Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis Financial Economics

Value at Risk forecasting

Abstract

The purpose of this paper is to investigate whether a dynamic Value at Risk model and high frequency realized volatility models can improve the accuracy of 1-day ahead VaR forecasting beyond the performance of frequently used models. As such, this paper constructs 60 conditional volatility forecasting models. Several extensions of the GARCH model are included, such as the nonlinear and asymmetric models. Moreover, several return distributions are assumed for the error term, in order to allow for more flexible modeling in the tails. A rolling Model Confidence Set is subsequently constructed, ensuring that only models with superior out-of-sample forecasting performance remain. A model averaging technique is applied to the remaining superior models, which generates the dynamic VaR forecasts. Moreover, several extensions of the HAR realized volatility model are included in this paper to forecast VaR. In a series of extensive back tests, this paper finds that the dynamic VaR model produces forecasts which are superior to traditional models and HAR models. The result hold for the 95% VaR, but are even more pronounced for the 99% VaR. The traditional models severely underestimate risk at higher confidence levels, whereas the applied dynamic VaR correctly accounts for it.

Keywords: Value at Risk, Model Averaging, GARCH, Realized Volatility, High Frequency, Model Confidence Set, VaR back testing

Thedo Linssen
383538

Supervisor: Rogier Quaedvlieg
Second assessor: Sjoerd van Bekkum

June, 2018

Contents

1 Introduction **1**

2 Literature Review **3**

3 Data **5**
3.1 Data transformation 5
3.2 GARCH subsamples 5
3.3 Descriptive statistics 6

4 Methodology **7**
4.1 Employed GARCH models 7
4.2 General assumptions 9
4.3 Probability density functions 9
4.4 Value at risk 10
4.5 Constructing dynamic Value at Risk forecasts 10
4.6 Realized Volatility models 12
4.7 Traditional methods 13
4.8 Back testing procedure 13

5 Results **15**
5.1 Optimal GARCH parameters 15
5.2 Realized volatility models and estimates 17
5.3 VaR back test results 19
5.4 Robustness analysis: MCS composition 28

6 Conclusion **31**

1 Introduction

Quantifying risk in the financial sector continues to prove to be a challenging and ever developing field of interest within finance and econometrics. The view to control financial risks more accurately has gained widespread acceptance, especially in light of the increasing number of financial disasters (Jorion, 1996). To this end, the Basel committee on banking supervision announced financial regulation that banks should adhere to, in an effort to impose a uniform measure of risk and stabilize the economy. The risk model that the financial regulators imposed became known as the Value at Risk (VaR) measure, which Jorion (1996) defines as the worst expected financial loss at a given confidence interval over a certain target horizon. Due to the fact that the VaR is expressed in monetary terms, risk managers can more easily evaluate whether they are comfortable with the level of risk inherent to the trading activities. (Jorion, 1996)

The RiskMetric model, developed by the risk management group of J.P. Morgan, became the official and most important benchmark in measuring market risk (So & Philip, 2006). Despite the risk measure's many strong points, the original RiskMetrics measure of calculating VaR has several shortcomings. For example, the common RiskMetrics model as well as other classical models, assume that the return distribution is conditionally normally distributed, which is often rarely the case in financial return data (Angelidis et al., 2004). So & Philip (2006) note that heavy tails in the form of excess kurtosis, and skewness as measured by the skew parameter are omnipresent, which in turn may lead to severe bias in the Value at Risk estimates, hereby inherently jeopardizing financial stability. Empirics have further proved that financial return data exhibits long term dependence on market volatility (Ding et al., 1993) and return data often shows severe volatility clustering (Bera & Higgins, 1993). The VaR risk measure itself is also characterized by a few shortcomings. If a hypothetical 1-day ahead 95% VaR of a portfolio is 1 million, a portfolio can be constructed such that daily loss is less than 1 million with 95.1% probability and 100 million with a 4.9% probability. In addition, Hull (2012) notices that the VaR is not a coherent risk measure, since it does not meet the subadditivity criterion. This implies that the sum of the risk measure of two merged portfolios may exceed their individuals sum[1].

Mitigation of some of these drawbacks have led to a search for more sophisticated models to forecast VaR more accurately (Angelidis et al., 2004). Several shortcomings of for instance the historical simulation and variance covariance model can be improved upon by modeling the conditional volatility with Generalized Autoregressive Heteroskedastic (GARCH) models. By employing conditional volatility estimates, a whole new class of GARCH-VaR models arises, potentially improving risk estimates. These models, among other improvements, feature the ability to account for time varying volatility and have long term memory of volatility (So & Philip, 2006). By using different variants of the GARCH models and by assuming various conditional return distributions,

[1]For this reason, various efforts have been made towards modeling market risk with alternative measures such as Expected Shortfall (ES), which aim at mitigating these issues.

academics and practitioners aim at further improving VaR forecasting.

Since access to reliable volatility forecasts is paramount for regulators and financial practitioners (Bollerslev et al., 2016), other avenues, such as exploiting high frequency data for risk forecasting purposes have recently been explored. Giot & Laurent (2004) find that the recent widespread availability of high frequency data has given rise to this promising field. Exploiting intra-day data allows financial risk management to observe the volatility, rather than treat it as a latent variable (Corsi et al., 2008). The Heterogenous Autoregressive (HAR) model are among the most popular and best performing in recent empirical works, since they capture long-memory of volatility and even outperform parametric GARCH models in forecasting ability (Corsi et al., 2008). Succinctly, Angelidis et al. (2004) recognize that the choice of an adequate and final risk forecasting model is far from resolved, which makes exploring these and alternative methods viable. The question that naturally arises is:

Can realized volatility models or a composite of advanced GARCH-VaR models provide more accurate 1-day ahead Value at Risk forecasts than frequently employed models?

This paper generates 95% and 99% 1-day ahead VaR forecasts for several GARCH-VaR models with returns originating from the S&P500 and AEX. Apart from a standard GARCH model, this paper will analyse a multitude of GARCH extensions, such as nonlinear asymmetric GARCH models and component GARCH models. Next to modeling different varieties of the GARCH family, this paper will also address the underlying return distributions for each model. To this end, not only the normal distribution will be analysed, but also the student t-distribution and the generalized error distribution will be assumed for the error term. This will allow for more flexible risk modeling in the tails (Angelidis et al., 2004). Analogous to Hansen & Lunde (2005), this paper constructs four types of GARCH variants by varying the lag length combinations of the constructed models. In total, 60 GARCH models are eventually constructed. Hansen & Lunde (2005) notice that an inferior model is likely to be lucky when many models are compared. For this reason, only GARCH models that generate superior out-of-sample VaR forecasts are considered, whereas the remaining models are excluded. By constructing a Model Confidence Set (MCS), as developed in Hansen et al. (2011), and employing the loss function as stipulated in González-Rivera et al. (2004), the accuracy of the models is compared, hereby taking into consideration the models interdependence. The MCS ultimately includes the best GARCH forecasting models with a specified probability. Once the models with superior predictive ability are obtained, a forecast combination technique similar to (Bernardi et al., 2014) is performed, because Stock & Watson (2004, 1998, 1999) find that model averaging produces superior results as compared to model selection. This entire process is executed on a rolling window in order to generate VaR forecast ex ante. The obtained composite model is eventually referred to as the 'dynamic VaR'.

Additionally, this paper employs three types of high frequency HAR models and estimates the out-of-sample realized volatility. The volatility estimates are then used to construct VaR forecasts. Finally, the dynamic VaR model and HAR models are extensively compared to traditional models in a full back test.

This paper finds that the dynamic VaR model produces forecasts which are superior to traditional models and HAR models. The result hold for the 95% VaR, but are even more pronounced for the 99% VaR. The traditional models severely underestimate risk at higher confidence levels, whereas the applied dynamic VaR correctly accounts for it. The remainder of this research paper is structured as follows. The following section provides an extensive overview of risk modeling in literature. Next, the data transformations as well as the descriptive statistics are shown in the data section. The methodology subsequently elaborates in detail on the construction of the dynamic VaR forecasting model as well as the HAR models. The results naturally follow, and conclusions are drawn in the final section of this paper.

2 Literature Review

Since the 1996 amendment from the Basel Committee, banks are legally required to hold capital reserves for the incurred market risk. For assets in the trading book, the Basel Committee allows companies to use the internal model based approach and the standardized approach. The fact that correlations between the assets in the portfolio are considered in the former allows banks to hold less capital reserves. This in turn makes the internal model based approach a favourable choice for especially big banks (Hull, 2012). Giot & Laurent (2003a) further notice that the popularity of VaR as a metric can be explained by that the fact that VaR is easy to understand and that it aggregates the loss of a portfolio with a certain probability in monetary terms.

The fact that the use of VaR as a major improvement over earlier risk management techniques spreads quickly, should not mask its shortcomings (Jorion, 1996). For instance, Angelidis et al. (2004) notice that many applications presume that asset returns are normally distributed, while in fact excess kurtosis and skewness are frequently observed (Giot & Laurent, 2003b), which in turn leads to unreliable VaR estimates. For this reason, researchers and academics have extensively tried to improve VaR forecasting. Venkataraman et al. (1997) investigated whether the use of a mixture of normal distributions could improve forecasting performance. They find that the employed model is better able to account for extreme events due to fatter tails, and that their model performs significantly better as compared to classical approaches. Giot & Laurent (2003b) model daily VaR for a range of ARCH models and assume a skewed student distribution for the error term. They show that when modeling in the tails, models that rely on a symmetrical density distribution underperform models that rely on their skewed counterpart. Under the same distributional assumption, Giot & Laurent (2003a) find in another paper that the skewed APARCH and ARCH model deliver reliable VaR forecasts. However, they favour the use of the skewed

ARCH due to the ease with which it is calculated since it does not require nonlinear parameter optimisation.

Another focal point of research is the specification of the conditional volatility forecasting models. The first GARCH model was developed by Bollerslev (1986) as a generalisation of the ARCH model that was originally developed by (Engle, 1982). Although the GARCH model can account for volatility clustering, it for example cannot deal with asymmetries in return data (Brooks, 2014). A vast amount of extensions to this model have ever since been developed in order to account for these potential deficiencies. Among the most popular asymmetric models are the EGARCH model (Nelson, 1991) and the GJR-GARCH Glosten et al. (1993) that are both capable to deal with the leverage effect. In the latter case, a term is simply added to the standard GARCH formula that captures the leverage effect (Brooks, 2014). Another noticeable model is the component GARCH model developed by (Lee & Engle, 1993). In their model, the long run volatility is allowed to be time varying. Consequently, the component model is able to capture autocorrelation patterns in variance which die out slower than what is possible in the simple shorter-memory models (Christoffersen, 2012). Other extensions include the Nonlinear Asymmetric GARCH (NAGARCH) developed by Engle & Ng (1993) and the threshold GARCH (TGARCH) developed by (Rabemananjara & Zakoian, 1993).

Bates & Granger (1969) find that a combination of forecasts yields better performance than individual models, provided each set of forecasts possesses independent information. This is an important finding, since Samuels & Sekkel (2011) recognize that econometricians nowadays have access to a great number of competing models. They construct various different methods to select models from a larger pool of candidates based on out-of-sample forecasting performance. Eventually, they find that substantial gains in forecasting performance can be achieved by omitting inferior models and that especially the MCS is appropriate in delivering robust results in picking models based on out-of-sample forecasting performance.

Another promising avenue in risk forecasting is exploiting high frequency data in order to forecast volatility. Since the availability of high frequency data allows the construction of realized volatility, rather than modeling latent daily volatility, various new models have been proposed in recent academic work. Arguably, the HAR model has arisen as the most popular form (Bollerslev et al., 2016). Corsi (2009) introduced the HAR model, which essentially is an AR-type of model that aggregates realized variance over different intuitive time horizons. Corsi (2009) notices that the model parsimoniously accounts for many empirical features of financial return series, such as the long-memory property, fat tails and self-similarity. Corsi (2009) also notices that due to its simplistic formulation, it can easily be extended to other more sophisticated models, such as the HARRVJ as developed in Andersen et al. (2007). By including a jump component, total variance is decomposed into a continuous and discontinuous part. Patton & Sheppard (2015) provide evidence that future volatility is more strongly related to past negative returns than to positive returns.

They incorporated this phenomenon and introduced the SHAR model. This semi-variance model decomposes variance based on past return. Giot & Laurent (2004) compared the VaR estimates obtained by a realized volatility model and an APARCH model using two different equity indexes. Although both deliver adequate performance, none of them stands out in their paper in terms of performance.

