Bachelor thesis draft - comparison of GARCH-family models with realized volatility models in estimating volatility of time series

František Vlach, 66851139

March 2024

1 Abstract

TODO: Abstrakt

There should be some abstract in English
a taky česky

2 Introduction

Measuring and forecasting latent volatility have many important applications in many areas of finance including asset allocation, option pricing and risk management. (Brownlees and Gallo, 2009) For the past 40 years, several methods were successfully utilized for this purpose within the (G)ARCH framework, developed by Engle (1982) and further expanded by Bollerslev (1987), Engle and Bollerslev (1986), Zakoian (1994), Engle et al. (1987), Nelson (1991) and others. Alternative measurements based on different assumptions and different information sets have been in use for some time, some of them use historical variances, range or implied volatilities.

Proper introduction

Obecná ekonomická otázka - downside risk

The idea of using proxies of volatility obtained from intra-daily data sampled at high frequency has been proposed by Merton (1980), but it was only two decades later until databases containing detailed information of transactions in financial markets became available (Brownlees and Gallo, 2009) and were subject to studies by authors including Andersen and Bollerslev (1998), Andersen et al. (2001a), Barndorff-Nielsen and Shephard (2002a), Bollerslev et al. (2003).

Under suitable assumptions, these volatility proxies converge with increasing sampling frequency to the integrated variance, i. e. the integral of an instantaneous or spot volatility of an underlying continuous time process over a short period. An open question is how to forecast volatility on the basis of existing information and the relationship to the latent underlying process (e. g. with or without jumps). In theory, it is possible to construct ex-post measures of return variability with arbitrary precision. (Brownlees and Gallo, 2009)

Value at risk (VaR) is one of the most important measures of the market risk that has been widely used for financial risk management by institutions including banks, regulators and portfolio managers. (So and Yu, 2006)

Downside risk in general - economic question

VaR only a specific way of representing Downside risk

My methods only a specific way to estimate VaR

3 Value at risk

Value at risk (VaR) is a commonly used statistic for measuring potential risk of economic losses in financial markets (So and Yu, 2006). It is one of the most important measures of market risk and has been used by banks, regulators and and portfolio managers. The RiskMetrics model for measuring VaR has become a benchmark for measuring market risk since its development in 1994 by J. P. Morgan's risk management group. Other methods have also been developed such as those based on extreme value theory, as studied, among others, by Danielsson and De Vries (2000) and Ho et al. (2000), high frequency data, as studied by for example Beltratti and Morana (1999) or GARCH conditional moments, as studied by Wong and So (2003). The RiskMetrics model assumes that returns of a financial asset follow a conditional normal distribution with zero mean and variance being expressed as an exponentially weighted moving average (EWMA) of squared returns. These assumptions are problematic and introduce a significant loss of accuracy to the forecasted values. First, it is well documented that a return distribution usually has a heavier tails than normal distribution. Therefore, the normality assumption may generate a significant bias, mainly concerning the tail properties of the return distribution. Secondly, recent empirical studies, for example Ding et al. (1993), among others, show that return series may exhibit long memory or long-term dependence on market volatility, which was found to have a significant influence on derivative pricing and volatility forecasting (So and Yu, 2006).

VaR is defined as the maximum loss over a given time horizon at a given confidence level. It can be used to get a sense of the minimum amount that a financial institution is expected to lose with a small probability α over a given time horizon k. For example an $\alpha = 5\%$ 1-day VaR of \$10 million states that in $\frac{1}{20}$ days, a realized loss of at least \$10 million can be expected, or vice versa, in $\frac{19}{20}$ days, the maximum expected loss is \$10 million. Let P_t be the price of a financial asset on day t. A k-day VaR on day t is defined by

$$P\left(P_{t-k} - P_t \leq VaR\left(t, k, \alpha\right)\right) = 1 - \alpha$$

Given a distribution of return, VaR can be determined and expressed in terms of percentile q_α of the return distribution, as shown by Dowd (1998) and Jorion (2006). This implies that good VaR estimates can only be produced with accurate forecasts of the percentiles q_α, which realized on appropriate volatility modeling. Since this volatility is time-varying, we need to use appropriate econometric models to incorporate it.

Since 1998, banks with substantial trading activity have been required to set aside capital in order to insure for the case of extreme portfolio losses. The set-aside capital, also called the market risk capital requirement, is linked to a measure of portfolio risk. Currently, portfolio risk is measured with the use of its VaR, which is defined to be the loss which is expected to be exceeded only with $\alpha\%$ probability, i. e. only $\alpha\%$ of the time over a fixed time interval. Current regulatory framework requires that financial institutions use their own internal risk models to calculate and report a 1% value-at-risk, the VaR(0.01) over a 10-day horizon. The VaR is defined as

$$VaR_t(\alpha) = -F^{-1}(\alpha|\Omega_t)$$

where $F^{-1}(\cdot|\Omega_t)$ represents the quantile function of the profit and loss distribution which varies over time as market conditions and the portfolio's composition, represented by Ω_t change. Accurate means of examining whether the reported VaR represents an accurate measure of actual risk level are necessary since financial institutions use their own internal risk models to determine the specific level of VaR, based on which they adhere to risk-based capital requirements. (Campbell, 2005)

3.1 Estimating VaR

There are multiple ways of estimating VaR, based on different mathematical cosntructions. For estimating VaR, we need to define the corresponding quantile of the assumed distribution. There is empirical evidence showing that if we assume normality, the produced results are often weak. We can test the Jarque-Bera test to test the hypothesis that the stock returns follow a normal distribution. Since financial time series tend to have heavy tails , it is natural to use the Students's t-distribution instead. We can fit the number of degrees of freedom of this distribution automatically so that it fits our data the best.

reference?

reference

The Delta-normal approach in estimating VaR assumes normality of stock returns. VaR can be defined with the use of variance of returns as

$$VaR(\alpha) = \mu + \sigma \cdot N^{-1}(a)$$

where mu is the mean stock return, σ is the standard deviation of returns, a is the selected confidence level and N^{-1} is the inverse PDF function generating the corresponding quantile of a normal distribution given a.

What about version without mean? Maybe mention that mean is negliglible.

This original model is rarely used in practice today since the results of such a model are often very poor due to the assumptions of normality of returns and constant daily variance which are typically false.

In order to account for time-varying volatility, we can use conditional variance given as the output of one of our models. For this approach, VaR is expressed as

$$VaR(a) = \mu + \hat{\sigma}_{t|t-1} \cdot F^{-1}(a)$$

where $\hat{\sigma}_{t|t-1}$ is the conditional standard deviation given the information at $t-1$ and F^{-1} is the inverse PDF function of a t-distribution.

references

3.2 Statistical framework of VaR backtests

A variety of tests have been proposed since 1990's in order to measure the accuracy of a VaR model. Many of these focus on a particular transformation of the reported VaR and realized profit or loss. We may consider the event that the loss on a portfolio exceeds its reported VaR, that is, $VaR_t(\alpha)$. Denoting the daily profit or loss on the portfolio over a fixed time interval, we can define the hit function as follows:

$$I_{t+1}(\alpha) \begin{cases} 1, & \text{if } x_{t,t+1} \leq -VaR_t(\alpha) \\ 0, & \text{if } x_{t,t+1} > -VaR_t(\alpha) \end{cases}$$

Christoffersen (1998) reduces the problem of determining the accuracy of a VaR model to determining whether the hit sequence

$$[I_{t+1}(\alpha)]_{t=1}^{t=T}$$

satisfies two properties:

1. The unconditional coverage property states that the probability of realizing a loss in excess of the reported VaR must be precisely $\alpha\%$. If the losses in excess of the reported VaR occur more frequently, then it is a suggestion that the VaR measure systematically understates the portfolio's actual level of risk and vice versa, finding too few VaR violations may signal a systematic overstating of the risk level.

2. The idependence property places a restriction on the ways in which these violations may occur. Specificially, any two elements of the hit sequence must be independent from each other. This condition requires that the previous history of VaR violations must not convey any information about whether an additional VaR violation will occur. If previous VaR violations presage a future VaR violation, it suggests a general inadequacy in the reported VaR measure. An example of such inadequacy may be bunching, that is, the occurence of violations of VaR is cumulated together. This represents a violation of the independence property that signals a lack of responsiveness in the reported VaR measure as changing market risks fail to be fully incorporated into the reported VaR measure which makes successive runs of VaR violations more likely.

These two properties are separate and distinct and must be both satisfied by an accurate VaR model. Only hit sequences that satisfy both properties can be described as evidence of an accurate VaR model. The two properties of the hit sequence, $[I_{t+1}(\alpha)]_{t=1}^{t=T}$, are often combined into a single statement:

$$I_t(\alpha) \overset{i.i.d}{\sim} B(\alpha)$$

i. e. the hit sequence is identically and independently distributed as a Bernoulli random variable with probability α.

3.3 Tests of VaR accuracy

There is an intense academic debate on the validity of risk measures in general and VaR in particular (Dumitrescu et al., 2012). Since VaR is unobservable, we have to rely upon the testing of the violations to test its validity. Three main issues need to be addressed when evaluating VaR sequences: First, the power, or the specificity of the model. It plays a key role especially in small samples, as in 250 or 500 observations, i. e. 1-year or 2-years ahead. It has been shown VaR tests generally have lower power as the backtesting procedure is too optimistic in terms of rejecting the validity of a model. Second, the backtesting methodology has to be model-free. Third, estimation risk must be taken into account, i. e. testing procedures can successfully answer the question of VaR validity only by taking into account estimation error, as the risk of estimation error as the risk of estimation error present in the estimates of the parameters pollutes VaR forecasts. Conditional on allowing for these errors, we should observe neither under-rejecting or over-rejecting.

3.3.1 Unconditional coverage tests

The earliest proposed VaR backtests focus only on the first property, that is, unconditional coverage. These tests test only whether the reported VaR level is violated more or less than $\alpha\%$ of the time.

The commonly used proportion of failure (POF) test developed by Kupiec (1995), has a null hypothesis of simple the probability of an exception being equal to the significance level. The Kupiec tests statistic has the form

$$POF = 2 \cdot \left(\left(\frac{1-\hat{\alpha}}{1-\alpha} \right)^{T-I(\alpha)} \cdot \left(\frac{\hat{\alpha}}{\alpha} \right)^{I(\alpha)} \right)$$

$$\hat{\alpha} = \frac{1}{T} \cdot I(\alpha)$$

$$I(\alpha) = \sum_{t=1}^{T} I_t(\alpha)$$

where T is the sample size. The POF is assumed to have a $\chi^2(1)$ distribution. We can see that if the proportion of VaR violations is exactly equal to $\alpha\%$, then the POF test statistic is equal to zero. As the proportion differs from $\alpha\%$, the POF test statistic grows, indicating an evidence that the portfolio's underlying level of risk is either systematically underestimated or overestimated by the proposed VaR measure.

Other tests also exist to assess the unconditional coverage property of a given VaR model. One alternative is to simply base a test directly on the sample average of the number of VaR violations over a given time period, $\hat{\alpha}$. Under the assumption that the VaR under consideration is accurate, then a scaled version of $\hat{\alpha}$,

$$z = \frac{\sqrt{T} \cdot (\hat{\alpha} - \alpha)}{\sqrt{\alpha} \cdot (1-\alpha)}$$

has an approximate $N(0,1)$ distribution and since the exact finite distribution of z is known and so hypothesis tests can be conducted in exactly the same way that hypothesis tests are conducted in the case of Kupiec's POF statistic.

