$ = Number in red zone within Basel II regulation
$M$ = Multiplication factor in the Basel II regulation
including an add-on factor due to the backtesting performance
$\overline{\text{VaR}}$ = Average 99% 1-day Value at risk
$\overline{\text{UPL}}$ = Average 90% bootstrap upper prediction limit of VaR
$\overline{\hat{p}}$ = Average backtesting value
$NP$ = Number of predictions

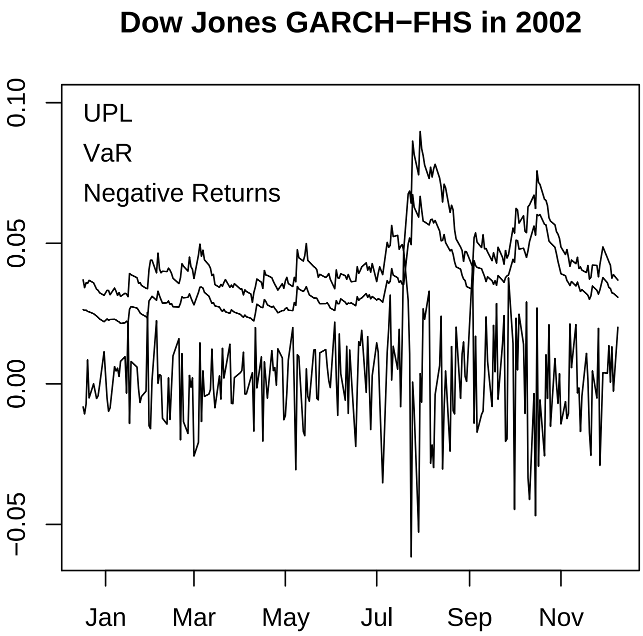



Figure 5.4: Basel II analysis of Dow Jones time series

The backtesting values $\overline{\hat{p}}$ in Table (5.5) are relatively high, especially for the FTSE time series. Here, the backtesting value is $\overline{\hat{p}} = 0.016$. According to our simulation results, the underestimation risk will be considerably high. The additional penalty factor on-top of the usual multiplication factor $M = 3$ only slightly increases the Basel II multiplicator $M$. This additional factor is introduced as "a built-in positive incentive to maintain the predictive quality" of an applied model, compare Basel Committee on Banking Supervision (2006, D. Market Risk - The Internal Models Approach, 4 Quantitative standards, p. 196). The impact of the underestimation risk on the penalty factor is presumably to weak. As a consequence, the incentives are not strong enough to maintain a high predictive quality.

# 6 Conclusion

To assess the reliability of forecasting models, it is crucial to know the degree of their potential risk underestimation. Therefore, we systematically analyze the effect of real market risk situations on the underestimation of risk for eight main univariate forecasting models. We calculate upper prediction limits for VaR and ES (based on bootstrapping) and their corresponding exceedance rates for measuring the degree of underestimation. To capture the historical relevant risk situations, we design our Monte Carlo study based on real financial time series, i.e. their volatility risk – measured by the volatility excess VE – and their heavy tail risk – measured by the quantile surplus QS.

The analysis reveals a systematic underestimation risk for all models, especially in the interesting cases with volatility and heavy tail risk. In these cases, the models based on the normal distribution have terribly high underestimation risk. They are not reliable to evaluate market risk. The HS and the EWMA-FHS models show a moderate underestimation risk with a relatively high estimation risk – their upper prediction limits are considerably high. The Non-Normal GARCH models have a low underestimation risk and a low estimation risk. Thus, their model risk is small and they can be regarded as robust. The GARCH-GPD and -FHS model are particulary reliable for predicting the VaR and the GARCH-Hill for forecasting the ES.

In the important cases with volatility and heavy tail risk, there is a strong positive correlation between the underestimation measured by the ex ante exceedance rates and the underestimation measured by the ex post backtesting within the Basel II regulation. This means, that the evaluation of risk underestimation in applied studies can be based on backtesting procedures. However, although we see a strong relationship between the ex ante and ex post risk evaluation, we conclude that the Basel II criteria within the backtesting procedures should be more restrictive, because we detect cases with 3 or 4 Basel II exceptions, that are within the Basel II green zone but show a high risk of underestimation according to the exceedance rates. Furthermore, the built-in positive incentive by the additional penalty factor in the Basel II regulation for capital charge of marked risk is not strong enough to ensure automatically a high quality in predicting the risk. Therefore, an adjustment of the Basel II guidelines has to be considered.