3 Data

3.1 Data transformation

Modeling and back testing is mainly performed in R and Matlab. Furthermore, the the price series of the S&P500 and the AEX index that will serve as the input for the GARCH models have been retrieved from the Datastream database. The price series are converted to continuously compounded returns by applying formula 1:

$$r_t = \ln(\frac{P_t}{P_{t_{-1}}}) \tag{1}$$

In this formula, r_t is the daily return, P_t is the price on day t and $P_{t_{-1}}$ is the price on the preceding day. All subsequent computations are performed using the obtained log returns, due to the benefits that come with this procedure. Log returns have the property of being additive, in contrast to simple returns. After converting the price series to additive returns, the sample size decreases by 1 observation, and spans from 14/03/2006 to 11/09/2017, which is, after correcting for non-trading days, equal to exactly 3000 observations. The reason for investigating this particular time period is that it incorporates the most recent available financial information.

For the intra-day measures which are required for the high frequency realized volatility models, various variables have been retrieved from the Oxford-Man realized library. More specifically, the bi-power variation, daily realized volatility and the corresponding daily returns were obtained for the S&P500 as well as the AEX index. The obtained variables are matched in terms of date with the return series originating from Thomson Reuters Datastream, in order to construct a proper framework for comparison of the different forecasting models.

3.2 GARCH subsamples

The total data sample that will be used for the GARCH models is divided in an in-sample period, an out-of-sample period and a second out-of-sample period. In order to obtain reliable GARCH parameter estimates, at least 1000 observations should be included (Christoffersen, 2012) in estimating the parameters. For this reason, the first 1000 observations of the in-sample period are used completely for calculating the required parameters of the GARCH models by employing the maximum likelihood estimation procedure. Subsequently, the two out-of-sample sub periods will be

used for 1-day ahead out-of-sample forecasting. As will be explained in detail in the methodology section, the 1000 observations in second out-of-sample period are vital. It is these 1000 observations that will be used for 1-day ahead dynamic VaR forecasting based on the set of superior models over a rolling window.

3.3 Descriptive statistics

Figure 1 shows the log returns of the S&P500 graphically. The figure is consistent with stylized facts of stock returns, such as volatility clustering. Especially during the financial crisis of 2008, a burst of volatility is exhibited. The same holds for the plotted returns of the AEX in appendix A (figure 5). Particularly in this case of the heteroskedastic volatility, employing GARCH and HAR models is desirable, since empirics prove they can appropriately account for it. Table 1 aims to

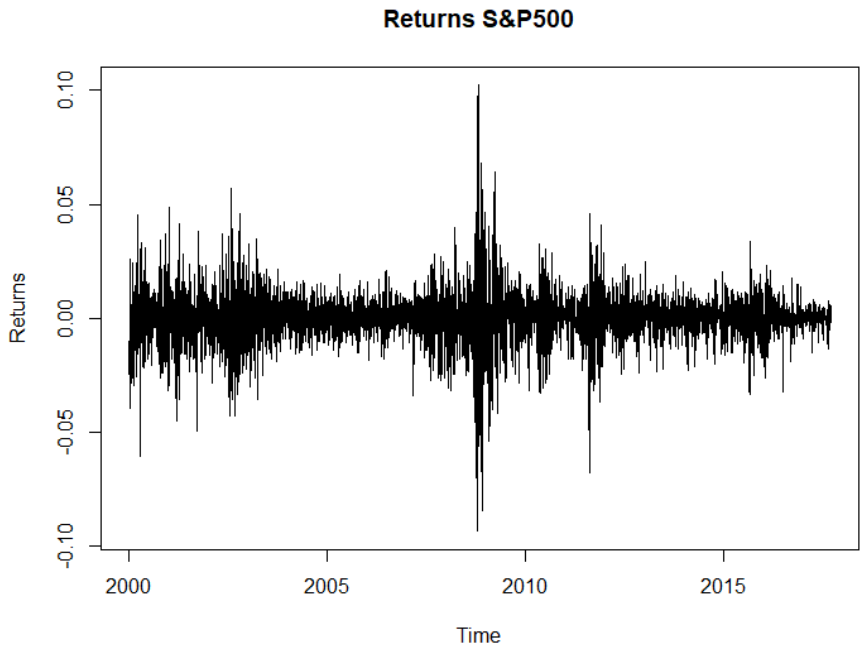


Figure 1: The continuously compounded returns from 14/03/2006 to 11/09/2017 of the S&P500 are plotted.

provide more insights into the data by listing the descriptive statistics of the datasets.

Both samples illustrate that the daily average return is 0%. Moreover, the standard deviation across the datasets is roughly equal. Nevertheless, the S&P500 was characterized by more extreme values as measured by the minimum and maximum recorded returns. In fact, the latter index has a higher kurtosis, meaning that its distribution has fatter tails as compared to the AEX index. Both indexes also exhibit negative skewness, where a skew parameter of 0 would be expected under a normal distribution. The excess kurtosis and negative skew jointly cause both datasets to fail the Jarque Bera test. In other words, the observed log return distributions are significantly different

Table 1: This table displays the descriptive statistics of the employed datasets.

	AEX	**S&P500**
Observations	3000	3000
Minimum	-0.084	-0.094
Median	0.000	0.001
Arithmetic Mean	0.000	0.000
Maximum	0.074	0.102
SE Mean	0.000	0.000
Variance	0.000	0.000
Stdev	0.011	0.012
Skewness	-0.424	-0.295
Kurtosis	6.681	11.266
Jarque Bera	5670	15910
p-val JB	0.000	0.000

from a normal distribution at any conventional significance level. In appendix A, figure 6a and 6b show the histograms of empirical return distributions. In the figures, leptokurtic features of the data are clearly visible, since there is a high concentration of returns around zero and somewhat fatter tails.

4 Methodology

This chapter will first elaborate on the employed models and the general assumptions that have been made with respect to the innovation distributions of the GARCH models. Subsequently, in order to calculate the VaR for the wide range of different forecasting models, the density functions of the assumed distributions will be presented, along with the formula to calculate the VaR itself. Once the VaR forecasts of all the analyzed GARCH-VaR models are obtained, this paper will construct a set of models with superior predictive ability according to the Model Confidence Set. It will further be explained how this is used on a rolling window in order to generate ex-ante out-of-sample VaR estimates of models belonging to the MCS. A model averaging technique is applied in order to construct the dynamic VaR forecasts on 95% and 99% confidence level. Subsequently, this paper explains how the various high frequency realized volatility models are constructed and used on a rolling window. The dynamic obtained VaR as well as the realized volatility models, on their turn, will be compared to traditional methods such as the EWMA, the historical simulation method and the normal distribution method in a sequence of historical back tests. The results must ultimately indicate whether more reliable VaR estimates can be obtained by either applying the dynamic VaR model or realized volatility models.

4.1 Employed GARCH models

A wide range of different models is used in order to forecast conditional volatility. Bollerslev (1986) introduced the first GARCH model as a generalization of the ARCH model. In the formula

which is shown in table 2, σ_t^2 is the estimated conditional variance and ω, α_i and β_j are estimated parameters from the data where the intercept, ω represents the long term variance. The model implies that today's volatility is the weighted sum of the long term variance, squared lags of the residuals and squared lags of estimated volatility.

The second model that is included is the Exponential GARCH model (EGARCH) developed by (Nelson, 1991). It gained popularity due to the fact that the model aims at mitigating some of the pitfalls that are inherent to the previously discussed GARCH model. Unlike the standard GARCH model, the EGARCH is able to account for asymmetry (Dutta, 2014). The model is specified such that positive and negative returns do not influence future conditional volatility in the same way, but they are assigned different weights. This can account for the leverage effect, where negative returns induce more volatility than positive returns (Christoffersen, 2012).

The third and fourth model that are included in the research are the Nonlinear Asymmetric GARCH (NAGARCH) developed by Engle & Ng (1993) and the threshold GARCH (TGARCH) developed by Rabemananjara & Zakoian (1993). By also relaxing assumptions, such as linearity on variance dynamics in the TGARCH case, the models aim at describing return variance more accurately.

The last reviewed model that received broad academic coverage is the component GARCH type of models, because of their ability to capture complex dynamics in a parsimonious model (Chen et al., 2011). This paper includes the Component Standard GARCH (CSGARCH) that Lee & Engle (1993) developed. Their model consists of a short run (transitory) component and long run component. In table 2, q_t in the csGARCH formula represents the long run component.

Analogous to Hansen & Lunde (2005), this paper estimates GARCH models with all combinations of a maximum lag length for p and q of 2 in order to keep the amount of models realistic. This means that for all the mentioned models in table 2, four different models will be estimated. Additionally, all of these models will be estimated using three different assumed innovation distributions, leading to a total number of 60 models[2]. Appendix B lists all the analyzed combinations. The first 1000 observations in the in-sample period is used to calculate the initial parameters of the employed GARCH models. Subsequently, the parameters of the GARCH models are estimated over a rolling window of 1000 observations and are re-estimated every 100 observations. This means that each model's parameters are re-estimated 20 times in total and ensures that the most recent information is incorporated repeatedly.

[2]Using five different forecasting models under three different assumed innovation distributions and four combinations of p and q per model leads to 5*3*4 = 60 models.

Table 2: This table shows the formulaic specifications of the models that are used throughout this paper in forecasting conditional volatility.

Employed forecasting models	
Name	
GARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
EGARCH	$log(\sigma_t^2) = \omega + \sum_{i=1}^{q} [\alpha_i e_{t-i} + \gamma_i(\lvert e_{t-i}\rvert - E\lvert e_{t-i}\rvert)] + \sum_{j=1}^{p} \beta_j log(\sigma_{t-j}^2)$
NAGARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (\varepsilon_{t-i} + \gamma_i \sigma_{t-i})^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
TGARCH	$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i [(1-\gamma_i)\varepsilon_{t-i}^{+} - (1+\gamma_i)\varepsilon_{t-i}^{-}] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$
csGARCH	$\sigma_t^2 = q_t + \sum_{j=1}^{q} \alpha_j (\varepsilon_{t-j}^2 - q_{t-j}) + \sum_{j=1}^{p} \beta_j (\sigma_{t-j}^2 - q_{t-j})$ $q_t = \omega + \rho q_{t-1} + \phi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2)$

4.2 General assumptions

As discussed, daily conditional volatility σ_t is calculated using various GARCH models. Throughout this paper, returns are assumed to have zero mean, which is one of the most prudent choices a risk manager can make. On average daily returns are insignificantly different from zero and dominated by their standard deviation (Christoffersen, 2012). The daily returns in this paper can further be described by:

$$r_t = \sigma_t z_t = \epsilon_t \tag{2}$$

$$z_t \sim N(0,1) \tag{3}$$

Formula 3 implies that the innovation term z_t is a series of independently and i.i.d. normally distributed variables that have zero mean and a unit variance. As will be shown in the next section, apart from the normal distribution, the student's t-distribution and the generalized error distribution will be assumed for z_t in equation 3.

4.3 Probability density functions

When Engle (1982) introduced the ARCH process, z_t in equation 3 was assumed to follow a normal distribution. Formula 4 shows the corresponding density function that z_t would follow under the normal distribution.

$$D(z_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z_t-\mu)^2}{2\sigma^2}} \tag{4}$$

Bollerslev (1987) on the other hand, suggested using the student's t-distribution. By following the notation as stipulated by Angelidis et al. (2004), the density function of z_t under the student's

t-distribution is given as in equation 5.

$$D(z_t) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}(1+\frac{z_t^2}{v-2})^{-}\frac{v+1}{2} \tag{5}$$

In equation 5, v stands for the degrees of freedom, and $\Gamma(v) = \int_0^{\inf} e^{-x}x^{v-1}dx$. Like the normal distribution, this distribution is also symmetric around the mean. Moreover, as v tends to infinity, the distribution will converge to a normal distribution. Nelson (1991) on the other hand favours the use of the generalized error distribution, as given in formula 6.

$$D(z_t) = \frac{e^{(-0.5|z_t/\lambda|^v)}v}{2^{(1+\frac{1}{v})}\Gamma(v^{-1})\lambda} \tag{6}$$

In this density function, v stands for the tail thickness parameter. Moreover, the λ is equal to $[2^{(-2/v)}\Gamma(1/v)/\Gamma(3/v)]^{1/2}$. Using the distributions, the parameters of the estimated GARCH models are calculated by maximizing the corresponding log likelihood functions.

