Comparing Expected Shortfall forecasts by using Filtered Historical Simulation combined with a HEAVY or GARCH model

Walter Ing	Student 415726
Sander Barendse	Supervisor
Francine Gresnigt	Co-reader

Erasmus University Rotterdam, Erasmus School of Economics

Econometrics and Operations Research

Bachelor Thesis, July 2, 2017

Abstract

We use Filtered Historical Simulation with a time-varying volatility model to forecast Expected Shortfall (ES) forecasts. Either the HEAVY or GARCH model is used as the volatility model. Shephard and Sheppard (2010) introduced the HEAVY model which has quicker adjustments than the conventional GARCH model to structural breaks in the volatility process. The aim of this paper is to compare the specification of the ES forecasts and their out-of-sample predictive ability. 34 assets including trades assets, indexes computed by MSCI and exchange rates are used in our analysis. We show that the HEAVY model is favoured for particular assets in terms of the correct specification of the model and the predictive performance.

1 Introduction

The use of downside risk measures of assets or portfolios is important for regulation of financial institutions. The Basel Committee on Banking Supervision initially proposed the Value-at-Risk (VaR) as the downside risk measure. The VaR is the quantile of the returns or losses of assets for a given horizon and a given quantile level. Since then, the VaR is widely used in the financial sector. However, more recently, the Basel Committee has approved the Expected shortfall (ES) to be used instead of the Value-at-risk (VaR). One of the reasons for this is that the VaR fails to control for "tail risk". More specifically, the VaR does not give information about the level of return if the return exceeds the VaR in the left tail (violation). The ES is related to the VaR, both require a quantile level and a horizon. The ES is the expected return over a given horizon of an asset given that the return is lower than the VaR with a specific quantile level. Thus the ES provides information if the return violates the VaR. Often the VaR and ES are viewed in terms of losses, such that these involve positive values. In our case, we consider VaR and ES in terms of returns.

We forecast the ES with Filtered Historal Simulation (FHS) combined with a time-varying volatility model. With historical data of returns and a certain volatility model, the returns can be 'filtered' or 'devolatised'. Based on these assumed i.i.d. (independently and identically distributed) standardized returns, the empirical quantile of the distribution of the standardized returns can be estimated. Then by 'revolatising' or multiplying the estimated quantile by the forecasted volatility, the VaR can be estimated. This is a semi-parametric method, because we use non-parametric estimators of the quantile of the standardized residuals. One advantage is that it is simple to apply FHS. However, FHS can be less accurate for very extreme confidence or coverage levels.

The focus of this paper is the choice of the volatility model to combine with FHS to create forecasts. One interesting option is the high frequence-based volatility (HEAVY) model introduced by Shephard and Sheppard (2010). The idea of HEAVY models were built upon the insights of the ARCH literature which were first developed by Engle (1982) and Bollerslev (1986), combined with the use of high frequent intraday data of assets. The work of Shephard and Sheppard (2010) has shown that HEAVY models have momentum and mean reversion effect, and have quick adjustments to structural breaks in the volatility process. The models were run in periods with a lot of change in the volatility level to assess the ability in stressful situations. In terms of out-of-sample performance, they found by using Giacomini and White (2006) tests that the HEAVY model is favoured compared to the standard GARCH(1,1) model. This is especially the case for shorter horizon forecasts. They claim that the use of realized measures in their HEAVY models which have a high signal-to-noise ratio may explain the strength of the HEAVY model. Furthermore, also in sample and pointwise the standard HEAVY model dominates the GARCH model, but the out-performance decreases as the horizon increases.

Naturally, the other choice for the volatility model is the GARCH(1,1) model. Besides the standard GARCH(1,1) model there are many extensions of the GARCH models that can be considered. For example GARCHX, IGARCH and GJR-GARCH models. We only use the GARCH(1,1) as the benchmark model to compare with the HEAVY model.

A 1%, 5% and 10% coverage level will be used for the ES forecasts with a horizon of one day. The forecasts will be based on univariate models, where all models are estimated by a moving window with parameters of the models updated daily. To compare the two choices of forecasting the ES, two features will be compared. Namely, the correct specification of the ES forecasts using the backtests of Du and Escanciano (2016), and a comparison of the out-of-sample predictive ability by applying the tests of Giacomini and White (2006) where the loss function of Fissler et al. (2015) is considered.

The conditional backtest of Du and Escanciano (2016) can be seen as the analog of the conditional backtest of Christoffersen (1998) for the Value-at-risk. It is based on the idea that violations should be unpredictable, i.e. a martingele difference sequence (mds). Du and Escanciano (2016) claim that the integral of the violations over the coverage level in the left tail also need to be a mds. Therefore, their conditional test check this mds property. The unconditional test is also based on this mds property and is the analoque of the unconditional test for the VaR proposed by Kupiec (1995).

Fissler et al. (2015) proposed a scoring function for the ES. The important idea is that the ES is jointly elicitable. "A functional v is elicitable relative to the class P, if there exists a scoring function s which is strictly consistent for v relative to P", defined by Ziegel (2016). This scoring or loss function can be implemented in the predictive ability tests of Giacomini and White (2006). These predictive ability tests can be used even when the model is misspecified. These tests were developed from the idea of Diebold and Mariano (1995) and West (1996) to evaluate the accuracy of a particular forecasting method rather than the accuracy of the forecasting model.

One of the findings of this paper is the quicker adjustments of ES forecasts when the HEAVY model is considered as the volatility model rather than the GARCH(1,1). The model specification from the unconditional perspective, seems to be better for a lower coverage level irrespective of the volatility model. Based on both conditional and unconditional backtests of Du and Escanciano (2016), the HEAVY model gives more often a correct model specification than the GARCH model. The HEAVY model is also preferred over the GARCH model to be applied in FHS for several assets in terms of predictive performance while the GARCH model is never favoured. The preference is more present in terms of conditional use.

The structure of the remainder is as follows. Section 2 describes the data. Section 3 elaborates the models to forecast ES and shows the methods to compare the forecasts. In section 4 the estimated models and the results of the comparisons between the ES forecasts are presented. Finally, section 5 concludes.

2 Data

In this paper the database 'Oxford-Man Institute's realised library' version 0.1 that was used by Shephard and Sheppard (2010) will be used. This database starts on 3 January 1996 and finishes on 1 March 2009. There are 34 assets in this database, including traded assets, MSCI indexes or exchange rates. For each asset, daily returns, daily subsampled realised variances and daily realised kernels are available. If the market is either closed or if the data is regarded as poor quality for a certain asset, then this data point will be treated as missing. Days are not recorded in the database when all markets are closed for these specific days.

The daily realised variance is a simple realized measure. It non-parametrically estimates the variation of the price path of an asset and ignores the variation of prices overnight. In this paper the realized variance will be used as a realized measure, defined as:

$$RM_t = \sum_{0 \leq t_{j-1,t} < t_{h,t}} x_{j,t}^2 \leq 1,\ x_{j,t} = X_{t+t_{j,t}} - X_{t+t_{j-1,t}}$$

where $t_{j,t}$ are the normalised times of trades of quotes on the tth day and X_t is the price on day t. If the prices are observed without noise and if $\min_j |t_{j,t} - t_{j-1,t}| \downarrow 0$, then RM_t consistently estimates the quadratic variation of the price on the tth day. This was formalized by Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002). In order to mitigate the noise, the realised variances are subsampled. For more details see the work by Shephard and Sheppard (2010).

Below in Table 1, a short description of the 'OMI's realised library is given. The start date and the length of available observations of each asset are shown. The assets in the table are split into three sections, namely the raw assets, the MSCI assets and the exchange rates which are all quoted against the US dollar.

Table 1: A short description of the 'OMI's realised library', version 0.1.

Asset	Start date	T	Asset	Start date	T
Dow Jones Industrials	3-1-1996	3261	MSCI Australia	5-12-1999	2314
CAC 40	3-1-1996	3301	MSCI Belgium	2-7-1999	2435
FTSE 100	21-10-1997	2844	MSCI Brazil	7-10-2002	1577
Spanish IBEX	3-1-1996	3270	MSCI Canada	13-2-2001	2003
Nasdaq	3-1-1996	3262	MSCI Switzerland	10-6-1999	2427
Italian MIBTEL	4-7-2000	2176	MSCI Germany	2-7-1999	2441
S&P 400 Midcap	3-1-1996	3258	MSCI Spain	2-7-1999	2416
Nikkei	8-1-1996	3160	MSCI France	2-7-1999	2448
Russell 3000	3-1-1996	3262	MSCI UK	9-6-1999	2443
Russell 1000	3-1-1996	3262	MSCI Italy	2-7-1999	2430
Russell 2000	3-1-1996	3264	MSCI Japan	5-12-1999	2231
Milan MIB	3-1-1996	3289	MSCI South Korea	6-12-1999	2253
German DAX	3-1-1996	3296	MSCI Mexico	7-10-2002	1602
S&P TSE	4-1-1999	2529	MSCI Netherlands	2-7-1999	2447
S&P 500	3-1-1996	3263	MSCI World	12-2-2001	2091
British pound	4-1-1999	2576			
Euro	4-1-1999	2592			
Swiss franc	4-1-1999	2571			
Japanese yen	4-1-1999	2590			

Note: the table shows the length of each time series in days denoted by T. Every start date of each time series is included in the table. All series end on either 27-2-2009 or 1-3-2009.