We suppose that our results should lead to a risk management approach, that is more integrated, i.e. inclusion of an estimate of the risk inherent in risk estimation not only for VaR predictions but also for ES forecasts. Especially in high-risk situations additional risk criteria add value for the risk manager by decreasing the risk of underestimation.

As there is a very strong positive correlation between the backtesting for VaR and for ES the empirical evaluation of risk can be narrowed to the backtesting of VaR.

# 7 References


Andreev, A., and Kanto, A. (2005). Conditional value-at-risk estimation using non-integer degrees of freedom in Student's $t$-distribution. *Journal of Risk*, **7**, 55–61.

Basel Committee on Banking Supervision (2006). International Convergence of Capital Measurement and Capital Standards: A revised Framework, Comprehensive Version. *Bank for International Settlements: Basel, Switzerland*.

Bauwens, L., and Laurent, S. (2005). A new class of Multivariate Skew Densities, With Application to Generalized Autoregressive Conditional Heteroscedasticity Models. *Journal of Business & Economic Statistics*, **23**, 346–354.

Bauwens, L., and Laurent, S. (2002). A new class of multivariate skew densities, with application to GARCH models. *CORE and Department of Economics*, Discussion Paper, Université catholique de Louvain.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, **31**, 307–327.

Christoffersen, P., and Gonçalves, S. (2005). Estimation Risk in Financial Risk Management. *Journal of Risk*, **7**, 55–61.

Davison, A.C., and Hinkley, D.V. (1997). Bootstrap methods and their applications. *Cambridge University Press: Cambridge*.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, **3**, 1163–74.

Johnson, M.E., Tietjen, G.L., and Beckman, R.J. (1980). A new Family of Probability Distributions With Applications to Monte Carlo Studies. *Journal of American Statistical Association*, **75**, 276–279.

Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). Financial Modeling under Non-Gaussian Distributions. *Springer Verlag: London*.

Jorion, P. (2007). Value At Risk: the New Benchmark for Managing Financial Risk. *McGraw-Hill: New York et. al.*. Third Edition.

McNeil, A.J., Frey, R., and Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. *Princeton University Press: Princeton and Oxford*.

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

Nelson, D. B. (1991). Conditional Heteroscedasticity in asset returns: A skewness approach. *Econometrica*, **59**, 371–396.

Theodossiou, P. (1998). Financial data and the skewed generalized $t$-distribution. *Management Science*, **44**, 1650-61.

# 8 Appendix

Table 8.1: **Case 1:** no $VE(h_t = \sigma^2 = 0.0002)$, no $QS(\mathcal{N}(0,1))$, Coverage rate C-R, 90% upper prediction limit UPL, exceedance rate E-R, backtesting $\overline{\hat{p}}$ and $\overline{\hat{p}}_E$, theoretical values: $p = 1, p_E = 0.38, VaR = 3.24$ and $ES = 3.72$[1], data generating process (DGP): HV

| | | VaR Properties | | | | | ES Properties | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | Method | PI-Width in % VaR | C-R | UPL | E-R | $\overline{\hat{p}}$ | PI-Width in % ES | C-R | UPL | E-R | $\overline{\hat{p}_E}$ |
| 500 | HS | 22.3 | 84.0 | 3.45 | 22.7 | 1.12 | 21.4 | 73.9 | 3.91 | 31.8 | 0.48 |
| | HV | 11.9 | 87.8 | 3.39 | 12.7 | 0.97 | 11.5 | 87.7 | 3.88 | 12.8 | 0.35 |
| | EWMA-N | 9.2 | 28.0 | 3.32 | 44.1 | 1.27 | 8.0 | 24.8 | 3.80 | 45.1 | 0.54 |
| | EWMA-FHS | 25.1 | 58.5 | 3.64 | 23.6 | 1.10 | 25.7 | 57.3 | 4.23 | 23.0 | 0.47 |
| 1000 | HS | 15.9 | 85.4 | 3.41 | 18.5 | 1.15 | 16.2 | 79.9 | 3.89 | 25.3 | 0.48 |
| | HV | 8.5 | 88.0 | 3.34 | 12.2 | 1.10 | 8.2 | 88.0 | 3.83 | 12.3 | 0.46 |
| | EWMA-N | 9.1 | 27.5 | 3.32 | 44.7 | 1.20 | 8.0 | 24.2 | 3.79 | 45.6 | 0.57 |
| | EWMA-FHS | 19.3 | 48.6 | 3.61 | 23.9 | 1.03 | 20.5 | 49.0 | 4.21 | 20.9 | 0.42 |
| 1000 | GARCH-N | 10.5 | 88.8 | 3.37 | 13.4 | 1.08 | 10.2 | 88.7 | 3.86 | 13.6 | 0.44 |
| | GARCH-FHS | 17.3 | 87.2 | 3.43 | 17.7 | 1.12 | 17.8 | 82.1 | 3.93 | 22.5 | 0.48 |
| | GARCH-GPD | 15.5 | 87.4 | 3.42 | 15.9 | 1.05 | 18.2 | 84.0 | 3.95 | 20.1 | 0.39 |
| | GARCH-Hill | 14.6 | 80.1 | 3.34 | 27.8 | 1.18 | 21.0 | 88.5 | 4.07 | 11.5 | 0.44 |