4.4 Value at risk

The Value at Risk is the worst expected financial loss at a given confidence interval (Jorion, 1996). The value at risk throughout this paper is calculated at a 95% confidence level as well as the 99% level. VaR is calculated by multiplying the conditional volatility forecast with the appropriate quantile in one of the assumed density functions. In this paper, the following formula is used to calculate 1-day ahead value at risk forecasts:

$$VaR_{t+1} = F(\alpha)\sigma_{t+1} \tag{7}$$

In formula 7, $F(\alpha)$ corresponds to the desired quantile of the distribution's density function, which are given in formula 4-6. The σ_{t+1} is the 1-day ahead forecast as provided by risk forecasting models. Formula 7 is used to calculate the 1-day ahead VaR for the GARCH models as well as the realized volatility models.

4.5 Constructing dynamic Value at Risk forecasts

Since Bates & Granger (1969), a lot of research has been devoted to combining various sets of forecasts in order to improve the performance of a single risk forecasting model. To this end, this paper constructs a dynamic VaR model, and largely follows the methodology as stipulated by Bernardi et al. (2014) in doing so. Before the VaR estimates of the large number of models are combined, a superior set of models is first created, following the MCS as developed by Hansen et al. (2011). Performing this procedure on the set of available risk forecasting models reduces the set to a set that contains the best model with a specified level of confidence.

A loss function must first be supplied before the MCS can be constructed. In this paper, the

loss function as stipulated in González-Rivera et al. (2004) is used and is given in formula 8. Using this loss function is especially appropriate, given the fact that it penalizes forecasted VaR above and below the observed return asymmetrically (Bernardi et al., 2014).

$$l(y_t, VaR_t^\tau) = (\tau - d_t^\tau)(y_t - VaR_t^\tau) \tag{8}$$

In this formula, τ represents the chosen confidence level of VaR. Subsequently, this paper fully replicates the construction of the model confidence set as described in (Hansen et al., 2011). The initial set of considered models is M^0 and it contains the evaluated models in terms of their loss function as specified in formula 8. The relative performance variables can then be given by $d_{ij,t} = l_{i,t} - l_{i,t}$ for all $i, j \in M^0$. It is further assumed that $\mu_{ij} \equiv E(d_{ij,t})$. By repetitively testing the null hypothesis, and deleting inferior models, the set of superior models is eventually defined by:

$$M^* \equiv \left\{ i \in M^0 : \mu_{ij} \leq 0 for j \in M^0 \right\} \tag{9}$$

In this paper, the model confidence set will be constructed using the past 1000 out-of-sample VaR forecasts. Due to the fact that the out-of-sample periods combined is 2000 observations, the MCS will be constructed using a rolling window. The MCS is re-estimated every 100 observations, implying that the MCS is constructed 10 times in total per confidence level and per equity index. The first out-of-sample period (1000 observations) is used entirely to construct the first MCS, and the included models of this MCS are used to predict the VaR for the first 100 observations in the second out-of-sample period. This process is repeated until the entire second out-of-sample period is forecasted. Using this procedure ensures that the forecasted dynamic VaR can be generated ex-ante. Moreover, due to the fact that the process is repeated 10 times, the most recent set of superior models is used repeatedly. The formula used to calculate the dynamic VaR in the second out-of-sample period is given in equation 10.

$$VaR_t^{dynamic} = \frac{1}{m} \sum_{i=1}^{m} VaR_t \tag{10}$$

In this equation, m represents the number of models that belong to MCS, which may vary over time. The equation computes the arithmetic average of the individual VaR forecasts at any given time t of the models that belong to the rolling MCS.

Lastly, in order to analyze whether or not using models with superior predictive ability does increase forecasting ability, a second dynamic model is introduced. In this case, the simple arithmetic average of all 60 GARCH-VaR models will be calculated at every t using equation 10. However, in this case m simply represents the 60 GARCH-VaR models instead of models belonging to the MCS. The model will be referred to as 'dynamic model average'.

4.6 Realized Volatility models

The realized volatility measures are retrieved from the Oxford-Man realized library, and realized variance is calculated as follows:

$$RV_t = \sum_{i=1}^{m} r_{t,i}^2 \tag{11}$$

The intra-day returns r are calculated as the first log difference between two consecutive prices. Overnight returns have been omitted, but are incorporated in the first observation of the following trading day. As the number of intra-day observations m increases, RV_t gives a consistent estimate of the realized volatility (Bollerslev et al., 2016). This paper uses 5-min return data to mitigate the effect of noise. The first HAR model was introduced by Corsi (2009) and is specified as follows:

$$RV_{t+1}^d = C + \beta_1 RV_t^d + \beta_2 RV_t^w + \beta_3 RV_t^m + \epsilon_t \tag{12}$$

In this equation, RV_t^d refers to the daily realized variance, the weekly variance $RV_t^w = \frac{1}{5}\Sigma_{i=1}^{5} RV_{t-i+1}$ and lastly, the monthly variance $RV_t^m = \frac{1}{22}\Sigma_{i=1}^{22} RV_{t-i+1}$. The model, which will subsequently be referred to as HARRV, considers different volatility components over various time horizons parsimoniously. Moreover, it achieves its purposes of taking into account fat tails and long memory (Corsi, 2009). Corsi (2009) further notices that due to its simplistic formulation, it can easily be extended to other more sophisticated models, which leads to the HARRV model with jumps included. More specifically, this paper follows the HARRVJ model developed in Andersen et al. (2007) and it is defined as:

$$RV_{t+1}^d = C + \beta_1 RV_t^d + \beta_2 RV_t^w + \beta_3 RV_t^m + \beta_4 J_t + \epsilon_t \tag{13}$$

In this equation, J_t is the jump component and it equals $max(RV_t - BPV_t, 0)$, where the Bi-Power Variation has been retrieved directly from the Oxford-Man realized library. In this specification, total variation has been split into a continuous part and a discontinuous part as represented by the jump. In this model, the HAR model is simply expanded by also including a jump component as an independent variable in the regression.

Lastly, analogous to Patton & Sheppard (2015), this paper includes a Semi-variance HAR model (SHAR) model. The model decomposes total variation at the first lag into two parts, each of which depends on the sign of returns on the previous trading day.

$$RV_{t+1}^d = C + \beta_1^+ RV_t^d + \beta_1^- RV_t^d + \beta_2 RV_t^w + \beta_3 RV_t^m + \epsilon_t \tag{14}$$

The square root of all obtained out-of-sample variance forecasts is calculated. Next, formula 7 in combination with an assumed normal distribution are employed jointly to calculate the VaR for all three realized volatility models. Analogous to the GARCH models, VaR will be calculated

at the 95% and 99% confidence level.

4.7 Traditional methods

As mentioned, the performance of the dynamic VaR forecast will be compared to three popular and widespread accepted models. To this end, the first included model is the historical simulation method, since Christoffersen (2012) recognizes this is the most frequently used method in practice. In this simple method, past data is used to predict the future volatility and no assumptions are made with respect to a distribution. The first step is to calculate the market variable using the following equation:

$$v_{t+1} = v_n \frac{v_i}{v_{i-1}} \tag{15}$$

In the equation above, v_n is the most recent closing price and v_i is the value on day i. By ranking the obtained variables in ascending order, the VaR is found by picking the loss associated with the employed confidence interval (Christoffersen, 2012). In this paper, the last 250 trading days are used, which corresponds roughly to a whole calender year.

The second approach is the variance-covariance method, also known as the normal distribution method and is the most basic way to calculate VaR. By assuming profits and losses follow a normal distribution, the VaR can easily be obtained by multiplying the Z-score at the desired confidence level by the standard deviation of the returns. In this paper, the past 250 trading days will be taken into account. The formula is as follows:

$$VaR_{t+1} = \sigma_t N^{-1} \tag{16}$$

In contrast to the historical simulation method and the variance-covariance method, JP Morgan's RiskMetrics model places more emphasize on the role of recent observations. The formula used for tomorrow's volatility is as follows:

$$\sigma^2_{t+1} = \lambda\sigma^2_t + (1-\lambda)r^2_t \tag{17}$$

In this model, tomorrow's volatility is simply the by λ weighted effect of today's volatility and today's squared return. Due to the fact that λ is smaller than 1, the effect of distant returns decrease exponentially. This paper uses the value 0.94 for the λ parameter, since this value gives the best forecasting performance of the EWMA model (Christoffersen, 2012). Another advantage of this is that no parameters have to be optimized.

4.8 Back testing procedure

Once the dynamic and realized volatility VaR estimates are obtained, results are back tested in order to see if the more sophisticated models improves VaR forecasting. To this end, a multitude

of conditional and unconditional coverage back tests will be performed. The first considered back test was introduced by the Basel Committee (1996) as a regulatory back test, and it known as the Traffic Light test. In their specified methodology, a VaR model is calculated over the last 250 trading days and its forecast is compared to the actual performance. By calculating the cumulative probability of obtaining the number of VaR exceedances, the likelihood of having an accurate VaR model is defined as green, yellow and red. This test however, has a few drawbacks. Due to the fact that VaR exceedances are simply added up, VaR models that significantly overestimate market risk will are likely to end up in the green zone. Another flaw is the fact that the interdependence of the VaR exceedances is not taken into account.

Another unconditional coverage test that is performed in order to statistically test whether the amount of VaR violations can reasonably be expected given the confidence level is the binomial test. If x is the number of violations, N the number of observations and $p = 1$ - confidence level, then the number of expected violations is Np. If the forecasting model is accurate, the observed failure rate x/N should act as an unbiased measure of p and thus converge to p (Jorion et al., 2010). The test statistic, which gives reliable estimates for sufficiently large samples, subsequently follows a normal distribution and is given by:

$$Z = \frac{x - Np}{\sqrt{Np(1-p)}} \tag{18}$$

Both the traffic light and the binomial test cannot account for clustering of violations. For this reason, the likelihood test for independence developed by Christoffersen (1998) is included. This conditional coverage test assesses whether failures occur on consecutive days and incorporates it in the test statistic. Ideally, VaR violations on any day do not influence the next day's likelihood of a VaR violation. The likelihood ratio is given by:

$$LR_{ind} = -2ln(\frac{(1-\pi)^{\eta_{00}+\eta_{10}}\pi^{\eta_{00}\eta_{10}}}{(1-\pi_0)^{\eta_{00}}\pi_0^{\eta_{01}}(1-\pi_1)^{\eta_{10}}\pi_1^{\eta_{11}}}) \tag{19}$$

An accurate VaR model must however meet both the criteria of independence and conditional coverage (Campbell, 2006). For this reason, the mixed conditional coverage test is performed, which is a joint test of conditional coverage and independence. The mixed conditional coverage tests incorporates both Christoffersen's (1998) aforementioned conditional coverage test and the proportion of failures test by Kupiec (1995). The test statistic of the mixed test equals the sum of the likelihood ratio in equation 19 and the likelihood ratio of the proportion of failures test by Kupiec (1995) in equation 20. By accounting for both independence of failures the number of violations, the joint test provides a comprehensive picture of the VaR model's performance.

$$LR_{pof} = -2log(\frac{(1-pVaR)^{N-x}pVaR^x}{(1-\frac{x}{N})^{N-x}(\frac{x}{N})^x}) \tag{20}$$

Analogous to equation 18, x is the number of violations, $p =$ (1 - confidence level) and is N the

number of observations.

5 Results

This first two sections will elaborate on the optimal parameters that have been found for the employed forecasting models in an in-sample setting. Subsequently, the full back testing results of all considered VaR models is discussed in great depth. Lastly, it is interesting to analyze how the composition of the MCS varies over time. For this reason, this chapter also contains a robustness analysis that will analyze whether models with superior predictive ability will continue to be superior in the future. In this section, changes in the MCS will be monitored and discussed.

5.1 Optimal GARCH parameters

Due to the vast number of GARCH models that have been estimated to calculate the dynamic VaR models, table 3 shows only an excerpt of the employed forecasting models. The table features all five GARCH variants and all assumed distributions, but all models are restricted to the first lag of p and q to keep the number of models analyzable. The table shows the optimal coefficient estimates, along with their robust standard errors and various statistics such as the information criteria. The models are estimated in-sample for the S&P500 and the sample runs from 12/11/2013 to 11/09/2017. The parameter estimates of all utilized models seem to fit the models well, as measured by the corresponding significance. This holds to a somewhat lesser degree for the component GARCH models where each time at least two parameter estimates do not meet the desired significance level. Additionally, it seems that the model specifications are rather robust to changing the assumed innovation distribution, since parameters remain quite stable.