The tests of unconditional coverage, while useful in providing a benchmark for assessing the accuracy of a given VaR model, have two disadvantages: First, they are known to have difficulty to detect VaR measures which systematically under report risk. (Kupiec, 1995). Second, they focus exclusively on the unconditional coverage property of an adequate VaR measure and do not examine the extent to which the independence property is satisfied. Therefore, they may naturally fail to detect VaR measures that exhibit correct unconditional coverage but dependent VaR violations, which may result in losses that exceed the reported VaR in clusters or streaks, which may result in even more stress on a financial institution than large unexpected losses that occur somewhat more frequently than expected but are spread out over time.

It is safe to say that as dependent VaR violations signal a lack of responsiveness to changing market conditions and inadequate risk reporting, relying solely on unconditional coverage tests appears problematic. (Campbell, 2005)

3.3.2 Independence tests

Since the unconditional coverage tests fail to detect violations of the independence property of an accurate VaR measure, new tests have been developed to examine the independence property. An early test of this type is the Markov test developed by Christoffersen (1998), which examines whether or not the likelihood of a VaR violation depends on whether or not a VaR violation occurred on the previous day. Its null hypothesis is that the exceedances of VaR level are independently distributed over time. If the VaR measure accurately reflects the underlying portfolio risk then the change of violating today's VaR should be independent of whether or not yesterday's VaR was violated (Campbell, 2005). This test utilizes the fact that if VaR violations are completely independent then the amount of time that elapses between VaR violations should be independent of the amount of time from the previous violation. In this sense, the time between VaR violations should not exhibit any kind of duration dependence. Performing the test requires estimating a statistical model for the duration of time between violations by maximum likelihood using numerical methods. The test creates a 2×2 contingency table table which records violations of the institution's VaR on adjacent days. If the VaR measure accurately reflects the portfolio's risk then the proportion of violations that occur after a day in which no violation occurred. If these proportions differ greatly from each other, then the validity of the VaR measure comes under question. Christoffersen (2004) provide evidence that this test is more powerful than the original Christoffersen's test.

One main drawback of independence tests is that they all start from the assertion that any accurate VaR measure will result in a series of independent hits $[I_t(\alpha)]_{t=1}^{t=T}$. Accordingly, any test of the independence property must fully describe the way in which violations of the independence property arises, such as in the case of the Christoffersen's test, by allowing for the possibility that the change of violating tomorrow's VaR depends on whether or not yesterday's VaR was violated. However, there are many other ways in which the independence property may be violated, for example, the likelihood of violating tomorrow's VaR may depend on violating or not violating the VaR a week ago rather than yesterday. In such situation, the Markov tests will not be able to detect this type of independence property violation.

In statistical terms, the alternative hypothesis that the independence property is being tested against needs to be completely specified by any independence test. Intuitively, an independence test must describe the anomalies that it is looking for while examining whether or not the independence property is satisfied. Other types of violations will not be systematically detected by the given test. Therefore, independence tests can only detect inaccurate VaR measures to the extent that they are designed to identify violations of the independence property in ways that are likely to arise when internal risk models fail to provide accurate VaR measures. This information may come from a thorough understanding of when and how common risk models fail to accurately describe portfolio risk. Thus, even though these models may not perform the best in terms of changing market conditions, tests that examine the amount of clustering in VaR violations such as the Markov test may be useful in identifying inaccurate VaR models (Campbell, 2005).

3.3.3 Joint tests of unconditional coverage and Independence

Since an accurate VaR measure must exhibit both independence and unconditional coverage property, tests that examine these properties jointly provide an opportunity to detect VaR measures which are deficient in one way or another. Both the Markov test (Christoffersen, 1998) and duration test (Christoffersen, 2004) can be extended to test independence and unconditional coverage jointly. For a Markov test, this is simple: The joint Markov test examines whether there is any difference in the likelihood of a VaR violation following a previous VaR violation or nonviolation and at the same time determines whether these proportions are significintly different from α. The ability of joint tests to detect the VaR measure which violates either of the two properties, their ability to detect a measure which only violates one of the properties is decreased, compared to tests which test only independence or unconditional coverage. This drawback comes down to the fact that as one of the properties is satisfied, it is more difficult for the fact to detect the inadequacy of a VaR measure. This fact indicates that either conditional coverage or independence tests alone are preferable to joint tests when prior considerations are informative about the source of the VaR measure's potential inaccuracy (Campbell, 2005).

3.3.4 Tests based on multiple VaR levels-α

The above discussed tests attempt to determine the adequacy of a VaR measure at a single level, α. However, there is no reason to restrict the attention at a single VaR level, since the unconditional coverage and independence property of an accurate VaR measure should hold for any level of α. Several backtests based on multiple VaR levels were suggestsed. They utilize the fact that a $\alpha\%$

VaR should be exploited $\alpha\%$ times and also VaR violations at all levels should be independent from each other.

3.3.5 Regression-based tests

Engle and Manganelli (2004) propose a novel approach to quantile estimation: Instead of modelling the whole distribution, they model the quantile directly. The volatility clustering may be translated to saying that the distribution of it is autocorrelated. Consequently, the VaR, which is tightly linked to the standard deviation, must exhibit similar behavior. A natural way to formalize this characteristic is to use some type of autoregressive specification. They introduced the Conditional autoregressive quantile specification (CAviAR) with the null hypothesis of the number of exceedances being equal to the confidence level of the VaR model and the timing not exhibiting clustering.

This test relies on a linear model with the general idea to project VaR violations onto a set of explanatory variables and test different restrictions on parameters of the regression model that corresponds to the consequences of the martingale assumption. Both linear and non-linear regression models can be considered. The Dynamic quantile test of Engle and Manganelli (2004) (DQ-test) consists in testing linear restrictions in a linear model that links the violations to a set of explanatory variables, with a binary dependent variable.

4 Methods used for estimating volatility

The goal of this thesis is to demonstrate and compare the efficiency of respective methods which cna be used for volatility forecasting. For that, we will use to commonly used families of methods: The GARCH models which are being used for more than 4 decades, and methods based on realized volatility, which undergone a recent development in the last few years due to the availability of high-frequency trading data.

4.1 GARCH-family models - to be improved

To be improved

Add FI-GARCH

Add FIE-GARCH

The generalized autoregressive conditional heteroskedasticity (GARCH) model was introduced by Bollerslev (1987) as a generalization of the earlier ARCH model defined by Engle (1982) and since then, it has been widely used for studying the volatility of time series. The original general specification of the GARCH model was

$$r_t = \phi_0 + \phi_1 \cdot r_{t-1} + a$$

$$a_t = \sigma_t \cdot \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i \cdot a_{i-1}^2 + \sum_{j=1}^{s} \beta_j \cdot \sigma_{t-j}^2$$

or for the commonly used GARCH(1,1)

$$\sigma^2 = \alpha_0 + \alpha_i \cdot a_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2 d$$

Elaborate more

Several extensions were proposed by other authors later. Examples include:

The GARCH in mean (GARCH-M) model which connects the return of an asset to its volatility was introduced by Engle et al. (1987)

$$r_t = \mu + c \cdot \sigma_t^2 + a_t$$

$$a_t = \sigma_t \cdot \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i \cdot a_{i-1}^2 + \sum_{j=1}^{s} \beta_j \cdot \sigma_{t-j}^2$$

Add some discussion comparigon the performance of various GARCH family models in estimating VaR

The integrated GARCH (IGARCH), defined by Engle and Bollerslev (1986) model which is a unit-root integrated GARCH model in which the past squared shock is persistent:

$$r_t = \phi_0 + \phi_1 \cdot r_{t-1} + a$$

$$a_t = \sigma_t \cdot \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} (1 - \beta_i) \cdot a_{i-1}^2 + \sum_{j=1}^{s} \beta_j \cdot \sigma_{t-j}^2$$

Zakoian (1994) introduces the threshold autoregressive GARCH (TAR-GARCH), which is able to take into account the asymetric response in the volatility equation to the sign of a shock, which is supported empirically:

$$r_t = \phi_0 + \phi_1 \cdot r_{t-1} + a$$

$$a_t = \sigma_t \cdot \epsilon_t$$

$$\sigma_t^2 = \begin{cases} \alpha_0 + \alpha_1 \cdot a_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2, & a_{t-1} \leq 0 \\ \alpha_2 + \alpha_3 \cdot a_{t-1}^2 + \beta_2 \cdot \sigma_{t-1}^2, & a_{t-1} > 0 \end{cases}$$

The Glosten-Jagannathan-Runkle GARCH (GJR-GARCH), which is a simple version of a threshold GARCH, as defined by GLOSTEN et al. (1993):

$$\sigma_t^2 + \alpha_0 + \beta_1 \cdot \sigma_{t-1}^2 + \alpha_1 \cdot a_{t-1}^2 + \alpha_1 \cdot a_{t-1}^2 + \gamma \cdot I\left(a_{t-1} \leq 0\right) \cdot a_{t-1}^2$$

$$I\left(a_{t-1} \leq 0\right) = 0 \text{ if } a_{t-1} \leq 0 \text{ and } I\left(a_{t-1} \leq 0\right) = 1 \text{ if } a_{t-1} > 0,$$

The Exponential GARCH (EGARCH) model as defined by Nelson (1991):

$$a_t = \sigma_t \epsilon_t$$

$$a_t = \sigma_t \cdot \epsilon_t$$

$$log\left(\sigma_t^2\right) = \omega + \sum_{k=1}^{q} \beta_k \cdot g\left(Z_t\right) + \sum_{k=1}^{p} \alpha_k \cdot log\left(\sigma_{t-k}^2\right)$$

$$g\left(Z_t\right) = \Theta \cdot Z_t + \lambda \cdot \left(|Z_t| - E\left(Z_t\right)\right)$$

where Z_t is a standard normal variable, so $g\left(Z_t\right)$ allows the sign and the magnitude of Z_t to have separate effects on the volatility, which is especially useful in asset pricing context.

Since their introduction, the models have been subject to thorough study. It has been shown that GARCH models are frequently used in estimating volatility for estimating the Value-at-risk.

todo something performs somehow, something has some properties etc.

"Daily returns standardized by the square root of the volatility measures are not always normally distributed but do not exhibit ARCH effects."

TODO examples, references

rephrase JB, LB tests

4.2 Realized volatility

Initial work on realized volatility includes Zhou (1996), Andersen and Bollerslev (1998), Andersen et al. (2001a), Barndorff-Nielsen and Shephard (2002b), Meddahi (2002) or Andersen et al. (2001b).

Recently, the early results have been refined and extended by Bandi and Russell (2008), Oomen (2005), Zhang (2006), Hansen and Lunde (2006), Hansen et al. (2008), Martens and van Dijk (2007) and Christensen and Podolskij (2007).

Stylized facts on equity ultra high frequency data (UHFD) are described by Andersen et al. (2001b), Ebens (1999) and Hansen and Lunde (2006).

Modelling of different frequencies in the evolution of volatility is an alternative to traditional approaches which take long-range dependence into account in the ARFIMA models and regression models mixing information at different frequencies, e. g. the so-called Heterogeneous AR (HAR) model as developed by Corsi (2009).

The intra-daily prices are used as building blocks of the UHFD volatility. Let $p_{i,t}$ be the i-the intradaily log-price of day t sampled at frequency θ.

