HAR(d) to beat?

Forecasting volatility: A comparison of the HAR model and actual volatility in the Netherlands

Radboud Universiteit Nijmegen

Master Thesis in Economics

Corporate Finance and Control

Author: Nasos Thanasoulas (s 1002576)

Supervisor: Dr. Dirk - Jan Janssen

Academic year 2018 – 2019

Acknowledgement

First of all, I would like to thank my supervisor Dr. Dirk-Jan Janssen for his support, motivation, and inspiration on the topic of my thesis. I am grateful for his knowledge, guidance, and patience.

With the completion of this thesis, one of my dreams is coming true. It was a two-year journey with many difficulties, from different aspects, that is coming to an end and makes me feel proud of myself. I went through this journey with my fellow teammates, and we all supported each other until the end. So special thanks to all of you guys!

I want to explicitly thank my parents, my dad Nikolaos and my mom Evanthia that raised me with good manners and love, and they unconditionally supported me in every choice of my life. Lastly, I would like to thank my two lovely sisters. The young of our family, Dimitra, who was always encouraging me, and the oldest one Andriani. Although Andriani was never able to understand and speak to me because she is a person with special needs, I am sure that she is proud of me too!

Contents

Abstract

This thesis is testing the forecasting performance of the heterogeneous autoregressive model for realized volatility (HAR-RV), against the generalized autoregressive conditionally heteroskedastic model (GARCH (1, 1)), and the volatility derived from the volatility index in the Netherlands. Using data from the AEX index for the period 2000 to 2018, it has been found that the HAR-RV model was able to better forecast volatility for this period against GARCH (1, 1) and VAEX. The same results were produced when the models were tested in the two periods of crisis, namely 2000 to 2002 and 2007 to 2009. The daily out of sample forecasting performance of the models was based on a 252 days training period. The mean squared error (MSE) and the mean absolute error (MAE) methods have been used to estimate the forecasting performance of the models against the actual realized volatility. The indices of the S&P 500 and Nikkei 225 and their respective volatility indices have been tested for the same periods as a robustness check. The performance of the HAR-RV model was again superior against the GARCH (1, 1) model and the respective volatility indices for all the periods.

1. Introduction

The topic of forecasting volatility has attracted a lot of attention in the last decade from many academics and also financial professionals. It has been a subject of great discussion over the years and a lot of research has already been done. Volatility is the most important component when it comes to derivative pricing, improvement of portfolio and value at risk analysis (Gospodinov, Gavala & Jiang, 2006). Being able to carefully model and forecast volatility is crucial when dealing with risk and asset management but also with option pricing. Volatility is a statistical measure that is widely implemented in the financial sector. More specifically, it is frequently used for hedging (Brenner, Ou & Zhang, 2006) the pricing of derivatives and risk management (Antonakakis & Darby, 2007; Christoffersen & Diebold, 2000). Until today the fluctuation of the volatility has not been able to be depicted in a proper pattern (Parasuraman & Ramudu, 2011).Commonly, volatility is considered as a way of measuring the risk or the variability of an underlying asset (Härdle & Silyakova, 2012). The higher the volatility, the riskier the market index or underlying security. If we are able to predict where volatility can go, then we can automatically strengthen the ability to take the correct financial decision. For example, the volatility index (VIX) is an index that shows what the expectation on thirty-day forward-looking volatility can be. This index or similar indices for other stock markets are often used by investors and analysts as a measure of market risk when they want to take financial decisions. Gonzalez & Novales (2007) found that the volatility index of the Spanish market was able to capture the level of risk and help the market participants on taking financial decisions. For example, many institutions want to know the current values of the portfolio that they manage but also to be able to have accurate predictions of the future values of them. So forecasting volatility is essential also for institutions that are involved with portfolio management and options trading. Volatility can also be used from new investors, because it can provide useful insights as to what are the differences between high risk and investment. It is common for new investors to wrongly assume that investment risk is a quantifiable and well-defined concept, while in reality there seems to be no consensus about what investment risk really is and how it should be measured (Balzer, 1994). Therefore, in order for investors to be able to make more accurate decisions, it is important to know which model that can be used in forecasting volatility is producing the best results. All these processes require a reliable measure of past and future volatility. Furthermore, volatility receives a great deal of concern from public policy makers in their continuous attempt to stabilize financial markets and the economy as a whole (Fatas & Mihov, 2005).

Oftentimes in finance there is confusion between volatility, standard deviation and risk. It is important to make these concepts clear. In finance, volatility is often used when we refer to standard deviation (σ) or when we talk about variance (σ^2). The variance is calculated from the formula below:

$$\sigma^2 = \frac{1}{N-1}\sum_{t=1}^{N}(R_t - \bar{R})^2 \qquad (1)$$

It is formulated as the squared sum of the return minus the average return divided by the number of returns minus one. Where R_t is the return in a specific period chosen at time t, and $\bar{R}$ is the average return over that period. The link between risk and volatility is subtle. When talking about risk it is most of the times associated with high or low returns and can be described as the chance that the actual return of an investment will diverse from the expected actual return (Rubaltelli, Ferretti & Rubichi, 2006). Since it has been depicted how the variance is calculated, the standard deviation can be derived from the square root of the above equation (1). As mentioned by Ding, Granger& Engle (1993), it is better to measure volatility directly from absolute returns[1].

There is a distinction between realized volatility, which is the volatility of a security in a specific period based on historical data, and implied volatility, which depicts the current market value of volatility over a particular period of time based on the expected movements of the market. There are models that can predict volatility based on realized volatility. However, realized volatility by itself is not a predictor of volatility, but just simply the actual historical volatility that can be used in a model, to predict future volatility. On the other hand, implied volatility is the future expected volatility that is derived from financial instruments such as options. It has been found that implied volatility is outperforming a model that predicts future volatility based on realized volatility when it comes to forecasting future volatility (Christensen & Prabhala, 1998). Jorion (1995) has found

1According to Davidian & Carroll (1987) absolute returns are more robust against non-normality and asymmetry.

that implied volatility for options on foreign currency futures is efficient but at the same time biased estimator of the realized volatility[2].

Besides measuring volatility using the VIX or the realized volatility, another way can be considered to model volatility. One of the most common approaches of modeling volatility indirectly is using ARCH or GARCH models, but nowadays with the realized measures, it became possible to directly model volatility. One of the models that directly uses the realized measures to forecast volatility is the HAR model. The major idea of this model is that investors with different time horizons perceive and react to different types of volatility. It is a model that has a simple structure; it is easy to estimate and is able to replicate the main features of financial data (Corsi, 2003). The HAR model is basically an additive cascade of realized volatilities, generated at different time horizons, that follows an autoregressive process.

Although there has been a lot of research regarding the predictability of volatility, the topic still remains controversial and there is still uncertainty on which one is the best forecasting model. As mentioned before the HAR model has the advantage of being a simple model to estimate and able to reproduce many of the features of volatility data. Several researchers have previously focused on pre-crisis periods (Deo, Hurvich & Lu, 2006; Corsi, 2009), while there has been some research on after crisis periods as well (Vortelinos, 2017). The majority of the studies (Sharma & Vipul, 2015; Chin, Lee & Yap, 2016; Wen, Gong & Cai, 2016; Ma, Wei, Huang & Chen, 2014; Seďa, 2013) cover both crisis periods; however either they compare only a limited amount of forecasting volatility models (Sharma & Vipul, 2015; Wen et al., 2016; Deo et al., 2006; Ma et al., 2014; Chin et al., 2016) in their analysis or use data from only one specific country stock index (Chinet al., 2016; Seďa, 2013; Ma et al., 2014). When studies compare several autoregressive models the results seem to show that the HAR model has an advantage over the others in forecasting realized volatility (Vortelinos, 2017; Seďa, 2013; Corsi, 2009). In general models are better able to explain volatility when there is a stable financial environment and they usually break up in crisis time when

[2]The reason for that according to Jorion (1995) are some measurement errors and statistical problems. To be clearer, the underlying assumptions of an OLS regression must hold in order for the OLS procedure to produce the best possible estimates. The estimators that produce unbiased results and have the smallest variance are meant to be efficient (Berry, 1993). Efficient means that the distribution around the actual value gets smaller and smaller. However, the word biased as referred above from Jorion (1995) means that the estimator does not have its mean centered around the actual value. So, the more efficient the estimator is the narrower the distribution will lie around the actual value.

volatility is higher (Angabini & Wasiuzzaman, 1997; Banulescu, Hansen, Huang & Matei, 2018). Nevertheless, the lack of testing a limited variety of models on data from a limited amount of stock indices constraints the generalizability of the results of previous studies.

The goal of this paper is to examine whether the Heterogeneous Autoregressive (HAR) model for realized volatility (RV) compared to the GARCH model, the observations from the volatility indices and the actual volatility can better forecast volatility. Furthermore, special consideration will be paid on how the HAR model will perform when using data from two representative indices of America and Asia, and more specifically it will be checked if the results are similar or not with the ones derived from the Netherlands. This extra analysis and comparison of the two other indices will be used as a robustness check for the performance of the model. The GARCH model has been used in many studies for forecasting volatility, but surprisingly the HAR model has not been used that often. This might not only be of academic interest but also of practical interest as it can be a source of motivation for practitioners to develop advanced pricing models or algorithms for trading purposes. Most of the studies focused on a specific country or continent which limits the ability to generalize the conclusion that the HAR model is a better estimator than the GARCH model or the actual volatility. In summary, finding a model that properly forecasts volatility has been a going concern for both academic researchers and market professionals. Because of this going concern, it is of paramount importance to keep using the latest data so that we can notice distinctions in the predicting ability of the models between past and current time periods. So after taking into consideration the above research, the following research question has been formulated:

How well does the HAR model forecast volatility compared to the GARCH model and VAEX?

To provide a small overview of the results, and by using the MSE and MAE methods to compare them, the HAR model was able to better forecast volatility against the GARCH (1, 1) and the VAEX for the whole period of 2000 to 2018. The same forecasting performance was observed when the two periods of crisis were tested. Following the introduction part (section 1) the remainder of this thesis is organized as follows. Section 2 starts with the research methods based on the two models and the analysis of the concepts of volatility indices, implied volatility and realized volatility. Section 3 is providing a small comparison of the performance of the two models according to the literature and the motivation and hypothesis of this thesis. Section 4 describes the data that have been used for this research, their treatment, and the stylized facts of financial data. Section 5 is providing the methods used to compare the forecasting performance of the models and gives a presentation and analysis of the empirical results. In section 6 all the limitations that this research has been encountered are presented. Section 7 provides a summary and the main conclusions and findings of the analysis. Furthermore section 8 concludes this thesis and provides suggestions for future research. Finally the bibliography and the appendices are presented.

2. Research Methods

2.1. The HAR-RV model

The heterogeneous autoregressive model (HAR) was first introduced by Corsi (2003), and the primary purpose was to directly model and forecast the behavior of volatility in time series data. In general, the model appears to have a simple structure and is able to replicate the main features of financial data (Corsi, 2003). The main inspiration for the creation of the model stems from the heterogeneous market hypothesis and the asymmetric reproduction of volatility between long and short term perspectives, which takes into consideration volatilities realized in different periods (Corsi, 2003). It is a short memory process model which can produce the scope of modeling the long-memory performance of volatility in a straightforward and prudent way[3]. In general it is a simple auto regressive model for realized volatility which takes into consideration volatilities that have been realized over several horizons. What Corsi (2003) did was to focus on the daily realized volatility and predict the volatility of the next day based on this. As mentioned before the model stems from the heterogeneous market hypothesis which simply means that agents are not identical. Because of this heterogeneity the reaction to news can be different and thus cause different volatility components. HAR model assumes that volatility can be depicted as the sum of volatilities created by specific groups of market players with each of them having different time boundaries.

[3] Short memory process is defined in terms of no perseverance of observed autocorrelations, in contrast with the long memory process where we have persistence of observed correlations (Rossi, 2012). "Given the long memory and relatively slow decay of a response to a lagged squared innovation, the effect of pre-sample values might be expected to have a bigger impact than with stationary GARCH processes" (Baillie, Bollerslev & Mikkelsen, 1996).

When referring to volatility one has to keep in mind that there are two different kinds of volatility terms. The realized volatility which is also called historical volatility, and the implied volatility which is the volatility derived from the options market. In the recent years, with many high frequency financial data being easily available, the concept of modeling realized volatility has become an innovative and interesting research direction (Corsi, Mittnik, Pigorsch & Pigorsch, 2008). The component that the HAR model is using to forecast future volatility is the realized volatility. Realized volatility basically measures what happened in the past. It is the sum of squared intraday returns and according to Hansen &Lunde (2006) it is an ideal estimation of volatility considering that prices are observed continuously and without any measurement error. According to Andersson & Bollerslev (1998) when using intraday squared returns of five minutes or higher intervals, a proper measure of the latent mechanism that characterizes volatility can be estimated. As Taylor (2005) mentioned, volatility during a specific horizon can be more precisely estimated if the frequency of the returns increases. Assuming that volatility is constant the formula to calculate realized variance is the one below:

$$RV_t = \sum_{j=1}^{n} r_{t,j}^2 \qquad (2)$$

Thus, the realized volatility is the square root of the above equation (2).