Table 3: This table shows the in-sample parameter estimates and the significance of all employed forecasting models for the S&P500. The amount of employed lags is restricted to 1.

	GARCH-norm(1,1)		**Statistics**			**GARCH-std(1,1)**		**Statistics**			**GARCH-ged(1,1)**		**Statistics**	
	Estimate	Std. Error	LogLikelihood	3554.201		Estimate	Std. Error	LogLikelihood	3593		Estimate	Std. Error	LogLikelihood	3608.02
ω	0.00	0.00	AIC	-7.1024	ω	0.00	0.00	AIC	-7.18	ω	0.00	0.00	AIC	-7.21
α_1	0.18	0.02	BIC	-7.0877	α_1	0.21	0.04	BIC	-7.16	α_1	0.20	0.03	BIC	-7.19
β_1	0.72	0.03			β_1	0.76	0.17			β_1	0.74	0.06		
	EGARCH-norm(1,1)		**Statistics**			**EGARCH-std(1,1)**		**Statistics**			**EGARCH-ged(1,1)**		**Statistics**	
	Estimate	Std. Error	LogLikelihood	3598.282		Estimate	Std. Error	LogLikelihood	3630.23		Estimate	Std. Error	LogLikelihood	3639.12
ω	-0.74	0.01	AIC	-7.19	ω	-0.67	0.01	AIC	-7.25	ω	-0.74	0.01	AIC	-7.27
α_1	-0.27	0.02	BIC	-7.17	α_1	-0.29	0.02	BIC	-7.23	α_1	-0.30	0.02	BIC	-7.24
β_1	0.93	0.00			β_1	0.93	0.00			β_1	0.93	0.00		
γ_1	0.09	0.02			γ_1	0.13	0.02			γ_1	0.12	0.01		
	NAGARCH-norm(1,1)		**Statistics**			**NAGARCH-std(1,1)**		**Statistics**			**NAGARCH-ged(1,1)**		**Statistics**	
	Estimate	Std. Error	LogLikelihood	3612.953		Estimate	Std. Error	LogLikelihood	3645.46		Estimate	Std. Error	LogLikelihood	3651.56
ω	0.00	0.00	AIC	-7.22	ω	0.00	0.00	AIC	-7.28	ω	0.00	0.00	AIC	-7.29
α_1	0.11	0.01	BIC	-7.20	α_1	0.10	0.01	BIC	-7.26	α_1	0.11	0.01	BIC	-7.27
β_1	0.55	0.05			β_1	0.52	0.17			β_1	0.52	0.07		
γ_1	1.77	0.06			γ_1	1.90	0.39			γ_1	1.86	0.14		
	TGARCH-norm(1,1)		**Statistics**			**TGARCH-std(1,1)**		**Statistics**			**TGARCH-ged(1,1)**		**Statistics**	
	Estimate	Std. Error	LogLikelihood	3597.924		Estimate	Std. Error	LogLikelihood	3632.38		Estimate	Std. Error	LogLikelihood	3639.9
ω	0.00	0.00	AIC	-7.19	ω	0.00	0.00	AIC	-7.25	ω	0.00	0.00	AIC	-7.27
α_1	0.14	0.02	BIC	-7.17	α_1	0.16	0.02	BIC	-7.23	α_1	0.16	0.02	BIC	-7.25
β_1	0.82	0.02			β_1	0.83	0.02			β_1	0.82	0.02		
γ_1	1.00	0.25			γ_1	1.00	0.12			γ_1	1.00	0.14		
	csGARCH-norm(1,1)		**Statistics**			**csGARCH-std(1,1)**		**Statistics**			**csGARCH-ged(1,1)**		**Statistics**	
	Estimate	Std. Error	LogLikelihood	3557.529		Estimate	Std. Error	LogLikelihood	3595.36		Estimate	Std. Error	LogLikelihood	3610.36
ω	0.00	0.00	AIC	-7.11	ω	0.00	0.00	AIC	-7.18	ω	0.00	0.00	AIC	-7.21
α_1	0.16	0.71	BIC	-7.08	α_1	0.15	0.40	BIC	-7.15	α_1	0.17	0.02	BIC	-7.18
β_1	0.66	0.75			β_1	0.71	0.62			β_1	0.68	0.07		
ρ	0.99	0.00			ρ	1.00	0.02			ρ	1.00	0.00		
ϕ	0.02	0.83			ϕ	0.03	0.10			ϕ	0.03	0.06		

5.2 Realized volatility models and estimates

In order to gain preliminary insights into the HAR model regression coefficients and performance, various in-sample regressions have been run and summarized in table 4. The period of the in-sample regression coincides with the period that will be estimated out-of-sample and on a rolling window. Similar to Corsi (2009), the HARRV models show that for equity the estimate for tomorrow's realized variance crucially and significantly depends past lags of average realized variance. The models show that this effect is even present for long lags of realized variance, as measured by the β_3 coefficient. Interestingly, this paper finds that in terms of size, the dependence on past information even seems to be increasing in time, whereas Corsi (2009) finds that the reverse holds for the S&P500. The HARRVJ model shows mixed results. When the model is estimated on S&P500 data, the jump component, as measured by β_4 is highly significant, whereas the opposite holds when the AEX index data is employed. Nevertheless, the HARRVJ model also illustrates that past average realized variance is a highly significant in estimating future variance. The SHAR models also provide interesting insights into dissecting variance estimates. The interesting feature where 1 day past realized volatility is estimated by two regression coefficients, illustrates that the findings are consistent with a leverage effect. Consistent with Patton & Sheppard (2015), this paper finds that in each case, the β_1^- coefficient is significant and higher than its counterpart, indicative that past negative returns induce a higher degree of future estimated volatility.

Put more concisely, for all models it holds that the long memory property of the HAR models is crucial in accurately modeling realized volatility. Additionally, the models' independent parameters are jointly highly significant as measured by the F-statistic and the corresponding p-values. Interestingly, the models perform better in modeling AEX variance as compared to the S&P500. Further statistical back testing in terms of VaR will indicate how this translates to a risk management perspective.

Table 4: In-sample regressions have been run for both indexes and for all covered realized volatility models. The period covered is the 1000 observation out-of-sample period and ranges from 12-11-2013 to 11-09-2017. ***,** and * respectively represent a significance level of 1%, 5% and 10%. Robust standard errors are in the parentheses below the estimated regression coefficients.

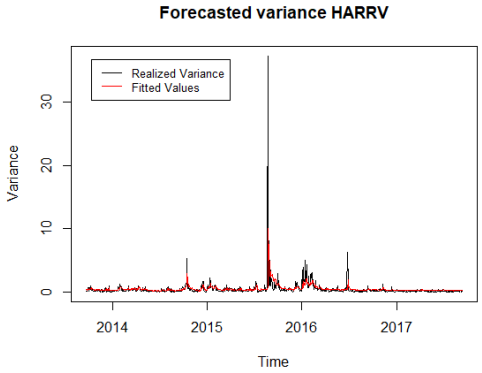
	HARRV	**HARRVJ**	**SHAR**	**HARRV**	**HARRVJ**	**SHAR**
		S&P500			AEX	
C	0.169	0.115	0.177	0.155	0.158	0.158
	(0.051)**	(0.052)*	(0.052)***	(0.049)**	(0.049)**	(0.049)**
β_1	0.223	0.718	0.261	0.205	0.217	0.278
	(0.037)***	(0.103)***	(0.075)***	(0.036)***	(0.041)***	(0.067)***
β_2	0.207	0.159	0.219	0.275	0.271	0.298
	(0.067)**	(0.067)*	(0.088)*	(0.067)***	(0.068)***	(0.078)***
β_3	0.208	0.177		0.299	0.302	
	(0.088)*	(0.087)*		(0.077)***	(0.077)***	
β_1^+			0.037			0.166
			(0.115)			(0.050)**
β_1^-			0.215			0.222
			(0.037)***			(0.040)**
β_4		-1.140			-0.187	
		(0.222)***			(0.293)	
R^2	0.138	0.160	0.141	0.220	0.220	0.221
adjusted R^2	0.136	0.157	0.137	0.218	0.217	0.218
F-stat	53.19***	47.52***	40.7***	93.7***	70.34***	70.63***
MSE	1.555	1.515	1.551	0.929	0.929	0.928

Due to the fact that the realized volatility models are estimated on a 1000 day rolling window with a refit every day, several thousand regressions are run in order to estimate the future volatility. Table 5 shows the mean square error for each model in the out-of-sample rolling window.

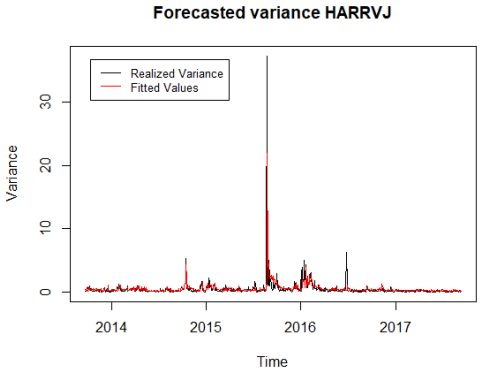
Table 5: This table shows the out-of-sample mean square errors of the regression models that are estimated on a 1000 day rolling window from 12-11-2013 to 11-09-2017.

	HARRV	**HARRVJ**	**SHAR**	**HARRV**	**HARRVJ**	**SHAR**
		S&P500			AEX	
MSE	1.610	1.756	1.651	0.989	1.006	1.039

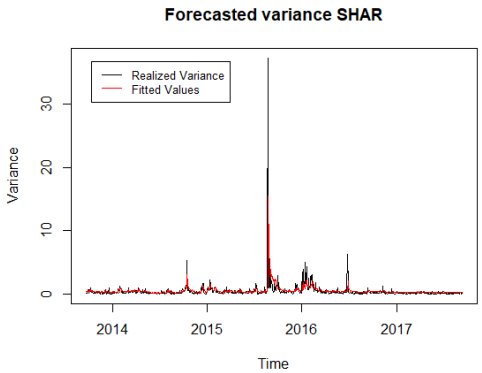
Lastly, figure 2 shows the out-of-sample plots of all employed realized volatility models. 2a-2c apply to the S&P500 and the last three plots apply to the AEX. The fitted values, as shown by the red line, are obtained by the 1-day ahead moving window regressions. The obtained fitted values are consequently converted to VaR estimates and its results are incorporated in the back tests, which is the topic of the next section.



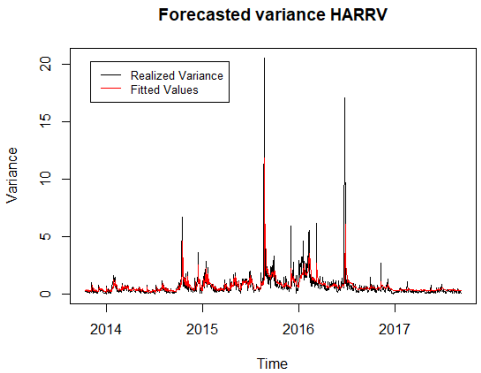
(a) Out-of-sample realized volatility S&P500



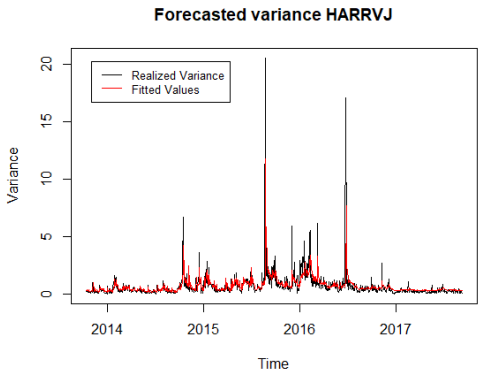
(b) Out-of-sample realized volatility S&P500



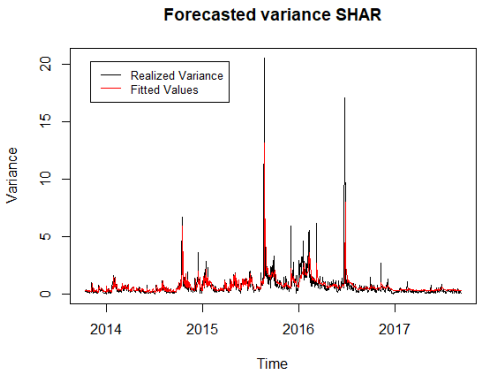
(c) Out-of-sample realized volatility S&P500



(d) Out-of-sample realized volatility AEX



(e) Out-of-sample realized volatility AEX



(f) Out-of-sample realized volatility AEX

Figure 2: Figure 2a-2f show the fitted values as obtained by moving window regressions plotted against the actual realized volatility.