Log price - I am using price this sentence does not fit into the rest of text - probably a relict

The intra-daily price series are constructed using either Tick Time Sampling (TTS) or Calendar Time Sampling (CTS). In TTS, the series is sampled every d ticks. In CTS, we take the last recorded tick-by-tick price every θ units of time starting from an initial time of the day (typically the opening) until market closing. In our case, sampling every minute delivers $n \cdot (5min) = 90$ for a market open between 9:30am and 5:00pm. Overnight information is not included in these series and this and this may have a consequence, as studied by Gallo (2001) who shows that the overnight squared return has a significant impact when used as a predetermined variable in a GARCH for the open-to-close returns. This problem is similarly present for realized volatility measures, as demonstrated by Martens (2002), Fleming et al. (2001) or Hansen and Lunde (2001), among others. The realized volatility has become the benchmark UHFD volatility measures, commonly used in applied work (Brownlees and Gallo, 2009). Under appropriate assumptions, including the absence of jumps and microstructure noise, the RV converges to the latent volatility as the sampling frequency increases. Realized variance can be computed as

Also how does this sentence fit into the context?

$$RV_{i,t} = \sum_{j=1}^{m} r^2_{i,t-1+j\cdot n}$$

and realized volatility as

$$RVol = \sqrt{RVar}$$

This variance or volatility is daily. We can further aggregate the daily realized variance to a longer period of time to have e. g. weekly realized variance :

$$RV^w_t = \frac{1}{5}\left(RV^d_t + RV^d_{t-1} + \ldots + RV^d_{t-4}\right)$$

or monthly realized variance:

$$RV^m_t = \frac{1}{22}\left(RV^d_t + RV^d_{t-1} + \ldots + RV^d_{t-21}\right)$$

We can further decompose the realized variance into positive semivariances, as proposed by Barndorff-Nielsen et al. (2010) as $RV_t = RS^-_t + RS^+_t$ where

$$RS^-_{i,t} = \sum_{j=1}^{m} r^2_{i,t-1+j\cdot n},\ \text{if } r_{i,t-1+j\cdot n} < 0$$

$$RS^+_{i,t} = \sum_{j=1}^{m} r^2_{i,t-1+j\cdot n},\ \text{if } r_{i,t-1+j\cdot n} > 0$$

In the same fashion, we can compute higher order realized moments: realized skewness as

$$RSkew_{i,t} = \frac{\sqrt{m}\sum_{i=1}^{m} r_{i,t-1+j\cdot n}^{3}}{RV_t^{\frac{3}{2}}}$$

and realized kurtosis as

$$RKurt_{i,t} = \frac{\sqrt{m}\sum_{i=1}^{m} r_{i,t-1+j\cdot n}^{4}}{RV_t^{2}}$$

Using the realized measures, we can estimate several types of models.

4.2.1 Autoregressive model of realized volatility

An AR(p) process as

$$RV_t = \beta_0 + \sum_{i=1}^{p} \beta_i \cdot RV_{t-i} + \epsilon_t$$

or

$$RVol_t = \beta_0 + \sum_{i=1}^{p} \beta_i \cdot RVol_{t-i} + \epsilon_t$$

Similarly, we could extend the AR(p) model to ARMA(p,q) model with the same logic as for the original returns. The problem of this approach is that a simple ARMA model neglects long-time memory. This could be dealt with using some ARFIMA model. However, different models have been developed to adress this problem specifically in the context of realized volatility.

4.2.2 Heterogeneous autoregression

Another method which we can use for studying conditional heteroskedasticity with the use of realized volatility is the Heterogeneous Autoregression (HAR), introduced by Corsi (2009):

$$RV_t = \alpha_0 + \beta_1 \cdot RV_{t-1} + \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + u_t$$

where the $RV_{t-1}^{(h)}$ is h-period realized variance, so $RV_{t-1}^{(5)}$ corresponds to 1 week and $RV_{t-1}^{(22)}$ corresponds to one month, and u_t is a normally distributed error term.

To this baseline model, we can add the realized semivariances, as proposed by Patton and Sheppard (2015):

$$RV_t = \alpha_0 + \beta_1^{+} \cdot RS_{t-1}^{+} + \beta_1^{-} \cdot RS_{t-1}^{-} + \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + u_t$$

or realized skewness and kurtosis , as proposed by Amaya et al. (2015):

$$RV_t = \alpha_0 + \beta_1^{\cdot} RV_{t-1} + \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + \beta_s \cdot RSkew_{t-1} + \beta_k \cdot RKurt_{t-1} + u_t$$

or

$$RV_t = \alpha_0 + \beta_1^{+} \cdot RS_{t-1}^{+} + \beta_1^{-} \cdot RS_{t-1}^{-} + \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + \beta_s \cdot RSkew_{t-1} + \beta_k \cdot RKurt_{t-1} + u_t$$

4.2.3 Realized GARCH TODO TODO

TODO

The realized GARCH model is a crossover between the traditional GARCH family models and models utilizing realized volatility. It was introduced by HANSEN et al. (2012) and can be constructed as

Improve equation, add theory

$$h_t = \omega + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \alpha \cdot r_{t-1}^2$$

It can be also extended to more advanced model in a similar fashion like the standard GARCH:

TODO (check if true

5 Forecast

In order to be able to make decisions related to future, we need to forecast, i. e. extrapolate into future with the use of currently available values. In general, for an observed value X forecasted by a model M, therefore

$$X_t = M\left(X_{t-1}\right)$$

where ξ are all the remaining variables in the model, the forecasting is

$$\widehat{X_{t+1|t}} = M\left(X_t\right)$$

where $\widehat{X_{t+1|t}}$ is the conditional expectation of X_{t+1} given the information available at time t. Similarly, for time $t+2$, the conditional expectation X_{t+2} is

$$\widehat{X_{t+2|t}} = M\left(\widehat{X_{t+1|t}}\right)$$

which is equivalent to

$$\widehat{X_{t+2|t}} = M\left(M\left(X_t\right)\right)$$

Since forecasts from any model will not be perfect, we need to consider the forecast error, which is the difference between the true value and the forecasted value:

$$e_{t+1|t} = X_{t+1} - \widehat{X_{t+1|t}}$$

which is

$$e_{t+1|t} = X_{t+1} - M\left(X_t\right)$$

similarly

$$e_{t+2|t} = X_{t+2} - \widehat{X_{t+2|t}}$$

which is

$$e_{t+2|t} = X_{t+2} - M\left(M\left(X_t\right)\right)$$

The forecast is unbiased if

$$E\left(e_{t+h|t}\right) = 0$$

For forecasting, there are several ways how to approach the forecasted values in terms of refitting the model. First, in a fixed scene, we estimate the model on the available data and keep it estimated for the whole forecast without refitting. In other words, the newly forecasted values area only inputted into the existing model, without them influencing the model itself. This is a naive approach because future values can certainly influence what the model should look like.

Second is the rolling scheme. For a fixed rolling window size, we estimate the model and forecast the future values and then roll the window of observations to be used in the estimation of the model, so that we use the newly forecasted estimation(s) and in the next model fitting, some amount of observations at the beginning of the time series is not used.

Third is the expanding scheme. We start with a given number of observations and after estimating the model and forecasting the new values, we expand the window of observations used in estimating the next model to the newly forecasted value.

For both rolling and expanding window, we can refit the model with different frequency. The more straightforward and most accurate way is to refit the model for every single forecasted value. However, this comes at the cost of high computational power necessary, especially with computationally intensive models and long time series. Therefore, we may choose to refit only after a certain period (e. g. monthly, that is, every 21 forecasted observations), weekly (every 5 observations) or similar in order to decrease running time.

References everywhere

5.1 Volatility forecasting

Join forecast error to model evaluation - MSE, MAE

In the VaR framework, volatility forecasting is an interesting discipline for comparing different volatility measures (Bollerslev et al., 2003). Although in this thesis the topic is limited to a single asset at time, it can be extended into a multivariete problem.

references

5.2 VaR forecasting

Expand, wording

In 1994, the risk management group in J.P. Morgan developed the RiskMetrics model for measuring VaR, which has since become a standard for measuring risk. (So and Yu, 2006) This model assumes that returns of a financial asset follow a conditional normal distribution with zero mean and variance is expressed as an exponentially weighted moving average (EWMA) of historical squared returns. The first drawback of RiskMetrics is that it was shown many times that a return distribution usually has heavy tails, so with the assumption of normality we may introduce bias which is mainly present in the tail properties of the return distribution. The second drawback is that many financial financial return series may exhibit long memory, as shown by Ding et al. (1993) or long-term dependence on market volatility, as studied by So (2000), who finds that such long term dependence has a significant impact on the pricing of financial derivatives as well as forecasting market volatility. Besides the traditional and extended GARCH models, several long memory GARCH models were proposed to incorporate the long memory volatility property, such as the FIGARCH developed by Baillie et al. (1996).

There is a wide variety of methods for forecasting VaR, the performance of many of them is compared by Kuester (2005). Giot and Laurent (2004) propose a following approach for forecasting VaR using realized volatility: Let r_t be the daily (close-to-close) return at time t on a single asset. Then we assume that

$$r_t = \sqrt{h_t \cdot \nu_t},\ \nu_t \sim F$$

where h_t is the conditional variance of the daily return art time t and ν_t is an i.i.d. unit variance and possibly skewed and leptokurtic random variable from a cumulative distribution F. The one-day-ahead $100 \cdot (1-p)\,\%$ VaR is defined as the maximum one day ahead loss, that is

$$VaR_{t|t-1}^{p} = -F^{-1}(p)\sqrt{h_t}$$

assuming that h_t is known, conditional on the information available at time $t-1$. In a GARCH framework, we can predict the one-day-ahead forecast of the conditional variance of returns and use a distributional assumption on F to provide the proper quantile of the distribution of the standardized residuals. We can depart from this procedure if a series for a return variance proxy is directly available. Let $rv_{(m,\theta)\cdot t}$ denote such a generic proxy computed according to definition m using intradaily data sampled at frequency θ on day t and let $rv_{(m,\theta)\cdot t|t-1}$ denote its expectation conditional on the information available at time $t-1$, using suitable model specification. Then we assume that the conditional variance of returns is some function of $rv_{(m,\theta)\cdot t|t-1}$ and a vector of parameters ϕ, for example, $h_t = f\left(rv_{m,\theta t|t-1}|\phi\right)$. To be able to work within this framework, we need to first specify a model capturing the dynamics of the volatility measures to obtain the conditional expectations of volatility, second a model that maps the conditional variance of returns with the conditional expectation of the volatility measures and third an appropriate distribution for the standardized return distribution.

5.3 Forecasting VaR

The out-of-sample VaR forecasting is performed with both rolling and expanding windows using the whole time series from its beginning until the start of forecast, which is assumed to be the information set as of time t at each step. Then the one-day-ahead VaR prediction at time $t+1$ is derived as

$$TODO$$

The prediction is based on the inverse of the cumulative distribution function of the Student's t distribution with the degrees of freedom estimated in order to fit the most appropriate distribution. We move the estimation window ahead by one day and repeat the procedure until we gather the series of one day ahead predictions, spanning across 252 days. TODO

6 Model evaluation

Several metrics can be used to evaluate the forecasting performance, such as the Mincer-Zarnowitz type regression which forecasts are contrasted against a target, in our case, all the models pairwise, as demonstrated by Bollerslev et al. (2003) or Aït-Sahalia and Mancini (2008), implied volatility measures such as VIX, as studied by Engle and Gallo (2006), or within a risk management management framework, the quality of the derived Value-at-risk (VaR) or Expected Shortfall (ES) which have emerged as promininent measures of market risk (Giot and Laurent, 2004).