As mentioned before the HAR model has a simple structure. The whole model has been built on three different time horizons which are daily, weekly (we account for 5 trading days) and monthly (we account for 22 trading days). The model is modeling the realized volatility of tomorrow based on the realized volatility of yesterday, the realized volatility of last week and the realized volatility of last month. Using this cascade framework, the model looks like this:

$$RV_{t+1d}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \varepsilon_{t+1d}^{(d)} \qquad (3)$$

After getting the realized volatility from the data, the realized volatility of yesterday, last week and last month will be used to estimate the realized volatility of tomorrow. The calculation of $RV_t^{(w)}$ and $RV_t^{(m)}$ is the following:

$$RV_t^{(w)} = \frac{1}{5}\left(RV_{t-1}^{(d)} + RV_{t-2}^{(d)} + RV_{t-3}^{(d)} + RV_{t-4}^{(d)} + RV_{t-5}^{(d)}\right) \tag{4}$$

$$RV_t^{(m)} = \frac{1}{22}\left(RV_{t-1}^{(d)} + RV_{t-2}^{(d)} + \cdots + RV_{t-22}^{(d)}\right) \tag{5}$$

So from the above equations, the formula for h step ahead forecast can be derived:

$$RV_{t+h}^{(d)} = c + \beta^{(d)} RV_{t+h-1}^{(d)} + \beta^{(w)} RV_{t+h-1}^{(w)} + \beta^{(m)} RV_{t+h-1}^{(m)} + \varepsilon_{t+h} \tag{6}$$

We do not know what is going to happen tomorrow, but we want to predict it with the equation (3). So as a first step the model has to be estimated on the in sample data. The model can easily be estimated by an Ordinary Least Squares (OLS) regression. The realized volatility of tomorrow will be used as the dependent variable and the realized volatility of yesterday, last week and last month as the independent variables. After running the regression, the betas for each independent variable will be derived. After deriving the betas, it will be tested whether the model can predict because until now nothing has been predicted but just fitted the model. To predict the volatility of the next day, we need at least a window of 22 trading days. The first day (after this window) from the data will be taken as a starting point plus the volatility of last week plus the volatility of last month. So in this case, the volatility of the next trading day will be predicted from the model until the last date of the data which is the 31st of December 2018. The betas that derived before will be multiplied by the volatility of each time period in the model to get the volatility of tomorrow. The estimation period will then be rolled forward by adding one new day every time and dropping the most far-off day. By doing this, the sample size that will be used to estimate the model will remain at a locked length, and there were be no overlap at the forecasting. Thus it will allow one day ahead (out of sample) volatility to be obtained. This procedure is called rolling regression. At this point, it should be mentioned that the new model has to be re-estimated and the new betas have to be derived before moving to the next forecasting day.

The same procedure will be followed to estimate the forecasting power of the model for the two different periods of crisis. The first crisis period is running from the beginning of 2000 until mid-2002, and the second period of crisis from 2007 to 2008 and maybe affected also 2009.

2.2. The GARCH model

The GARCH model was first introduced by Bollerslev (1986) & Taylor (1987). Back in these days the concept of realized volatility modeling was not even introduced. At that period the daily volatility was calculated as the squared daily return without taking into consideration any subintervals. For example if an asset had a lot of fluctuation during one day, and its opening price happened to equal its closing price, then the volatility of the underlying asset was estimated to be zero. The GARCH model is a conditional volatility model which allows the conditional variance to depend on the previous lags[4]. It is based on the ARCH model by Engle (1982), who used it to show that the conditional volatility is affected by volatility clustering[5]. An autoregressive conditionally heteroskedastic (ARCH) model is a time series model with econometric applications that consider the variance of the current error term as a function of the variance of the error conditions of the previous time periods. One of the drawbacks of the ARCH model was that it responds slowly to large unusual shocks. Thus, the need of an improvement of this model was crucial. Assuming an autoregressive moving average (ARMA) model for the error variance, then the model is a generalized autoregressive conditionally heteroskedastic (GARCH) model. Many different versions of the GARCH model have been developed such as the, EGARCH, GJR-GARCH, TGARCH, NGARCH, and FIGARCH models. Each of these models has its own strengths and weaknesses since there are many assumptions and parameters involved. There will be no further analysis for them since the main focus of this paper will be on the simple GARCH (p, q) version. GARCH models were designed to deal with the problem of volatility clustering which is the phenomenon where large changes in prices tend to cluster together. As a result, there

[4] Most time series econometric models are operating with the assumption of variance being constant, in contrast with the GARCH process model that allows conditional variance to change over time and thus being a function of past errors (Bollerslev, 1986). In general, conditional variance can be described as the variance of a variable based on the value of one or more other different variables.
[5] The problem of volatility clustering is described in section 4.4.4.

is a persistence of the amplitudes of price changes. Although returns are not correlated in general, the absolute returns or the squared returns are showing a positive correlation (Cont, 2007). By using the GARCH model, we can model the conditional heteroscedasticity and the heavy-tailed distributions of financial markets data.

Before describing the GARCH model, the ARCH specification has to be introduced. The following return process has to be specified:

$$r_t = \mu_t + \varepsilon_t \text{ , with } \varepsilon_t = z_t\sqrt{\sigma_t} \qquad (7)$$

Where, μ_t is a drift term that is explained by the structural model and $z_t$is an independent shock with zero mean and unit variance signifying that ε_t is normally distributed ε_t~ Z(0,σ_t). The conditional variance in (7), can be transformed into time-varying by specifying the ARCH (q) process:

$$\sigma_t = c + \sum_{i=1}^{q} a_i \, \varepsilon_{t-1}^2 \qquad (8)$$

Where c is a constant and a_i is the coefficient for the past squared shocks (ε_t^2). Then the GARCH (p,q) model is derived by adding p lagged conditional variances, with orders p ≥ 1 and q ≥ 1 :

$$\sigma_t = c + \sum_{i=1}^{q} a_i \, \varepsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j \, \sigma_{t-j} \qquad (9)$$

Where β_j is the coefficients for the past conditional variances, p is the past squared error terms and q is the past estimated volatility terms. When q=0 then the above equation (9) reduces to an autoregressive conditional heteroskedastic (ARCH) model.

Given a distribution of ε_t in equation (7) and setting p=q=1 then the GARCH (1, 1) is derived:

$$\sigma_t = c + a_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1} \qquad (10)$$

For which the condition $c \geq 0$, $a_1 \geq 0$ and $\beta_1 \geq 0$ should stand for every positive value of σ_t.

Since the GARCH model is non-linear, it cannot be estimated by an OLS regression like the HAR model. Thus, the Gaussian maximum likelihood (GMLE) method should be used for parameter estimation[6]. When assuming normally distributed errors and starting from some parameter vector θ and a time series of size T (y_1,y_2.... y_T) the GMLE method calculates the probability density for this specific sample by taking the product over all the marginal conditional probability densities of the observed data. In general, the GARCH model is using the returns to forecast volatility, and it depicts that today's return consists of yesterday's return plus some volatility part and this volatility is what we need. This model is also using a rolling regression method to forecast volatility, by moving one day ahead and leaving one day behind for every forecast, which means that the data window size remains stable. This volatility can be derived from the GARCH model using Time Series Modelling 4 (TSM4) software.

2.3. The Volatility Indices

The third parameter that this paper is going to look at and compare with the actual volatility is the data taken from the volatility indices. A volatility index is measuring the expectations of the market on future volatility of the underlying equity index (Siriopoulos & Fassas, 2009). As mentioned before the VIX is an index that shows what the expectation on thirty-day forward-looking volatility can be based on the S&P 500 index, and it is a real-time market index. Thus by having an accepted

[6] The maximum likelihood method is analyzed in appendix 10.1

quantitative measure for volatility, we have an advantage when contrasting price moves and potential risk correlated with different securities and markets. With the same principle like in VIX index, the VAEX and the JNIV are measuring the thirty-day forward-looking volatility on AEX and Nikkei 225 indices respectively. Bluhm & Yu (2001) found that the German volatility index, named VDAX, is better at forecasting volatility compared to GARCH model. Fleming, Ostdiek & Whaley (1995) have also concluded that VIX is better at forecasting volatility in comparison to other historical measures. On the other hand, Kambouroudis, McMillan & Tsakou (2016) tested the forecasts of implied and realized volatility against ten GARCH models. Although both implied and realized volatility encompass significant information regarding future volatility, the GARCH models were able to better predict volatility for in and out of sample data. Kumar & Verma & Gupta (2016) tested the forecasting ability of GARCH model against implied volatility in option pricing and they found that the GARCH model is better than implied volatility. Chung, Sun & Shih (2008) tested if the HAR model and the mixed data sampling (MIDAS) regression can outperform implied volatility model. After checking their results based on the S&P 500 index for the period 1995 to 2005, which encompasses the financial crisis of 2000 to mid-2002, they concluded that implied volatility has more information content and as a result higher forecasting capacity than the out of sample volatility forecasts from the HAR and the MIDAS.

Market makers are facing the issue of hedging the volatility risk (McDonald, 2013). As mentioned before implied volatility can be the most accurate estimation of the volatility for an asset. Implied volatility is derived from the options market[7]. Commonly, with stocks we just have the realized volatility, which describes how much a stock has changed in percentage terms. If we want to price an option, then volatility has to be used as an input, but volatility is not observable and this can raise the question of how is possible to price an option in practice. This can be done by calculating historical volatility based on historical returns. If we follow this procedure then we will probably face the problem of expected future volatility being different than historical volatility (McDonald, 2013). The reason for that is that there might be some periods that investors are expecting high uncertainty due to political turmoil or some government information releases. So it is not possible to always rely on history in order to get the most reliable estimation s of future volatility.

[7] An option is a claim that an investor can use to speculate and hedge on the future value of a stock price (McDonald, 2013). There are two types of options: call option and put option.

The price of an option should be able to demonstrate the expectations of the market regarding the distribution of the future stock price. The best way to derive the price of an option is to calculate the option's implied volatility. It is the volatility that will be extracted by using the Black and Scholes formula of pricing options assuming that volatility is constant. Usually implied volatility is deviating from the historical volatility values (Parasuraman & Ramudu, 2011). If we know the price of a put or a call option the Black Scholes formula implied volatility is the unique parameter of volatility for which the formula reclaims the price of the option (Lee, 2005). One of the main components of the Black Scholes formula is the strike price of the underlying asset which also has to be assumed to be constant[8]. In case there is a variation in the exercise prices (or strike price, as mentioned above), then different implied volatilities will be produced (Guo & Su, 2004). This phenomenon is generally known as volatility skew and has a pattern which also vary depending on the status of the option. An option can be in the money or out of the money. When the out of the money and the in the money options are having higher implied volatility than the at the money option then it is called volatility smile. In the opposite scenario, when in and out of the money options are having lower implied volatility than the at the money option, it is called volatility sneer (Guo & Su, 2004). In theory volatility skew scenarios are a bit ambiguous since implied volatility should not be dependent on the options' exercise prices (Guo & Su, 2004).

The Black and Scholes formula is discussed and analyzed below. The Chicago Board Options Exchange (CBOE) started reporting in 1993 an index regarding implied volatility which is called 'VIX'. Since 2003 this index has been named as 'Old VIX' and began reporting the implied volatility for the S&P 500 index (McDonald, 2013). In general implied volatility varies over time.

The volatility indices are directly computed by the options exchange. The volatility index (VIX) is based on the Black and Scholes (1973) option valuation formula which is built up from price inputs from the S&P 500 index (Siriopoulos & Fassas, 2009). The volatility of the market can be observed through the volatility index. Given the Black and Scholes formula and saying that there is volatility in there we get the price for the options. The formula is based on some strong and unrealistic assumptions, but some extensions of the formula according to (Wilmott, Howison, & Dewynne, 1995) were able to overcome these constraints. Only the assumption that the volatility is constant is the strongest. Under the same principle, the volatility indices for the Amsterdam Exchange

[8] Strike price is defined as the price in which a derivative contract can be sold or bought

(VAEX) and Nikkei 225 (N225) are also obtained. After getting the price for the options the implied volatility can be derived.

The formula of Black and Scholes (1973) to find the price of a European call option (c) is:

$$c = S_0 e^{-\delta t} N(d_1) - K e^{-\mathrm{r}t} N(d_2) \qquad (13)$$

Where: K is the strike price, $S_0$is the price of the underlying asset, T is the time to maturity, δ is the dividend yield, r is the rate and N(x) is the cumulative probability distribution function of the normal distribution, where:

$$d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r - \delta + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \qquad \text{and} \qquad d_2 = d_1 - \sigma\sqrt{T}$$

After obtaining the value for the call option the value of a European put option can be easily derived based on the put-call parity which states that the price of the call plus the present value of the strike price is equal to the price of the put plus the value of the underlying discounted at the dividend yield:

$$c + K e^{-\mathrm{r}t} = p + S_0 e^{-\delta t} \qquad (14)$$

Given the assumption that options markets are efficient, implied volatility should also be an efficient estimator of future volatility (Christensen & Prabhala, 1998).

3. Literature review

3.1. Introduction

As already mentioned in the introduction, there is an ongoing debate in the forecasting volatility literature regarding which is the best model or method to use. By analyzing the findings from other studies in the next section, a better idea of the underlying issue will be provided in order to give clearer view of the topic. There are three different ways to model volatility, and these are the stochastic volatility models, the Autoregressive Conditional Heteroscedasticity models, and the realized volatility, in this case, the HAR model. The main focus will be on the HAR model as it is the focal point on this thesis. The distinction between realized and implied volatility is already depicted at the previous section. The HAR model will be compared based on the literature review with the GARCH (1, 1) model and will end with the motivation and hypothesis of this thesis.