5.3 VaR back test results

The generic S&P500 summary statistics of the conducted back test are shown in table 6 and it provides preliminary insights into the models' performance. For instance, due to the fact that the forecasting models in the top panel are evaluated at a 95% confidence interval, the models should have approximately 50 violations. Most of the models perform well and do not exceed the

threshold. From the first three classical models, the EWMA performs worst and its ratio of failures to the expected number of violations equals 1.14 meaning that the model fail up to 14% more than expected.

Dynamic VaR in the table refers to the created VaR model that consists of the GARCH-VaR models that are included in the most recent rolling MCS. Hence it is based on forecasting models that are ought to have superior forecasting ability. The Dynamic VaR average is the simple arithmetic average of the 60 GARCH-VaR models as described in appendix B. The two aforementioned VaR models, do not deviate much from the expected number of violations. In fact, the Dynamic VaR models have respectively 52 and 51 violations, indicative of accurate forecasting abilities. These numbers correspond to what is expected from 94.8% and 94.9% confidence level. The realized volatility models, which use high frequency data as input, also perform relatively well and at first sight do not appear to significantly underestimate or overestimate market risk.

When focusing on the bottom panel, a vastly different picture arises. In this case, all models underestimate market risk. As a matter of fact, the ratio of failures over expected ranges from 1.1 to 2.5. In other words, forecasting models exhibit up to 150% more violations of what can be reasonably expected given the 99% VaR level. The models that at first sight appear to suffer most from failing to correctly account for risk in the extreme tails are the classical models and the realized volatility models, with the exception of historical simulation. Nevertheless, both the dynamic VaR and the dynamic VaR average model continue to perform relatively well.

Table 6: This table shows the generic back test summary of the 1-day ahead VaR forecast. The period covered ranges from 12-11-2013 to 11-09-2017. Dynamic VaR refers to the created VaR model that consists of the GARCH-VaR models that are included in the most recent rolling MCS. The Dynamic VaR average is the simple arithmetic average of the 60 GARCH-VaR models as described in appendix B1. The results apply to the S&P500.

Model	**Level**	**Observed**	**Obs.**	**Failures**	**Expected**	**Ratio**
Variance Covariance	95%	95.1%	1000	49	50	0.98
Historical Simulation	95%	94.9%	1000	51	50	1.02
EWMA	95%	94.3%	1000	57	50	1.14
Dynamic VaR	95%	94.8%	1000	52	50	1.04
Dynamic VaR average	95%	94.9%	1000	51	50	1.02
HARRV	95%	94.7%	1000	53	50	1.06
HARRVJ	95%	94.6%	1000	54	50	1.08
SHAR	95%	94.5%	1000	55	50	1.1
Variance Covariance	99%	97.5%	1000	25	10	2.5
Historical Simulation	99%	98.9%	1000	11	10	1.1
EWMA	99%	97.5%	1000	25	10	2.5
Dynamic VaR	99%	98.7%	1000	13	10	1.3
Dynamic VaR average	99%	98.8%	1000	12	10	1.2
HARRV	99%	98.0%	1000	20	10	2
HARRVJ	99%	97.7%	1000	23	10	2.3
SHAR	99%	98.0%	1000	20	10	2

In figure 3, the visual performance of most models in table 6 is shown by plotting the 95% VaR

against the empirical returns. In this figure, the realized volatility models have been excluded in order to avoid clutter. The figure confirms that the variance covariance method is slow in adapting to changes in volatility. The historical simulation model shows a stepwise pattern, and also is slow in adapting to changes in volatility. The EWMA however, is responsive to changes in returns. Nevertheless, it appears to be somewhat lagging, which results in its poor performance. Analogous to the EWMA, the dynamic VaR models react very quickly to changes in volatility. In figure 5, they can hardly be distinguished from each other since their forecasts appear to be overlapping.

In order to save space, the 99% VaR forecasts and empirical returns of the S&P500 are plotted in figure 7 in the appendix. A surprising difference in this plot is the fact that the dynamic model and the dynamic average model estimates tend to diverge between January 2015 an July 2015. Nevertheless, visual inspection points out that in retrospect, this divergence cannot empirically be justified. During this period, volatility is not excessive and the dynamic model average appears to be extremely conservative during this time frame. In this model, the rolling MCS appears to correctly exclude models with inferior forecasting ability, potentially leading to improved risk management applications.

Lastly, figure 4 allows for visual inspection of the VaR violations on the 99% VaR confidence level for the high frequency HAR models. The 95% chart has been omitted in this paper, due to the fact that excessive violations of VaR tend to occur at higher confidence levels. The figure indicates that the out-of-sample VaR forecasts of the high frequency models tend to overlap throughout the testing period. Additionally, the graph reveals that violations tend to be clustered for all HAR models. Periods of many violations seem to follow periods of relatively few violations.

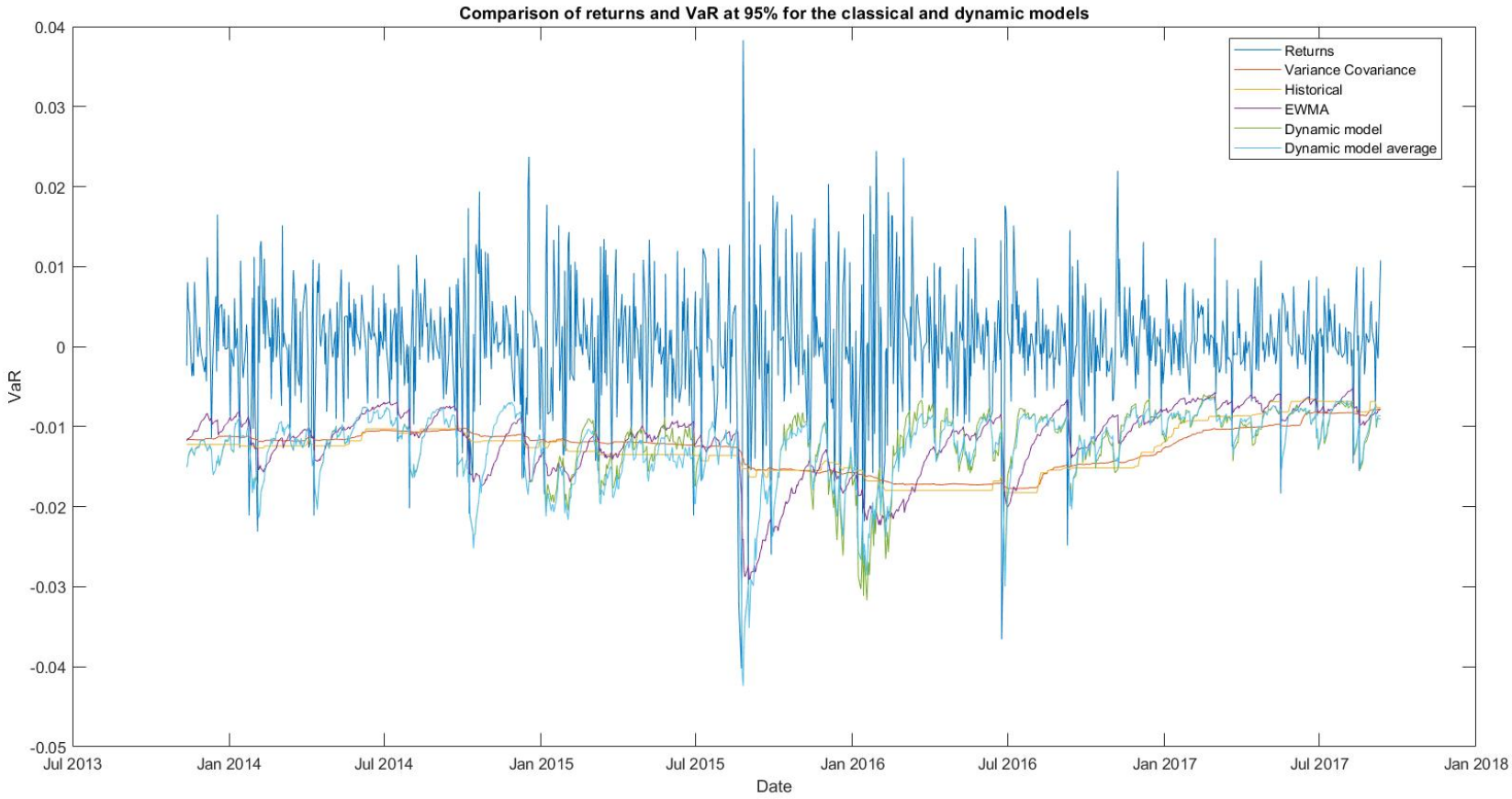


Figure 3: This figure shows the S&P500 returns as well as the 1-day ahead 95% VaR forecast of all considered models from 12-11-2013 to 11-09-2017.

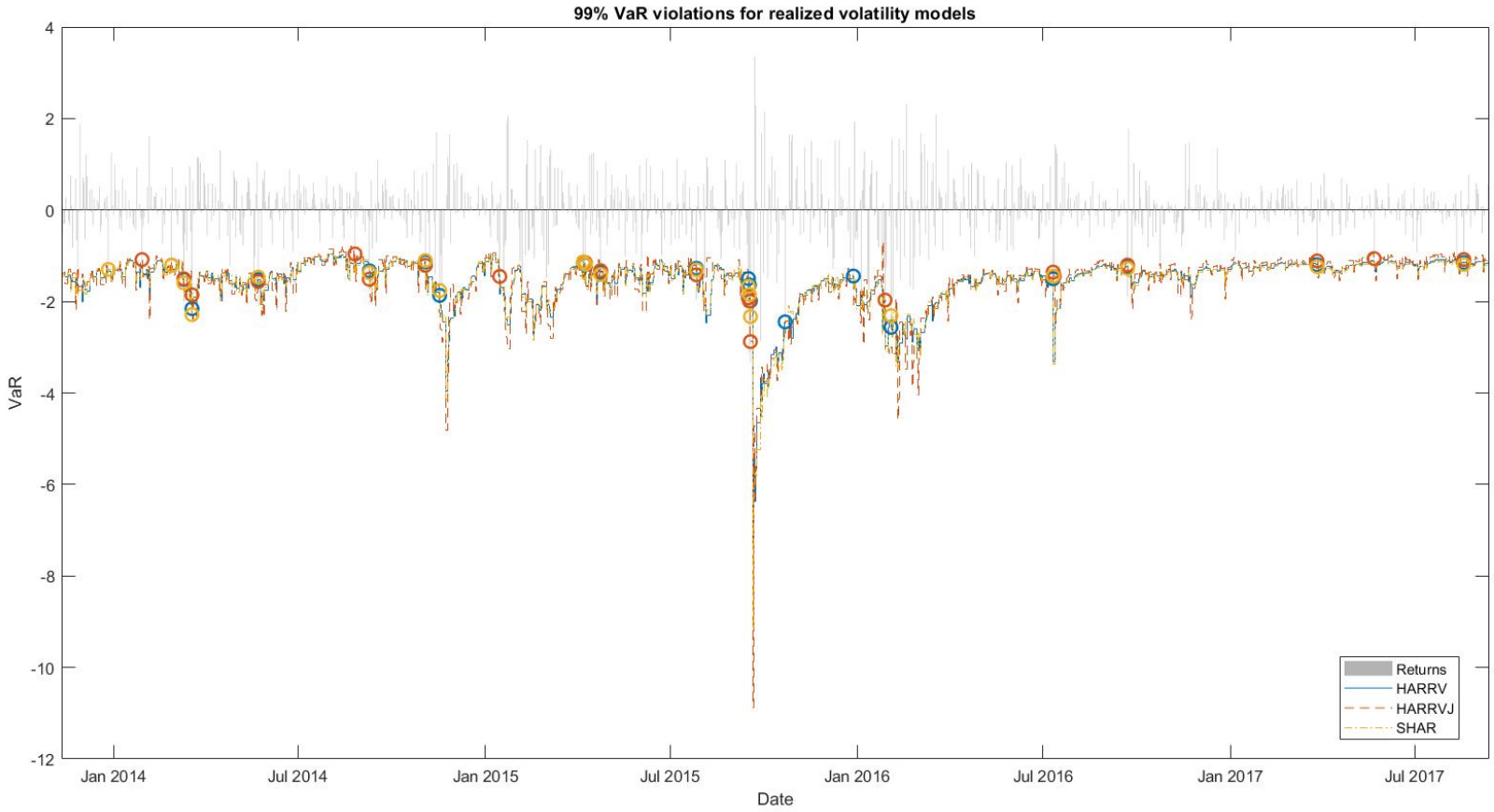


Figure 4: This figure shows the S&P500 returns as well as the 1-day ahead 99% VaR forecast of all realized volatility models from 12-11-2013 to 11-09-2017.