In Table 2, summary statistics of the data are presented. The Avol is the square root of the mean of 252 times the squared returns or the realised measure, such that Avol is on a scale of an annualised volatility. It shows that the Avol is higher for the squared returns, while the avol is roughly the same for the different realized measures. For example, Avol is typically just over 20% for the raw indexes, while it is around 16% for the realized measures. Realized measures do not have overnight returns effects which explains this difference. SD is the standard deviation of either percentage daily squared returns or realised measures. In contrast to Avol, this is not annualised. From the table it follows that the standard deviation is higher for the squared returns and lower for the realised measures. The summary statistics are expected. Note that the statistics of the realized variance are very similar to the realized kernel. In this paper the realized variance will be used as the realized measure.

Table 2: Summary statistics of the data

Asset	r_t^2		Realized variance		Realized kernel	
	Avol	SD	Avol	SD	Avol	SD
Dow Jones Industrials	19.5	4.81	15.2	1.94	15.1	1.95
CAC 40	23.7	5.95	18.1	2.18	18.4	2.21
FTSE 100	20.7	4.66	15.2	1.62	15.6	1.74
Spanish IBEX	23.8	6.57	16.8	1.76	16.6	1.73
Nasdaq	28.1	8.35	17.9	2.22	18.8	2.52
Italian MIBTEL	20.2	5.07	13.2	1.34	13.8	1.52
S&P 400 Midcap	21.7	5.68	13.5	1.90	13.8	1.96
Nikkei	25.0	6.96	16.0	1.37	16.6	1.48
Russell 3000	20.4	5.32	14.3	1.86	14.6	1.90
Russell 1000	20.4	5.38	14.7	1.91	15.0	1.94
Russell 2000	23.3	6.02	13.2	1.85	13.5	1.96
Milan MIB	23.2	5.69	16.5	1.84	17.0	1.99
German DAX	25.2	6.57	21.2	3.10	21.4	3.22
S&P TSE	21.0	5.54	14.2	1.82	14.4	1.89
S&P 500	20.8	5.46	15.6	2.09	15.9	2.14
MSCI Australia	16.4	3.05	8.8	0.53	9.1	0.57
MSCI Belgium	23.5	10.53	16.5	1.66	16.2	1.84
MSCI Brazil	43.7	24.35	28.6	6.30	29.6	7.21
MSCI Canada	19.5	5.05	12.6	1.67	13.1	1.88
MSCI Switzerland	20.7	5.25	14.6	1.44	14.6	1.56
MSCI Germany	25.7	6.94	21.1	3.10	20.9	2.99
MSCI Spain	24.0	6.08	17.6	1.84	17.6	1.92
MSCI France	24.0	6.29	18.3	2.23	18.5	2.32
MSCI UK	20.0	4.95	15.6	1.84	15.8	1.89
MSCI Italy	21.4	5.35	16.1	1.82	16.3	1.93
MSCI Japan	23.7	6.40	14.3	1.27	14.5	1.26
MSCI South Korea	32.1	9.63	21.6	2.61	21.9	2.80
MSCI Mexico	29.6	11.81	16.3	2.59	17.5	2.87
MSCI Netherlands	23.9	6.14	17.7	2.09	17.9	2.25
MSCI World	17.8	4.22	13.1	1.44	13.7	1.68
British pound	9.2	0.75	9.8	0.51	9.5	0.51
Euro	10.5	0.79	11.2	0.45	10.6	0.45
Swiss franc	11.1	0.91	11.6	0.39	10.9	0.38
Japanese yen	10.9	1.32	11.7	0.64	11.2	0.63

Note: 100 times differences of the log price (i.e. roughly percent change) are used for the calculations. Avol is the square root of the mean of 252 times of squared returns or the realised measure. This is approximately the annualised volatility. SD is the standard deviation of either the percent daily returns or realised measure.

3 Methodology

First, we show our method to obtain the one-step ahead VaR and ES forecasts with the two volatility models in the FHS. Secondly, the method of Du and Escanciano (2016) to evaluate the ES forecasts will be shown. Lastly, the ES forecasts obtained from the different models will be compared to each other by using the predictive tests along the lines of Giacomini and White (2006).

3.1 Expected shortfall forecast

Let r_t denote the daily return on day t of an asset and assume that

$$r_t = \mu + \epsilon_t$$

or

$$r_t - \mu = \epsilon_t = \sigma_t z_t$$
$$z_t \sim G(0,1)$$

where $G(0,1)$ is an unknown distribution with mean 0 and variance 1.

If we also assume $\mu = 0$, then $\sigma_t^2 = \mu_G + \alpha_G r_{t-1}^2 + \beta_G \sigma_{t-1}^2$ can be considered as a constant mean GARCH(1,1) model. The one-step ahead forecast from this model for the variance of the returns is as follows:

$$\hat{\sigma}_{t+1|t}^2 = \hat{\omega}_G + \hat{\alpha}_G r_t^2 + \hat{\beta}_G \hat{\sigma}_t^2$$

where the coefficients are estimated by QMLE where we thus assume a simplified distribution for the standardized residuals. In our case, we use the Gaussian Quasi-likelihood. These coefficients will be consistent even if the standardized residuals are misspecified. Recall that the formal definition of the VaR at time t is as follows:

$$VaR_t^{1-q} = \inf\{r_t : G_t(r_t) \geq q\},$$

where $G_t(r_t)$ is the cumulative density distribution of r_t and q is the confidence or coverage level. The squared root of $\hat{\sigma}_{t+1}^2$ will be used for the VaR forecasts as shown below.

$$V\hat{a}R_{t+1|t}^{1-q} = \hat{\sigma}_{t+1|t} \hat{G}_{t,q}^{-1}$$

Here $\hat{G}_{t,q}^{-1}$ is the q-th empirical quantile of $z_t = \frac{\epsilon_t}{\sigma_t}$. To estimate this quantile we define $\hat{z}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$ which will be calculated in the in-sample period. We order these residuals in an increasing order:

$$\hat{z}_1^* < \hat{z}_2^* < \cdots < \hat{z}_{m-1}^* < \hat{z}_m^*$$

such that $\hat{z}_1^*$ is the smallest residual and $\hat{z}_m^*$ is the largest residual. Here m is the length of the moving window. Then, $\hat{G}_{t,q}^{-1}$ can be estimated as $\hat{z}_{\lceil qm \rceil}^*$ where $\lceil x \rceil$ denotes the smallest integer not smaller than x.

Recall that

$$ES_{t+1|t}^{1-q} = E_t(r_{t+1} | r_{t+1} < VaR_{t+1|t}^{1-q})$$

such that the ES can be estimated as

$$E\hat{S}_{t+1|t}^{1-q} = \hat{\sigma}_{t+1|t} \frac{\sum_{i=1}^{\lceil qm \rceil} \hat{z}_i^*}{\lceil qm \rceil}.$$

Next, consider the HEAVY model that has two equations:

$$var(r_t | F_{t-1}^{HF}) = h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad \omega, \alpha \geq 0, \quad \beta \in [0,1) \tag{1}$$
$$E(RM_t | F_{t-1}^{HF}) = \mu_t = \omega_R + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \quad \omega_R, \alpha_R, \beta_R \geq 0, \alpha_R + \beta_R \in [0,1) \tag{2}$$

Equation 1 is referred to as the HEAVY-r model and equation 2 is referred to as the HEAVY-RM model. Only the HEAVY-r model is needed for one-step ahead forecasts while the HEAVY-RM model is used for multistep-ahead forecasts. See Shephard and Sheppard (2010) for the different methods for multi-step ahead forecasts. Each equation will be estimated separately by using a Gaussian quasi-likelihood along the lines of Shephard and Sheppard (2010). For convenience, assume for now that the in-sample starts from the first observation available. The first equation will be estimated as follows:

$$logQ_1(\omega, \phi) = \sum_{t=2}^{m} -\frac{1}{2} log(h_t + r_t^2/h_t), \; \phi = (\alpha, \beta)'$$

where we take $h_1 = T^{-1/2} \sum_{t=1}^{\lfloor m \rfloor^{1/2}} r_t^2$ as proxy for the first latent variable h_1 in the estimation sample. Here $\lfloor x \rfloor$ denotes the largest integer not larger than x. The same structure is used for the second equation:

$$logQ_2(\omega_R, \phi_R) = \sum_{t=2}^{m} -\frac{1}{2}log(\mu_t + RM_t/\mu_t),\ \phi_r = (\alpha_R, \beta_R)'$$

where we take $\mu_1 = T^{-1/2}\sum_{t=1}^{\lfloor m \rfloor^{1/2}} RM_t$ as the proxy for the first latent variable μ_1 in the estimation sample. See Shephard and Sheppard (2010) for more details.