1) VaR and ES calculated with DGP-parameters; VaR, ES, UPL, $p$, $p_E$ and $\overline{\hat{p}}$, $\overline{\hat{p}}_E$ multiplied by 100.

Table 8.2: **Case 3:** no $VE(h_t = \sigma^2 = 0.0002)$, high $QS(t(4, 1.1))$, Coverage rate C-R, 90% upper prediction limit UPL, exceedance rate E-R, backtesting $\overline{\hat{p}}$ and $\overline{\hat{p}}_E$, theoretical values: $p = 1, p_E = 0.32, VaR = 3.96$ and $ES = 5.60$[1]

| | | VaR Properties | | | | | ES Properties | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | Method | PI-Width in % VaR | C-R | UPL | E-R | $\overline{\hat{p}}$ | PI-Width in % ES | C-R | UPL | E-R | $\overline{\hat{p}_E}$ |
| 500 | HS | 45.2 | 84.2 | 4.58 | 22.8 | 1.15 | 50.1 | 67.3 | 6.39 | 39.0 | 0.41 |
| | HV | 18.6 | 14.1 | 3.52 | 90.1 | 2.05 | 14.9 | 1.8 | 4.03 | 98.8 | 1.26 |
| | EWMA-N | 13.8 | 13.2 | 3.30 | 85.4 | 2.54 | 10.1 | 3.7 | 3.77 | 96.5 | 1.83 |
| | EWMA-FHS | 51.1 | 64.3 | 4.98 | 23.6 | 1.31 | 61.2 | 61.8 | 7.21 | 30.0 | 0.54 |
| 1000 | HS | 31.7 | 85.4 | 4.43 | 18.6 | 1.17 | 40.7 | 74.8 | 6.34 | 30.5 | 0.41 |
| | HV | 14.1 | 7.2 | 3.45 | 95.3 | 2.03 | 11.2 | 0.9 | 3.96 | 99.5 | 1.28 |
| | EWMA-N | 23.0 | 24.0 | 3.37 | 83.6 | 2.44 | 17.5 | 9.4 | 3.85 | 95.8 | 1.70 |
| | EWMA-FHS | 49.0 | 59.5 | 4.93 | 22.0 | 1.18 | 61.8 | 62.3 | 7.23 | 24.7 | 0.42 |
| 1000 | GARCH-N | 16.0 | 8.8 | 3.46 | 94.7 | 1.92 | 12.9 | 1.4 | 3.97 | 99.2 | 1.27 |
| | GARCH-FHS | 34.4 | 86.9 | 4.48 | 17.3 | 1.11 | 53.6 | 75.9 | 6.33 | 29.4 | 0.39 |
| | GARCH-GPD | 30.1 | 86.8 | 4.42 | 16.3 | 1.05 | 53.3 | 76.7 | 6.35 | 28.4 | 0.47 |
| | GARCH-Hill | 28.8 | 83.4 | 4.33 | 22.3 | 1.11 | 53.5 | 86.3 | 6.97 | 16.4 | 0.37 |

1) VaR and ES calculated with DGP-parameters; VaR, ES, UPL, $p$, $p_E$ and $\overline{\hat{p}}$, $\overline{\hat{p}}_E$ multiplied by 100.