Table 7: This table summarizes the full out-of-sample back test results that were run on the S&P500 at both a 95% and 99% VaR level for 1000 observations. T.L. Cum Prob. stands for the Traffic Light cumulative probability. CC L.R. and MCC. L.R. respectively stand for the Conditional Coverage test and the Mixed Conditional Coverage Likelihood Ratio. Panel c and d show the final results where each of the tests (except the TL) is evaluated with an alpha of 5%.

Test	**Normal**	**Historical**	**EWMA**	**Dynamic**	**Dynamic average**	**HARRV**	**HARRVJ**	**SHAR**
				VaR level: 95%				
panel a		**Classical**			**Dynamic**		**Realized Volatility**	
T.L. Cum. Prob.	0.48	0.59	0.86	0.65	0.59	0.70	0.75	0.79
Binomial test Z	-0.15	0.15	1.02	0.29	0.15	0.44	0.58	0.73
P-value	0.44	0.44	0.15	0.39	0.44	0.33	0.28	0.23
CC L.R.	4.37	1.98	0.18	0.03	0.06	0.01	0.41	0.32
P-value	0.04	0.16	0.67	0.85	0.80	0.91	0.52	0.57
MCC L.R.	4.39	2.00	1.17	0.12	0.09	0.20	0.73	0.83
P-value	0.11	0.37	0.56	0.94	0.96	0.90	0.69	0.66
panel b				**VaR level: 99%**				
T.L. Cum. Prob.	1.00	0.70	1.00	0.87	0.79	1.00	1.00	1.00
Binomial test Z	4.77	0.32	4.77	0.95	0.64	3.18	4.13	3.18
P-value	0.00	0.38	0.00	0.17	0.26	0.00	0.00	0.00
CC L.R.	2.06	8.15	2.06	2.00	2.29	3.51	6.04	3.51
P-value	0.15	0.00	0.15	0.16	0.13	0.06	0.01	0.06
MCC L.R.	18.10	8.25	18.10	2.83	2.67	11.33	18.52	11.33
P-value	0.00	0.02	0.00	0.24	0.26	0.00	0.00	0.00
panel c				**Final results: VaR level: 95%**				
T.L.	green	green	green	green	green	green	green	green
Binomial test	accept	accept	accept	accept	accept	accept	accept	accept
CC	reject	accept	accept	accept	accept	accept	accept	accept
MCC	accept	accept	accept	accept	accept	accept	accept	accept
panel d				**Final results: VaR level: 99%**				
T.L.	red	green	red	green	green	yellow	yellow	yellow
Binomial test	reject	accept	reject	accept	accept	reject	reject	reject
CC	accept	reject	accept	accept	accept	accept	reject	accept
MCC	reject	reject	reject	accept	accept	reject	reject	reject

The formal back testing results are summarized in table 7. Panel a and b show the obtained statistics as well as the corresponding p-values for the S&P500 at both the 95 and 99% VaR. Panel c and d show the final results of the conducted tests. As mentioned in the methodology section, the null hypotheses of the binomial, conditional coverage and mixed conditional coverage have been specified such that a 'pass' in the table implies that the model meets the criteria of being adequate and an alpha of 5% is utilized as a threshold. The binomial test is a two-sided test, such that a p-value ≤ 0.025 would classify as the rejection area.

As expected from the previously obtained generic test results, the fact that all models in panel a pass the traffic light test and the binomial test is not a surprise. In the worst case of the EWMA, the cumulative probability of 57 violations given the threshold equals 86% and is well below the 95% level. As for the conditional coverage test, all models except for the normal method pass the test. This model just fails the test by falling slightly below the alpha of 0.05 and evidence suggests that violations of the model are not independent on consecutive days. Nevertheless, when evaluated using the mixed conditional coverage test, the normal method, along with all other models, pass the test.

More interestingly, evaluating the models' performance using a 99% VaR magnifies the differences in forecasting accuracy of the considered models and allows for more insights into the modeling of VaR in the extreme tails. When focusing in panel b on the Traffic Light test, a dispersion of accuracy become clearly visible. According to the Basel committee's criteria, the Normal and EWMA method would be classified as red which is the worst possible category and this classification is only assigned when the cumulative probability exceeds 0.9999. In other words, these models significantly underestimate the market risk. To a somewhat lesser extent, the same holds for all realized volatility models. Nevertheless, both dynamic VaR models pass the traffic light test and are considered adequate according to this criterion. The results of the binomial test are exactly in line with aforementioned findings. When analyzing the dependence of violations, the p-values indicate that the dynamic VaR models perform the best. The classical and realized volatility models show mixed results and some classify as adequate. Nevertheless, all the classical and realized volatility models that just pass the CC test subsequently fail the the mixed conditional coverage and that the results are highly significant.

In order to correctly account for robustness of the results, the exact same analysis is also performed for the AEX index. The generic summary statistics are shown in table 8 and the findings appear to be consistent with the results for the S&P500. Once again, they reveal that when higher degrees of confidence are required, the models have a greater tendency to fail, as measured per the ratio column. In the top panel, the subgroup of realized volatility models overall appears to perform best. The dynamic models on the other hand underestimate market risk, and this is most pronounced in the dynamic VaR average model. The binomial test will eventually indicate whether a ratio of 0.78 can be classified as a significant overestimation of market risk.

Table 8: This table shows the generic back test summary of the 1-day ahead VaR forecast. The period covered ranges from 12-11-2013 to 11-09-2017. Dynamic VaR refers to the created VaR model that consists of the GARCH-VaR models that are included in the most recent rolling MCS. The Dynamic VaR average is the simple arithmetic average of the 60 GARCH-VaR models as described in appendix B1. The results apply to the AEX index

Model	**Level**	**Observed**	**Obs.**	**Failures**	**Expected**	**Ratio**
Variance Covariance	95%	95.4%	1000	46	50	0.92
Historical Simulation	95%	94.8%	1000	52	50	1.04
EWMA	95%	94.3%	1000	57	50	1.14
Dynamic VaR	95%	95.3%	1000	47	50	0.94
Dynamic VaR average	95%	96.1%	1000	39	50	0.78
HARRV	95%	94.8%	1000	52	50	1.04
HARRVJ	95%	94.8%	1000	52	50	1.04
SHAR	95%	95.0%	1000	50	50	1
Variance Covariance	99%	98.0%	1000	20	10	2
Historical Simulation	99%	98.7%	1000	13	10	1.3
EWMA	99%	97.8%	1000	22	10	2.2
Dynamic VaR	99%	98.4%	1000	16	10	1.6
Dynamic VaR average	99%	99.2%	1000	8	10	0.8
HARRV	99%	98.4%	1000	16	10	1.6
HARRVJ	99%	98.3%	1000	17	10	1.7
SHAR	99%	98.4%	1000	16	10	1.6

In the bottom panel, a wide dispersion of accurate 1-day ahead VaR forecasting is exhibited by the models, since the ratios range from 0.8 to 2.2. Interestingly, the historical simulation method performs, despite its simplistic calculation, quite well at both VaR levels. The dynamic VaR models show an interesting pattern, since at the 99% VaR, the dynamic model underestimates market risk, but the dynamic VaR average overestimates market risk. Lastly, analogous to the S&P500 case, the realized volatility models overestimate market risk, albeit to a lesser extent.

Table 9: This table summarizes the full out-of-sample back test results that were run on the AEX index at both a 95% and 99% VaR level for 1000 observations. T.L. Cum Prob. stands for the Traffic Light cumulative probability. CC L.R. and MCC. L.R. respectively stand for the Conditional Coverage test and the Mixed Conditional Coverage Likelihood Ratio. Panel c and d show the final results where each of the tests (except the TL) is evaluated with an alpha of 5%.

Test	**Normal**	**Historical**	**EWMA**	**Dynamic**	**Dynamic average**	**HARRV**	**HARRVJ**	**SHAR**
				VaR level: 95%				
panel a		**Classical**			**Dynamic**		**Realized Volatility**	
T.L. Cum. Prob.	0.31	0.65	0.86	0.37	0.06	0.65	0.65	0.54
Binomial test Z	-0.58	0.29	1.02	-0.44	-1.60	0.29	0.29	0.00
P-value	0.28	0.39	0.15	0.33	0.06	0.39	0.39	0.50
CC L.R.	3.22	3.43	8.14	0.28	1.23	1.77	0.61	0.10
P-value	0.07	0.06	0.00	0.60	0.27	0.183	0.44	0.75
MCC L.R.	3.57	3.51	9.13	0.47	3.98	1.86	0.69	0.10
P-value	0.17	0.17	0.01	0.79	0.14	0.40	0.71	0.95
panel b					**VaR level: 99%**			
T.L. Cum. Prob.	1.00	0.87	1.00	0.97	0.33	0.97	0.99	0.97
Binomial test Z	3.18	0.95	3.81	1.91	-0.64	1.91	2.22	1.91
P-value	0.00	0.17	0.00	0.03	0.26	0.03	0.01	0.03
CC L.R.	3.51	2.00	0.44	0.52	0.13	0.52	0.59	0.52
P-value	0.06	0.16	0.51	0.47	0.72	0.47	0.44	0.47
MCC L.R.	11.33	2.83	11.28	3.60	0.56	3.60	4.68	3.60
P-value	0.00	0.24	0.00	0.17	0.75	0.17	0.10	0.17
panel c					**Final results: VaR level: 95%**			
T.L.	green	green	green	green	green	green	green	green
Binomial test	accept	accept	accept	accept	accept	accept	accept	accept
CC	accept	accept	reject	accept	accept	accept	accept	accept
MCC	accept	accept	reject	accept	accept	accept	accept	accept
panel d					**Final results: VaR level: 99%**			
T.L.	yellow	green	yellow	yellow	green	yellow	yellow	yellow
Binomial test	reject	accept	reject	accept	accept	accept	reject	accept
CC	accept	accept	accept	accept	accept	accept	accept	accept
MCC	reject	accept	reject	accept	accept	accept	accept	accept

The formal back testing results of the AEX index are summarized in table 9. Panel a shows that none of the models has significantly more violations than can be attributable to chance. Interestingly, despite the fact that the Dynamic average model appears to underestimate market risk, it still passes the binomial test. Only in 6% of the cases, a more extreme deviation from the threshold is expected. Moreover, due to the fact that the traffic light does not consider underestimations of risk, the traffic light test also classifies it as 'green'. Moreover, all other models are also classified as such. When analyzing the conditional coverage test results, the traditional models perform relatively poorly as measured by the corresponding p-values. Nevertheless, the normal and historical method still manage to pass the test. Once again, the estimates of the dynamic models are not clustered. The same applies to the realized volatility models. Lastly, all models except for EWMA perform well as measured by the joint MCC test.

More interestingly, panel b successfully highlights the points where the forecasting ability of the different analyzed models diverges. As can be inferred from the generic back test results in table 8, many models fail the unconditional coverage tests. For instance, all classical models are classified as 'yellow', except for the historical method. In addition, the dynamic model classifies as 'yellow', whereas the dynamic average model classifies as 'green'. This finding is somewhat surprising, given the fact that the dynamic model's out-of-sample forecasts are based on models that are ought to have superior forecasting ability. All realized volatility models are classified as 'yellow'. The findings of the Traffic Light test are in line with the binomial testing results, where a large number of models fail to accurately estimate the number of VaR violations. Nevertheless, as a group, the dynamic models pass the binomial test, whereas the classical and realized volatility models either fail or just pass the test. Moreover, the violations of the models again do not appear to be clustered in consecutive days as all models are not rejected by the conditional coverage test. Nevertheless, when the joint MCC test is performed, the classical models perform poorly. Both the EWMA as well as the normal method are considered inadequate.