With the GARCH model a similar way will be used to forecast the volatility of the returns. Namely,

$$\hat{h}_{t+1|t} = \hat{\omega} + \hat{\alpha}RM_t + \hat{\beta}\hat{h}_t.$$

By taking the squared root of $\hat{h}_{t+1|t}$, we have the conditional standard deviation forecast. Then in the same way the ES can be forecasted:

$$\hat{ES}_{t+1|t}^{1-q} = \sqrt{\hat{h}_{t+1|t}}\frac{\sum_{i=1}^{\lceil qm \rceil} \hat{z}_i^{r*}}{\lceil qm \rceil}$$

where $\hat{z}_i^{r*}$ is the series of $\hat{z}_t^r = \frac{\hat{\epsilon}_t}{\sqrt{\hat{h}_{t+1|t}}}$ ordered in an increasing order.

A moving window will be used for our forecasts. Let T denote the total number of observations available of a particular asset and define $\lambda = n/m$ where n is the number of observations of the out-of-sample and where m is the number of observations in-sample. Since every asset has different T, the restrictions

$$\frac{n}{m} \leq c \in (0,1) \tag{3}$$

$$n + m = T \tag{4}$$

will be imposed. If we use restriction 4 in restriction 3 we get $\frac{n}{m} \leq c \Leftrightarrow n \leq c\cdot(T-n) \Leftrightarrow (1+c)\cdot n \leq c \cdot T$. To satisfy the restrictions we use $n = \left\lfloor \frac{c}{(1+c)} \cdot T \right\rfloor$ and $m = T - n$ for the forecasts for each asset. On one hand we want λ close to zero to minimize the estimation error which is needed for the evaluation tests of Du and Escanciano (2016). On the other hand n needs to be large to use the predictive ability tests of Giacomini and White (2006). In our cases we choose $c = 0.2$. As the average number of total observations equals 2664, the number of out-of-sample observations will be around 444 observations for most assets.

3.2 Evaluation Expected shortfall forecast

The basic backtests proposed by Du and Escanciano (2015) to evaluate the expected shortfall forecasts will be used. Note that the basic backtests are recommended when the estimation effects are limited.

The first test with the null hypothesis

$$H_{0u} : E(H_t(q,\theta)) = q/2$$

can be seen as the analoque of the unconditional backtest for VaR proposed by Kupiec (1995). Here $H_t(q) = \frac{1}{q}\int_0^q 1(u_t \leq u)du = \frac{1}{q}(q - u_t)1\{u_t \leq q\}$ and $u_t(\theta_0) = G(r_t, \Omega_{t-1}, \theta_0) = P(R_t \leq r_t|\Omega_{t-1}, \theta_o)$ where $1\{\cdot\}$ is the indicator function and $G(\cdot, \Omega_{t-1}, \theta_0)$ is the conditional cumulative distribution of the returns.

By using the standardized residuals $\hat{z}_t$ or $\hat{z}_t^r$ depending on the model which are not ordered by size,

$$u_t(\theta_0) = G(r_t, \Omega_{t-1}, \theta_0) = P(R_t \leq r_t|\Omega_{t-1}, \theta_o) = P(\frac{R_t - \hat{\mu}}{\hat{\sigma}_t} \leq \hat{z}_t)$$

can be estimated as $\hat{u}_t = \frac{\sum_i^m 1\{\hat{z}_i < \hat{z}_t\}}{m}$. Then $H_t(q)$ can be estimated as $\hat{H}_t(q) = \frac{1}{q}(q - \hat{u}_t)1\{\hat{u}_t \leq q\}$.

A simple t-test can be used for this null hypothesis as shown below

$$U_{ES} = \frac{\sqrt{n}(\bar{H}(q) - q/2)}{\sqrt{q(1/3 - q/4)}},$$

where $\bar{H}(q) = \frac{1}{n}\sum_1^n \hat{H}_t(q)$ and n is the number of observations out-of-sample. Under the assumption that $\lambda = 0$, this t-test converges in distribution to the standard normal distribution.

Next, the conditional backtest with the following null hypothesis is considered

$$H_{0c} : E(H_t(q, \theta_0) - q/2|\Omega_{t-1}) = 0$$

which can be seen as the analogue of the conditional backtest of the VaR, Christoffersen (1998) for example. To test this hypothesis, we first need to define

$$\gamma_{nj} = \frac{1}{n-j}\sum_{t=1+j}^{n}(H_t(q) - q/2)(H_{t-j}(q) - q/2) \quad \text{and} \quad \rho_{nj} = \frac{\gamma_{nj}}{\gamma_{n0}}.$$

This can be estimated by

$$\hat{\gamma}_{nj} = \frac{1}{n-j}\sum_{t=1+j}^{n}(\hat{H}_t(q) - q/2)(\hat{H}_{t-j}(q) - q/2) \quad \text{and} \quad \hat{\rho}_{nj} = \frac{\hat{\gamma}_{nj}}{\hat{\gamma}_{n0}}.$$

A simple conditional test proposed by Du and Escanciano (2016) using $\hat{\rho}_{nj}$ is the Box-Pierce test statistic

$$C_{ES}(k) = n\sum_{j=1}^{k}\hat{\rho}_{nj}^2$$

where k is the number of different autocorrelation we consider. Under the assumption $\lambda = 0$ it holds that this Box-pierce test converges into a chi-squared distribution with k degrees of freedom, that is

$$n\sum_{j=1}^{k}\hat{\rho}_{nj}^2 \xrightarrow{d} \chi_k^2.$$

We will take $k = 5$ as this is suggested by Du and Escanciano (2016). In their Monte Carlo simulations they have shown that this conditional backtests with $k = 5$ performs better than with k equal to 1 or 3 in terms of power.

3.3 Comparison of the ES forecasts of the different models

After evaluating the specification of the different models, we focus on the difference between the forecasts of the different models despite their correct or misspecification. To compare the forecasts we first introduce the loss function as described in Fissler et al. (2015). This loss function is given as follows:

$$\begin{aligned} L_M(V\hat{a}R_{t+1}, \hat{ES}_{t+1}, r_{t+1})_{t+1} = &(1\{r_{t+1} \leq V\hat{a}R_{t+1}\} - q)(V\hat{a}R_{t+1} - r_{t+1}) \\ &+ \frac{1}{q}\exp(\hat{ES}_{t+1})1\{r_{t+1} \leq V\hat{a}R_{t+1}\}(V\hat{a}R_{t+1} - r_{t+1}) \\ &+ \exp(\hat{ES}_{t+1})(\hat{ES}_{t+1} - V\hat{a}R_{t+1}) - \exp(\hat{ES}_{t+1}) \end{aligned}$$

Here M indicates whether the GARCH or HEAVY model is used in the FHS to forecast $V\hat{a}R_{t+1}$ and $\hat{ES}_{t+1}$. The difference of the losses obtained from the two different models is denoted by:

$$\Delta L_{m,t+1} = L_H(V\hat{a}R_{t+1}, \hat{ES}_{t+1}, r_{t+1})_{t+1} - L_G(V\hat{a}R_{t+1}, \hat{ES}_{t+1}, r_{t+1})_{t+1}$$

will be used for the two predictive tests that are proposed by Giacomini and White (2006), here m is the number of observations in the in-sample period which depends on the particular asset that is considered. The two predictive tests include a conditional and an unconditional test. The unconditional tests has the following null hypothesis:

$$H_0 : E[\Delta L_{m,t+1}] = 0$$

which claims that the models are equally accurate on average. To test this hypothesis, the following simple t-statistic is applied

$$t_{m,n} = \frac{\Delta \bar{L}_{m,n}}{\hat{\sigma}_n \sqrt{n}},$$

where $\Delta \bar{L}_{m,n} = n^{-1} \sum_{t=m}^{t-1} \Delta L_{m,t+1}$ and where $\hat{\sigma}_n$ is the HAC estimator of the asymptotic variance of $\sigma_n^2 = var[\sqrt{n} \Delta \bar{L}_{m,n}]$.

The conditional null hypothesis is as follows

$$H_0 : E[\Delta L_{m,t+1} | \mathcal{F}_t] = 0$$

which says that one cannot predict which model will predict more accurate at time $t+1$, given the information in time t. This null is equivalent to stating that $E[\tilde{h}_t \Delta L_{m,t+1}] = 0$ for all measurable functions $\tilde{h}_t$ in $\mathcal{F}_t$, see Giacomini and White (2006). The test statistic for the conditional hypothesis that is considered by Giacomini and White (2006) is as follows:

$$T^h_{m,n} = n \left(n^{-1} \sum_{t=m}^{T-1} h_t \Delta L_{m,t+1} \right) \hat{\Sigma}_n^{-1} \left(n^{-1} \sum_{t=m}^{T-1} h_t \Delta L_{H,G,t+1} \right) = n \bar{Z}_{m,n} \hat{\Sigma}_n^{-1} \bar{Z}_{m,n}$$

where $\bar{Z}_{m,n} \equiv n^{-1} \sum_{t=m}^{T-1} Z_{m,t+1}$, $Z_{m,t+1} \equiv h_t \Delta L_{m,t+1}$ and $\hat{\Sigma}_n \equiv n^{-1} \sum_{t=m}^{T-1} Z_{m,t+1} Z'_{m,t+1}$ is a p x p matrix estimator of the covariance matrix of $Z_{m,t+1}$. Under H_0 $T^h_{m,n} \xrightarrow{d} \chi^2_p$ as $n \to \infty$.