3.2. HAR and GARCH model comparison

Taking a step forward to the predictive power of the above mentioned models and looking specifically to the HAR model, Corsi (2009) found that when testing the HAR model on FX rates USD/CHF data, the model showed exceptionally favorable out of sample forecasting performance against standard models. Seda (2013) used the HAR-RV model to test its performance on data taken from the Czech stock market (PX) index for the period 2004 to 2012. He split his dataset in three sub periods, namely pre-crisis, crisis and post-crisis periods. He tested the HAR model against the simple autoregressive (AR) and GARCH (1, 1) models. He concluded that the HAR model showed excellent in-sample forecasting performance against the AR (1) and GARCH (1, 1) models for all the three periods that he tested. On the other hand, Chung, Sun & Shih (2008) tested if the HAR model and the mixed data sampling (MIDAS) regression can outperform implied volatility model. After checking their results based on the S&P 500 index for the period 1995 to 2005, which encompasses the financial crisis of 2000 to mid-2002, they concluded that implied volatility has more information content and as a result higher forecasting capacity than the out of sample volatility forecasts from the HAR and the MIDAS. So their results were not similar to what has been observed until now from previous studies. Seda (2012) used the HAR model to test its performance on the S&P 500 index in the U.S., but he made a small change on the model. He constructed the realized volatility to estimate the model both in standard and logarithmic form. After running his analysis, he found that the logarithmic form performs similarly to the standard version of the model and produces even better results. Audrino & Knaus (2016) tested the HAR model against the least absolute shrinkage and selection operator (Lasso). For their analysis they used data from several companies like Nike, Citigroup, Harley Davidson and Exxon Mobil Corporation from 2001 to 2010. They found that the HAR model was showing equal in and out of sample performance compared to the Lasso approach.

McAleer & Medeiros (2008) used a multiple regimes smooth transition extension of the HAR model which can approximate the long memory behavior of the volatility. They named this model as HARST model, and they tested it against the HAR and the alternative latent volatility models. In some cases, the HARST was performing better than the HAR model, and in other cases, the HAR was outperforming the HARST. When they tested against other volatility models, using data from 1994 to 2003, both the HAR and the HARST models were better at forecasting volatility and especially when they were combined. So the original HAR model and its extension, the HARST model, were both better able to forecast volatility in normal times but also during period of turmoil. Something similar was done by Huang, Gong, Chen & Wen (2013), where they converted the realized volatility into adjusted realized volatility and by that they created the HAR-ARV model. Using data from the Shanghai and Shenzhen stock markets for, 2007 to 2012 they concluded that the HAR-ARV model was better at forecasting that the original HAR-RV model. Ma, Wei, Huang & Chen (2014) used the HAR model for high-frequency data of the Shanghai Composite index for the period 2000 to 2013. They split their dataset into two subgroups, the in-sample and the out of sample. They compared and contrasted the forecasting performance of the HAR model with multifractal volatility, realized volatility, realized bipower variation and their analogous short memory model, and they found that the HAR model outperformed all other models[9]. Thus, their inferences were based on a long period with strong stock market performance, but also including two periods with financial downturns. According to Vortelinos (2017), which used a dataset from seven U.S financial markets from 2002 to 2011, the HAR model produced the best accurate

[9] The way of measuring the scale of the returns to change together with time in a stochastic manner is called realized power variation (Barndorff-Nielsen & Shephard, 2003). The extension of this is called bipower variation.

forecasts against the Principal Components Combining (PCC) model, Neutral Networks (NN) and Generalized Autoregressive Heteroskedasticity (GARCH) models. So, again the HAR model was able to better forecast volatility during normal and downturn economic periods.

Beside the original HAR model some researchers created and tested some other versions of the model. Chin et. al (2016) created a different version of the HAR model introducing a structural break heavy-tailed heterogeneous autoregressive model by improving it with jump-robust estimators[10]. The reason for doing that was that possible structural breaks often cause problems when we want to estimate volatility. They applied this model in data from the blue-chip stock market index of the 30 major German companies (DAX) for the period of 2008 to 2015, and they found that this version of the HAR model is performing better than the standard model and in general outperformed the other forecasting models. Thus the extension of the HAR model was performing better during stable economic conditions and during economic crisis. Jou, Wang & Chiu (2013) used the HAR model for option pricing against the NGARCH, which is considered as the best model in pricing options among the GARCH family models. They introduced the logarithmic HAR (log-HAR), which is more beneficial compared to other option pricing model that use realized volatility. The reason for that is that the log-HAR model is following a simpler estimation procedure in comparison with the other models. His analysis was based on data retrieved from the S&P 500 index for the period of 2007 to 2008, which is exactly the financial crisis period. He concluded that the HAR type models were able to better predict out of sample option prices than the GARCH type models. He also mentioned that this can be due to the reason that HAR models are closer to VIX index in financial markets since their base is realized volatility. Another version of the HAR model was introduced by Cubadda, Guardabascio & Hecq (2017), which was the Vector Heterogeneous Autoregressive Index Model (VHARI). The practicality of this model is that it can keep the same temporal cascade structure as the original HAR model while using a common index structure[11]. Applying this version of the HAR model they found that it outperforms the univariate HAR models for the S&P 500 and Nikkei 225 indices. Tian, Yang & Chen (2017) developed a time varying version of the HAR-RV model. By doing that they allowed to the predictors and to the coefficients from the regression to change over time. They used data

[10] By saying structural break we refer to unexpected changes at a point in time of a time series.

[11] In this case common index structure is that the weekly or monthly index is equal to the weekly or monthly moving average of the daily index (Cubadda, Guardabascio&Hecq, 2017).

from the Chinese market for agricultural commodity futures and they found that the model that they introduced was performing better at forecasting than the simple HAR-RV model. Andersen, Bollerslev & Diebold (2007) also slightly modified the HAR model to allow and control for jumps introducing the HAR-RV-CJ model. Using data from U.S for the period running from 1990 to 2002, which include the financial crisis of 2001-2002, they also found that their model outperformed the famous GARCH model and other related stochastic volatility models for the out of sample forecasting window.

Although GARCH models have been frequently used in the volatility forecasting literature their capability to forecast has not been unchallenged. According to Blair, Poon & Taylor (2010), the GARCH model was producing significant coefficients at the in-sample data but was performing quite poor for the out of sample data. Kat & Heynen (1994) found that the GARCH model performs better when it comes to modeling exchange rates but not that good for stock indices. On the other hand Awartani & Corradi (2005), after running their models for predicting the volatility of the S&P 500 stock index, they concluded that the GARCH model is performing better than the Exponential GARCH and the Asymmetric GARCH. They also found that asymmetries are playing a significant role in predicting volatility. Luo, Pairote & Chatpatanasiri (2017) tested the forecasting performance of the GARCH model against EGARCH and TGARCH using data from the SSE380 index. They produced a comprehensive analysis for the mean return and the conditional variance based on their data and they found that the GARCH model was the best at making volatility predictions among the others. Ekong & Onye (2017) have also used GARCH-family models to estimate the volatility using data from the Nigerian stock exchange. They compared the results based on the root square mean error method, and they found that the GARCH (1, 1) and the EGARCH (1, 1) were able to possess the best forecasting results. On the other hand Goyal (2000) found that a simple ARMA model can perform better that a GARCH-M model. For his analysis he used daily and monthly data from 1962 to 1998 of the CRSP value weighted index. An extensive comparison for the out of sample predicting ability of the ARCH/GARCH models was given by Hansen & Lunde (2004). In their research, they clearly state that for exchange rate data a simple GARCH (1, 1) model is performing better than any other version of the model, but for return data, the conclusion was a bit different. Other GARCH specifications outperformed the GARCH (1,1) and ARCH models, but still without clear evidence. Thus, GARCH (1, 1) can be a good starting point for forecasting volatility, and it can always be enriched with other parameters which can produce even better results. One example can be to include long memory, which was also confirmed by some papers.

Another finding from the Poon & Granger (2003) study was that the more sophisticated non-latent models based on realized volatility were outperforming the GARCH models and the simple non-latent approaches[12]. Although the HAR model is treating volatility as non-latent, GARCH is treating volatility as latent and at the same time it has shown weakness in being able to capture volatility directly. This direct approach of modeling volatility has been an innovation for the volatility forecasting world. Thus simple time series models were able to be used and outperform the traditional indirect approaches. The HAR model is one of these models that forecasts volatility by directly using the realized measures.

3.3. Motivation and Hypothesis

The need to account for the above-mentioned limitations in the introduction and possibly generate more trustworthy results signifies the motivation behind this thesis. More specifically, the thesis mainly contributes to the current literature by conducting an out-of-sample forecasting comparison of the HAR model on three of the major equity indexes worldwide including both periods of crisis after 2000. Based on the research of McAleer & Medeiros (2008), Seda (2013), Ma, Wei, Huang & Chen (2014), Vortelinos (2017), Jou et al., (2013), Chin, et al. (2016) and Andersen et al., (2007) which all found that the HAR model outperformed all the other models using data which included at least one period of financial crisis, this is my hypothesis:

H0: The HAR model is expected to perform better at forecasting volatility during the time of crisis.

[12] When we talk about latent variables we are referring to variables that are not directly observed but usually are deduced from other observable variables. In the case on a non-latent model we have exactly the opposite.

4. Data description

4.1. Introduction

In this section, the data that have been used for this thesis will be described and analyzed. A big part of the data for the research have been retrieved from the Eikon database. The daily adjusted closing prices from Amsterdam Exchange (AEX) index for the years 2000 to 2018 which is nineteen years period has been used. Daily adjusted closing prices from the S&P 500 index and the Nikkei 225 index have been also used for the same period. The time window will be from 1/1/2000 to 31/12/2018 except from the Nikkei 225 index that starts from 1/2/2000 until 31/12/2018 (due to not availability of data for the month of January 2000). For the AEX index there will be 4845 observations, for S&P 500 index there will be 4768 observations and for the Nikkei 225 there will be 4627 observations. The time series plot of the adjusted closing prices of all indices are shown in Figures 1, 2 and 3 of the appendix. The sample size is adequate to get reliable results and the differences in the number of observations between the indices has to do with different local holidays, trading days per country and some missing data in the beginning of the 2000 for Nikkei 225. The data for the realized volatility for these years have been retrieved from the library of the Oxford-Man Institute of Quantitative Finance. The realized volatility is normally calculated with equation (2), but in this case it is already derived directly from this database. The data for the volatility indices for S&P 500, AEX and Nikkei 225 have been also extracted from Eikon database for the same time periods as the adjusted closing prices.

From the daily adjusted closing prices of all three indices the log returns have been calculated from the formula:

$$\ln\left(\frac{r_t}{r_{t-1}}\right) \tag{15}$$

Where r_t is the return of the current day and r_{t-1} is the return of the previous day at their logarithmic (ln) version. Log returns are time additive and assuming that the log returns are normally distributed for short periods (daily in this case), then adding these normally distributed variables produces an n period log return that is also normally distributed. The time series plot of the log returns of all indices are shown in Figures 4, 5 and 6 of the appendix. The realized volatility for Nikkei 225 starts from February 2000 and not January, like in AEX and S&P 500 indices, which is compatible with the starting period for the data of adjusting closing prices.

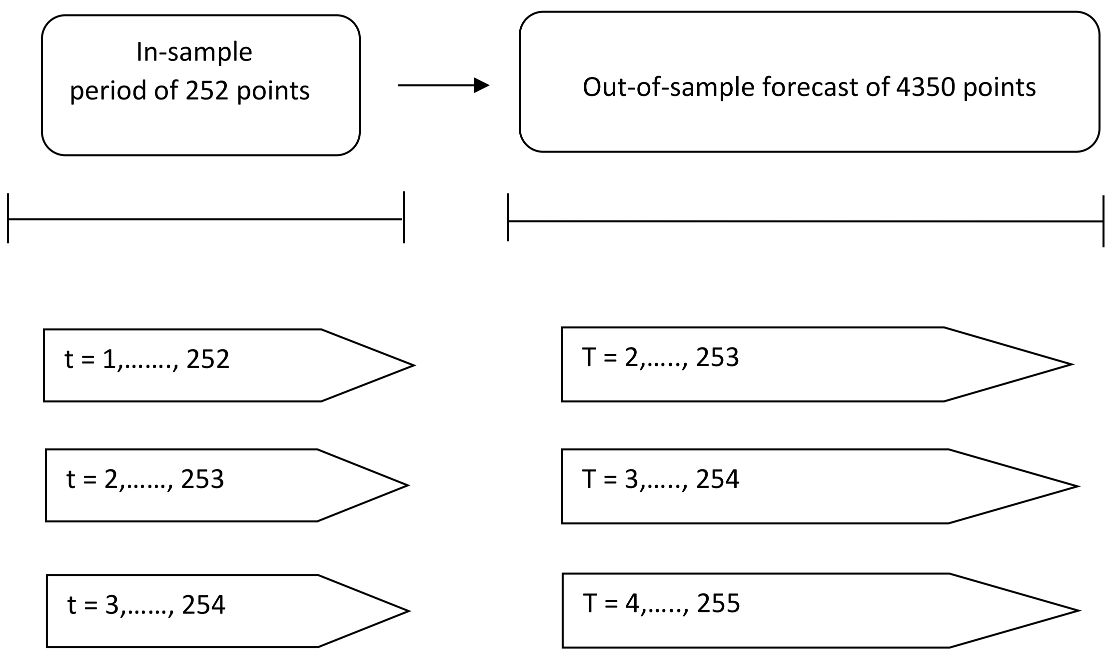
4.2. The raw data treatment of the three indices

By looking at Figures 7, 8 and 9 it can be easily observed that the volatility has not been stable through the period of investigation. Specifically, there are two periods that standing out. These are the periods between the beginning of 2000 until mid-2002, and the period between 2007 and beginning of 2009. They are both periods of crisis with volatility reaching extremely high values and some clear volatility clustering. The first period of crisis was due to the subsequent collapse of the internet bubble which started around 1996. The second crisis which is considered as the biggest after 1930, started with the crisis of the subprime mortgages in the US and continued as an international crisis when the investment bank of "Lehman Brothers" collapsed on 15th of September 2008. From the figures 7, 8 and 9 it can be seen that the impact of the first crisis was less for the Nikkei 225 index compared with the other two indices, since the spike of the volatility was lower for that period. All this turbulence in the economy is normally embedded at the closing prices of the indices which can be faced as outliers. Normally, it would be ideal to exclude all these outliers, but in this case this is not the point of this thesis. It is actually mostly focusing on the forecasting performance of the models and the volatility indices within these periods of crisis. As already mentioned above the data are running from January 2000 to December 2018 except from Nikkei 225 that starts from February 2000. The business calendar for each country has been used. All three data sets have been tested for normality with checking if the returns are normally distributed or suffer from kurtosis or skewness, see Figures 10, 11 and 12 of the appendix. For all three indices the phenomenon of leptokurtosis is visible which is normal for daily index returns[13].