Table 8.3: **Case 4:** medium $VE(h_t$ with $\alpha_1 = 0.06)$, no $QS(\mathcal{N}(0,1))$, Coverage rate C-R, 90% upper prediction limit UPL, exceedance rate E-R, backtesting $\overline{\hat{p}}$ and $\overline{\hat{p}}_E$, theoretical values: $p = 1, p_E = 0.38$, average values from simulation: $\overline{VaR} = 3.22$ and $\overline{ES} = 3.70$[1], data generating process (DGP): GARCH-N

| | | VaR Properties | | | | | ES Properties | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | Method | PI-Width in % VaR | C-R | UPL | E-R | $\overline{\hat{p}}$ | PI-Width in % ES | C-R | UPL | E-R | $\overline{\hat{p}_E}$ |
| 500 | HS | 23.7 | 66.1 | 3.50 | 26.0 | 1.18 | 23.2 | 63.0 | 3.99 | 29.7 | 0.49 |
| | HV | 12.2 | 44.4 | 3.39 | 28.6 | 1.03 | 11.8 | 43.4 | 3.88 | 28.9 | 0.42 |
| | EWMA-N | 9.1 | 34.9 | 3.30 | 44.3 | 1.24 | 8.0 | 31.0 | 3.77 | 45.2 | 0.56 |
| | EWMA-FHS | 24.1 | 66.9 | 3.58 | 22.9 | 1.21 | 24.5 | 65.1 | 4.14 | 23.6 | 0.50 |
| 1000 | HS | 17.0 | 55.3 | 3.46 | 24.7 | 1.06 | 17.9 | 55.8 | 3.98 | 25.5 | 0.44 |
| | HV | 8.7 | 33.0 | 3.34 | 30.6 | 1.04 | 8.4 | 32.2 | 3.84 | 30.9 | 0.42 |
| | EWMA-N | 9.0 | 34.7 | 3.29 | 44.8 | 1.26 | 7.9 | 31.0 | 3.76 | 49.0 | 0.51 |
| | EWMA-FHS | 18.6 | 58.0 | 3.54 | 22.6 | 1.07 | 19.6 | 58.2 | 4.12 | 20.8 | 0.43 |
| 1000 | GARCH-N | 13.1 | 88.2 | 3.38 | 12.1 | 1.04 | 12.9 | 88.1 | 3.88 | 12.3 | 0.43 |
| | GARCH-FHS | 19.0 | 87.9 | 3.42 | 16.6 | 1.17 | 19.0 | 82.8 | 3.91 | 22.0 | 0.47 |
| | GARCH-GPD | 17.3 | 87.5 | 3.42 | 15.8 | 1.10 | 19.4 | 84.0 | 3.93 | 20.0 | 0.40 |
| | GARCH-Hill | 16.5 | 81.4 | 3.34 | 25.9 | 1.21 | 22.0 | 88.4 | 4.04 | 12.9 | 0.40 |

1) $\overline{VaR}$ and $\overline{ES}$ calculated with DGP-parameters; VaR, ES, UPL, $p$, $p_E$ and $\overline{\hat{p}}$, $\overline{\hat{p}}_E$ multiplied by 100.

Table 8.4: **Case 6:** medium $VE(h_t$ with $\alpha_1 = 0.06)$, high $QS(t(4, 1.1))$, Coverage rate C-R, 90% upper prediction limit UPL, exceedance rate E-R, backtesting $\overline{\hat{p}}$ and $\overline{\hat{p}_E}$, theoretical values: $p = 1, p_E = 0.32$, average values from simulation: $\overline{VaR} = 3.88$ and $\overline{ES} = 5.49$[1]