$T^h_{m,n}$ can also be estimated as mR^2 where R^2 is the squared correlation coefficient for the artificial regression of the constant unity on $(h_t \Delta L_{m,t+1})'$, see Giacomini and White (2006). In our case the scoring function h_t is chosen as $[1, \Delta L_{m,t-1}]$.

When the null hypothesis is rejected, it implies that the test functions h_t can predict $\Delta L_{m,t+1}$. To decide which model is better when the null is rejected, we use the decision rule of Giacomini and White (2006). This rule consists of two steps. The first step is to regress $\Delta L_{m,t+1}$ on h_t over the out-of-sample period. Let $\hat{\delta}_n$ denote the vector of the regression coefficients. The idea of the second step is based on the approximation of $\hat{\delta}'_n h_t \approx E[\Delta L_{m,t+1} | \mathcal{F}_t]$ for $t = m+1, m+2, \cdots T-2, T-1$. Therefore we can calculate $\hat{\delta}'_n h_t$ over this sample and calculate the proportion of $1\{\hat{\delta}'_n h_t > 0\}$. If the proportion is larger than 50%, it means that the HEAVY model is preferred over the GARCH(1,1) model.

4 Results

4.1 Estimated models

In this section, the fits of the different models used by Shephard and Sheppard (2010), the likelihood changes by comparing different models due to restrictions and iterative forecasts are replicated. Note that the results are not identical to the results of Shephard and Sheppard (2010). The results of the fits often differ in the third decimal. This difference could be explained due to small differences between different local minimizers that are used for the QMLE. For the iterative forecasts, the results can even differ in the first decimal. The tables in this subsection only contain information for most raw indexes. The complete tables are shown in the Appendix.

Table 3: Fit of the GARCH and HEAVY models of various indexes.

Asset	Heavy-R		GARCHX			GARCH		HEAVY-RM		Integrated	
	α	β	α_X	β_X	γ_X	α_G	β_G	α_R	β_R	α_g	α_R
Dow Jones Industrials	0.406	0.738	0.407	0.737	0.000	0.082	0.913	0.411	0.567	0.063	0.336
CAC 40	0.528	0.673	0.525	0.675	0.000	0.081	0.916	0.417	0.573	0.068	0.350
FTSE 100	0.608	0.658	0.614	0.656	0.000	0.105	0.892	0.434	0.562	0.085	0.369
Spanish IBEX	0.631	0.674	0.479	0.714	0.035	0.112	0.886	0.393	0.604	0.085	0.343
Nasdaq	0.723	0.661	0.439	0.744	0.051	0.082	0.915	0.427	0.568	0.063	0.349
Italian MIBTEL	0.806	0.630	0.806	0.631	0.000	0.107	0.889	0.512	0.486	0.080	0.436
S&P 400 Midcap	0.847	0.641	0.269	0.795	0.083	0.100	0.885	0.392	0.603	0.073	0.333
Nikkei	0.506	0.773	0.508	0.772	0.000	0.079	0.906	0.346	0.641	0.065	0.295
Russell 3000	0.449	0.746	0.446	0.748	0.000	0.081	0.911	0.403	0.574	0.059	0.313
Russell 1000	0.398	0.767	0.396	0.769	0.000	0.078	0.916	0.402	0.577	0.058	0.315
Russell 2000	0.946	0.679	0.243	0.812	0.102	0.107	0.885	0.387	0.622	0.077	0.322
Milan MIB	0.498	0.746	0.342	0.779	0.047	0.102	0.895	0.484	0.518	0.076	0.417
German DAX	0.446	0.673	0.446	0.674	0.000	0.093	0.903	0.457	0.536	0.075	0.376

Note: the complete sample is used for each estimation. The complete table including all 34 assets is shown in the Appendix.

Above in Table 3, the parameters of the models that were considered are shown. The complete sample is used to estimate the parameters. In this paper we focus on the GARCH and the standard HEAVY model. For one-step ahead forecasts, only the HEAVY-r model is relevant. It can be seen that β is typically around 0.6, except for the exchange rates. ω which is not shown in the table, is typically very small around 0. For the GARCH model, β_G is higher around 0.9. Therefore the GARCH model has more memory, since it averages more observations that are further away from the present. The conditional variance obtained from the HEAVY-r model can roughly be seen as a constant plus a weighted sum of recent realised measures.

Below in Table 4 twice the likilihood change is shown by imposing restrictions. It shows that the difference between the standard HEAVY-r model and the extended HEAVY-r model or the GARCHX model is rather small. However, the likelihood change between the GARCH and the GARCHX model is much larger. This implies that the standard HEAVY model can be favoured compared to the traditional GARCH model.

Table 4: Twice the likelihood change between different models by imposing restrictions.

Asset	Compare to extended HEAVY-r		Impose unit root		No momentum
	HEAVY-r	GARCH	GARCH	HEAVY-RM	$\beta = 0$
Dow Jones Industrials	0.0	-199.5	-48.5	-19.5	-56.5
CAC 40	0.0	-148.9	-30.9	-14.5	-67.3
FTSE 100	0.0	-125.7	-32.5	-12.3	-55.1
Spanish IBEX	-9.3	-113.7	-59.2	-12.1	-78.4
Nasdaq	-15.8	-108.4	-31.2	-14.4	-72.9
Italian MIBTEL	0.0	-141.2	-40.6	-9.9	-38.1
S&P 400 Midcap	-64.5	-61.8	-61.5	-11.0	-89.1
Nikkei	0.0	-116.5	-64.5	-9.9	-84.6
Russell 3000	0.0	-187.2	-49.9	-21.2	-61.3
Russell 1000	0.0	-186.2	-45.4	-20.0	-61.8
Russell 2000	-163.1	-64.9	-57.5	-12.7	-134.7
Milan MIB	-16.5	-100.7	-48.5	-13.0	-75.6
German DAX	0.0	-153.3	-47.2	-16.0	-63.7

Note: left-hand side is compared to the GARCHX model and right-hand side compares the unconstrained GARCH and HEAVY-RM model with constrained models that impose an unit root. The final column compares the likelihood of the unconstrained HEAVY-RM model with the constrained HEAVY-RM model with $\beta = 0$. The complete table including all 34 assets is shown in the Appendix.

In Table 5, the in-sample forecasts are compared between the HEAVY and GARCH model by evaluating the in-sample likelihood ratio tests for losses of both models. See Shephard and Sheppard (2010) for the corresponding t-test and losses in details. Negative LR tests favour the HEAVY model. Note that the negative t-statistic is smallest for horizon of 1. The negative t-statistics decrease as the horizon increases. This means that the outperformance of the forecasts for in-sample generated by the HEAVY model decrease as the horizon increases.

Table 5: In sample likelihood ratio tests for losses obtained from the HEAVY and GARCH models.

Asset	t-statistic for non-nested LR tests for h					
	1	2	3	5	10	22
Dow Jones Industrials	-5.72	-3.80	-3.06	-2.96	-2.01	0.88
CAC 40	-4.39	-3.06	-2.34	-0.68	0.04	1.73
FTSE 100	-5.17	-3.39	-2.69	-1.70	-0.14	-0.14
Spanish IBEX	-2.83	-2.58	-1.48	-0.60	-1.11	-0.57
Nasdaq	-2.47	-0.47	-0.33	-0.71	1.12	-0.33
Italian MIBTEL	-4.11	-3.29	-3.35	-1.78	-0.78	-0.75
S&P 400 Midcap	0.07	1.11	1.06	0.08	0.27	-0.41
Nikkei	-3.87	-2.72	-2.24	-0.68	0.28	0.64
Russell 3000	-5.73	-4.04	-3.26	-4.04	-1.67	0.03
Russell 1000	-5.44	-3.90	-3.24	-3.88	-1.51	0.45
Russell 2000	1.66	2.23	2.14	1.20	1.34	0.01
Milan MIB	-1.89	-1.00	-0.91	-0.10	0.03	-0.02
German DAX	-5.12	-3.44	-2.86	-1.10	-0.81	-0.34

Note: negative values favour HEAVY models. Both models are estimated using the quasi-likelihood. Here h denotes the forecast horizon of the iterative forecasts. The complete table including all 34 assets are shown in the Appendix.

4.2 Comparison of the ES forecasts

In this section the conventional significance level of 0.05 is used for all the tests. In Table 6 the p-values of the Du and Escanciano (2016) backtests considering ES forecasts with a coverage level of $q = 0.01$ are displayed. Recall that there are 34 assets considered. The unconditional test is rejected for 3 and 8 occasions for the case of using the HEAVY and GARCH model, respectively. The number of rejections for the conditional test is 7 and 11 for the use of the HEAVY as GARCH model, respectively. Both models perform quite well for most of the cases. There is slight advantage for use of the HEAVY model in the FHS method compared to the use of the GARCH model.