[13] The phenomenon of leptokurtosis is showing fat tails and a greater peak for the mean compared to the normal distribution, although the mean and the variance are still the same (Brooks, 2008).

As already mentioned at the research methods section, the HAR-RV and the GARCH (1, 1) model have to be first estimated on a specific training period of data. This period, which is the in sample period, was selected to be at a range of 252 points (which is exactly the whole year of 2000) for both models. Thus, the out of sample period will run from 1st of January 2001 to 31st of Dec 2018. The estimation period will then be rolled forward by adding one new day every time and dropping the most far-off day. By doing this, the sample size that will be used to estimate the model will remain at a locked length, and there were be no overlap at the forecasting. Thus it will allow one day ahead (out of sample) volatility to be obtained. The forecasting relies on a daily sampling frequency. A visual representation of this procedure can be seen at the figure below.

Figure 11: The 1 day ahead forecasting procedure



An alternative forecasting method would be to use weekly returns for the GRACH model, which are constructed based on the daily returns or even monthly. Then the sampling frequency would have to be set to 5 for weekly frequency and to 22 for monthly frequency. That will change the sampling frequency and will affect also the forecast. Similar procedure can be also applied to the HAR-RV model where the weekly realized volatility or the monthly realized volatility can become the dependent variables with the rest of the model remaining as it is in equation (3). For the purpose of this thesis only the daily frequency will be tested. For reporting reasons the variables produced from the MSE method have been scaled up by 100.000 and the variables produced from the MAE have been scaled up by 100.

For the first period of crisis the models have been tested on 327 points for AEX, 322 for S&P 500 and 324 for Nikkei 225. The training window for the models remained the same as before, counting for 252 points. The second period had been tested for 574, 574, and 560 for AEX, S&P 500 and Nikkei 225 respectively. The two models have used the same training period for both crisis periods.

4.3. The treatment of the volatility indices

It has already been indicated that for the purpose of this thesis three different volatility indices will be used. Unlike the volatility retrieved from the GARCH (1, 1) and the HAR model, the use of VIX, VAEX and JNIV does not require the use any econometric model to forecast since the prices are already computed. However, all three volatility indices (VIX, VAEX and JNIV) are reported as an average daily volatility which is annualized by using 365 days. Thus, the values have been first multiplied with the square root of (252/365) to make sure that that they will be compatible with the realized measures that are annualized using 252 business days. Then this value has been divided by 365 and the result has been squared to get the monthly variance. As a last step the monthly variance has been divided by 22 business days to get the daily volatility. Now, the volatility that has been derived with the above mentioned procedure from the volatility indices is ready to be compared with the forecasting results taken from the two model and the actual realized volatility, since this is the scope of this thesis. The mean squared method, and the mean absolute error methods, which will be analyzed in section 5.1, will be also used to compare the volatility from the volatility indices with the actual realized volatility. Table 1a below is providing a summary of the difference between these two variables for the three volatility indices, but only for the mean squared error method. The table 1b for the other method can be find at the appendix.

Table 1a: Summary of Volatility indices MSE method

Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
VAEX MSE	4592	.009	.026	0	.352
VIX MSE	4516	.005	.046	0	2.841
JNIV MSE	4341	.012	.043	0	.811

So what we see on this table is that the variable with the lowest mean is for the VIX (0.005), which means that it is closer to the actual realized volatility than the other two volatility indices that they report a mean of 0.09 for VAEX and 0.012 for JNIV.

4.4. Stylized facts for financial data

When dealing with statistical analysis of financial time series data, it can be expected to face a set of stylized facts that might emerge from this analysis. The detection and knowledge of these facts can be helpful to derive better statistical models and produce more reliable forecasting results. It can also be helpful to decide which model to choose to get better forecasting results. Financial time series data are always sensitive to fluctuations and in this case it's the volatility fluctuation that we are dealing with. The most common stylized facts are: Fat tails, stationarity and autocorrelation.

4.4.1. Fat tails

If the distribution of stock returns has fat tails then this can have some influential implications in the financial time series analysis. When talking about fat tails we mean all these extreme values that are observed on the very right or left side of the normal distribution bell curve and this can lead to an underestimation of potential risk (Jilla, Nayak & Bathula, 2017). In simple words fat tails are a sign that the stock market has unexpected large and small outcomes under the normal distribution. The fat tail can be observed with a graphical method named Quantile – Quantile (QQ) plot. Figures 10, 11 and 12 in appendix provide an overview, but the phenomenon of extreme values is not so visible in these three cases which is good.

4.4.2. *Stationarity*

In time series, stationarity is the phenomenon where the statistical properties of financial data are remaining constant even when the time origin changes (Jilla, Nayak & Bathula, 2017). Stationarity as a concept is very essential for the time series analysis and its always necessary to make the data stationary before running any regression. While testing for stationarity we check whether the time series maintains a unit root[14]. There are two tests to check if our data are stationary or not. The first one is the Augmented Dickey-Fuller (ADF) test to check for unit roots and the second is the Phillips–Perron (PP) test. In this thesis the Augmented Dickey-Fuller (ADF) test will be used. There are also some other tests for examining stationarity which are setting the stationarity as the null hypothesis. The standard Dickey-Fuller test is testing the following assumptions by using two hypotheses. The null hypothesis (Ho) is that the time series has a unit root, so it is not stationary, and the alternative hypothesis (Ha) which says that the time series does not have a unit root, thus it is stationary. After running the Augmented Dickey-Fuller (ADF) test for all three indices returns we can conclude that they are all stationary. The test statistic value for AEX is -70.245 which is way higher than the 1% critical value. The test statistic value for S&P 500 is -74.637 which is way higher than the 1% critical value. Lastly The test statistic value for Nikkei 225 is -70.455 which is way higher than the 1% critical value. Thus the null hypothesis of presence of unit root can be rejected for all of them. Table 2 summarizes all the test statistic and critical values for the three indices.

Table 2: Summary test statistic Dickey – Fuller for unit root

Dickey-Fuller	---------- Augmented Dickey-Fuller ---------			
Index	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
AEX	-70.245	-3.430	-2.860	-2.570
S&P 500	-74.637	-3.445	-2.980	-2.650
NIKKEI 225	-70.455	-3.325	-2.730	-2.420

[14] A unit root process is a stochastic movement in a time series which many times is called random walk with drift. In the case that a time series has a unit root it displays an unpredictable standardized pattern.

4.4.3. *Autocorrelation*

Autocorrelation or serial correlation is the phenomenon in which past returns are influencing future returns. According to Figlewski (1997), positive serial correlation can be often found in the daily closing prices of securities and equities. The analysis of the autocorrelation can show the exact impact of past returns on the future returns based on a shock that happened or an announcement. If the returns are correlated then there is a strong confirmation for predictability. By using lags when running a regression it can be observed for how long this impact is statistically significant. Another stylized fact is the volatility clustering.

4.4.4. *Volatility clustering*

Markets are unpredictable and as Mandelbrot& Hudson (2007) have mentioned, they are like 'roiling seas'. Like the sea can have calm and turbulent periods with flows and backflows, same happens with the markets. Some days prices are stable, or might move in tiny increments, and other days they might leap or plunge. Market risk cannot be easily captured. The prices of the stocks are characterized by discontinuously movement and this is one of the reasons why markets are more riskier than many financial professionals think (Mandelbrot & Hudson, 2007). In general price movements can be unpredictable, but they can also be dependent to each other, and in this case is called volatility clustering. The phenomenon of volatility clustering has attracted the attention of researchers and this had led to the development of some stochastic models. Models such as the GARCH (1, 1) and stochastic volatility models have been created to model this phenomenon and a big debate has been running whether there is a long range dependency in volatility (Cont, 2007). Many assets can go through periods of turbulence or stable periods. When studying the behavior of volatility for these periods, it can give a signal of high volatility which will be followed by high volatility or low volatility which will be followed by low volatility. Usually econometricians call this autoregressive conditional heteroscedasticity (ARCH). Thus, we will have periods with many daily squared returns being large and periods with many daily squared returns being small which is called volatility clustering. Volatility clustering is an often problem in the financial time series data. Figures 7, 8 and 9 at the appendix are giving an overview of this phenomenon.

5. Empirical results and analysis

5.1. *Forecasting performance*

The next step after taking the outcomes from the two models and the volatility indices, is to assess their forecasting performance. According to Lunde & Hansen (2001) using the mean squared error method (MSE) is one way of appraising the forecasting performance of the GARCH (1, 1) and the HAR models. The criterion of selecting the best model is not only one, and it can be expressed in terms of a loss function or utility function (Lunde & Hansen, 2001)[15]. This loss function is defined as:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(RV - \hat{\sigma})^2 \qquad (16)$$

Where N stands for the total number of forecasts, RV is the realized volatility of one-day horizon and $\hat{\sigma}$ is the estimated volatility over the same horizon from the models and the respective volatility indices. There is one drawback of the MSE method which was mentioned from Wilhelmsson (2006), and that is that the loss function can be sensitive to some outliers. Lunde & Hansen (2001) on their paper, are also referring to other methods that can be used for assessing the forecasting performance of the models. One of them is the mean absolute error method (MAE). The difference with the MSE method is that instead of taking the squared mean distance, MAE is using the absolute distance of the realized volatility with the forecasted one. The MAE loss function is equally treating all loses where on the other hand the MSE method is punishing the big losses. For this thesis only the MSE and MAE methods will be used. The MAE loss function is defined as:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|RV - \hat{\sigma}| \qquad (17)$$

[15] Loss function can define the distance between the result taken from a model and the expected result. Can simply indicate the degree of error between the model and the prediction.

Where N stands for the total number of forecasts, RV is the realized volatility of one day horizon and $\hat{\sigma}$ is the estimated volatility over the same horizon from the models and the respective volatility indices.

5.2. In-sample results

The main objective of this thesis is to evaluate the performance of the HAR-RV and the GARCH (1, 1) models and determine which of these two models is producing the best out of sample variance. Nevertheless, investigating the in-sample fit measures of the models can provide a good indication of the forecasting performance. The in-sample period for all indices is running for the whole year of 2000, which is 252 points. Since the GARCH model is modeling the mean return and the volatility, starting with the AEX index the GARCH (1, 1) model had a coefficient for the moving average term for the mean return (MA1) of 0.0576, but not significant, and coefficients for modeling the volatility (GARCH AR1 and GARCH MA1) of 0.71803 and 0.56998, which were both significant at 1% level of significance. The table 3a below is providing an overview of the regression for the GARCH (1, 1) model.

Table 3a: Regression for AEX of the initial sample

Numer of obs 252 Strong convergence iteration time: 0.11					
	Estimate	Std. Err.	T Ratio	p-Value	Sig
MA1	0.0576	0.06049	0.952	0.342	
[2]GARCH Intercept^(1/2)	0.00597	0.0018	-------	--------	--------
GARCH AR1	0.71803	0.15774	4.552	0	***
GARCH MA1	0.56998	0.1558	3.658	0	***
R-Squared = 0.0012					

Moving now to the in-sample performance of the HAR-RV model, first some tests have to be done. Often times in time series we observe the phenomena of multicollinearity and heteroskedasticity for the residuals[16]. In order to test for multicollinearity we have to run a variance inflation factor test (VIF)[17]. Looking at table 3b we see that we do not have multicollinearity since all values are below 5. Testing for heteroskedasticity we have to use a Breusch-Pagan test. Using this test it was found that there is indeed heteroskedasticity. For this reason robust standard errors was used in the regression to correct for heteroskedasticity. After conducting these two tests for the HAR-RV model the realized volatility of yesterday (RV1) was significant at a level of 5%, the realized volatility of last week (RV5) was not significant, and the realized volatility of last month (RV22) was again significant at a level of 5%. All the coefficients were positive and the r squared (R^2) was 21,4 %. The coefficient for RV22 is more than double than that of the RV1, and this is consistent with the findings from Corsi (2009). The table 3c below is providing an overview of the coefficients of the estimation of the initial sample.

Table 3b: VIF AEX

	VIF	1/VIF
RV5	2.624	.381
RV22	1.994	.502
RV1	1.638	.611
Mean VIF	2.085	.