| | | VaR Properties | | | | | ES Properties | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | Method | PI-Width in % VaR | C-R | UPL | E-R | $\overline{\hat{p}}$ | PI-Width in % ES | C-R | UPL | E-R | $\overline{\hat{p}_E}$ |
| 500 | HS | 47.9 | 72.9 | 4.64 | 24.9 | 1.08 | 51.3 | 63.3 | 6.45 | 36.3 | 0.42 |
| | HV | 19.1 | 27.3 | 3.49 | 75.8 | 2.04 | 15.2 | 5.3 | 4.00 | 95.7 | 1.35 |
| | EWMA-N | 14.3 | 12.5 | 3.23 | 91.1 | 2.51 | 10.5 | 2.8 | 3.69 | 98.8 | 1.72 |
| | EWMA-FHS | 50.9 | 70.3 | 4.85 | 22.4 | 1.24 | 61.5 | 65.9 | 7.03 | 30.8 | 0.52 |
| 1000 | HS | 33.9 | 66.3 | 4.48 | 20.3 | 1.07 | 42.5 | 66.4 | 6.45 | 27.9 | 0.38 |
| | HV | 14.8 | 23.4 | 3.45 | 79.1 | 1.87 | 11.8 | 3.3 | 3.95 | 97.4 | 1.28 |
| | EWMA-N | 22.1 | 22.8 | 3.28 | 89.4 | 2.22 | 16.8 | 7.4 | 3.75 | 98.2 | 1.54 |
| | EWMA-FHS | 47.2 | 67.2 | 4.76 | 20.6 | 1.08 | 59.0 | 68.9 | 6.98 | 24.6 | 0.40 |
| 1000 | GARCH-N | 20.0 | 13.2 | 3.45 | 91.3 | 2.04 | 16.1 | 1.3 | 3.96 | 99.2 | 1.38 |
| | GARCH-FHS | 37.2 | 87.3 | 4.43 | 16.5 | 1.18 | 42.9 | 75.0 | 6.19 | 30.2 | 0.49 |
| | GARCH-GPD | 33.2 | 86.4 | 4.38 | 16.0 | 1.12 | 42.7 | 75.9 | 6.20 | 29.8 | 0.50 |
| | GARCH-Hill | 32.1 | 83.4 | 4.29 | 21.5 | 1.20 | 53.6 | 85.2 | 6.77 | 17.7 | 0.44 |

1) $\overline{VaR}$ and $\overline{ES}$ calculated with DGP-parameters; VaR, ES, UPL, $p$, $p_E$ and $\overline{\hat{p}}$, $\overline{\hat{p}}_E$ multiplied by 100.

Table 8.5: **Case 7:** high $VE(h_t$ with $\alpha_1 = 0.012)$, no $QS(\mathcal{N}(0,1))$, Coverage rate C-R, 90% upper prediction limit UPL, exceedance rate E-R, backtesting $\overline{\hat{p}}$ and $\overline{\hat{p}_E}$, theoretical values: $p = 1, p_E = 0.38$, average values from simulation[1]: $\overline{VaR} = 3.07$ and $\overline{ES} = 3.53$[2], data generating process (DGP): GARCH-N

| | | VaR Properties | | | | | ES Properties | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | Method | PI-Width in % VaR | C-R | UPL | E-R | $\overline{\hat{p}}$ | PI-Width in % ES | C-R | UPL | E-R | $\overline{\hat{p}_E}$ |
| 500 | HS | 34.1 | 31.8 | 3.80 | 24.1 | 1.51 | 34.5 | 30.0 | 4.50 | 23.1 | 0.67 |
| | HV | 14.6 | 17.6 | 3.37 | 32.1 | 1.69 | 14.2 | 17.1 | 3.86 | 32.2 | 0.98 |
| | EWMA-N | 9.4 | 30.0 | 3.15 | 47.7 | 1.38 | 8.2 | 26.7 | 3.60 | 49.0 | 0.59 |
| | EWMA-FHS | 24.7 | 62.1 | 3.44 | 24.6 | 1.29 | 25.0 | 61.0 | 3.98 | 24.7 | 0.53 |
| 1000 | HS | 26.3 | 22.5 | 3.80 | 21.4 | 1.35 | 29.7 | 21.0 | 4.64 | 18.4 | 0.55 |
| | HV | 10.8 | 12.1 | 3.34 | 31.0 | 1.63 | 10.6 | 11.8 | 3.83 | 31.1 | 0.98 |
| | EWMA-N | 9.3 | 30.1 | 3.14 | 49.1 | 1.36 | 8.1 | 26.6 | 3.59 | 50.3 | 0.60 |
| | EWMA-FHS | 19.1 | 54.3 | 3.39 | 25.4 | 1.46 | 19.9 | 54.2 | 3.94 | 23.3 | 0.44 |
| 1000 | GARCH-N | 14.3 | 88.3 | 3.22 | 12.1 | 1.11 | 14.1 | 88.3 | 3.70 | 12.8 | 0.47 |
| | GARCH-FHS | 19.8 | 87.8 | 3.26 | 16.9 | 1.19 | 19.8 | 83.4 | 3.73 | 22.0 | 0.50 |
| | GARCH-GPD | 18.1 | 87.2 | 3.25 | 15.9 | 1.14 | 20.1 | 84.6 | 3.74 | 20.0 | 0.41 |
| | GARCH-Hill | 17.3 | 81.6 | 3.19 | 25.8 | 1.29 | 22.4 | 88.8 | 3.85 | 12.2 | 0.44 |