The dataset we use includes 35 time series with daily granularity in the period starting on 2010-01-05 and ending on 2016-01-22 on stock returns with precalculated Realized Volatility, positive realized semi-volatility, negative realized semi-volatility, realized skewness and realized kurtosis. The specific stocks which we use are listed in the table in appendix.

The process of comparing the models will be by following steps for each model:

First, a full-sample size model will be used for fitting the in-sample model. Then there will be a discussion of the fits and comparison of qualitative differences of the estimates. Since the true conditional variance is latent, it needs to be substituted by some ex-post estimator based on observed quantities as they become available. Possible candidates to serve as unbiased proxies for volatility are, for example, the daily squared returns or RV. One way this can be done is visually by plotting the in-sample fits compared to the selected proxy. The loss function measures the difference between realization and the forecast. Some of the commonly used loss functions are Mean absolute error (MAE):

$$\mathcal{L}\left(RV_{t+1}, \hat{RV}_{t+1|t}\right) = \left|RV_{t+h} - \hat{RV}_{t+h|t}\right|$$

Mean squared error (MSE):

$$\mathcal{L}\left(RV_{t+1}, \hat{RV}_{t+1|t}\right) = \left(RV_{t+h} - \hat{RV}_{t+h|t}\right)^2$$

Quasi-likelihood (QLIKE):

$$\mathcal{L}\left(RV_{t+1}, \hat{RV}_{t+1|t}\right) = \left(log\left(\hat{RV_{t+h|t}}\right) + \frac{RV_{t+h}}{\hat{RV}_{t+h|t}}\right)$$

Furthermore, we need to check whether the ARMA-GARCH and RGARCH models have normally distributed residuals by graphical representation, the Q-Q plot and formally by the Jarque-Bera test with the null hypothesis of normal distribution. The test statistics of the Jarque-Bera test, as defined by Jarque and Bera (1980) is:

Double check, why would we want to check the residuals of volatility models?

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^3\right)$$

where n is the number of observations, or degrees of freedom, S is the sample skewness:

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right)^{\frac{3}{2}}}$$

and K is the sample kurtosis:

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right)^{2}}$$

where

$$JB \sim \chi^2\left(2\right)$$

in case that the data is normally distributed.

Second, we compare the out-of-sample forecasting performance of the selected models using different forecasting schemes: Expanding window, in which the sample used to estimate the parameters of the model grows as the forecaster makes predictions for successive observations, and Rolling window , in which the sequence of forecasts is based on parameters estimated using a rolling sample of fixed size. Then we can plot and compare the forecast errors from all models, compute the loss functions (MSE and MAE), compare model performance by the Diebold-Mariano test, introduced by Diebold and Mariano (2002):

$$DM - T = \frac{\sqrt{T}\cdot\overline{d}}{\sqrt{\omega}} \overset{a}{\sim} N\left(0,1\right)$$

where $d_t = \mathcal{L}_{1,t} - \mathcal{L}_{2,t}$ is the loss differential, which is assumed to be stationary, and the null hypothesis of the DM test is $E\left[d_t\right] = 0$, and $\overline{d} = T^{-1}\sum_t d_t$ and ω is its asymptotic variance, estimated as a sample variance of d_t. We can also utilize the Mincer-Zarnowitz regression, which regresses realized values on forecasts, as defined by Mincer and Zarnowitz (1969):

$$\sigma_{t+1} = \beta_0 + \beta_1 \cdot \hat{\sigma}_{t+1}$$

and test the joint hypothesis, where not rejecting it means that the forecast is unbiased.

$$\beta_0 = 0,\ \beta_1 = 1$$

Third, the results from the above described methods shall be summarised and we will observe whether the comparison of results of performance of respective models is consistent across time series, or whether they differ. If they differ, we will discuss what may be determining the performance of respective models for each respective time series.

7 Model estimation and evaluation

We estimate the following models, described in section **??** (The text in **bold** at the beginning of each item specified how the model will be referred to henceforth):

- **AR(1)-RV**: the first-order autoregressive model of realized volatility

  $$RV = \alpha + \beta_1 \cdot RV_{t-1} + \epsilon_t,\ \epsilon_t \sim N\left(0, \sigma^2\right)$$

  where RV is the realized variance.

- **HAR**, heterogeneous autoregression, specified as

  $$RV_t = \alpha_0 + \beta_1 \cdot RV_{t-1} + \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + \mu_t$$

  where $RV_{t-1}^{(h)}$ is a h-period realized variance, i. e. $RV_{t-1}^{(5)}$ is weekly realized variance and $RV_{t-1}^{(22)}$ is monthly realized variance.

- **HAR-AS**, asymmetric heterogeneous autoregression, specified as

  $$RV_t = \alpha_0 + \beta_1^+ \cdot RS_{t-1}^+ + \beta_1^- \cdot RS_{t-1}^- \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + \mu_t$$

  where RS_{t-1}^+ is the positive realized semivolatility and RS_{t-1}^- is the negative realized semivolatility.

- **HAR-RS** - heterogeneous autoregression with realized skewness, specified as

  $$RV_t = \alpha_0 + \beta_1 \cdot RV_{t-1} + \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + \beta_4 \cdot RSkew_{t-1} \mu_t$$

  where $Rskew$ is the realized skewnes.

- **HAR-RSRK**, heterogeneous autoregression with realized skewness and realized kurtosis, with the specification

  $$RV_t = \alpha_0 + \beta_1 \cdot RV_{t-1} + \beta_2 \cdot RV_{t-1}^{(5)} + \beta_3 \cdot RV_{t-1}^{(22)} + \beta_4 \cdot RSkew_{t-1} + \beta_5 \cdot RKurt_{t-1} \mu_t$$

  where $Rkurt$ is the realized kurtosis.

- **RGARCH**: realized GARCH(1,1), specified as

  $$h_t = \omega + \alpha \cdot\ r_{t-1}^2 + \beta \cdot h_{t-1} + \gamma \cdot x_{t-1}$$

  where x_t represents the noisy measurement of realized volatility.

- **GARCH**: ARMA(1,1)-GARCH(1,1), the baseline GARCH with the specification

  $$a_t = \sigma_t \cdot \epsilon_t$$
  $$\sigma_t^2 = \alpha_0 + \alpha_i \cdot a_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2$$

  where ϵ_t is the error from an ARMA model.

For the sake of parsimony, we chose not to include any more advanced GARCH-family models such as the aforementioned EGARCH, IGARCH, GJR-GARCH, GARCH-M, FIGARCH, FIEGARCH and similar. However one needs to keep in mind that it has been shown that these models in general outperform the base GARCH(1,1) in forecasting volatility due to asymmetries and long memory of financial time series.

reference

We performed the Jarque-Bera test on the residuals of all models for each stock and its p-values are always < 0.01, so we can reject the null hypothesis of normality of residuals.

so what?

Ljung-Box test?

8 Data

To be able to observe a trend in performance in each model, we are using a set of daily data for 80 stocks. From the 100 most traded US stocks as of 7th May 2024, we eliminated these which had a long break of missing value in the middle, did not have a sufficient number of observations prior to outbreak of covid (at least 1000 observations before 29th November 2019), and several stocks in which some of the models failed to converge, even with careful parameter tuning. The used stocks and their basic properties are shown in Table 1.

Source for 1000 obs

source

For each stock we have the following values for every date:

- Close price
- Realized variance
- Realized positive semivariance
- Realized negative semivariance
- Realized skewness
- Realized kurtosis

We use close price to compute returns as simply

$$r_{i,t} = \frac{P^{close}_{i,t} - P^{close}_{i,t-1}}{P^{close}_{i,t}}$$

The data is extracted from the database. All the realized measures time series use the 5-minute intra-day returns between 9:30am and 17:00pm and follows CTS. Each of the time series has a different starting point, depending on the availability of data, the earliest starting point being 5th January 1998. The last day of the time series is the 9th December 2022.

Which?

For each stock, we run the Augmented Dickey-Fuller (ADF) test for a presence of unit root on the returns. For all stocks, the ADF p-value is < 0.01 as expected, thus rejecting the null hypothesis of unit root in the return time series. We also performed a Jarque-Bera test on returns also with < 0.01 for all stocks, thus rejecting the null hypothesis of normality of returns, as expected. Therefore, we will assume Student's t-distribution to model the expected returns.

We also perform the Ljung-Box (LB) test for autocorrelation and Jarque-Bera (JB) test for normality. The p-values of these tests are summarized in Figure 1.

We can see, the p-values of the Ljung-Box suggest that there is a small amount of return time series which do exhibit serial correlation, however the most of them are serially independent.

How is serial correlation/ independence relevant for the methods? What if the results differ as here?

Stock	Start	End date	Obs	Obs bef 2019-05-24	Stock	Start	End date	Obs	Obs bef 2019-05-24
AAPL	1998-01-05	2022-12-09	6276	5510	LRCX	1998-01-05	2022-12-12	6277	5510
ABBV	2013-01-03	2022-12-09	2503	1737	MA	2006-05-26	2022-12-12	4166	3399
ACN	2001-07-20	2022-12-09	5383	4617	MCD	1998-01-05	2022-12-12	6277	5510
ADBE	1998-01-05	2022-12-09	6276	5510	MCHP	1998-01-05	2022-12-12	6277	5510
AMAT	1998-01-05	2022-12-09	6276	5510	MELI	2007-08-13	2022-12-12	3862	3095
AMD	1998-01-05	2022-12-09	6276	5510	META	2012-05-21	2022-12-12	2659	1892
AMGN	1998-01-05	2022-12-09	6276	5510	MRK	1998-01-05	2022-12-12	6277	5510
AMZN	1998-01-05	2022-12-09	6276	5510	MSFT	1998-01-05	2022-12-12	6277	5510
ANET	2014-06-09	2022-12-09	2144	1378	MSTR	2002-08-29	2022-12-12	5108	4341
AVGO	2009-08-07	2022-12-09	3360	2594	MU	1998-01-05	2022-12-12	6277	5510
BA	1998-01-05	2022-12-09	6276	5510	NEE	1998-01-05	2022-12-12	6277	5510
BAC	1998-01-05	2022-12-09	6276	5510	NFLX	2002-05-24	2022-12-12	5175	4408
BKNG	2007-05-01	2022-12-09	3932	3166	NKE	1998-01-05	2022-12-12	6277	5510
C	1998-01-05	2022-12-09	6276	5510	NOW	2012-07-02	2022-12-12	2630	1863
CAT	1998-01-05	2022-12-09	6276	5510	NVDA	1999-01-25	2022-12-12	6012	5245
CMCSA	1998-01-05	2022-12-09	6276	5510	NXPI	2010-08-09	2022-12-12	3109	2342
CMG	2006-01-27	2022-12-09	4248	3482	ORCL	1998-01-05	2022-12-12	6277	5510
COP	2002-09-04	2022-12-09	5104	4338	PANW	2012-07-23	2022-12-12	2616	1849
COST	1998-01-05	2022-12-09	6276	5510	PEP	1998-01-05	2022-12-12	6277	5510
CRM	2004-06-24	2022-12-09	4650	3884	PFE	1998-01-05	2022-12-12	6277	5510
CSCO	1998-01-05	2022-12-09	6276	5510	PG	1998-01-05	2022-12-09	6276	5510
CVS	1998-01-05	2022-12-09	6276	5510	QCOM	1998-01-05	2022-12-12	6277	5510
CVX	2001-10-11	2022-12-09	5329	4563	SBUX	1998-01-05	2022-12-12	6277	5510
DIS	1998-01-05	2022-12-09	6276	5510	SMCI	2007-03-30	2022-12-12	3955	3188
EMR	1998-01-05	2022-12-09	6276	5510	SO	1998-01-05	2022-12-12	6277	5510
FCX	1998-01-05	2022-12-09	6276	5510	SPGI	2007-04-30	2022-12-12	3935	3168
FTNT	2009-11-19	2022-12-09	3287	2521	SYK	1998-01-05	2022-12-12	6277	5510
GE	1998-01-05	2022-12-09	6276	5510	TJX	1998-01-05	2022-12-12	6277	5510
GME	2002-05-13	2022-12-09	5183	4417	TMO	1998-01-05	2022-12-12	6277	5510
GOOG	2014-03-28	2022-12-09	2193	1427	TMUS	2007-04-20	2022-12-12	3941	3174
GS	1999-05-05	2022-12-09	5941	5175	TSN	1998-01-05	2022-12-12	6277	5510
HD	1998-01-05	2022-12-09	6276	5510	TXN	1998-01-05	2022-12-12	6277	5510
HES	2007-03-14	2022-12-09	3966	3200	UNH	1998-01-05	2022-12-12	6277	5510
IBM	1998-01-05	2022-12-09	6276	5510	V	2008-03-20	2022-12-12	3710	2943
INTC	1998-01-05	2022-12-09	6276	5510	VRTX	1998-01-05	2022-12-12	6277	5510
JNJ	1998-01-05	2022-12-09	6276	5510	WDAY	2012-10-15	2022-12-12	2557	1790
JPM	1998-01-05	2022-12-09	6276	5510	WFC	1998-01-05	2022-12-12	6277	5510
KO	1998-01-05	2022-12-09	6276	5510	WMT	1998-01-05	2022-12-12	6277	5510