Table 3c: Linear regression for AEX of the initial sample

Number of obs 252

RV	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
RV1	0.172	0.077	2.23	0.027	0.020	0.324	**
RV5	0.166	0.147	1.13	0.260	-0.124	0.456	
RV22	0.410	0.162	2.52	0.012	0.090	0.730	**
Constant	0.000	0.000	2.15	0.033	0.000	0.000	**

Mean dependent var	0.000	SD dependent var	0.000
R-squared	0.214	Number of obs	230.000
F-test	16.275	Prob > F	0.000
Akaike crit. (AIC)	-3568.049	Bayesian crit. (BIC)	-3554.297

**** p<0.01, ** p<0.05, * p<0.1*

[16] We call multicollinearity the phenomenon where some independent variable of a regression model are correlated with each other. This can cause problems at fitting the model.
Heteroskedasticity is referring to the fact that some subpopulations of random variables have different variability from others.
[17] Variance inflation test is assessing the degree that the variance of an estimated regression is increasing if the predictors are correlated. If the VIF test gives a value of between 5 and 10 then the independent variables might be high correlated.

For the S&P 500 index the GARCH (1, 1) model had a coefficient of 0,0096 for (MA1), but not significant, and coefficients for modeling the volatility (GARCH AR1 and GARCH MA1) of 0.8103 and 0.642, which were both significant at 1% and 5% levels of significance respectively. All VIF values where under 5 so there was no multicollinearity and the robust standard errors have been used at the regressions to correct for heteroskedasticity. Thus, for the HAR-RV model the realized volatility of yesterday (RV1) was significant at a level of 1%, the realized volatility of last week (RV5) was not significant, and the realized volatility of last month (RV22) was again significant at a level of 10%. All the coefficients were positive and the r squared (R^2) was 22,2 %.

Moving to Nikkei 225 index the GARCH (1, 1) model had a coefficient of 0,0439 for (MA1), but not significant, and coefficients for modeling the volatility (GARCH AR1 and GARCH MA1) of 0.9505 and 0.9448, which were both significant at 1% level of significance. All VIF values where under 5 so there was no multicollinearity and the robust standard errors have been used at the regressions to correct for heteroskedasticity. Thus, for the HAR-RV model the realized volatility of yesterday (RV1) was significant at a level of 5%, the realized volatility of last week (RV5) was significant at a level of 1%, and the realized volatility of last month (RV22) was not significant. All the coefficients were positive and the r squared (R^2) was 17,2 %. The tables 4a, 4b, 4c, 5a, 5b and 5c for S&P 500 and Nikkei 225 for the regressions and the VIF results are presented at the appendix.

5.3. Out of sample results

Through the whole statistical analysis, the empirical implications of the HAR-RV and the GARCH (1, 1) models, together with the volatility forecasted from volatility indices have been tested. The actual realized volatility is the benchmark of comparing all the results. The closer the predictions from the models and the volatility indices forecast to the actual volatility, the better the forecasting result. Using the MSE and the MAE methods, the inferences can be made based on the mean. The lowest the mean the better the forecasting performance. According to the hypothesis, the HAR-RV model is expected to produce better forecasting results during the two periods of crisis. The forecasting performance has been also tested for the whole period of data and then specifically for the two separate periods.

Since this thesis focuses mainly on the Netherlands and subsequently AEX index, it is optimal to start with the analysis of these results first. Consistent with the finding from Corsi (2009), Seda (2013), and Ma, et.al (2014) in the literature review, the HAR-RV model is outperforming the GARCH (1, 1) model and the volatility forecasted from the VAEX for the whole period of 2000 to 2018 when testing the forecasting performance with the MSE method. Specifically the mean for the HAR-RV model is 0.002, for the GARCH (1, 1) 0.007, and for the VAEX 0.009. The HAR-RV model was able to outperform the GARCH (1, 1) and the VAEX also when the results where compared with the MAE method. The mean for the HAR-RV model, the GARCH (1, 1), and the VAEX were 0.004, 0.011 and 0.019 respectively. Tables 3a and 3b below are providing an overview.

Table 3a: AEX Descriptive Statistics of MSE method for 2001-2018

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	4592	.007	.042	0	1.012
HAR MSE	4594	.002	.017	0	.731
VAEX MSE	4592	.009	.026	0	.352

Table 3b: AEX Descriptive Statistics of MAE method for 2001-2018

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	4592	.011	.238	0	3.181
HAR MAE	4594	.004	.116	0	2.703
VAEX MAE	4592	.019	.22	0	1.876

The findings from the two crisis periods are also consistent with the findings from McAleer & Medeiros (2008), Seda (2013), Ma, et.al (2014), Vortelinos (2017), Jou, et.al (2013), Chin, et al. (2016) and Andersen, et. Al (2007) as mentioned in the literature review. The HAR-RV model was again able to produce better forecasting results against the GARCH (1, 1) model and the VAEX. Namely during the first crisis the mean from the MSE method for the HAR-RV model is 0.003, for the GARCH (1, 1) is 0.007 and for the VAEX 0.007 too. Analogous results were provided from the MAE method. The mean for the HAR-RV model was 0.007, for the GARCH (1, 1) 0.014 and for the VAEX 0.022. Tables 4a and 4b below are providing an overview.

Table 4a: AEX Descriptive Statistics of MSE method for 2001 to mid-2002

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	327	.007	.026	0	.279
HAR MSE	326	.003	.022	0	.31
VAEX MSE	327	.007	.015	0	.165

Table 4b: AEX Descriptive Statistics of MAE method for 2001 to mid-2002

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	327	.014	.221	0	1.669
HAR MAE	326	.007	.161	0	1.761
VAEX MAE	327	.022	.163	.032	1.284

Moving to the results of the forecasting performance for the second period of crisis, 2007 until the beginning of 2009, the HAR-RV model was again superior to its competitors. However, it can be highlighted that the mean was slightly higher for all the models due to the bigger magnitude of crisis. This means that the volatility was much higher for this period compared with the first period of crisis. Using the MSE method of comparison the mean for the HAR-RV model was 0.005, for the GARCH (1, 1) 0.029 and for the VAEX 0.025. The outcome was the same with the MAE method with the HAR-RV model producing a mean of 0.009, the GARCH (1, 1) a mean of 0.025 and the VAEX a mean of 0.034. Thus, all the values for the mean for this period are slightly higher compared to the first period of crisis and the whole period of analysis. Tables 5a and 5b below are providing an overview.

Table 5a: AEX Descriptive Statistics of MSE method for 2007 to early 2009

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	574	.029	.107	0	1.012
HAR MSE	573	.005	.036	0	.731
VAEX MSE	574	.025	.051	0	.352

Table 5b: AEX Descriptive Statistics of MAE method for 2007 to early 2009

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	574	.025	.474	0	3.181
HAR MAE	573	.009	.207	0	2.703
VAEX MAE	574	.034	.364	.01	1.876

Andersen, Bollerslev & Meddahi (2011) have tested the effect of the volatility that has been forecasted from the models on the actual volatility. They suggested that the higher the R^2 from the regression the better the degree of predictability from the respective model. The same test was run by Hansen & Lunde (2006). Thus, the R^2 which has been obtained by regressing the realized volatility on the forecasted volatility from the respective models will be reported. This R^2 can be also called coefficient of determination because it can provide info for which model is better at predicting. The results are reported at the tables 6a, 6b and 6c below. It can be observed that for all three indices the R^2 is always higher for the volatility predicted from HAR-RV model than the other two variables which is another positive indication for the forecasting performance of this model.

Table 6a: R^2 and the regression coefficients for AEX data

Linear regression

RV	Coef.	St.Err.	R^2	p-value	[95% Conf	Interval]	Sig
HARVAR	0.899	0.057	0.496	0.000	0.786	1.012	***
GARCHVAR	0.506	0.035	0.451	0.000	0.436	0.576	***
VIXVAR	0.682	0.031	0.494	0.000	0.621	0.742	***

Table 6b: $R^2$and the regression coefficients for S&P 500 data

Linear regression

RV	Coef.	St.Err.	R^2	p-value	[95% Conf	Interval]	Sig
HARVAR	0.920	0.010	0.634	0.000	0.900	0.940	***
GARCHVAR	0.647	0.009	0.507	0.000	0.628	0.666	***
VIXVAR	0.726	0.009	0.581	0.000	0.709	0.744	***

Table 6c: R^2 and the regression coefficients for NIKKEI data

Linear regression

RV	Coef.	St.Err.	R^2	p-value	[95% Conf	Interval]	Sig
HARVAR	0.925	0.012	0.591	0.000	0.902	0.948	***
GARCHVAR	0.295	0.005	0.408	0.000	0.284	0.305	***
VIXVAR	0.361	0.005	0.487	0.000	0.350	0.372	***

5.4. Robustness check of forecasting

As already mentioned, the main focus of this thesis is the Netherlands and the AEX index. Thus, after presenting the results from the AEX index on section 5.2., the outcomes from the other two indices will be also presented and used as a robustness check. The HAR-RV model was remarkably better at forecasting volatility for this index for all the periods, and it would be interesting to see if the same holds for the other two indices. The same forecasting periods and the same comparison methods were used as for the AEX index.

Starting with the S&P 500 index and the period of 2001 to 2018 (recall that the year of 2000 was used as a training period for the models), the following results were produced. The mean for the HAR-RV model using the MSE method was 0.002, for the GARCH (1, 1) model was 0.004 and for the VIX 0.005. With the MAE method, the mean for the HAR-RV was 0.005, for the GARCH (1, 1) model was 0.007 and for the VIX 0.014. Moving to the first period of crisis the forecasting dominance of the HAR-RV model did not change. The mean for the HAR-RV model using the MSE method was 0.0009, for the GARCH (1, 1) model was 0.0013 and for the VIX 0.0045. Similar results were produced with the MAE method were the mean for the HAR-RV model was 0.005, for the GARCH (1, 1) model 0.008 and for the VIX 0.019. The forecasting results remained the similar also for the second period of crisis. Looking first at the MSE method, the mean for the HAR-RV was 0.012, for the GARCH (1, 1) model was 0.024 and for the VIX 0.019. The mean that was produced from the MAE method was 0.014 for the HAR-RV model, 0.023 for the GARCH (1, 1) model and 0.027 for the VIX.

The forecasting performance of the models and the volatility indices was lastly tested on the Nikkei 225 index. The first period is again the 2001 to 2018, which is the whole data period used for this thesis. Looking at the results from the MSE method the HAR-RV model had a mean of 0.001, the

GARCH (1, 1) a mean of 0.010 and the JNIV 0.012. With the MAE method the mean for the HAR-RV model, the GARCH (1, 1) model and the JNIV were 0.004, 0.014 and 0.025 respectively. For the first period of crisis the mean for the HAR-RV model using the MSE method was 0.001, for the GARCH (1, 1) model was 0.003 and for the JNIV 0.014. From the MAE method the HAR-RV model reported a mean of 0.006, the GARCH (1, 1) model a mean of 0.014 and the JNIV a mean of 0.033. The last period is the second period of crisis. The mean for the HAR-RV model using the MSE method was 0.003, for the GARCH (1, 1) model was 0.055 and for the JNIV 0.050. With the MAE method, the mean for the HAR-RV model was 0.008, for the GARCH (1, 1) model was 0.033 and for the JNIV 0.049. Since the main index of analysis is the AEX, all the results from the S&P 500 index are demonstrated in tables 7a to 7f of the appendix, and all the tables for the Nikkei 225 index in tables 8a to 8f of the appendix.

After conducting the analysis for all the indices and comparing the results, it can be easily observed that the HAR-RV model is unbeatable for all periods and under both comparison methods. Now that the analysis had finished, it can be concluded that the HAR-RV model is superior against its competitors, but still the inference cannot be totally generalized since different methods of comparison or testing more indices can produce different results.

6. Conclusion

Forecasting volatility is a topic that has been discussed and analyzed a lot in the last years. A lot of models have been used for this purpose and still there is no clear view of which one of them is better able to forecast volatility. This thesis has been focused on the forecasting performance of the HAR-RV and the GARCH (1, 1) models as well as the volatility derived from the volatility indices. The realized volatility has been used as a benchmark to evaluate the forecasting performance of the above metrics. The HAR-RV model is relatively new model in comparison with the GARCH (1, 1) model and its simple structure has made it a very reliable model at forecasting volatility. The AEX index was the main index of analysis for the whole period of 2000 to 2018 with an extra analysis of the performance of the models at the two periods of financial crisis of 2000 to 2002 and 2007 to 2009. As a robustness check of the forecasting performance two more indices have been tested by the models for the same main period and the other two periods of crisis.

The HAR-RV model was superior to its competitors for all the periods of the analysis and for all three different indices. First of all, the results that have been taken from the HAR-RV model were in line with the findings from Seda (2013), where he also tested the model on a stock market index, including the 2007 to 2009 period of crisis. The strategy was similar to this thesis and he also got better forecasting results from the HAR-RV model for all periods against the GARCH (1, 1) model. The results are also in line with the analysis from McAleer & Medeiros (2008), which also found better predicting power at the HAR-RV model against other volatility models. The same forecasting performance of the HAR-RV has been seen also at the analysis of Ma, Wei, Huang & Chen (2014) and Vortelinos (2017). Finally the findings of this analysis can be confirmed from the findings from Audrino & Knaus (2016), which they compared the HAR-RV model with other volatility models for a period in and out of crisis. On the other hand the results are contrasting with the results from Chung, Sun & Shih (2008), which they found better predictive power from the VIX index. The GARCH (1, 1) was either the second best closer to the actual realized volatility and sometimes even in the third place. According to the literature review the poor performance of the GARCH (1, 1) model against the HAR-RV model is basically due to the fact that is not able to directly capture the volatility. None of the literature until this point has mentioned that the GARCH (1, 1) model has been able to outperform the HAR-RV model. Thus, this thesis is confirming the predictive power of that model.