1) According to the Cauchy-Schwarz Inequality $\sqrt{\overline{h}} \geq \overline{\sqrt{h}}$. The inequality increases with higher VE. As a consequence, the $\overline{VaR}$, which is based on $\overline{\sqrt{h}}$, will decrease in comparison to Case 1 and Case 4.
2) $\overline{VaR}$ and $\overline{ES}$ calculated with DGP-parameters; VaR, ES, UPL, $p$, $p_E$ and $\overline{\hat{p}}$, $\overline{\hat{p}_E}$ multiplied by 100.

Table 8.6: **Case 9:** high $VE(h_t$ with $\alpha_1 = 0.12)$, high $QS(t(4,1.1))$, Coverage rate C-R, 90% upper prediction limit UPL, exceedance rate E-R, backtesting $\overline{\hat{p}}$ and $\overline{\hat{p}_E}$, theoretical values: $p = 1, p_E = 0.32$, average values from simulation[1]: $\overline{VaR} = 3.31$ and $\overline{ES} = 4.69$[2]

| | | VaR Properties | | | | | ES Properties | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | Method | PI-Width in % VaR | C-R | UPL | E-R | $\overline{\hat{p}}$ | PI-Width in % ES | C-R | UPL | E-R | $\overline{\hat{p}_E}$ |
| 500 | HS | 59.4 | 43.6 | 4.60 | 22.9 | 1.33 | 59.2 | 44.8 | 6.38 | 29.7 | 0.54 |
| | HV | 22.4 | 27.1 | 3.28 | 50.7 | 2.24 | 18.0 | 18.2 | 3.76 | 74.7 | 1.56 |
| | EWMA-N | 18.1 | 13.8 | 2.68 | 88.8 | 2.55 | 13.5 | 3.9 | 3.27 | 97.6 | 1.73 |
| | EWMA-FHS | 55.4 | 70.6 | 4.29 | 23.2 | 1.29 | 67.4 | 64.4 | 6.26 | 31.9 | 0.53 |
| 1000 | HS | 45.0 | 30.5 | 4.52 | 19.9 | 1.15 | 53.0 | 35.3 | 6.65 | 21.9 | 0.42 |
| | HV | 18.6 | 22.9 | 3.30 | 46.2 | 2.06 | 14.9 | 18.2 | 3.78 | 72.2 | 1.46 |
| | EWMA-N | 28.4 | 24.9 | 2.92 | 86.8 | 2.54 | 21.8 | 9.7 | 3.34 | 96.9 | 1.76 |
| | EWMA-FHS | 54.9 | 68.7 | 4.22 | 20.7 | 1.14 | 68.8 | 68.8 | 6.23 | 25.3 | 0.42 |
| 1000 | GARCH-N | 21.0 | 12.1 | 2.88 | 92.3 | 2.02 | 16.9 | 1.0 | 3.31 | 99.9 | 1.32 |
| | GARCH-FHS | 37.3 | 86.8 | 3.70 | 19.0 | 1.13 | 39.8 | 72.3 | 5.07 | 35.4 | 0.48 |
| | GARCH-GPD | 33.4 | 85.8 | 3.65 | 19.4 | 1.08 | 39.7 | 73.4 | 5.09 | 34.3 | 0.50 |
| | GARCH-Hill | 32.3 | 81.9 | 3.58 | 25.6 | 1.15 | 49.7 | 84.8 | 5.35 | 20.3 | 0.40 |

1) According to the Cauchy-Schwarz Inequality $\sqrt{\overline{h}} \geq \overline{\sqrt{h}}$. The inequality increases with higher VE. As a consequence, the $\overline{VaR}$, which is based on $\overline{\sqrt{h}}$, will decrease in comparison to Case 3 and Case 6.
2) $\overline{VaR}$ and $\overline{ES}$ calculated with DGP-parameters; VaR, ES, UPL, $p$, $p_E$ and $\overline{\hat{p}}$, $\overline{\hat{p}_E}$ multiplied by 100.