5.4 Robustness analysis: MCS composition

As explained in the methodology section, the Model Confidence Set is constructed 10 times per equity index and per confidence level by using the past 1000 out-of-sample observations. Subsequently, the included models are used for calculating the dynamic VaR for the coming 100 periods, and this process is repeated. In total, the MCS is constructed 40 times, and table 10 and 11 dissect the excluded models based on three characteristics in order to provide more insights into the models' forecasting abilities. This robustness analysis will analyze the composition in depth. The best case scenario is when a single model remains and all other 59 models are excluded (Bernardi & Catania, 2015). Nevertheless, Bernardi & Catania (2015) further note that when this is not the case, consequences are mitigated by pooling the estimates, since models are subject to various levels of misspecification and observations are subject to structural breaks.

Table 10: This table shows the decomposition of the models that are excluded in the rolling Model Confidence Sets. The results apply to the S&P500 VaR estimates at a 95% confidence level. Moreover, the relative proportion of excluded model, distribution and lag order is provided. Due to rounding, the percentages may not add up to 100%.

	Number of Model Confidence Set									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Model										
GARCH	0	0	0	0	0	11	11	4	5	2
%	0	0	0	0	0	50	50	21	25	13
eGARCH	3	3	2	2	3	2	2	3	3	3
%	100	100	100	100	100	9	9	16	15	19
cGARCH	0	0	0	0	0	9	9	12	12	11
%	0	0	0	0	0	41	41	63	60	69
NAGARCH	0	0	0	0	0	0	0	0	0	0
%	0	0	0	0	0	0	0	0	0	0
TGARCH	0	0	0	0	0	0	0	0	0	0
%	0	0	0	0	0	0	0	0	0	0
Distribution										
normal	0	0	0	0	0	7	6	4	4	4
%	0	0	0	0	0	32	27	21	20	25
student T	0	0	0	0	0	7	7	8	7	4
%	0	0	0	0	0	32	32	42	35	25
generalized error	3	3	2	2	3	8	9	7	9	8
%	100	100	100	100	100	36	41	37	45	50
Lag length										
(1,1)	1	1	0	0	0	6	6	4	5	4
%	33	33	0	0	0	27	27	21	25	25
(1,2)	1	0	0	0	1	7	7	5	6	5
%	33	0	0	0	33	32	32	26	30	31
(2,1)	0	1	1	1	1	7	6	5	4	4
%	0	33	50	50	33	32	27	26	20	25
(2,2)	1	1	1	1	1	2	3	5	5	3
%	33	33	50	50	33	9	14	26	25	19
Sum of deleted models	**3**	**3**	**2**	**2**	**3**	**22**	**22**	**19**	**20**	**16**

Table 10 shows the 95% VaR MCS for the S&P500. The table clearly shows that no single model is ever superior to all other models at any given point in time. As a matter of fact, during the construction of the first 5 MCS, the econometric procedure points out that as much as 57 of the remaining models are ought to have equal predictive power given a 70% confidence level. Nevertheless, during the last 5 MCS constructions, the procedure consistently excludes approximately one third of the all models, indicative that predictive abilities of the models is subject to time and not constant. Moreover, table 10 shows that when dissecting on well performing GARCH models, the NAGARCH and TGARCH models perform extremely well, and as a matter of fact, they are never excluded and are ought to have superior predictive ability at the used 70% confidence level. The remaining GARCH models are ought inferior in certain cases. Another interesting finding is that the normal distribution appears relatively appropriate in describing the distribution of the innovation term. It accounts for a relatively low percentage of the omitted models. The same applies to the student's t-distribution, albeit to a lesser degree. The generalized

error distribution, however, appears inadequate and does not fare well.

Table 13 in appendix E lists the models that are ought inferior at the 99% VaR confidence level. The results are consistent with table 10 and its shows that the eGARCH, GARCH and component GARCH models are excluded solely. Most noticeably, employing an eGARCH model in combination with an assumed generalized error distribution results in models that according to the provided loss function and MCS do not model future conditional VaR estimates appropriately.

Table 11: This table shows the decomposition of the models that are excluded in the rolling Model Confidence Sets. The results apply to the AEX index VaR estimates at a 95% confidence level. Moreover, the relative proportion of excluded model, distribution and lag order is provided. Due to rounding, the percentages may not add up to 100%.

	Number of Model Confidence Set									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Model										
GARCH	12	10	11	9	10	1	5	8	3	6
%	55	45	50	41	45	7	36	73	43	60
eGARCH	1	3	2	5	3	3	3	3	3	2
%	5	14	9	23	14	21	21	27	43	20
cGARCH	9	9	9	8	9	10	6	0	1	2
%	41	41	41	36	41	71	43	0	14	20
NAGARCH	0	0	0	0	0	0	0	0	0	0
%	0	0	0	0	0	0	0	0	0	0
TGARCH	0	0	0	0	0	0	0	0	0	0
%	0	0	0	0	0	0	0	0	0	0
Distribution										
normal	7	9	7	8	8	3	2	2	2	3
%	32	41	32	36	36	21	14	18	29	30
student T	7	6	7	7	6	4	5	4	1	2
%	32	27	32	32	27	29	36	36	14	20
generalized error	8	7	8	7	8	7	7	5	4	5
%	36	32	36	32	36	50	50	45	57	50
Lag length										
(1,1)	7	8	7	9	8	3	2	2	1	0
%	32	36	32	41	36	21	14	18	14	0
(1,2)	6	7	7	8	7	4	4	4	1	5
%	27	32	32	36	32	29	29	36	14	50
(2,1)	6	6	6	5	6	2	1	1	0	0
%	27	27	27	23	27	14	7	9	0	0
(2,2)	3	1	2	0	1	5	7	4	5	5
%	14	5	9	0	5	36	50	36	71	50
Sum of deleted models	**22**	**22**	**22**	**22**	**22**	**14**	**14**	**11**	**7**	**10**

Table 11 shows the decomposition of the omitted VaR forecasting models in the construction of the MCS that apply to the AEX index. This table also shows that no single model is ever superior to all other models at any given point in time. Moreover, it can be inferred that again the NAGARCH and TGARCH specifications in any combination perform extremely well, and in fact, are classified as models with superior predictive ability with a high degree of certainty. On the other hand, the remaining GARCH specifications, and most noticeably the standard GARCH model, account for the largest share of omitted models. The fact that the regular GARCH model is excluded

relatively often is consistent with the findings of Hansen & Lunde (2005). In their extensive research of 330 different models, they find that GARCH models in generally are outperformed by more sophisticated models in modeling conditional volatility estimates for equity. In contrast to the model exclusions of the AEX index, all assumed distributions seem to account for a relatively high share to omitted models. However, the generalized error distribution does stand out, albeit less pronounced than when applied to the S&P500. When considering the employed lag length, including more parameters as measured by the models' p and q does not seem to contribute nor worsen to the forecasting performance.

Table 14 in the appendix summarizes the excluded models at the 99% VaR level for the AEX index. Once again, the MCS algorithm excludes fewer models when risk in the extreme tails of the distribution is to be modeled. The table substantiates earlier findings that especially eGARCH models with an assumed generalized error distribution generate inferior VaR estimates. Moreover the TGARCH and NAGARCH models continue to contribute to enhanced VaR forecasting.

6 Conclusion

The purpose of this paper is to investigate whether a dynamic Value at Risk model and high frequency realized volatility models can improve the accuracy of 1-day ahead VaR forecasting beyond the performance of frequently used models. As such, this paper constructs 60 conditional volatility forecasting models. Several extensions of the GARCH model are included, such as the nonlinear and asymmetric models. Moreover, several return distributions are assumed for the error term, in order to allow for more flexible modeling in the tails. A rolling Model Confidence Set is subsequently constructed, ensuring that only models with superior out-of-sample forecasting performance remain. A model averaging technique is applied to the remaining superior models, which generates the dynamic VaR forecasts. Moreover, this paper exploits high frequency data in order to forecast VaR. Several extensions of the HAR realized volatility model are included in this paper.

In a series of extensive back tests, the results provide evidence in favour of the fact that all considered models are relatively adequate in modeling VaR for the S&P500 and the AEX at the 95% confidence level. Nevertheless, a clear divergence in performance occurs when the VaR is required at higher degrees of confidence. The dynamic VaR models produce forecasts which are superior to traditional models and HAR models. The result hold for the 95% VaR, but are even more pronounced for the 99% VaR. The traditional and HAR models severely underestimate risk at higher confidence levels, whereas the applied dynamic VaR correctly accounts for it. Moreover, the violations provided by the dynamic models have a greater tendency to be independent, as measured by the conditional coverage tests. This paper provides little evidence that the dynamic model produces estimates superior to the 'dynamic VaR average' model. In short, due to the computational requirements of constructing a rolling MCS, implementing the simple dynamic average

model seems more feasible.

References

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, *89*(4), 701–720.

Angelidis, T., Benos, A., & Degiannakis, S. (2004). The use of garch models in var estimation. *Statistical methodology*, *1*(1-2), 105–128.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468.

Bera, A. K., & Higgins, M. L. (1993). Arch models: properties, estimation and testing. *Journal of economic surveys*, *7*(4), 305–366.

Bernardi, M., & Catania, L. (2015). The model confidence set package for r.

Bernardi, M., Catania, L., & Petrella, L. (2014). Are news important to predict large losses? *arXiv preprint arXiv:1410.6898*.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542–547.

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*(1), 1–18.

Brooks, C. (2014). *Introductory econometrics for finance*. Cambridge university press.

Campbell, S. D. (2006). A review of backtesting and backtesting procedures. *The Journal of Risk*, *9*(2), 1.

Chen, X., Ghysels, E., & Wang, F. (2011). Hybrid garch models and intra-daily return periodicity. *Journal of Time Series Econometrics*, *3*(1).

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

Christoffersen, P. F. (2012). *Elements of financial risk management*. Academic Press.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, *27*(1-3), 46–78.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, *1*(1), 83–106.

Dutta, A. (2014). Modelling volatility: Symmetric or asymmetric garch models. *Journal of Statistics: Advances in Theory and Applications*, *12*(2), 99–108.

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

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance*, *48*(5), 1749–1778.

Giot, P., & Laurent, S. (2003a). Market risk in commodity markets: a var approach. *Energy Economics*, *25*(5), 435–457.

Giot, P., & Laurent, S. (2003b). Value-at-risk for long and short trading positions. *Journal of Applied Econometrics*, *18*(6), 641–663.

Giot, P., & Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and arch type models. *journal of Empirical Finance*, *11*(3), 379–398.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, *48*(5), 1779–1801.

González-Rivera, G., Lee, T.-H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting*, *20*(4), 629–645.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, *20*(7), 873–889.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hull, J. (2012). *Risk management and financial institutions,+ web site* (Vol. 733). John Wiley & Sons.

Jorion, P. (1996). Risk2: Measuring the risk in value at risk. *Financial analysts journal*, *52*(6), 47–56.

Jorion, P., et al. (2010). *Financial risk manager handbook: Frm part i* (Vol. 625). John Wiley & Sons.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models.

Lee, G., & Engle, R. (1993). A permanent and transitory component model of stock return volatility.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, *97*(3), 683–697.

Rabemananjara, R., & Zakoian, J.-M. (1993). Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics*, *8*(1), 31–49.

Samuels, J. D., & Sekkel, R. M. (2011). *Forecasting with large datasets: trimming predictors and forecast combination* (Tech. Rep.). Working paper.

So, M. K., & Philip, L. (2006). Empirical analysis of garch models in value at risk estimation. *Journal of International Financial Markets, Institutions and Money*, *16*(2), 180–197.

Stock, J. H., & Watson, M. W. (1998). *A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series* (Tech. Rep.). National Bureau of Economic Research.

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*(2), 293–335.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Venkataraman, S., et al. (1997). Value at risk for a mixture of normal distributions: the use of quasi-bayesian estimation techniques. *Economic Perspectives-Federal Reserve Bank of Chicago*, *21*, 2–13.