Table 6: The p-values of the backtests of Du and Escanciano (2016).

Asset	HEAVY		GARCH	
	Unconditional	Conditional	Unconditional	Conditional
Dow Jones Industrials	0.067	0.997	0.275	0.997
CAC 40	0.659	0.654	0.718	0.013*
FTSE 100	0.067	0.034*	0.063	0.159
Spanish IBEX	0.207	0.071	0.110	0.000***
Nasdaq	0.345	1.000	0.513	0.942
Italian MIBTEL	0.818	1.000	0.932	0.457
S&P 400 Midcap	0.202	1.000	0.280	0.134
Nikkei	0.000***	0.004**	0.009*	0.048*
Russell 3000	0.825	1.000	0.017*	0.995
Russell 1000	0.980	1.000	0.022*	0.991
Russell 2000	0.257	1.000	0.127	0.994
Milan MIB	0.955	1.000	0.469	0.012*
German DAX	0.758	0.877	0.476	0.999
S&P TSE	0.688	0.781	0.255	0.850
S&P 500	0.667	0.999	0.092	0.994
MSCI Australia	0.679	1.000	0.021*	0.000***
MSCI Belgium	0.094	0.999	0.046*	0.042*
MSCI Brazil	0.604	1.000	0.548	0.999
MSCI Canada	0.691	0.125	0.059	0.047*
MSCI Switzerland	0.865	0.942	0.950	0.980
MSCI Germany	0.830	0.205	0.551	1.000
MSCI Spain	0.086	0.000***	0.696	0.000***
MSCI France	0.259	0.001**	0.700	0.013*
MSCI UK	0.506	0.000***	0.110	0.005*
MSCI Italy	0.564	1.000	0.732	1.000
MSCI Japan	0.093	0.999	0.167	1.000
MSCI South Korea	0.636	1.000	0.679	0.999
MSCI Mexico	0.951	0.037*	0.060	0.938
MSCI Netherlands	0.250	0.000***	0.099	0.002**
MSCI World	0.006**	0.606	0.000***	0.819
British pound	0.169	0.991	0.008**	0.824
Euro	0.706	1.000	0.801	0.999
Swiss franc	0.015*	0.058	0.001***	0.060
Japanese yen	0.173	0.998	0.092	0.998
Number of rejections	3	7	8	11

Note: ES forecasts with $q = 0.01$ are considered. In the last row the number of rejections for each hypothesis is shown. Furthermore,* represents a p-value smaller than 0.05, ** represents a p-value smaller than 0.01, *** represents a p-value smaller than 0.001

The p-values of the backtests of Du and Escanciano (2016) concerning ES forecasts with a coverage level of $q = 0.05$ are shown in Table 7. The unconditional backtest rejects the null hypothesis for 14 assets at the conventional significance level when the HEAVY model is used in the FHS. The number of rejections is somewhat larger when the GARCH model is used, namely 23 cases. The unconditional test seems to favour the HEAVY model. For the conditional test, the number of rejections are 8 amd 10 for the case of using HEAVY and GARCH model respectively.

Table 7: The p-values of the backtests of Du and Escanciano (2016).

Asset	HEAVY		GARCH	
	Unconditional	Conditional	Unconditional	Conditional
Dow Jones Industrials	0.004*	0.616	0.026*	0.228
CAC 40	0.272	0.518	0.090	0.067
FTSE 100	0.643	0.006**	0.033*	0.025*
Spanish IBEX	0.015*	0.031*	0.029*	0.002**
Nasdaq	0.314	0.603	0.015*	0.261
Italian MIBTEL	0.732	0.362	0.259	0.356
S&P 400 Midcap	0.067	0.479	0.000***	0.045*
Nikkei	0.000***	0.000***	0.062	0.000***
Russell 3000	0.124	0.660	0.002**	0.100
Russell 1000	0.039*	0.589	0.002**	0.070
Russell 2000	0.089	0.729	0.000***	0.036*
Milan MIB	0.217	0.317	0.127	0.405
German DAX	0.085	0.935	0.259	0.019*
S&P TSE	0.628	0.291	0.002**	0.087
S&P 500	0.011*	0.501	0.000***	0.033*
MSCI Australia	0.834	0.115	0.009**	0.001***
MSCI Belgium	0.000***	0.004**	0.002**	0.128
MSCI Brazil	0.078	0.612	0.057	0.316
MSCI Canada	0.257	0.199	0.000***	0.070
MSCI Switzerland	0.052	0.614	0.010**	0.824
MSCI Germany	0.012*	0.831	0.009**	0.218
MSCI Spain	0.004 **	0.038*	0.040*	0.003**
MSCI France	0.031*	0.734	0.052	0.127
MSCI UK	0.044*	0.034*	0.051	0.071
MSCI Italy	0.222	0.724	0.039*	0.714
MSCI Japan	0.010*	0.025*	0.553	0.062
MSCI South Korea	0.526	0.913	0.379	0.773
MSCI Mexico	0.081	0.172	0.026*	0.058
MSCI Netherlands	0.239	0.003**	0.029*	0.101
MSCI World	0.001**	0.280	0.000***	0.274
British pound	0.057	0.844	0.002**	0.891
Euro	0.045*	0.564	0.105	0.165
Swiss franc	0.123	0.772	0.004**	0.669
Japanese yen	0.003**	0.485	0.000***	0.031*
Number of rejections	14	8	23	10

Note: ES forecasts with $q = 0.05$ are considered. In the last row the number of rejections for each hypothesis is shown. Furthermore,* represents a p-value smaller than 0.05, ** represents a p-value smaller than 0.01, *** represents a p-value smaller than 0.001

In Table 8, the p-values of the backtests are shown for ES forecasts at a coverage level of $q = 0.1$. A quite similar pattern is shown with respect to ES forecasts at $q = 0.01$. Both the unconditional and conditional test favour the HEAVY model. However, the number of rejections for the unconditional test is larger for both models.

Table 8: The p-values of the backtests of Du and Escanciano (2016).

Asset	HEAVY		GARCH	
	Unconditional	Conditional	Unconditional	Conditional
Dow Jones Industrials	0.001**	0.279	0.006**	0.152
CAC 40	0.176	0.556	0.099	0.147
FTSE 100	0.408	0.123	0.102	0.264
Spanish IBEX	0.039*	0.107	0.069	0.019*
Nasdaq	0.455	0.610	0.007**	0.045*
Italian MIBTEL	0.167	0.924	0.006**	0.586
S&P 400 Midcap	0.065	0.664	0.001***	0.052
Nikkei	0.000***	0.013*	0.004**	0.234
Russell 3000	0.039*	0.116	0.001**	0.011*
Russell 1000	0.012*	0.112	0.001***	0.015*
Russell 2000	0.086	0.073	0.001***	0.018*
Milan MIB	0.046*	0.977	0.007	0.305
German DAX	0.118	0.964	0.339	0.183
S&P TSE	0.297	0.380	0.000***	0.099
S&P 500	0.006**	0.074	0.001***	0.014*
MSCI Australia	0.299	0.472	0.006	0.136
MSCI Belgium	0.000***	0.002**	0.001***	0.077
MSCI Brazil	0.050*	0.488	0.011*	0.414
MSCI Canada	0.034*	0.300	0.000***	0.330
MSCI Switzerland	0.026*	0.980	0.007**	0.994
MSCI Germany	0.036*	0.942	0.027*	0.205
MSCI Spain	0.005**	0.119	0.038*	0.061
MSCI France	0.031*	0.731	0.035*	0.524
MSCI UK	0.021*	0.650	0.012*	0.930
MSCI Italy	0.060	0.971	0.003**	0.638
MSCI Japan	0.000***	0.012*	0.025*	0.338
MSCI South Korea	0.132	0.861	0.156	0.171
MSCI Mexico	0.351	0.130	0.034*	0.065
MSCI Netherlands	0.114	0.067	0.041*	0.107
MSCI World	0.000***	0.110	0.000***	0.150
British pound	0.007**	0.871	0.000***	0.604
Euro	0.014*	0.243	0.021*	0.031**
Swiss franc	0.202	0.378	0.030*	0.451
Japanese yen	0.000***	0.162	0.000***	0.000***
Number of rejections	20	3	29	8

Note: ES forecasts with $q = 0.10$ are considered. In the last row the number of rejections for each hypothesis is shown. Furthermore, * represents a p-value smaller than 0.05, ** represents a p-value smaller than 0.01, *** represents a p-value smaller than 0.001

The p-values of the predictive ability tests of Giacomini and White (2006) are shown below in Table 9. The ratio between the numbers of assets where the HEAVY model is preferred and the total number of assets is shown in the last row. This ratio is for both coverage levels $q = 0.05$ and $q = 0.10$ similar, conditionally and unconditionally. Consider the unconditional test, here the null hypothesis is rejected at the conventional significance level for 10 out of 34 assets due to negative t-statistics. For these 10 assets, the HEAVY model is favoured to be used in our FHS method. For the rest of the 24 assets, the loss functions of the different models do not seem to be significantly different from each other. Note that there are no preferences for the exchange rates indexes. Now for the conditional test, rejection occurs for 16 out of 34 assets. In these cases, the decision rule tells us that the HEAVY models is preferred. This can be seen by the proportion that is lower than 0.5 for the rejected cases.