For the evaluation of the forecasting performance of the models and the volatility indices, the methods of MSE and MAE have been used. The volatility was higher for the second period of crisis compared to the first one and that was a sign of the magnitude of that crisis. To conclude, this thesis has mainly contributed to the literature by conducting an out-of-sample forecasting comparison of the HAR-RV model on three of the major equity indexes worldwide including both periods of crisis after 2000. The results though are giving extra power to the HAR_RV model since the analysis was extended in two periods of crisis and not only in one as it was done by McAleer & Medeiros (2008), Seda (2013), Ma, Wei, Huang & Chen (2014), Vortelinos (2017), Jou et al., (2013), Chin, et al. (2016) and Andersen et al., (2007). Following the title of this thesis it can be confirmed that the HAR-RV model was Har (d) to beat.

7. Limitations

Although the HAR model is a simple model to use and produces a perfect estimation of the volatility, there are a few studies about it. Hence we cannot generalize our inferences regarding its results. The whole year of 2000 is excluded from the out of sample data. The reason is that approximately one year period is needed to estimate the HAR and the GARCH (1, 1) models, thus this year will not be included at the results section. Specifically 252 points from the starting date of the data has been used for this purpose. The data for the NIKKEI 225 index had a starting date of 1st of February 2000 instead of 1st of January 2000 as with AEX and S&P 500 indices, due to unavailability of data for this specific month. It is a minor difference which did not really affected the analysis and forecasting performance of the models. For the first period of crisis the out of sample forecast window was smaller compared with the second crisis due to the reason that the whole year of 2000 was used as a training period. The low R^2 of the HAR model regressions is mainly due to the small training period of 252 points of the model. Finally, the models that were discussed in this thesis do not take into consideration macroeconomic variables that might affect volatility. Several studies from Yogaswari, Nugroho & Astuti (2012), Nkoro & Uko (2014) and Khalid & Khan (2017) found that macroeconomic variables like inflation, interest rates, and exchange rates affect volatility but we don't account for them on this research.

8. Future research suggestion

Since this thesis has mainly focused on the original version of the HAR-RV model and the 1,1 order of the GARCH model, a good future research suggestion would be that the extensions of these two models could be tested for the same periods. For example, a log HAR-RV model and the 1,2 or 1,3 order of the GARCH model. As already mentioned in the literature review EGARCH and TGARCH are two extensions of the GARCH model that can be also tested for their forecasting performance. It has been clear that the whole analysis was based on daily window forecasting, thus it might be interesting to have a look at the weekly and monthly forecasting performance of the models to see if the results will be similar or not. Except the MSE and the MAE methods of comparison there are other methods that can be implemented to test the forecasting performance. Lunde & Hansen (2001) in their paper are mentioning 6 other methods that can be used. Lastly, conducting the same analysis in a different forecasting horizon and using a bigger training period for the models might provide contrasting results. The reason for that could be that the data are referring to a period without turmoil and the bigger in-sample window can make the models more accurate at forecasting.

9. Bibliography

- 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.
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International economic review, 885-905.
- Andersen, T. G., Bollerslev, T., & Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. Journal of Econometrics, 160(1), 220-234.
- Angabini, A., & Wasiuzzaman, S. (1997). GARCH models and the financial crisis-A study of the Malaysian stock market. Studies, 2007 (2008).
- Antonakakis, N., & Darby, J. (2007). Forecasting nominal exchange rate volatility in developing and industrialized countries. University of Strathclyde.
- Audrino, F., & Knaus, S. D. (2016). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. Econometric Reviews, 35(8-10), 1485-1521.
- Awartani, B. M., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. International Journal of Forecasting, 21(1), 167-183.
- Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 74(1), 3-30.
- Balzer, L. A. (1994). Measuring investment risk: A review. The Journal of Investing, 3(3), 47-58.
- Banulescu, D., Hansen, P. R., Huang, Z., & Matei, M. (2018). Volatility during the financial crisis through the lens of high frequency data: a realized GARCH approach.
- Barndorff-Nielsen, O. E., & Shephard, N. (2003). Realized power variation and stochastic volatility models. Bernoulli, 9(2), 243-265.
- Berry, W. D. (1993). Understanding regression assumptions (Vol. 92). Sage Publications.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of political economy, 81(3), 637-654.

- Blair, B. J., Poon, S. H., & Taylor, S. J. (2010). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. In Handbook of Quantitative Finance and Risk Management (pp. 1333-1344). Springer, Boston, MA.
- Bluhm, H., & Yu, J. (2001). Forecasting volatility: Evidence from the German stock market.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327.
- Brenner, M., Ou, E.Y., Zhang, J.E., 2006. Hedging volatility risk. Journal of Banking and Finance 30, 811–821.
- Brooks, C. (2008). Introductory Econometrics for Finance 2nd edition, Cambridge University Press
- Chin, W. C., Lee, M. C., & Yap, G. L. C. (2016). Heterogeneous autoregressive model with structural break using nearest neighbor truncation volatility estimators for DAX. SpringerPlus, 5(1), 1883.
- Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. Journal of financial economics, 50(2), 125-150.
- Christoffersen, P., & Diebold, F. (2000). How Relevant is Volatility Forecasting for Financial Risk Management? The Review of Economics and Statistics, 82(1), 12-22.
- Chung, H., Sun, E. Y., & Shih, K. C. (2008). Do HAR and MIDAS models outperform implied volatility model? Evidence from range-based realized volatility. Manuscript, Oversea Chinese Institute of Technology.
- Cont, R. (2007). Volatility clustering in financial markets: empirical facts and agent-based models. In Long memory in economics (pp. 289-309). Springer, Berlin, Heidelberg.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174-196.
- Corsi, F. (2003). A Simple Long Memory Model of Realized Volatility. Manuscript, University of Southern Switzerland.
- Corsi, F., Mittnik, S., Pigorsch, C., &Pigorsch, U. (2008). The volatility of realized volatility. Econometric Reviews, 27(1-3), 46-78.

- Cubadda, G., Guardabascio, B., & Hecq, A. (2017). A vector heterogeneous autoregressive index model for realized volatility measures. International Journal of Forecasting, 33(2), 337-344.
- Davidian, M., & Carroll, R. J. (1987). Variance function estimation. Journal of the American Statistical Association, 82(400), 1079-1091.
- Day, T., Lewis, C., (1988). The behavior of the volatility implicit in option prices. Journal of Financial Economics 22, 103—122.
- Deo, R., Hurvich, C., & Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. Journal of Econometrics, 131(1-2), 29-58.
- 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.
- Ekong, C. N., & Onye, K. U. (2017). Application of Garch Models to Estimate and Predict Financial Volatility of Daily Stock Returns in Nigeria.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.
- Fatás, A. & I. Mihov (2005). Policy volatility, institutions, and economic growth, Discussion Paper No. 5388, CEPR, London.
- Figlewski, S. (1997). Forecasting volatility. Financial markets, institutions & instruments, 6(1), 1-88.
- Fleming, J., Ostdiek, B., & Whaley, R. E. (1995). Predicting stock market volatility: A new measure. Journal of Futures Markets, 15(3), 265-302.
- Gonzalez, M. T., & Novales, A. (2007). Why a volatility index can be useful in the Spanish financial market. XV Foro de Finanzas.
- Gospodinov, N., Gavala, A., & Jiang, D. (2006). Forecasting volatility. Journal of Forecasting, 25(6), 381-400.
- Goyal, A. (2000). Predictability of stock return volatility from GARCH models. Anderson Graduate School of Management, UCLA, 110, 90095-1481.
- Guo, W., & Su, T. (2004). Factors in implied volatility skew in corn futures options. Mimeo, 1-24.

- Hansen, P. R., & Lunde, A. (2004). Realized variance and IID market microstructure noise. Brown University.
- Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. Journal of Business & Economic Statistics, 24(2), 127-161
- Härdle, W. K., & Silyakova, E. (2012). Volatility Investing with Variance Swaps. In Handbook of Computational Finance (pp. 203-219). Springer, Berlin, Heidelberg.
- Huang, C., Gong, X., Chen, X., & Wen, F. (2013). Measuring and forecasting volatility in Chinese stock market using HAR-CJ-M model. In Abstract and Applied Analysis (Vol. 2013). Hindawi.
- Jilla, K., Nayak, C., & Bathula, A. (2017). Stylized facts of financial time series: A comprehensive analysis, 302-307
- Jorion, P. (1995). Predicting volatility in the foreign exchange market. The Journal of Finance, 50(2), 507-528.
- Jou, Y. J., Wang, C. W., & Chiu, W. C. (2013). Is the realized volatility good for option pricing during the recent financial crisis?. Review of Quantitative Finance and Accounting, 40(1), 171-188.
- Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2016). Forecasting stock return volatility: a comparison of GARCH, Implied volatility, and realized volatility models. Journal of Futures Markets, 36(12), 1127-1163.
- Kat, H. M., & Heynen, R. C. (1994). Volatility Prediction: A Comparison of the Stochastic Volatility, GARCH (1, 1) and Egarch (1, 1) Models. Journal of Derivatives, 2(2).
- Khalid, W., & Khan, S. (2017). Effects of macroeconomic variables on the stock market volatility: the Pakistan experience. Global Journal of Management And Business Research.
- Kumar, Nand & Verma, Rishabh & Gupta, Puneet. (2016). Forecasting Volatility Using GARCH: A Case Study. International Journal of Advanced Production and Industrial Engineering. 1. 39-49.
- Lamoureux, C.G., Lastrapes, W., (1993). Forecasting stock return variance: towards understanding stochastic implied volatility. Review of Financial Studies 6, 293—326.
- Lee, R. W. (2005). Implied volatility: Statics, dynamics, and probabilistic interpretation. In Recent advances in applied probability (pp. 241-268). Springer, Boston, MA.

- Lunde, A., & Hansen, P. R. (2001). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH (1, 1)? (No. 2001-04).
- Luo, L., Pairote, S., & Chatpatanasiri, R. (2017). GARCH-type forecasting models for volatility of stock market and MCS test. Communications in Statistics-Simulation and Computation, 46(7), 5303-5312.
- Ma, F., Wei, Y., Huang, D., & Chen, Y. (2014). Which is the better forecasting model? A comparison between HAR-RV and multifractality volatility. Physica A: Statistical Mechanics and its Applications, 405, 171-180.
- Mandelbrot, B., & Hudson, R. L. (2007). The Misbehavior of Markets: A fractal view of financial turbulence. Basic books.
- McAleer, M., & Medeiros, M. C. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. Journal of Econometrics, 147(1), 104-119.
- McDonald, R. L. (2013). Derivatives Markets: Pearson New International Edition. Pearson Higher Ed.
- Nkoro, E., & Uko, A. K. (2014). Impact of the Volatility of Macroeconomic Variables on the Volatility of Stock Market Returns: Nigeria’s Experience. In Global Strategies in Banking and Finance (pp. 218-230). IGI Global.
- Parasuraman, N. R., & Ramudu, P. J. (2011). Historical and implied volatility: an investigation into NSE NIFTY futures and options. Australian Journal of Business and Management Research, 1(7), 112.
- Poon, S. H., & Granger, C. W. (2003). Forecasting volatility in financial markets: a review. In Journal of Economic Literature.
- Rossi, E. (2012). Long Memory processes: Econometric specifications. University of Pavia.
- Rubaltelli, E., Ferretti, R., & Rubichi, S. (2006). The risk-return trade off: Expected and required return. Available at SSRN 913807.
- Seďa, P. (2012). Performance of Heterogeneous Autoregressive Models of Realized Volatility: Evidence from US Stock Market. World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 6(12), 3421-3428.

- Seďa, P. (2013). Heterogeneous Autoregressive Model of the Realized Volatility: Evidence from Czech Stock Market.
- Sharma, P., & Vipul, V. (2015). Forecasting stock index volatility with GARCH models: international evidence. Studies in Economics and Finance, 32(4), 445-463.
- Siriopoulos, C., & Fassas, A. (2009). Implied volatility indices–a review.
- Taylor, S. J. (1987). Forecasting the volatility of currency exchange rates. International Journal of Forecasting, 3(1), 159-170.
- Taylor, S. J. (2005). Asset Price Dynamics, Volatility, and Prediction, volume 1. Princeton University press, 41 William Street, Princeton, New Jersey 08540.
- Tian, F., Yang, K., & Chen, L. (2017). Realized volatility forecasting of agricultural commodity futures using the HAR model with time-varying sparsity. International Journal of Forecasting, 33(1), 132-152.
- Vortelinos, D. I. (2017). Forecasting realized volatility: HAR against Principal Components Combining, neural networks, and GARCH. Research in international business and finance, 39, 824-839.
- Wen, F., Gong, X., & Cai, S. (2016). Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. Energy Economics, 59, 400-413.
- Wilhelmsson, A. (2006). GARCH forecasting performance under different distribution assumptions. Journal of Forecasting, 25(8), 561-578.
- Wilmott, P., Howson, S., Howison, S., & Dewynne, J. (1995). The mathematics of financial derivatives: a student introduction. Cambridge university press.
- Xie, Y. (2007). Maximum likelihood estimation and forecasting for GARCH, Markov switching, and locally stationary wavelet processes (Vol. 2007, No. 107).
- Yogaswari, D. D., Nugroho, A. B., & Astuti, N. C. (2012). The Effect of Macroeconomic Variables on Stock Price Volatility: Evidence from Jakarta Composite Index, Agriculture, and Basic Industry Sector. Journal of Business and Management, 46, 96-100.