Appendix A

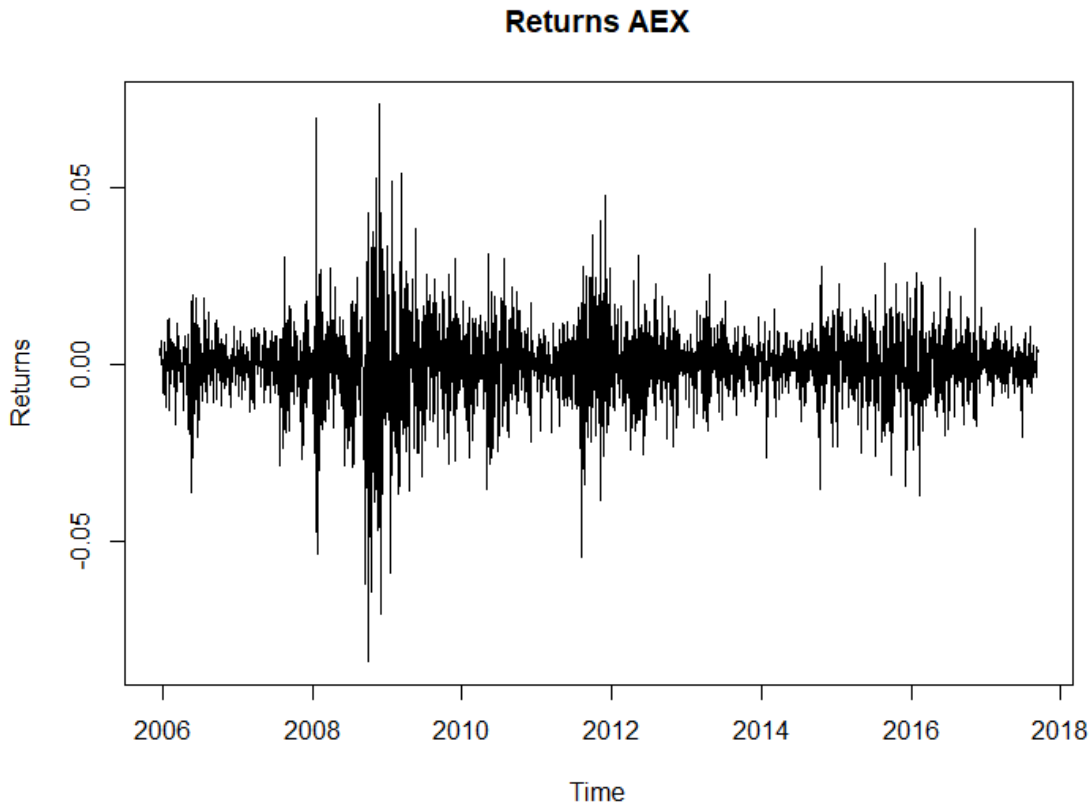
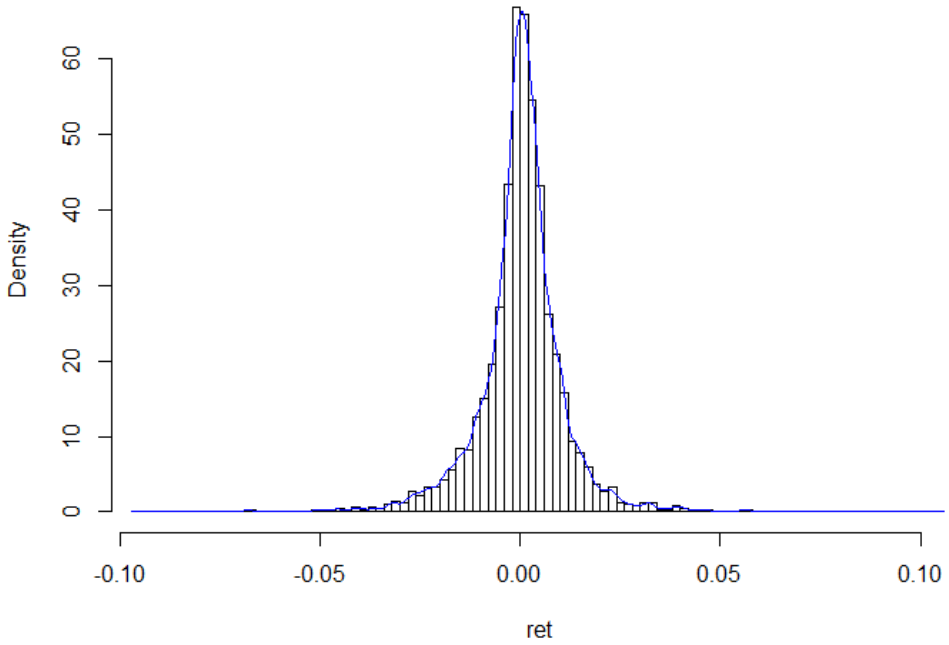
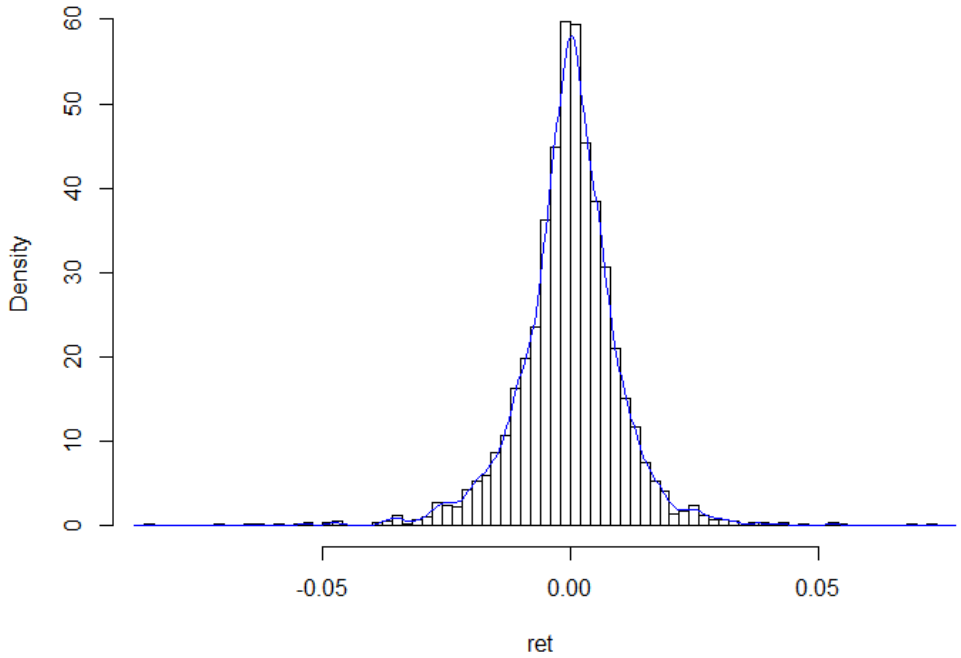


Figure 5: The continuously compounded returns from 14/03/2006 to 11/09/2017 of the AEX are plotted.



(a) Histogram S&P500



(b) Histogram AEX

Figure 6: The return distribution of the full data sample of the S&P500 and the AEX are plotted figure 6a and 6b respectively. A normal distribution is superimposed on the graphs in order to highlight deviations from the normality assumption.

Appendix B

Table 12: This table shows all the combinations of p,q, the model and the assumed innovation distribution that are analyzed throughout the paper. The formulas of the model names are specified in table 2. P and q refer to the amount of lags as described in that table and norm, std and ged refer respectively to the normal, the student and the generalized error distribution.

Model	**p**	**q**	**distribution**	**Model**	**p**	**q**	**distribution**
GARCH	1	1	norm	NAGARCH	1	1	norm
GARCH	1	2	norm	NAGARCH	1	2	norm
GARCH	2	1	norm	NAGARCH	2	1	norm
GARCH	2	2	norm	NAGARCH	2	2	norm
GARCH	1	1	std	NAGARCH	1	1	std
GARCH	1	2	std	NAGARCH	1	2	std
GARCH	2	1	std	NAGARCH	2	1	std
GARCH	2	2	std	NAGARCH	2	2	std
GARCH	1	1	ged	NAGARCH	1	1	ged
GARCH	1	2	ged	NAGARCH	1	2	ged
GARCH	2	1	ged	NAGARCH	2	1	ged
GARCH	2	2	ged	NAGARCH	2	2	ged
eGARCH	1	1	norm	TGARCH	1	1	norm
eGARCH	1	2	norm	TGARCH	1	2	norm
eGARCH	2	1	norm	TGARCH	2	1	norm
eGARCH	2	2	norm	TGARCH	2	2	norm
eGARCH	1	1	std	TGARCH	1	1	std
eGARCH	1	2	std	TGARCH	1	2	std
eGARCH	2	1	std	TGARCH	2	1	std
eGARCH	2	2	std	TGARCH	2	2	std
eGARCH	1	1	ged	TGARCH	1	1	ged
eGARCH	1	2	ged	TGARCH	1	2	ged
eGARCH	2	1	ged	TGARCH	2	1	ged
eGARCH	2	2	ged	TGARCH	2	2	ged
csGARCH	1	1	norm				
csGARCH	1	2	norm				
csGARCH	2	1	norm				
csGARCH	2	2	norm				
csGARCH	1	1	std				
csGARCH	1	2	std				
csGARCH	2	1	std				
csGARCH	2	2	std				
csGARCH	1	1	ged				
csGARCH	1	2	ged				
csGARCH	2	1	ged				
csGARCH	2	2	ged				

Appendix C

Figure 7: This figure shows the S&P500 returns as well as the 1-day ahead 99% VaR forecast of all classical and dynamic models from 12-11-2013 to 11-09-2017.

Appendix D

Table 13: This table shows the models that are excluded in the rolling Model Confidence Sets. The results apply to the VaR estimates at a 99% confidence level. Norm, std and ged respectively refer to the normal, standard t-, and generalized error distribution. Moreover, the numbers between brackets refer to p and q respectively as defined in table 2. Results apply to S&P500.

Excluded models in MCS				
MCS 1	**MCS 2**	**MCS 3**	**MCS 4**	**MCS 5**
eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH ged (2,1)	eGARCH ged (1,2)
eGARCH ged (1,2)	eGARCH ged (2,1)	eGARCH ged (2,1)	eGARCH ged (2,2)	eGARCH ged (2,1)
eGARCH ged (2,2)	eGARCH ged (2,2)	eGARCH ged (2,2)		eGARCH ged (2,2)
MCS 6	**MCS 7**	**MCS 8**	**MCS 9**	**MCS 10**
eGARCH ged (1,2)	GARCH norm (1,1)	GARCH norm (1,1)	GARCH norm (1,1)	eGARCH ged (1,2)
cGARCH norm (1,2)	cGARCH norm (1,1)	cGARCH norm (1,1)	cGARCH norm (1,1)	eGARCH ged (2,1)
cGARCH ged (1,2)	GARCH norm (1,2)	GARCH norm (1,2)	GARCH norm (1,2)	eGARCH ged (2,2)
eGARCH ged (2,1)	eGARCH ged (1,2)	eGARCH ged (1,2)	eGARCH ged (1,2)	
eGARCH ged (2,2)	cGARCH norm (1,2)	cGARCH norm (1,2)	eGARCH ged (2,1)	
GARCH norm (1,1)	eGARCH ged (2,1)	eGARCH ged (2,1)	eGARCH ged (2,2)	
cGARCH norm (1,1)	eGARCH ged (2,2)	eGARCH ged (2,2)		
cGARCH ged (1,1)				
GARCH norm (1,2)				
GARCH ged (1,2)				

Table 14: This table shows the models that are excluded in the rolling Model Confidence Sets. The results apply to the VaR estimates at a 99% confidence level. Norm, std and ged respectively refer to the normal, standard t-, and generalized error distribution. Moreover, the numbers between brackets refer to p and q respectively as defined in table 2. Results apply to the AEX index.

Excluded models in MCS				
MCS 1	**MCS 2**	**MCS 3**	**MCS 4**	**MCS 5**
eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH std (1,1)
eGARCH ged (2,2)	eGARCH ged (2,2)	eGARCH ged (1,2)	eGARCH ged (1,2)	eGARCH ged (1,1)
		eGARCH ged (2,2)	cGARCH norm (2,1)	cGARCH norm (1,1)
			eGARCH ged (2,2)	eGARCH ged (1,2)
			cGARCH norm (2,2)	cGARCH norm (1,2)
				cGARCH norm (2,1)
				eGARCH ged (2,2)
				cGARCH norm (2,2)
MCS 6	**MCS 7**	**MCS 8**	**MCS 9**	**MCS 10**
eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH ged (1,1)	eGARCH ged (1,2)
eGARCH ged (1,2)	eGARCH ged (1,2)	eGARCH ged (1,2)	eGARCH ged (1,2)	eGARCH ged (2,2)
cGARCH norm (2,1)	eGARCH ged (2,2)	eGARCH ged (2,2)	eGARCH ged (2,2)	
eGARCH ged (2,2)				
cGARCH norm (2,2)				