Table 9: The p-values of the unconditional and conditional predictive ability tests of Giacomini and White (2006).

Asset	q=0.05			q=0.10		
	Unc	Con	Proportion	Unc	Con	Proportion
Dow Jones Industrials	0.126	0.062	-	0.144	0.024*	0.107
CAC 40	0.079	0.243	-	0.211	0.109	-
FTSE 100	0.096	0.079	-	0.116	0.034*	0.165
Spanish IBEX	0.271	0.045*	0.125	0.496	0.014*	0.529
Nasdaq	0.075	0.046*	0.041	0.063	0.029*	0.087
Italian MIBTEL	0.019*	0.107	-	0.021*	0.017*	0.105
S&P 400 Midcap	0.040*	0.116	-	0.060	0.295	-
Nikkei	0.120	0.006**	0.941	0.149	0.444	-
Russell 3000	0.010**	0.036*	0.020	0.036*	0.009**	0.076
Russell 1000	0.012*	0.042*	0.017	0.033*	0.004**	0.076
Russell 2000	0.147	0.045*	0.053	0.187	0.228	-
Milan MIB	0.022*	0.098	-	0.032*	0.003**	0.139
German DAX	0.096	0.032*	0.027	0.153	0.035*	0.106
S&P TSE	0.091	0.027*	0.045	0.028*	0.010**	0.100
S&P 500	0.003**	0.013*	0.018	0.007**	0.042*	0.044
MSCI Australia	0.068	0.167	-	0.077	0.221	-
MSCI Belgium	0.223	0.427	-	0.303	0.870	-
MSCI Brazil	0.399	0.103	-	0.084	0.189	-
MSCI Canada	0.038*	0.009**	0.045	0.014*	0.008**	0.081
MSCI Switzerland	0.104	0.261	-	0.259	0.138	-
MSCI Germany	0.268	0.015*	0.104	0.205	0.037*	0.183
MSCI Spain	0.424	0.040*	0.289	0.302	0.009**	0.756
MSCI France	0.105	0.232	-	0.220	0.136	-
MSCI UK	0.137	0.023*	0.039	0.176	0.156	-
MSCI Italy	0.047*	0.018*	0.069	0.106	0.004**	0.176
MSCI Japan	0.486	0.925	-	0.414	0.930	-
MSCI South Korea	0.007**	0.043*	0.011	0.015*	0.042*	0.096
MSCI Mexico	0.363	0.946	-	0.396	0.240	-
MSCI Netherlands	0.005**	0.069	-	0.036*	0.197	-
MSCI World	0.194	0.021*	0.066	0.092	0.055	-
British pound	0.051	0.192	-	0.023*	0.101	-
Euro	0.473	0.883	-	0.352	0.919	-
Swiss franc	0.106	0.459	-	0.253	0.712	-
Japanese yen	0.055	0.239	-	0.147	0.455	-
Ratio	0.29	-	0.47	0.29	-	0.47

Note: if the conditional test is rejected, the decision rule is used. The proportion of $1\{\hat{\delta}_n' h_t > 0\}$ over the out-sample period is shown below. A proportion smaller than 0.5 favours the HEAVY model. On the left hand side, the ES forecasts at a coverage level of 0.05 are evaluated while the right hand side considers ES forecasts at a coverage level of 0.1. Ratio in the last row is the number of assets where the HEAVY model is favoured divided by the total number of assets. Furthermore, * represents a p-value smaller than 0.05, ** represents a p-value smaller than 0.01

Interestingly, the results of the predictive ability tests are somewhat different for the coverage level of $q = 0.01$. These results are displayed in Table 13 in the Appendix. For the unconditional test all 8 rejected cases were based on negative t-statistics which leads to a ratio of roughly 0.24 which is rather the same compared to other coverage levels. However the Ratio for the conditional test is roughly 0.24 which is smaller compared to the ratio of 0.47 obtained with the other coverage levels. Note that there are again no preferences for the exchange rates indexes. This observation is reasonable as Shephard and Sheppard (2010) also found that the HEAVY fit is not much better than the GARCH fit for exchange rates. This can be explained by the fact that the HEAVY model holds on average much more memory for exchange rates. This can be confirmed by their higher

β around 0.85, see Table 10 in the Appendix. The difference between the HEAVY and GARCH model is therefore smaller for exchange rates.

The one-step ahead ES forecasts of the Dow Jones Industrial returns with a coverage level of $q = 0.05$ are shown in Figure 1. The forecasts obtained with the FHS in combination with the HEAVY model seem to adjust in greater magnitude. Consequently, the variance of the ES forecasts is larger compared to the 'GARCH' forecasts. This feature is observed for most assets and different coverage levels.

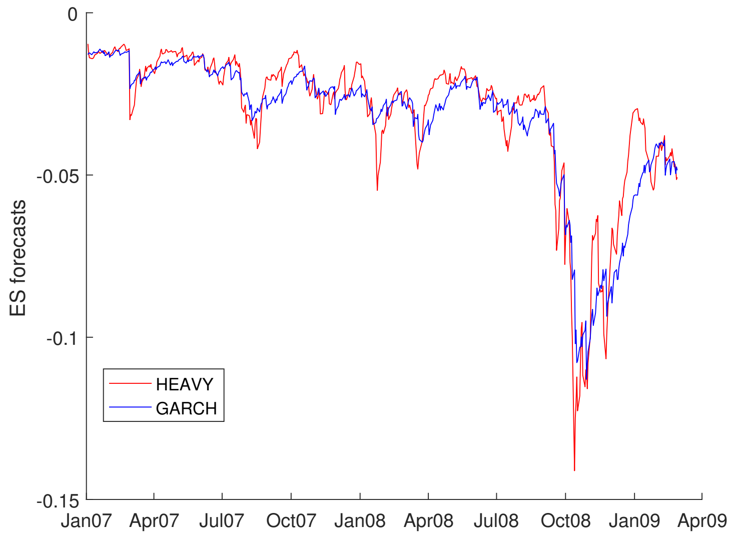

Figure 1: The ES forecasts for the Dow Jones Industrial index at a coverage level of $q = 0.05$

Figure 2 gives the loss functions with a coverage level of $q = 0.05$ of both models for Dow Jones Industrial index. It is difficult to see from the figure, but the loss function is on average lower for the HEAVY model due to several periods with lower losses. These periods are paired with lower ES forecasts. For example, consider the most extreme period, from December 2008 till mid January 2009 where the loss function is lower for the HEAVY model. The ES forecasts are larger when the HEAVY model is used compared to the GARCH model in this period as shown in Figure 3.

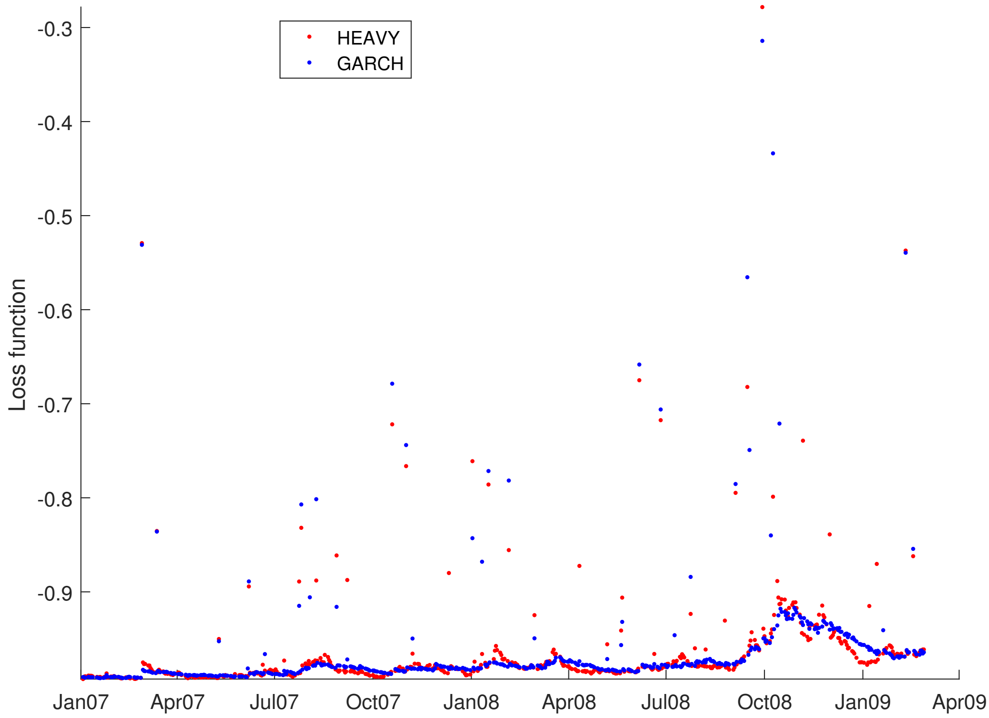

Figure 2: The loss functions for the Dow Jones Industrial index at a coverage level of $q = 0.05$