10. Appendix

10.1. Maximum likelihood

As mentioned above the GARCH (1, 1) model cannot be estimated by a simple OLS regression and the use of maximum likelihood estimation should be used. Assuming that we have normally distributed errors, and starting from the parameter vector which is denoted by the Greek letter 'θ', and a time series observations that has the size of $(y_1, \dots., y_n)$, the maximum likelihood estimation method is able to calculate the probability density for this specific sample. It can also be written as L(θ|$(y_1, \dots., y_n)$. This can be done by taking the product of all the conditional probability frequencies of the observed data, keeping in mind as mentioned before that the returns should be independent from each other. Unfortunately in the GARCH (1, 1) model the returns are usually not independent, but still the joint density function can be written as following:

$$f\big((y_1, \dots., y_n), |\theta\big) = \mathrm{f}(y_n|I_{n-1})\mathrm{f}(y_{n-1}|I_{n-2}) \dots.\mathrm{f}(y_n) \dots.\mathrm{f}(y_1) \quad (11)$$

the likelihood equation is maximized with respect to the unknown parameters and is written as:

$$L(\theta|I_{n-1}) = \Pi_{t=1}^{n} f(y_t|I_{t-1}) \qquad (12)$$

Where I_t is the information that is available at the time t and f is the density function of y_t. In general the maximum likelihood estimation is selecting the values of the parameters of the model under which the data have the greater probability of being generated (Xie, 2007). There might be though a difficulty when using the maximum likelihood estimation in time series, and this has to do with the derivation of the likelihood function. Since the observations of time series are usually dependent, the maximization of the likelihood can often be complex.