Table 1: This table shows the overview of start date, end date, number of observations and number of observations for the before covid training set, i. e. observations prior to 2019-05-24

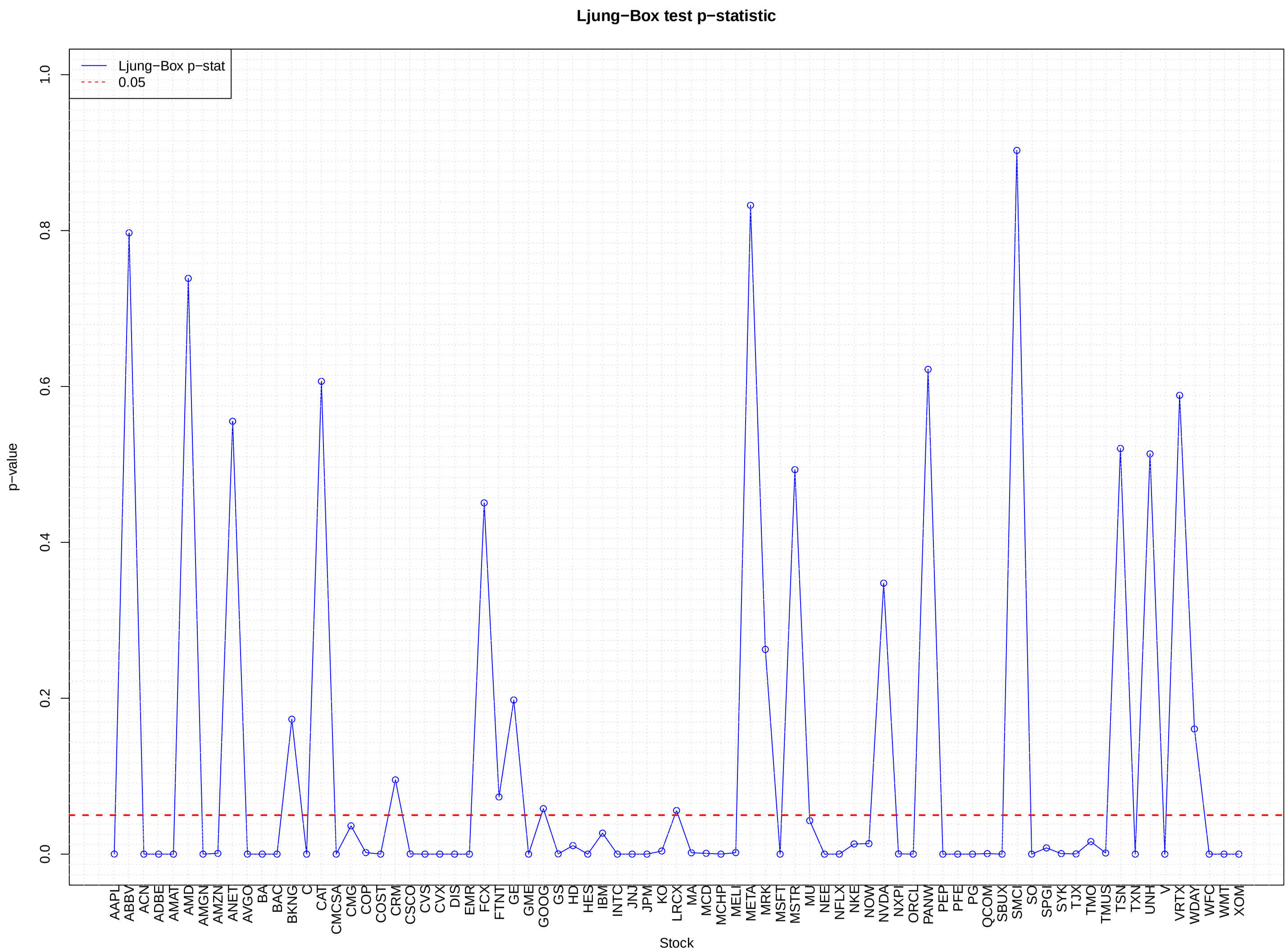

Figure 1: This graph shows the p-stats of the Ljung-Bos test for serial autocorrelation on returns for each stock.

9 Forecasting

We perform a fitting and forecasting of each model for all the available stocks. The forecasting is performed with both expanding and rolling window to be able to compare whether there are different results depending on the selected refitting scheme. For GARCH and RGARCH model, we refit every 21 observations, i. e. monthly because these models are very computationally intensive. For all the remaining models, we for refit every observations. The window length is always from the beginning of the stock time series until the specified date. The number of observations in a forecast window is shown in table 1. The stocks were selected such that for each time series, there are at least 1000 observations available, as per the finding by .

reference GARCH 1000 obs. But probably in the theory section

Table 2 shows an overview of how many forecasts of respective stocks perform better with expanding or rolling window for each model according to mean absolute error. Table 3 shows an overview of how many forecasts of respective stocks perform better with expanding or rolling window for each model according to mean square error. Altogether we can say that there are only small nuances and we cannot observe a clear pattern stating that one type of forecasting window performs better for specific type of model, the only exception being the HAR-RSRK, in which rolling forecasting scheme outperforms expanding forecasting scheme for all the stocks we study. Therefore, for the volatility estimation, we will use results obtained using both schemes.

Doublecheck whether this really makes sense and try to reason why

Figures 2, 3, 4, 5 show the MAE and MSE for both forecast types for each type of model for all stocks. Tables 4 and 5 summarize the mean, standard deviation and ranking of the respective error function for each model across all stocks. From these tablers we can see that based on error functions, realized GARCH has the best forecasting performance while plain GARCH is the worst and a first-order autoregressive model on realized volatility the second worst.

What can we see?

Methodology how to compute the errors - RV as true value

Tables 6 and 7 show the mean value and standard deviation of the p-values of Diebold-Mariano test used to compare the forecasting performance of different models. Table 8 shows the percentage of tests for each stock in which the p-value was lower than 0.05, i. e. that the models differ in forecasting performance for the respective stock. In a Diebold-Mariano test, the null hypothesis states that both models are equivalently good, so a high p-value means that we do not have evidence that the models are different, whereas a small (¡0.05) p-value suggests that the models perform differently. Since table 8 shows the proportion of cases in which the p-value is lower than 0.05, we can assume that for the combinations in which this proportion is relatively high, the model performance is indeed different, whereas for combinations with a low ratio, there is no such evidence. We can definitely see that the performance of RGARCH differs from the performance of the AR(1)-RV, HAR and GARCH models, GARCH furthermore differs from the AR(1)-RV model, and HAR differs frm the AR(1)-RV model. However, the results of the DM test do not state anything about which model performs better than the other.

Table 9 shows the means and standard deviations of the p-values of Mincer-

Zarnowitz regression, and in the third row the percentage of in how many cases the p-value was < 0.05. We can clearly see that on 5% significance level, the estimates are mostly biased. However, a closer look at the mean p-values shows that while the mean p-values of the autoregressive realized volatility and GARCH models are very close to zero, the means of Mincer-Zarnowitz p-values for the HAR models are just above 0.05 and that of realized GARCH models are above 0.1. From that we can tell that even though most of the models are biased according on 5% significance level, we can again see that the realized GARCH performs the best, followed by HAR models, and the autoregressive realized volatility and GARCH models perform the worst.

	AR(1)-RV	HAR	HAR-AS	HAR-RS	HAR-RSRK	RGARCH	GARCH
Rolling	58.00	59.00	74.00	71.00	77.00	49.00	54.00
Expanding	19.00	18.00	3.00	6.00	0.00	28.00	23.00

Table 2: This table shows a summary of how many stocks for each model perform better with expanding or rolling forecasting scheme according to mean absolute error.

	AR(1)-RV	HAR	HAR-AS	HAR-RS	HAR-RSRK	RGARCH	GARCH
Rolling	60.00	54.00	74.00	73.00	77.00	45.00	52.00
Expanding	17.00	23.00	3.00	4.00	0.00	32.00	25.00

Table 3: This table shows a summary of how many stocks for each model perform better with expanding or rolling forecasting scheme according to mean square error.

	AR(1)-RV	HAR	HAR-AS	HAR-RS	HAR-RSRK	RGARCH	GAR
MSE expanding	0.00016 (6)	0.00013 (2)	0.00013 (3)	0.00013 (5)	0.00013 (4)	0.00009 (1)	0.000
MSE rolling	0.00016 (6)	0.00013 (3)	0.00013 (2)	0.00013 (4)	0.00013 (5)	0.00010 (1)	0.000
MAE expanding	0.00940 (6)	0.00746 (3)	0.00744 (2)	0.00764 (5)	0.00760 (4)	0.00554 (1)	0.009
MAE rolling	0.00924 (6)	0.00740 (3)	0.00737 (2)	0.00755 (4)	0.00760 (5)	0.00556 (1)	0.009

Table 4: This table shows the means of error measures for each model with its order (smallest to largest) for each respective error measure and forecasting window in parentheses.