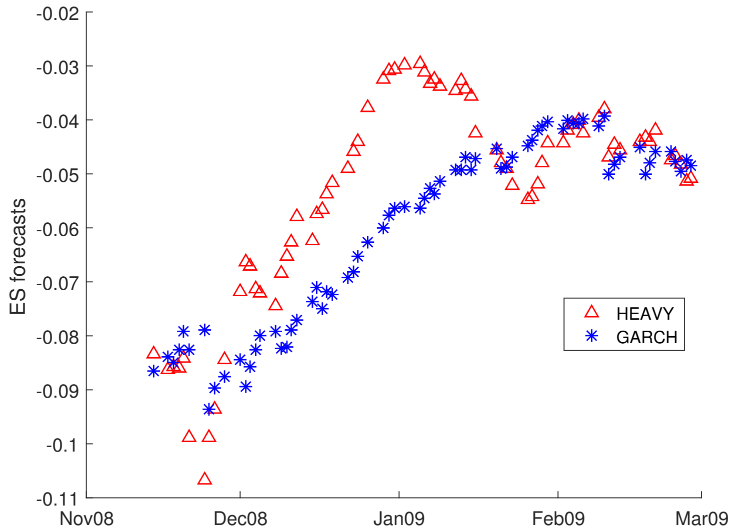

Figure 3: The loss functions for the Dow Jones Industrial index at a coverage level of $q = 0.05$

5 Conclusion

The backtests of Du and Escanciano (2016) indicate that the ES forecasts are correctly specified for both FHS combinations in most cases when the coverage level of $q = 0.01$ is used. The number of correct unconditional specifications decreases as the coverage level increases. This applies for both applications of the HEAVY and GARCH model. However the conditional specification is correct for most of the cases independent of the coverage level. When the HEAVY model is used rather than the GARCH model as the time-varying volatility model for the FHS method the correct specification occurs more often.

Based on the predictive ability tests of Giacomini and White (2006), conditionally wise the favour goes to the use of the HEAVY model for almost half of the cases for coverage levels of $q - 0.05$ and $q = 0.10$. Differently is when a coverage level of $q = 0.01$ is used. Then the preference from the conditional perspective for the use of the HEAVY model is only 24% of the cases. Unconditionally, the preference for the HEAVY model occurs in 29% of the cases for coverage levels of 0.05 and $q = 0.10$. When the coverage level of 0.01 is considered, the preference from the unconditional view for the HEAVY model is present in 24% of the assets. So the preference of the HEAVY model is present for all coverage levels, but for the lowest coverage level of 0.01 this outperformance is in smaller extent. The preferences for the use of the HEAVY model is only present for the raw indexes and the MSCI indexes. The preference is not observed for the exchange rates indexes. This may be due to the larger memory of the HEAVY model for the exchange rates indexes.

The greater adjustments of the ES forecasts when the HEAVY model is considered might play an important role for the preference for the HEAVY model in terms of specification and predictive ability. One could investigate whether the favour is still present when less stressful times are used for forecasts.

For further research, it could be interesting to investigate multi-step ahead ES forecasts. Also besides using empirical data, it might be reasonable to assume distributions of returns to estimate the quantiles.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), The distribution of realized exchange rate volatility, *Journal of the American statistical association* 96(453), 42–55.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(2), 253–280.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of econometrics* 31(3), 307–327.

Christoffersen, P. F. (1998), Evaluating interval forecasts, *International economic review* pp. 841–862.

Diebold, F. X. and Mariano, R. S. (1995), Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13(3), 253–263.

Du, Z. and Escanciano, J. C. (2016), Backtesting expected shortfall: accounting for tail risk, *Management Science* .

Engle, R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica: Journal of the Econometric Society* pp. 987–1007.

Fissler, T., Ziegel, J. F. and Gneiting, T. (2015), Expected shortfall is jointly elicitable with value at risk-implications for backtesting, *arXiv preprint arXiv:1507.00244* .

Giacomini, R. and White, H. (2006), Tests of conditional predictive ability, *Econometrica* 74(6), 1545–1578.

Kupiec, P. H. (1995), Techniques for verifying the accuracy of risk measurement models, *The journal of Derivatives* 3(2), 73–84.

Shephard, N. and Sheppard, K. (2010), Realising the future: forecasting with high-frequency-based volatility (heavy) models, *Journal of Applied Econometrics* 25(2), 197–231.

West, K. D. (1996), Asymptotic inference about predictive ability, *Econometrica: Journal of the Econometric Society* pp. 1067–1084.

Ziegel, J. F. (2016), Coherence and elicitability, *Mathematical Finance* 26(4), 901–918.

6 Appendix

Table 10: Fit of the GARCH and HEAVY models of various indexes and exchange rates.

Asset	Heavy-R		GARCHX			GARCH		HEAVY-RM		Integrated	
	α	β	α_X	β_X	γ_X	α_G	β_G	α_R	β_R	α_g	α_R
Dow Jones Industrials	0.406	0.738	0.407	0.737	0.000	0.082	0.913	0.411	0.567	0.063	0.336
CAC 40	0.528	0.673	0.525	0.675	0.000	0.081	0.916	0.417	0.573	0.068	0.350
FTSE 100	0.608	0.658	0.614	0.656	0.000	0.105	0.892	0.434	0.562	0.085	0.369
Spanish IBEX	0.631	0.674	0.479	0.714	0.035	0.112	0.886	0.393	0.604	0.085	0.343
Nasdaq	0.723	0.661	0.439	0.744	0.051	0.082	0.915	0.427	0.568	0.063	0.349
Italian MIBTEL	0.806	0.630	0.806	0.631	0.000	0.107	0.889	0.512	0.486	0.080	0.436
S&P 400 Midcap	0.847	0.641	0.269	0.795	0.083	0.100	0.885	0.392	0.603	0.073	0.333
Nikkei	0.506	0.773	0.508	0.772	0.000	0.079	0.906	0.346	0.641	0.065	0.295
Russell 3000	0.449	0.746	0.446	0.748	0.000	0.081	0.911	0.403	0.574	0.059	0.313
Russell 1000	0.398	0.767	0.396	0.769	0.000	0.078	0.916	0.402	0.577	0.058	0.315
Russell 2000	0.946	0.679	0.243	0.812	0.102	0.107	0.885	0.387	0.622	0.077	0.322
Milan MIB	0.498	0.746	0.342	0.779	0.047	0.102	0.895	0.484	0.518	0.076	0.417
German DAX	0.446	0.673	0.446	0.674	0.000	0.093	0.903	0.457	0.536	0.075	0.376
S&P TSE	0.640	0.694	0.637	0.693	0.002	0.067	0.930	0.362	0.635	0.054	0.324
S&P 500	0.379	0.773	0.377	0.774	0.000	0.076	0.918	0.417	0.564	0.054	0.340
MSCI Australia	0.213	0.646	0.976	0.668	0.043	0.098	0.894	0.323	0.671	0.069	0.291
MSCI Belgium	0.768	0.568	0.374	0.692	0.093	0.143	0.854	0.399	0.608	0.105	0.359
MSCI Brazil	0.663	0.652	0.660	0.653	0.001	0.096	0.877	0.431	0.538	0.072	0.375
MSCI Canada	0.497	0.772	0.481	0.771	0.009	0.076	0.911	0.364	0.630	0.061	0.329
MSCI Switzerland	0.699	0.638	0.699	0.638	0.000	0.131	0.860	0.474	0.508	0.093	0.425
MSCI Germany	0.567	0.592	0.568	0.592	0.000	0.107	0.885	0.461	0.530	0.084	0.388
MSCI Spain	0.589	0.659	0.589	0.660	0.000	0.090	0.907	0.415	0.580	0.068	0.365
MSCI France	0.595	0.628	0.595	0.628	0.000	0.090	0.908	0.453	0.543	0.074	0.386
MSCI UK	0.580	0.618	0.582	0.617	0.000	0.110	0.886	0.456	0.543	0.086	0.393
MSCI Italy	0.581	0.660	0.583	0.659	0.000	0.100	0.896	0.536	0.463	0.075	0.467
MSCI Japan	0.736	0.722	0.740	0.721	0.000	0.088	0.901	0.458	0.534	0.075	0.387
MSCI South Korea	0.761	0.662	0.763	0.662	0.000	0.071	0.928	0.432	0.564	0.059	0.393
MSCI Mexico	0.873	0.711	0.720	0.726	0.032	0.095	0.885	0.363	0.625	0.068	0.328
MSCI Netherlands	0.538	0.678	0.538	0.678	0.000	0.105	0.889	0.453	0.542	0.084	0.396
MSCI World	0.337	0.799	0.337	0.799	0.000	0.085	0.909	0.377	0.610	0.068	0.340
British pound	0.162	0.811	0.161	0.815	0.000	0.042	0.950	0.282	0.699	0.035	0.264
Euro	0.056	0.934	0.034	0.949	0.013	0.031	0.968	0.247	0.746	0.028	0.223
Swiss franc	0.049	0.944	0.045	0.948	0.002	0.030	0.966	0.237	0.750	0.026	0.220
Japanese yen	0.171	0.774	0.172	0.775	0.000	0.048	0.934	0.396	0.554	0.036	0.341

Note: the whole sample is used for each estimation.