Figure 1: AEX adj. closing prices plot

Figure 2: S&P 500 adj. closing prices plot

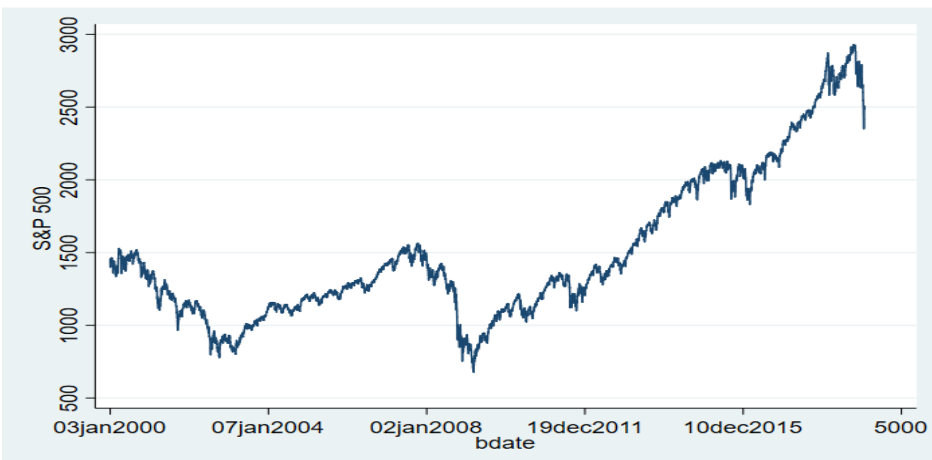


Figure 3: NIKKEI 225 adj. closing prices plot

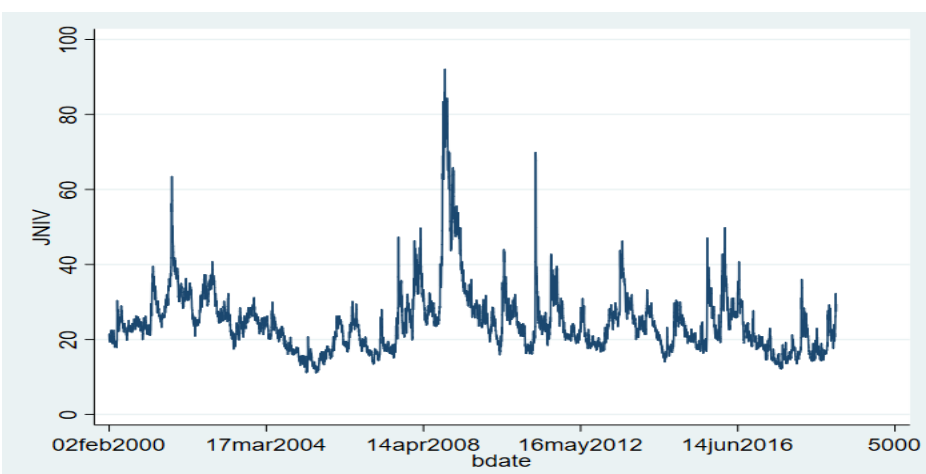


Figure 4: AEX log return plot

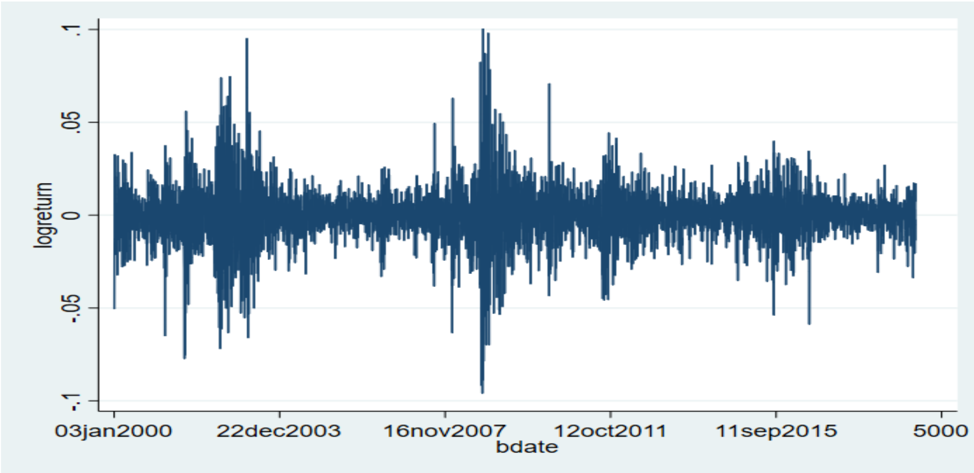


Figure 5: S&P 500 log return plot

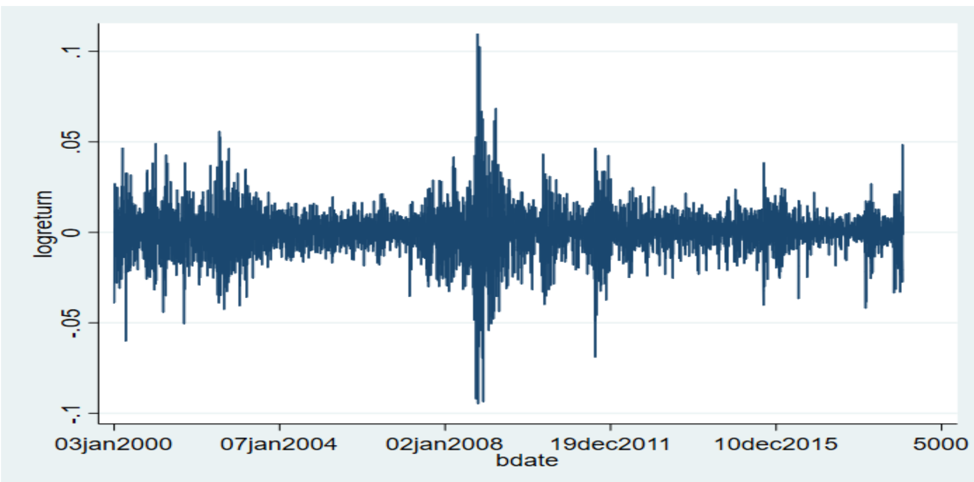


Figure 6: NIKKEI 225 log return plot

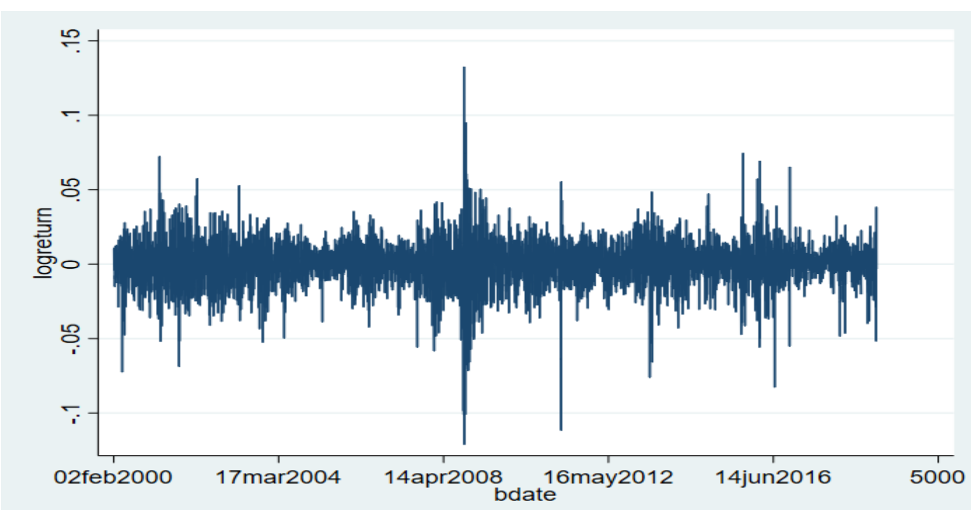


Figure 7: AEX conditional variance plot

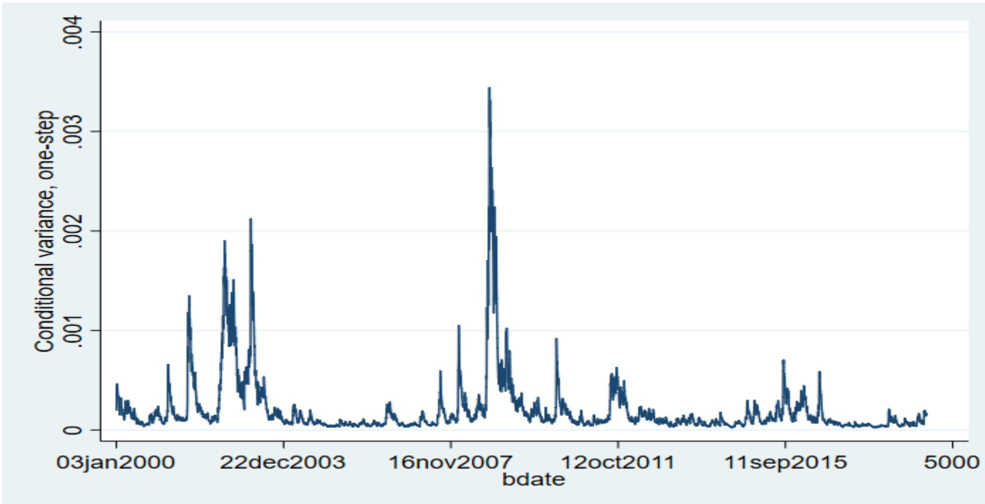


Figure 8: S&P 500 conditional variance plot

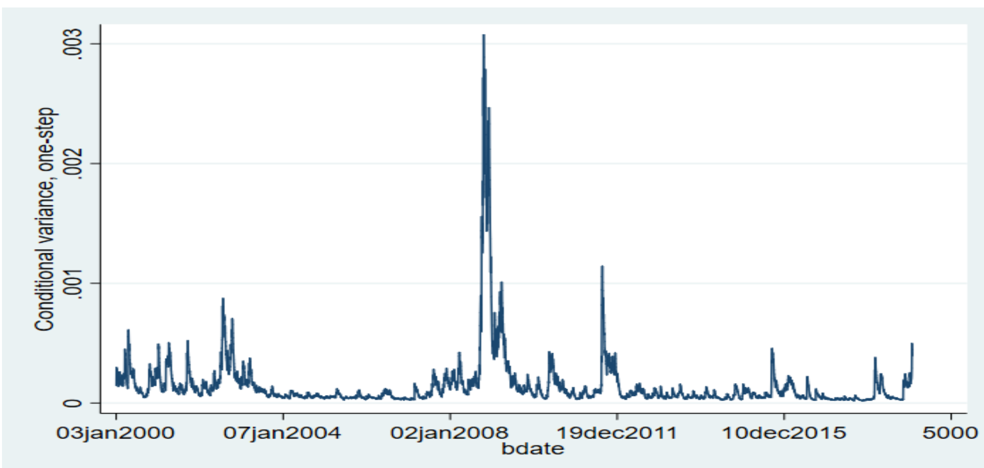


Figure 9: NIKKEI 225 conditional variance plot

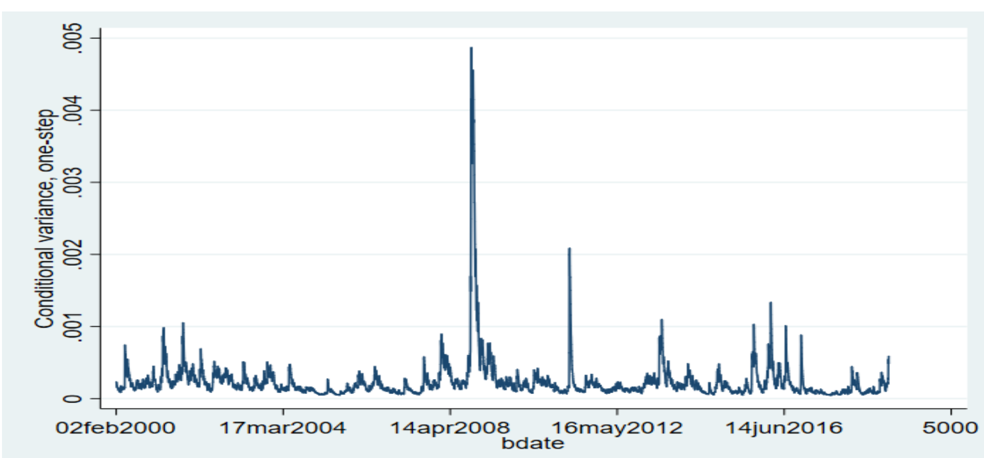


Figure 10 : AEX Normal distribution

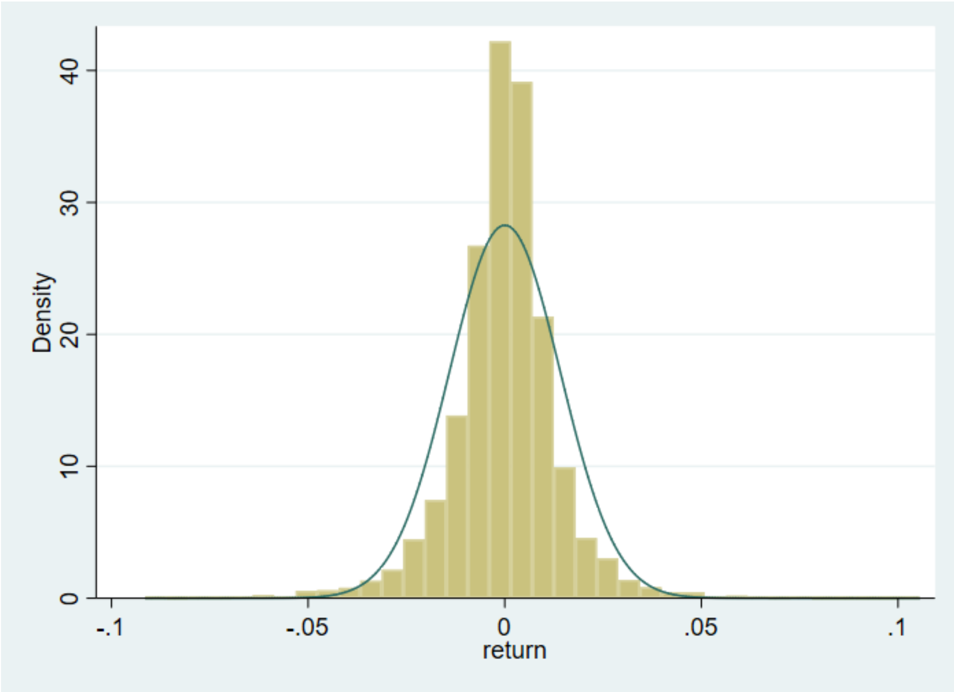


Figure 11 : S&P 500 Normal distribution

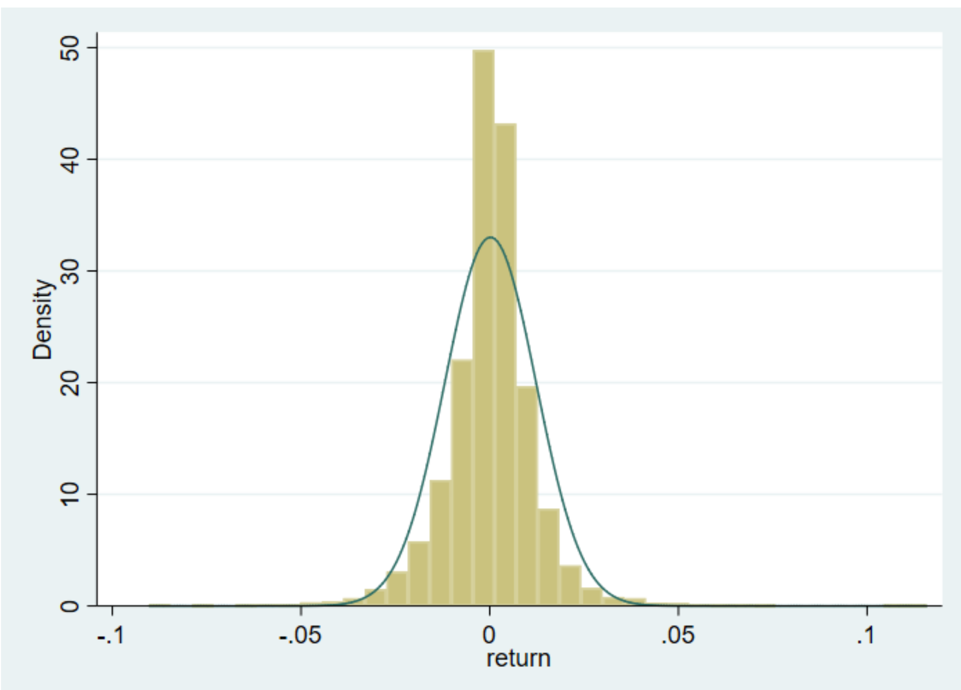


Figure 11 : NIKKEI 225 Normal distribution

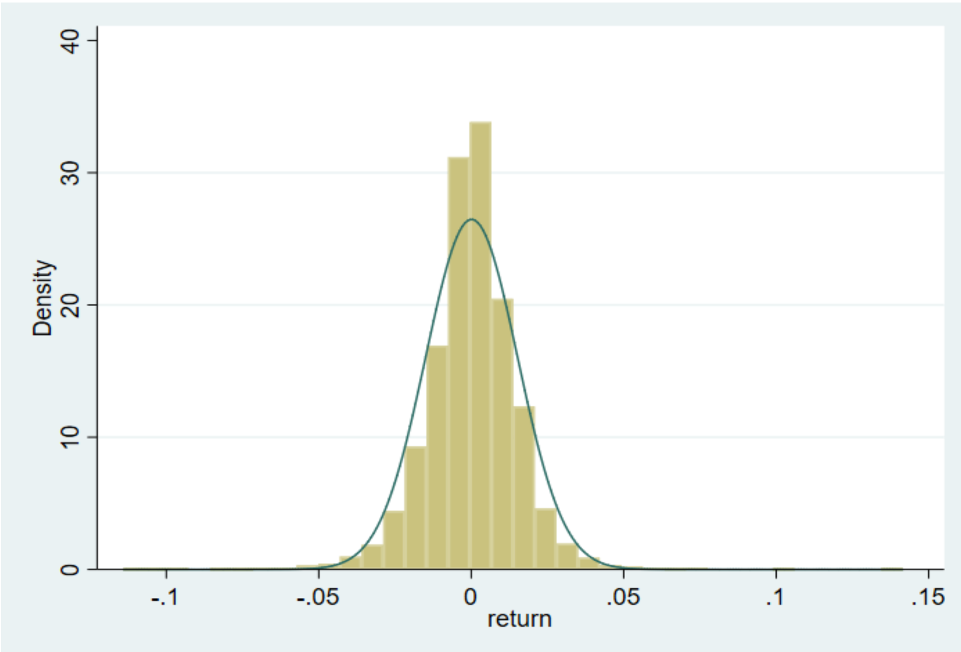


Table 1b: Summary of Volatility indices MAE method

Variable	Obs	Mean	Std.Dev.	Min	Max
VAEX MAE	4592	.019	.022	0	.188
VIX MAE	4516	.014	.016	0	.533
JNIV MAE	4341	.024	.024	0	.285

Table 4a: Regression for S&P 500 of the initial sample

Numer of obs 252 Strong convergence iteration time: 0.07					
	Estimate	Std. Err.	T Ratio	p-Value	Sig
MA1	0.00959	0.07263	-0.132	0.895	
[2]GARCH Intercept^(1/2)	0.00619	0.0045	-------	--------	--------
GARCH AR1	0.81025	0.30679	2.641	0.009	***
GARCH MA1	0.642	0.26474	2.425	0.016	**

R-Squared = 0.0014

Table 4b: Linear regression for S&P 500 of the initial sample

Number of obs 252

RV	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
RV1	0.358	0.074	4.85	0.000	0.213	0.504	***
RV5	0.068	0.141	0.48	0.631	-0.210	0.346	
RV22	0.282	0.157	1.80	0.074	-0.027	0.592	*
Constant	0.000	0.000	2.48	0.014	0.000	0.000	**

Mean dependent var	0.000	SD dependent var	0.000
R-squared	0.222	Number of obs	252
F-test	21.419	Prob > F	0.000
Akaike crit. (AIC)	-3453.721	Bayesian crit. (BIC)	-3439.986

**** p<0.01, ** p<0.05, * p<0.1*

Table 4c: VIF S&P 500

	VIF	1/VIF
RV5	2.519	.397
RV22	1.902	.526
RV1	1.581	.633
Mean VIF	2.001	.

Table 5a: Regression for Nikkei 225 of the initial sample

Numer of obs 252

Strong convergence
iteration time: 0.07

	Estimate	Std. Err.	T Ratio	p-Value	Sig
MA1	-0.04396	0.04814	-0.913	0.362	
[2]GARCH Intercept^(1/2)	0.0032	0.0009	-------	--------	--------
GARCH AR1	0.95051	0.02671	35.586	0	***
GARCH MA1	0.94483	0.0601	15.721	0	***
R-Squared = 0.0011					

Table 5b: Linear regression for Nikkei 225 of the initial sample

Number of obs	252

RV	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
RV1	0.159	0.077	2.06	0.041	0.007	0.311	**
RV5	0.460	0.140	3.29	0.001	0.185	0.735	***
RV22	0.019	0.184	0.10	0.920	0.381	0.344	
Constant	0.000	0.000	2.81	0.005	0.000	0.000	***

Mean dependent var	0.000	SD dependent var	0.000
R-squared	0.172	Number of obs	252
F-test	16.393	Prob > F	0.000
Akaike crit. (AIC)	-3881.342	Bayesian crit. (BIC)	-3867.130

**** p<0.01, ** p<0.05, * p<0.1*

Table 5c: VIF Nikkei 225

	VIF	1/VIF
RV5	2.291	.437
RV1	1.757	.569
RV22	1.45	.69
Mean VIF	1.833	.

Table 7a: S&P 500 Descriptive Statistics of MSE method for 2001-2018

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	4516	.004	.056	0	3.235
HAR MSE	4517	.002	.042	0	2.453
VIX MSE	4516	.005	.046	0	2.841

Table 7b: S&P 500 Descriptive Statistics of MAE method for 2001-2018

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	4516	.007	.019	0	.569
HAR MAE	4517	.005	.015	0	.495
VIX MAE	4516	.014	.016	0	.533

Table 7c: S&P 500 Descriptive Statistics of MSE method for 2001 to mid-2002

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	322	.0013	.004	0	.052
HAR MSE	321	.0009	.005	0	.072
VIX MSE	322	.0045	.005	0	.042

Table 7d: S&P 500 Descriptive Statistics of MAE method for 2001 to mid-2002

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	322	.008	.008	0	.072
HAR MAE	321	.005	.008	0	.085
VIX MAE	322	.019	.01	0	.065

Table 7e: S&P 500 Descriptive Statistics of MSE method for 2007 to early 2009

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	574	.024	.148	0	3.235
HAR MSE	573	.012	.108	0	2.453
VIX MSE	574	.019	.122	0	2.841

Table 7f: S&P 500 Descriptive Statistics of MAE method for 2007 to early 2009

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	574	.023	.043	0	.569
HAR MAE	573	.014	.032	0	.495
VIX MAE	574	.027	.034	0	.533

Table 8a: Nikkei Descriptive Statistics of MSE method for 2001-2018

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	4341	.010	.096	0	2.501
HAR MSE	4376	.001	.013	0	.523
JNIV MSE	4341	.012	.043	0	.811

Table 8b Nikkei Descriptive Statistics of MAE method for 2001-2018

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	4341	.014	.028	0	.5
HAR MAE	4341	.004	.01	0	.229
JNIV MAE	4341	.025	.024	0	.285

Table 8c: Nikkei Descriptive Statistics of MSE method for 2001 to mid-2002

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	324	.003	.003	0	.034
HAR MSE	324	.001	.002	0	.036
JNIV MSE	324	.014	.021	0	.259

Table 8d: Nikkei Descriptive Statistics of MAE method for 2001 to mid-2002

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	324	.014	.008	0	.058
HAR MAE	324	.006	.006	0	.06
JNIV MAE	324	.033	.018	.001	.161

Table 8e: Nikkei Descriptive Statistics of MSE method for 2007 to early 2009

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MSE	560	.055	.255	0	2.501
HAR MSE	559	.003	.02	0	.368
JNIV MSE	560	.050	.107	0	.811

Table 8f: Nikkei Descriptive Statistics of MAE method for 2007 to early 2009

Variable	Obs	Mean	Std.Dev.	Min	Max
GARCH MAE	560	.033	.066	0	.5
HAR MAE	559	.008	.017	0	.192
JNIV MAE	560	.049	.05	.002	.285

Whole period of data AEX

Variable	Obs	Mean	Std.Dev.	Min	Max
RV	4844	.0001162	.0002	.00000016	.004
GARCH VAR	4592	.0002082	.0003587	.00000017	.005
HAR VAR	4594	.0000932	.0001639	-.00000041	.002
VAEX VAR	4844	.0002982	.000338	.000042	.003