	AR(1)-RV	HAR	HAR-AS	HAR-RS	HAR-RSRK	RGARCH	GARCH
MSE expanding	0.00015	0.00009	0.00009	0.00009	0.00009	0.00009	0.00015
MSE rolling	0.00013	0.00009	0.00009	0.00009	0.00009	0.00009	0.00015
MAE expanding	0.00367	0.00241	0.00243	0.00241	0.00235	0.00203	0.00293
MAE rolling	0.00338	0.00237	0.00241	0.00236	0.00235	0.00210	0.00290

Table 5: This table shows the standard deviations of error measures for each model.

	AR(1)-RV	HAR	HAR-AS	HAR-RSV	HAR-RSRK	RGARCH	GARCH
AR(1)-RV	-	0.37	0.55	0.53	0.53	0.11	0.11
HAR	0.37	-	0.92	0.77	0.77	0.09	0.09
HAR-AS	0.56	0.93	-	0.13	0.19	0.29	0.18
HAR-RSV	0.55	0.79	0.1	-	0.33	0.23	0.19
HAR-RSRK	0.54	0.77	0.1	0.28	-	0.23	0.2
RGARCH	0.11	0.09	0.3	0.24	0.24	-	0.01
GARCH	0.12	0.1	0.19	0.2	0.2	0.02	-

Table 6: This table shows the means of p-values of the Diebold-Mariano test for respective combinations of models. The values below the diagonal are for rolling window forecast, the values above the diagonal are for the expanding window forecast.

	AR(1)-RV	HAR	HAR-AS	HAR-RSV	HAR-RSRK	RGARCH	GARCH
AR(1)-RV	-	0.33	0.32	0.31	0.31	0.22	0.23
HAR	0.33	-	0.14	0.16	0.17	0.22	0.22
HAR-AS	0.32	0.14	-	0.22	0.25	0.29	0.29
HAR-RSV	0.31	0.16	0.15	-	0.33	0.28	0.29
HAR-RSRK	0.31	0.17	0.19	0.29	-	0.28	0.29
RGARCH	0.21	0.21	0.3	0.29	0.29	-	0.06
GARCH	0.25	0.23	0.3	0.3	0.29	0.07	-

Table 7: This table shows the standard deviations of p-values of the Diebold-Mariano test for respective combinations of models. The values below the diagonal are for rolling window forecast, the values above the diagonal are for the expanding window forecast.

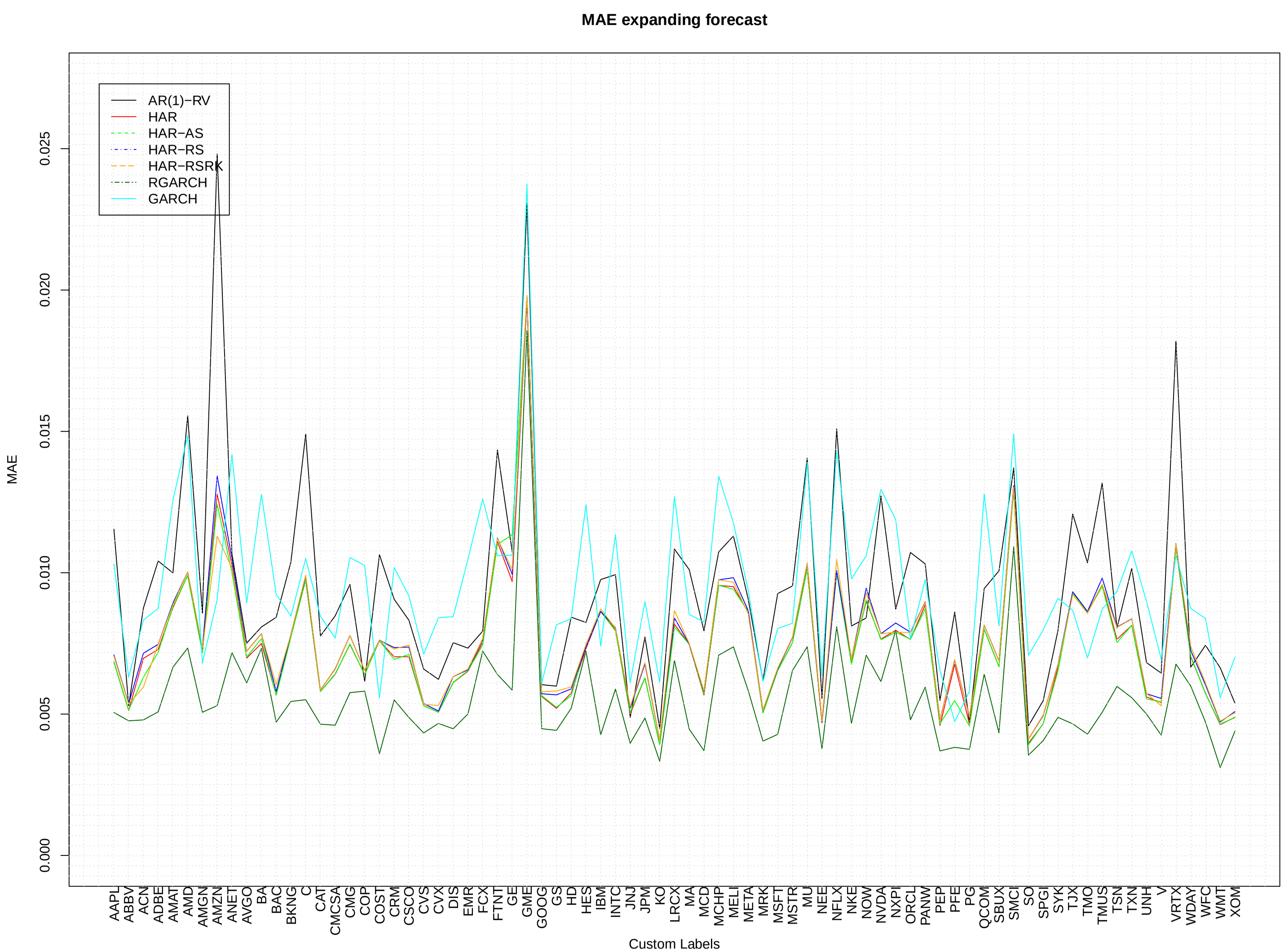

Figure 2: This graph shows the mean absolute error from expanding forecast for each stock for all models.

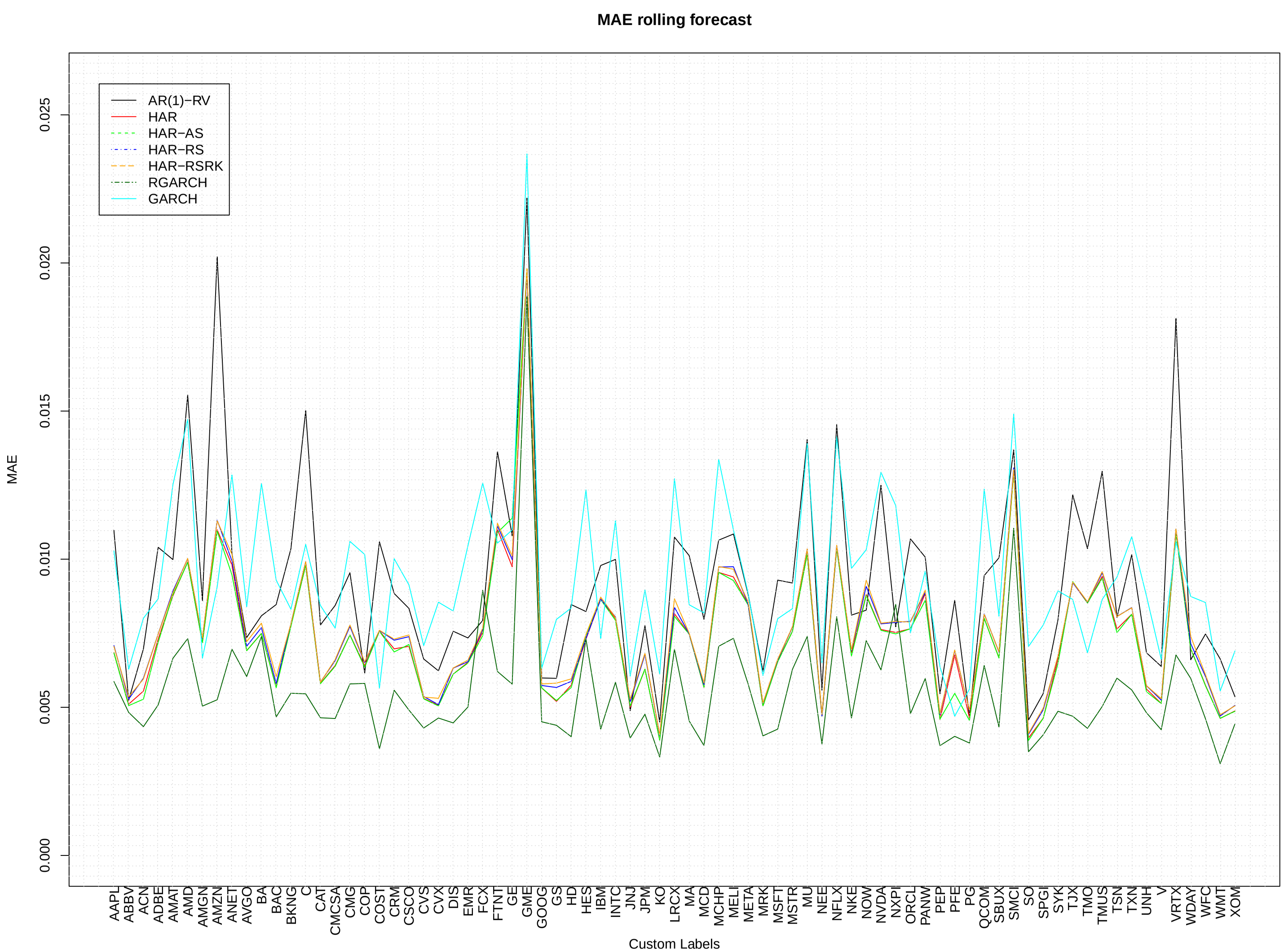

Figure 3: This graph shows the mean absolute error from rolling forecast for each stock for all models.