Table 11: Twice the likelihood change between different models by imposing restrictions.

Asset	Compare to extended HEAVY-r		Impose unit root		No momentum
	HEAVY-r	GARCH	GARCH	HEAVY-RM	$\beta = 0$
Dow Jones Industrials	0.0	-199.5	-48.5	-19.5	-56.5
CAC 40	0.0	-148.9	-30.9	-14.5	-67.3
FTSE 100	0.0	-125.7	-32.5	-12.3	-55.1
Spanish IBEX	-9.3	-113.7	-59.2	-12.1	-78.4
Nasdaq	-15.8	-108.4	-31.2	-14.4	-72.9
Italian MIBTEL	0.0	-141.2	-40.6	-9.9	-38.1
S&P 400 Midcap	-64.5	-61.8	-61.5	-11.0	-89.1
Nikkei	0.0	-116.5	-64.5	-9.9	-84.6
Russell 3000	0.0	-187.2	-49.9	-21.2	-61.3
Russell 1000	0.0	-186.2	-45.4	-20.0	-61.8
Russell 2000	-163.1	-64.9	-57.5	-12.7	-134.7
Milan MIB	-16.5	-100.7	-48.5	-13.0	-75.6
German DAX	0.0	-153.3	-47.2	-16.0	-63.7
S&P TSE	0.0	-120.8	-17.4	-5.6	-72.3
S&P 500	0.0	-211.0	-50.7	-17.9	-67.2
MSCI Australia	-7.9	-96.6	-31.2	-3.9	-55.8
MSCI Belgium	-22.7	-66.2	-60.2	-4.0	-57.4
MSCI Brazil	0.0	-60.1	-35.6	-7.1	-23.6
MSCI Canada	-0.4	-74.9	-23.2	-4.4	-57.9
MSCI Switzerland	0.0	-153.4	-65.8	-9.1	-32.7
MSCI Germany	0.0	-136.9	-45.0	-10.7	-44.5
MSCI Spain	0.0	-106.7	-31.5	-7.5	-44.5
MSCI France	0.0	-158.3	-27.7	-9.4	-47.1
MSCI UK	0.0	-134.3	-37.2	-9.3	-44.5
MSCI Italy	0.0	-154.7	-38.4	-8.7	-35.4
MSCI Japan	0.0	-111.8	-33.8	-6.2	-28.0
MSCI South Korea	0.0	-118.8	-15.2	-4.1	-43.5
MSCI Mexico	-3.4	-61.0	-36.4	-3.4	-43.2
MSCI Netherlands	0.0	-117.8	-40.9	-7.6	-46.8
MSCI World	0.0	-93.0	-25.7	-6.4	-104.0
British pound	0.0	-50.6	-15.7	-1.8	-28.3
Euro	-2.7	-18.8	-5.7	-1.6	-44.6
Swiss franc	-0.1	-33.2	-5.6	-1.7	-40.4
Japanese yen	0.0	-67.5	-38.6	-8.4	-26.1

Note: left-hand side is compared to the GARCHX model and right-hand side compares the unconstrained GARCH and HEAVY-RM model with those which impose an unit root. The final column compares the likelihood of the unconstrained HEAVY-RM model with the constrained HEAVY-RM model with $\beta = 0$

Table 12: In-sample likelihood ratio tests for losses generated by HEAVY and GARCH models.

Asset	t-statistic for non-nested LR tests for h					
	1	2	3	5	10	22
Dow Jones Industrials	-5.72	-3.80	-3.06	-2.96	-2.01	0.88
CAC 40	-4.39	-3.06	-2.34	-0.68	0.04	1.73
FTSE 100	-5.17	-3.39	-2.69	-1.70	-0.14	-0.14
Spanish IBEX	-2.83	-2.58	-1.48	-0.60	-1.11	-0.57
Nasdaq	-2.47	-0.47	-0.33	-0.71	1.12	-0.33
Italian MIBTEL	-4.11	-3.29	-3.35	-1.78	-0.78	-0.75
S&P 400 Midcap	0.07	1.11	1.06	0.08	0.27	-0.41
Nikkei	-3.87	-2.72	-2.24	-0.68	0.28	0.64
Russell 3000	-5.73	-4.04	-3.26	-4.04	-1.67	0.03
Russell 1000	-5.44	-3.90	-3.24	-3.88	-1.51	0.45
Russell 2000	1.66	2.23	2.14	1.20	1.34	0.01
Milan MIB	-1.89	-1.00	-0.91	-0.10	0.03	-0.02
German DAX	-5.12	-3.44	-2.86	-1.10	-0.81	-0.34
S&P TSE	-5.16	-4.52	-3.62	-2.28	-0.84	-0.17
S&P 500	-6.13	-4.44	-3.97	-4.08	-1.79	1.01
MSCI Australia	-3.15	-2.01	-2.65	-1.90	-2.45	-2.93
MSCI Belgium	-1.19	-1.29	-1.16	-1.86	-2.13	-2.11
MSCI Brazil	-3.56	-2.35	-1.59	-1.40	-1.52	-0.22
MSCI Canada	-3.92	-3.20	-3.16	-2.53	-1.66	-0.96
MSCI Switzerland	-4.30	-3.21	-2.43	-2.03	-0.35	-1.53
MSCI Germany	-5.27	-4.71	-4.17	-2.56	-1.16	-1.46
MSCI Spain	-3.71	-2.62	-2.12	-1.19	-0.29	-0.46
MSCI France	-5.70	-4.69	-3.45	-1.68	-0.53	0.09
MSCI UK	-5.60	-4.19	-3.37	-2.31	-0.33	-0.39
MSCI Italy	-5.36	-3.82	-3.35	-2.72	-0.92	-0.18
MSCI Japan	-5.29	-3.01	-2.28	-0.63	-0.02	0.79
MSCI South Korea	-4.82	-2.55	-2.23	-2.25	-0.39	2.90
MSCI Mexico	-2.47	-1.89	-1.94	-1.30	-2.00	-1.23
MSCI Netherlands	-4.78	-3.46	-2.42	-2.20	-1.35	-1.33
MSCI World	-5.51	-4.38	-3.44	-2.03	-1.27	-0.32
British pound	-3.30	-3.04	-2.07	-1.88	-1.50	-2.22
Euro	-1.14	-0.80	-0.70	-0.44	-0.27	-0.16
Swiss franc	-2.58	-2.87	-2.87	-2.11	-2.15	-2.28
Japanese yen	-3.02	-2.46	-1.32	-0.18	-0.89	0.68

Note: negative values favour HEAVY models. Both models are estimated using the quasi-likelihood, i.e. tuned to one-step-ahead predictions. Here h denotes the forecast horizon of the iterative forecasts.

Table 13: The p-values of the unconditional and conditional predictive ability tests of Giacomini and White (2006).

Asset	q=0.01		
	Unconditional	Conditional	Proportion
Dow Jones Industrials	0.194	0.007**	0.009
CAC 40	0.167	0.652	-
FTSE 100	0.164	0.167	-
Spanish IBEX	0.012*	0.008**	0.000
Nasdaq	0.305	0.408	-
Italian MIBTEL	0.073	0.011*	0.000
S&P 400 Midcap	0.133	0.456	-
Nikkei	0.307	0.129	-
Russell 3000	0.026*	0.000***	0.000
Russell 1000	0.022*	0.000***	0.000
Russell 2000	0.446	0.317	-
Milan MIB	0.016*	0.000***	0.000
German DAX	0.013*	0.000***	0.000
S&P TSE	0.470	0.975	-
S&P 500	0.050	0.000***	0.000
MSCI Australia	0.103	0.239	-
MSCI Belgium	0.276	0.530	-
MSCI Brazil	0.421	0.382	-
MSCI Canada	0.236	0.508	-
MSCI Switzerland	0.033*	0.135	-
MSCI Germany	0.051	0.003**	0.000
MSCI Spain	0.097	0.149	-
MSCI France	0.041*	0.375	-
MSCI UK	0.185	0.096	-
MSCI Italy	0.074	0.067	-
MSCI Japan	0.021*	0.137	-
MSCI South Korea	0.118	0.070	-
MSCI Mexico	0.074	0.215	-
MSCI Netherlands	0.095	0.350	-
MSCI World	0.255	0.075	-
British pound	0.231	0.554	-
Euro	0.346	0.196	-
Swiss franc	0.204	0.226	-
Japanese yen	0.455	0.350	-
Ratio	0.24	-	0.26

Note: if the conditional test is rejected, the decision rule is used. The proportion of $1\{\hat{\delta}_n' h_t > 0\}$ over the out-sample period is shown below. A proportion smaller than 0.5 favours the HEAVY model. On the left hand side, the ES forecasts at a coverage level of 0.05 are evaluated while the right hand side considers ES forecasts at a coverage level of 0.1. Ratio in the last row is the number of assets where the HEAVY model is favoured divided by the total number of assets. Furthermore, * represents a p-value smaller than 0.05, ** represents a p-value smaller than 0.01