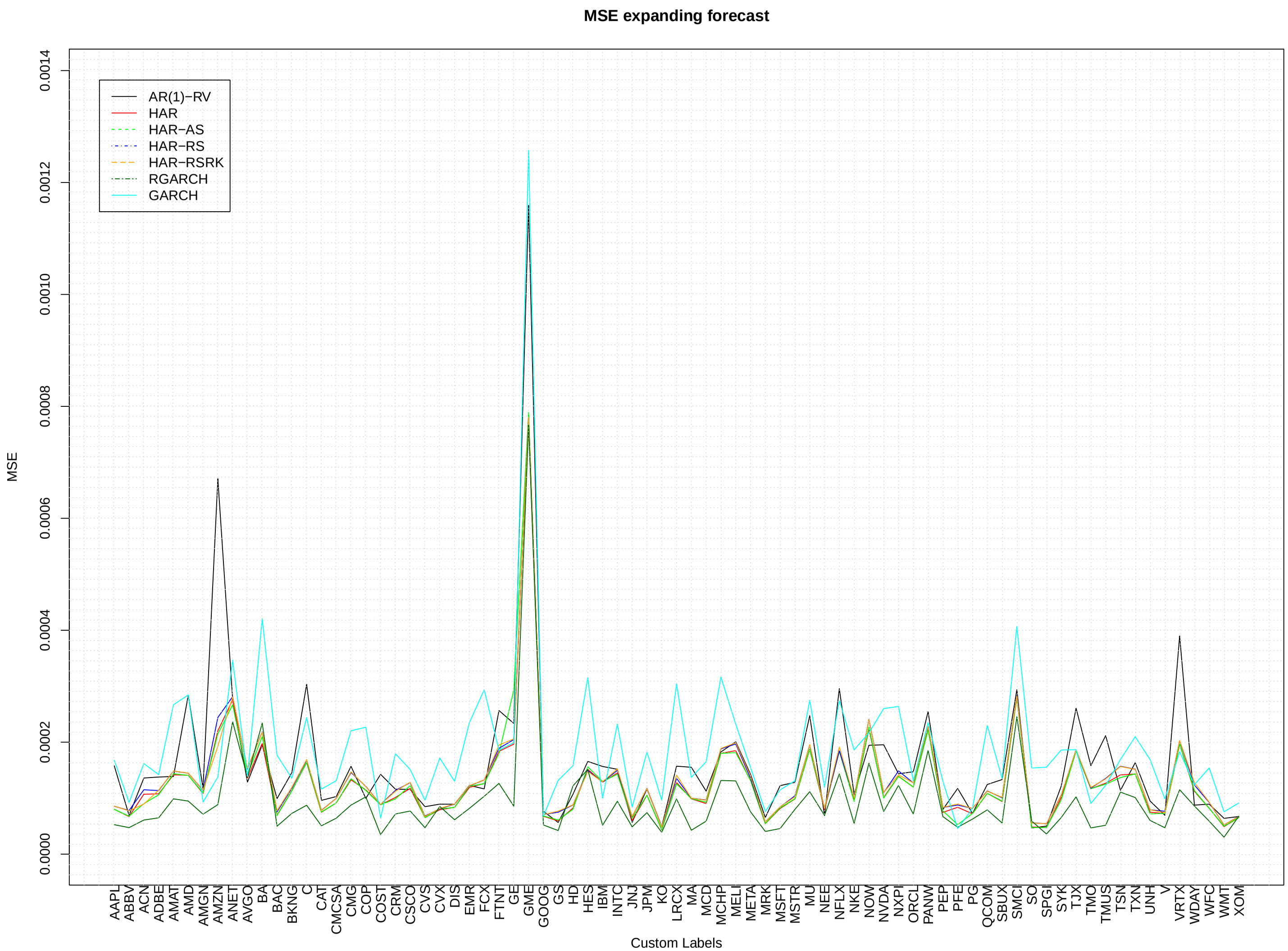

Figure 4: This graph shows the mean square error from expanding forecast for each stock for all models.

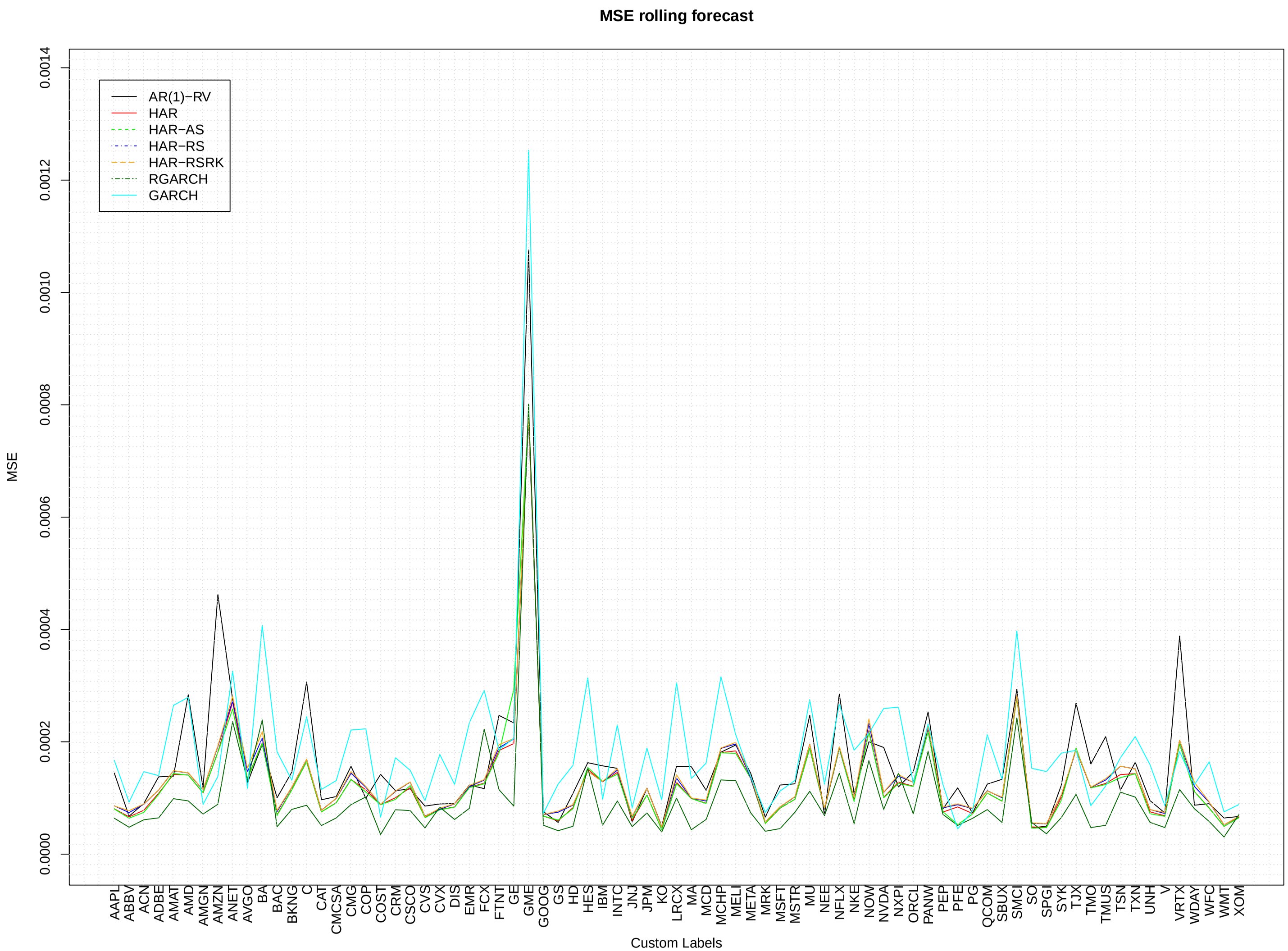

Figure 5: This graph shows the mean square error from expanding forecast for each stock for all models.

	AR(1)-RV	HAR	HAR-AS	HAR-RSV	HAR-RSRK	RGARCH	GARCH
AR(1)-RV	-	0.26	0.1	0.1	0.1	0.73	0.77
HAR	0.27	-	0.01	0	0	0.81	0.79
HAR-AS	0.09	0.01	-	0.55	0.43	0.3	0.61
HAR-RSV	0.09	0	0.53	-	0.27	0.36	0.56
HAR-RSRK	0.09	0	0.58	0.32	-	0.38	0.56
RGARCH	0.7	0.79	0.25	0.35	0.38	-	0.94
GARCH	0.75	0.78	0.61	0.57	0.57	0.94	-

Table 8: This table shows the the percentage for how many stocks the p-value of the Diebold-Mariano test was below 0.05 for respective combinations of models. The values below the diagonal are for rolling window forecast, the values above the diagonal are for the expanding window forecast.

	Mean	StdDev	% p-vals <0.05
AR(1)-RV Expanding	0.00013	0.00114	1.00000
AR(1)-RV Rolling	0.00013	0.00114	1.00000
HAR Expanding	0.06273	0.14018	0.76623
HAR Rolling	0.06883	0.14394	0.75325
HAR-AS Expanding	0.06468	0.14820	0.77922
HAR-AS Rolling	0.06961	0.14874	0.75325
HAR-RSV Expanding	0.05455	0.12424	0.77922
HAR-RSV Rolling	0.06312	0.13373	0.75325
HAR-RSRK Expanding	0.05455	0.12150	0.75325
HAR-RSRK Rolling	0.05455	0.12150	0.75325
Realized GARCH Expanding	0.10753	0.21313	0.66234
Realized GARCH Rolling	0.12506	0.23889	0.63636
ARMA-GARCH Expanding	0.00013	0.00114	1.00000
ARMA-GARCH Rolling	0.00000	0.00000	1.00000

Table 9: This table shows the means and stddevs of Mincer-Zarnowitz p-values, as well as the ratio of cases in which the t-value is lower than 0.05 (i. e. the ratio of biased tests).

10 Value at risk estimation

Table 10 shows the means of Kupiec's test p-values for each level of α. Since the null hypothesis of the Kupiec's test is that the probability of an exception is similar to the significance level α, the bigger the p-value, the more accurate the VaR measure. We can clearly see that for all levels of α, GARCH(1,1) performs the worst, followed by realized volatility autoregressive model. The results for Realized GARCH and HAR models change depending on the VaR level. However, it is important to keep in mind that with a yearly forecasting period, that is, $n = 252$, any model suffers from small sample period for the higher levels of α since, for example the expected number of violations of the VaR level for the 99% level is $1\% \cdot 252 = 2.52$, so there is a huge sensitivity for tiny changes in results.

Table 11 shows the means of Christoffersen's test p-values for each model and all levels of α. The ranking of models is identical to that of Kupiec's test for all significance levels, which suggests that the VaR methods perform equally well in terms of unconditional coverage and independence.

Table 12 shows the means of the Dynamic quantile test of Engle and Manganelli (2004).

More discussion

Table 13 shows the hit rate of each model for all significance levels. For a "perfect" VaR forecast model, the hit rate should be exactly equal to $1-\alpha$, i. e. for a $\alpha = 90\%$, the hit rate should be 10%, for $\alpha = 95\%$, hit rate should be 5% and for $\alpha = 99\%$, the hit rate of a perfect model should be 1%. We can clearly see that hit rates are lower than expected for all models, which means that the VaR models are overly conservative and the estimated intervals of returns are too wide. We can see that as stated in the theoretical discussion, In terms of ranking the performance of each model, since all the hit rates are lower than expected, a bigger hit rate signifies a better model. Therefore, we can see that, on average, GARCH and autoregressive realized volatility model perform the worst and realized GARCH model the best with the HAR measures in between. This is true for the 90% and 90% significance levels. The results for 99% significance are slightly different, for example showing a great difference in performance of realized GARCH model depending on the forecasting window, however, we need to keep in mind that on this significance level for a yearly forecast period the results suffer from small sample problem, as mentioned before.

	90% Kup.	95% Kup.	99% Kup.
AR(1)-RV expanding	0.10002 (3)	0.04706 (3)	0.05991 (3)
AR(1)-RV rolling	0.10420 (4)	0.05176 (4)	0.07234 (4)
HAR expanding	0.16606 (8)	0.15284 (7)	0.18888 (7)
HAR rolling	0.17008 (12)	0.16668 (9)	0.19808 (9)
HAR-AS expanding	0.15937 (5)	0.15830 (8)	0.19960 (11)
HAR-AS rolling	0.16538 (7)	0.17205 (12)	0.20176 (12)
HAR-RSV expanding	0.16510 (6)	0.17071 (10)	0.19306 (8)
HAR-RSV rolling	0.16821 (11)	0.17075 (11)	0.19903 (10)
HAR-RSRK expanding	0.16693 (10)	0.18845 (14)	0.21889 (14)
HAR-RSRK rolling	0.16693 (10)	0.18845 (14)	0.21889 (14)
RGARCH expanding	0.20837 (14)	0.12720 (6)	0.09496 (6)
RGARCH rolling	0.20588 (13)	0.12468 (5)	0.09496 (6)
GARCH expanding	0.01880 (2)	0.01095 (1)	0.03087 (2)
GARCH rolling	0.01655 (1)	0.01166 (2)	0.02764 (1)

Table 10: This table shows the means of Kupiec's test p-values for each model and forecasting window for each respective VaR level. In parentheses is the rank of the value across all models.

	90% Chr.	95% Chr.	99% Chr.
AR(1)-RV expanding	0.08825 (3)	0.05586 (3)	0.13892 (3)
AR(1)-RV rolling	0.09283 (4)	0.06260 (4)	0.15601 (4)
HAR expanding	0.15281 (10)	0.14165 (7)	0.31898 (7)
HAR rolling	0.15835 (12)	0.15195 (8)	0.33000 (8)
HAR-AS expanding	0.15112 (6)	0.15595 (9)	0.33620 (10)
HAR-AS rolling	0.15429 (11)	0.16691 (12)	0.34092 (12)
HAR-RSV expanding	0.14514 (5)	0.15648 (10)	0.33215 (9)
HAR-RSV rolling	0.15124 (7)	0.15807 (11)	0.33711 (11)
HAR-RSRK expanding	0.15175 (8)	0.18203 (14)	0.35321 (14)
HAR-RSRK rolling	0.15175 (8)	0.18203 (14)	0.35321 (14)
RGARCH expanding	0.22610 (14)	0.12982 (6)	0.19846 (6)
RGARCH rolling	0.22215 (13)	0.12934 (5)	0.19846 (6)
GARCH expanding	0.02398 (2)	0.02112 (1)	0.09157 (2)
GARCH rolling	0.02063 (1)	0.02170 (2)	0.08551 (1)

Table 11: This table shows the means of Christoffersen's test p-values for each model and forecasting window for each respective VaR level. In parentheses is the rank of the value across all models.

	90% DQ	95% DQ	99% DQ
AR(1)-RV expanding	0.02164 (1)	0.08909 (2)	0.88147 (10)
AR(1)-RV rolling	0.03213 (2)	0.07896 (1)	0.88002 (9)
HAR expanding	0.17511 (8)	0.19455 (10)	0.87989 (8)
HAR rolling	0.18470 (12)	0.19383 (9)	0.86790 (7)
HAR-AS expanding	0.17970 (10)	0.19367 (8)	0.84439 (1)
HAR-AS rolling	0.18332 (11)	0.20782 (12)	0.85628 (4)
HAR-RSV expanding	0.17034 (5)	0.19077 (7)	0.86785 (6)
HAR-RSV rolling	0.17897 (9)	0.19922 (11)	0.86512 (5)
HAR-RSRK expanding	0.17286 (6)	0.18964 (6)	0.85202 (2)
HAR-RSRK rolling	0.17286 (6)	0.18964 (6)	0.85202 (2)
RGARCH expanding	0.26334 (14)	0.22584 (13)	0.90812 (11)
RGARCH rolling	0.25349 (13)	0.22693 (14)	0.91187 (12)
GARCH expanding	0.07639 (4)	0.15306 (3)	0.92574 (13)
GARCH rolling	0.07455 (3)	0.15655 (4)	0.92734 (14)

Table 12: This table shows the means of Dynamic quantile test p-values for each model and forecasting window for each respective VaR level. In parentheses is the rank of the value across all models.

	90% Hit rate	95% Hit rate	99% Hit rate
AR(1)-RV expanding	0.05550 (3)	0.02128 (1)	0.00294 (12)
AR(1)-RV rolling	0.05628 (4)	0.02159 (2)	0.00309 (14)
HAR expanding	0.06380 (6)	0.02850 (8)	0.00278 (7)
HAR rolling	0.06401 (7)	0.02891 (10)	0.00278 (7)
HAR-AS expanding	0.06406 (8)	0.02773 (5)	0.00283 (9)
HAR-AS rolling	0.06468 (11)	0.02798 (6)	0.00278 (7)
HAR-RSV expanding	0.06354 (5)	0.02829 (7)	0.00247 (3)
HAR-RSV rolling	0.06406 (8)	0.02860 (9)	0.00253 (4)
HAR-RSRK expanding	0.06468 (11)	0.02912 (12)	0.00289 (10)
HAR-RSRK rolling	0.06468 (11)	0.02912 (12)	0.00289 (10)
RGARCH expanding	0.07782 (14)	0.03242 (14)	0.00273 (5)
RGARCH rolling	0.07782 (14)	0.03195 (13)	0.00294 (12)
GARCH expanding	0.05344 (1)	0.02262 (4)	0.00180 (2)
GARCH rolling	0.05401 (2)	0.02257 (3)	0.00175 (1)

Table 13: This table shows the means of hit rate for each model and forecasting window for each respective VaR level. In parentheses is the rank of the value across all models.

11 Areas for further study

For the sake of parsimony, we used only the base GARCH(1,1) model. However, since it has been shown that other GARCH-family models outperform GARCH in both simple volatility forecasting as well as in VaR forecasting, we should also include these models for a more reliable comparison. Furthermore, we can could also include other used VaR estimation models than the volatility-based.

It has been shown that financial time series express long memory, therefore, our focus should also aim towards long memory methods such as the FIGARCH or FIEGARCH. For the same reason, rather than using the simple first-order autoregressive model for realized volatility, we could attempt to study an ARFIMA model which addresses the long memory property of financial time series.

As a next point, for the realized volatility, we only used the data computed on one time granularity (5-minutes). There is a relevant literature which compares the performance of volatility models, depending on the granularity which the realized volatility is computed with. Therefore, we can also take this approach and check whether there is a difference in results depending on the selected granularity.

Reference

A more novel downside risk which is gaining popularity in the recent years is the expected shortfall which can also be estimated in various ways utilizing estimated volatility, so new research can focus on estimating estimating the expected shortfall rather than value-at-risk and we can see whether the ranking of comparable methods is similar or it changes depending on what risk measure we use.

Check whether ES can be really estmiated using volatility

In this thesis, we used a subsample of 77 stocks. For a more realiable outcome, we could perform the same study on stock indices like for example So and Yu (2006), since stock indices better capture the overall behavior of market and are more resistant to shocks in individual stocks. Alternatively, we could use data on all available stocks to study the whole market. However, this would be computationally very intensive which is why only a subsample was selected for the purpose of this study.

We took all the data without any smoothing for jumps or outlier values. A common method when studying fnancial time series is to use jump-corrected data using e. g. MedRV of Bipower variation in order to mitigate the influence of these outliers on results.

In the GARCH and RGARCH model, we chose monthly refitting scheme due to high computational intensity. It would be interesting to study each model in terms of how the results change with different refitting frequencies, like for example

Reference - there was an article which does that.

12 Conclusion

References

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1):135–167.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). Modeling and Forecasting Realized Volatility. NBER Working Papers 8160, National Bureau of Economic Research, Inc.

Aït-Sahalia, Y. and Mancini, L. (2008). Out of sample forecasts of quadratic variation. *Journal of Econometrics*, 147(1):17–33. Econometric modelling in finance and risk management: An overview.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

Bandi, F. M. and Russell, J. R. (2008). Microstructure Noise, Realized Variance, and Optimal Sampling. *The Review of Economic Studies*, 75(2):339–369.

Barndorff-Nielsen, O. E., Kinnebrouk, S., and Shephard, N. (2010). *Measuring downside risk: realised semivariance*, pages 117–136. Oxford University Press, (edited by t. bollerslev, j. russell and m. watson) edition.

Barndorff-Nielsen, O. E. and Shephard, N. (2002a). 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.

Barndorff-Nielsen, O. E. and Shephard, N. (2002b). 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.

Beltratti, A. and Morana, C. (1999). Computing value at risk with high frequency data. *Journal of empirical finance*, 6(5):431–455.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547.

Bollerslev, T., Diebold, F., Andersen, T., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579–625.

Brownlees, C. T. and Gallo, G. M. (2009). Comparison of Volatility Measures: a Risk Management Perspective. *Journal of Financial Econometrics*, 8(1):29–56.

Campbell, S. (2005). A review of backtesting and backtesting procedures. *Journal of Risk*, 9.

Christensen, K. and Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141(2):323–349.

Christoffersen, P. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2:84–108.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Danielsson, J. and De Vries, C. G. (2000). Value-at-risk and extreme returns. *Annales d'Economie et de Statistique*, pages 239–270.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business Economic Statistics*, 20(1):134–144.

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

Dowd, K. K. (1998). *Beyond value at risk*. Frontiers in Finance Series. John Wiley & Sons, Nashville, TN.

Dumitrescu, E.-I., Hurlin, C., and Pham, V. (2012). Backtesting value-at-risk: From dynamic quantile to dynamic binary tests. *Finance*, 33.

Ebens, H. (1999). Realized Stock Volatility. Economics Working Paper Archive 420, The Johns Hopkins University,Department of Economics.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1):3–27.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica*, 55(2):391–407.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business Economic Statistics*, 22(4):367–381.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56.

Gallo, G. M. (2001). Modelling the Impact of Overnight Surprises on Intra-daily Volatility. Econometrics Working Papers Archive wp2001-02, Universita' degli Studi di Firenze, Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti".

Giot, P. and 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., and 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.

Hansen, P. and Lunde, A. (2001). A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics, 20 (7), 873-889*, 20.

Hansen, P., Lunde, A., Shephard, N., and Barndorff-Nielsen, O. (2008). Designing realised kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica*, 76:1481–1536.

HANSEN, P. R., HUANG, Z., and SHEK, H. H. (2012). Realized garch: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Ho, L.-C., Burridge, P., Cadle, J., and Theobald, M. (2000). Value-at-risk: Applying the extreme value approach to asian markets in the recent financial turmoil. *Pacific-Basin Finance Journal*, 8(2):249–275.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259.

Jorion, P. (2006). *Value at risk, 3rd ed.* McGraw-Hill Professional, New York, NY, 3 edition.

Kuester, K. (2005). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4:53–89.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. Finance and Economics Discussion Series 95-24, Board of Governors of the Federal Reserve System (U.S.).

Martens, M. (2002). Measuring and forecasting sp 500 index-futures volatility using high-frequency data. *Journal of Futures Markets*, 22:497 − 518.

Martens, M. and van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, 138(1):181–207. 50th Anniversary Econometric Institute.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17:479–508.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Mincer, J. A. and Zarnowitz, V. (1969). *The Evaluation of Economic Forecasts*, pages 3–46. NBER.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Oomen, R. C. A. (2005). Properties of Bias-Corrected Realized Variance Under Alternative Sampling Schemes. *Journal of Financial Econometrics*, 3(4):555–577.

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

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

So, M. K. P. (2000). Long-term memory in stock market volatility. *Applied Financial Economics*, 10(5):519–524.

Wong, C.-M. and So, M. K. (2003). On conditional moments of garch models, with applications to multiple period value at risk estimation. *Statistica Sinica*, pages 1015–1044.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12(6):1019 – 1043.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business Economic Statistics*, 14(1):45–52.